

TÓPICOS

REVISTA DE FILOSOFÍA

SEPTIEMBRE-DICIEMBRE 2026 | ISSN 2007-8498 (EN LÍNEA)

La “extraña inversión del razonamiento” sobre el lenguaje en la obra de Dennett

Carolina Scotto

Por el camino de Dennett: una genealogía de la moralidad

Antoni Gomila

¿Hay diseño en la naturaleza? Daniel Dennett y el problema de la teleología: cuestiones epistemológicas y didácticas

Leonardo González Galli

Cuando el “ácido universal” darwiniano entra en contacto con la creatividad humana

Malena León

Human Free Will and Determinism: Critical Considerations on Daniel Dennett’s Philosophy and the Dangerous Idea of Freedom in Evolutionary Epistemology

Vicente Claramonte Sanz y Rodolfo Guarinos Rico

Dennett, Gould y la polémica sobre el adaptacionismo cuatro décadas después

Rodrigo Sebastián Braicovich

Disfunción, desajuste y subdeterminación empírica

Federico Burdman, Juan R. Loaiza y Francisco Pereira

Dennett y el materialismo cultural

Fernando Broncano

Simulación y justificación en la era del deep learning: AlphaFold como ejemplar paradigmático de modelos opacos

Erick Manuel Rubio, Leandro Giri, Miguel Fuentes y Elkin Navarro

Análisis de algunas críticas contemporáneas en torno a la eventual propuesta política de Alasdair MacIntyre

Alberto Díaz Rivero y Sebastián Buzeta Undurraga

El reconocimiento identitario en la filosofía de Hannah Arendt: la primacía del espectador y los límites del agente

María Camila Sanabria, Jesús María Carrasquilla Ospina y Tatiana Saavedra Flórez

La interpretación dialéctica de la hermenéutica

Luis Eduardo Gama Barbosa

Interpretation of Space and Understanding of Self: Reading Foucault and Ricoeur in One Frame

Vidya Mary George

La ausencia de la necesidad en Husserl como concepto explicativo del fenómeno del valor. Una propuesta desde Carl Menger y Ludwig Feuerbach

Diego Colomé

Descartes, un comienzo absoluto. La génesis del pensamiento moderno en la lectura heideggeriana de las *Regulae* de Descartes

José Francisco Ramírez Olvera

La interrupción en el sistema de la absoluta objetividad de Friedrich Heinrich Jacobi

Federico Vicum

El liberalismo como religión y su repercusión global. Las interpretaciones críticas de John Gray y John Milbank en perspectiva comparada

Luis Aránguiz Kahn

RESEÑAS

76

TÓPICOS, REVISTA DE FILOSOFÍA

UNIVERSIDAD PANAMERICANA
México 2026

Tópicos, Revista de Filosofía es una publicación cuatrimestral que aparece en los meses de enero, mayo y septiembre. Se distribuye internacionalmente mediante intercambio y donación. Para mayor información consulte nuestra página de Internet en: <https://revistas.up.edu.mx/topicos/>

Tópicos, Revista de Filosofía aparece en los siguientes servicios de indización y resúmenes: The Philosopher's Index, Répertoire bibliographique de la philosophie, SCImago Journal Rank, DIALNET, Latindex, Filos, Redalyc, Clase, SCOPUS, Elsevier, Scielo, Scielo Citation Index, REDIB, DOAJ y Fuente Académica EBSCO.

Tópicos, Revista de Filosofía acepta artículos originales no solicitados y no publicados previamente. Las especificaciones del formato y calidad de estos están explicados en la sección *Para los colaboradores*, al final de este ejemplar. Ahí puede leerse también nuestro código de ética.

Tópicos, Revista de Filosofía Año 36, número 76, septiembre-diciembre, es una publicación cuatrimestral editada y publicada por Centros Culturales de México, A.C., propietaria de la Universidad Panamericana, Facultad de Filosofía, Augusto Rodin #498, Col. Insurgentes Mixcoac, Del. Benito Juárez, C.P. 03920, México, DF. Tel. 54821649, topicos@up.edu.mx; topicosojs@up.edu.mx. Editor Responsable: Dr. Luis Xavier López Farjeat. Reservas de derechos al Uso Exclusivo No. 04 - 2013 - 101810182700 - 203. ISSN impreso: 0188 - 6649, ambos otorgados por el Instituto Nacional del Derecho de Autor. Licitud de título y contenido No. 16092, otorgado por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación. Reservas de derechos al Uso Exclusivo electrónico: 04-2013-102110203400-102. ISSN electrónico: 2007 - 8498; otorgados por el Instituto Nacional del Derecho de Autor. Impresa por Editorial Ducere, S.A. de C.V. Rosa Esmeralda num3-bis. Col. Molino de Rosas, C.P. 01470, Deleg. Álvaro Obregón, México, D.F. Este número se terminó de imprimir el 1 de septiembre de 2026 con un tiraje de 250 ejemplares.

Diseño de Portada: Litholred, Rúbrica Contemporánea.

Diseño de Caja: J. Luis Rivera N.

Compuesto en Palatino Linotype y Computer Modern Roman, con InDesign.

All TradeMarks are the property of their respective owners.

Toda correspondencia deberá enviarse a:

Tópicos, Revista de Filosofía

UNIVERSIDAD PANAMERICANA

Facultad de Filosofía

Augusto Rodin #498

Insurgentes Mixcoac

03920 México, D.F. México

<https://revistas.up.edu.mx/topicos>

© 2026 UNIVERSIDAD PANAMERICANA

Las opiniones expresadas por los autores en los artículos y reseñas son de su exclusiva responsabilidad.

TÓPICOS, REVISTA DE FILOSOFÍA

UNIVERSIDAD PANAMERICANA

CONSEJO EDITORIAL

Atocha Aliseda Llera, *Universidad Nacional Autónoma de México, México*

Juan Arana, *Universidad de Sevilla, España*

Virginia Aspe, *Universidad Panamericana, México*

Mauricio Beuchot, *Universidad Nacional Autónoma de México, México*

Marcelo Boeri, *Pontificia Universidad Católica de Chile, Chile*

Pierre Destrée, *Université Catholique de Louvain, Bélgica*

Alexander Fidora, *Universitat Autònoma de Barcelona, España*

Alejandro Herrera, *Universidad Nacional Autónoma de México, México*

Guillermo Hurtado, *Universidad Nacional Autónoma de México, México*

Gustavo Leyva, *Universidad Autónoma Metropolitana, México*

Hendrik Lorenz, *Princeton University, EUA*

Alfredo Marcos, *Universidad de Valladolid, España*

Carlos Pereda, *Universidad Nacional Autónoma de México, México*

Teresa Santiago, *Universidad Autónoma Metropolitana, México*

Richard C. Taylor, *Marquette University, EUA*

Alejandro Vigo, *Universidad de los Andes, Chile*

†Ignacio Angelelli, *University of Texas at Austin, EUA*

†Enrico Berti, *Università di Padova, Italia*

†Paulette Dieterlen, *Universidad Nacional Autónoma de México, México*

†Alfonso Gómez Lobo, *Georgetown University, EUA*

†Jorge Gracia, *State University of New York at Buffalo, EUA*

†Franco Volpi, *Università di Padova, Italia*

CONSEJO DE REDACCIÓN

María Elena García Peláez Cruz

Eduardo Charpenel Elorduy

EDITOR: Luis Xavier López Farjeat

SECRETARIA EDITORIAL: Karen González Fernández

ADJUNTO EDITORIAL: Luis Bazet

Índice general

Artículos

- Treinta años de *La peligrosa idea de Darwin*: estudios en torno a la obra de Daniel Dennett. Introducción de los editores del número
(Thirty Years of *Darwin's Dangerous Idea*: Studies on the Work of Daniel Dennett. Introduction by the Issue's Editors) 11
Joaquín Suárez-Ruíz y Rodrigo López-Orellana
- La “extraña inversión del razonamiento” sobre el lenguaje en la obra de Dennett
(The “Strange Inversion of Reasoning” about Language in Dennett's Work) 21
Carolina Scotto
- Por el camino de Dennett: una genealogía de la moralidad
(On Dennett's Path: A Genealogy of Morality) 67
Antoni Gomila
- ¿Hay diseño en la naturaleza? Daniel Dennett y el problema de la teleología: cuestiones epistemológicas y didácticas
(Is There Design in Nature? Daniel Dennett and the Problem of Teleology: Epistemological and Didactic Issues) 87
Leonardo González Galli
- Cuando el “ácido universal” darwiniano entra en contacto con la creatividad humana
(When Darwin's “Universal Acid” Comes into Contact with Human Creativity) 119
Malena León

Human Free Will and Determinism: Critical Considerations on Daniel Dennett's Philosophy and the Dangerous Idea of Freedom in Evolutionary Epistemology (El libre albedrío humano y el determinismo: reflexiones críticas sobre la filosofía de Daniel Dennett y la peligrosa idea de la libertad en la epistemología evolutiva)	159
<i>Vicente Claramonte Sanz y Rodolfo Guarinos Rico</i>	
Dennett, Gould y la polémica sobre el adaptacionismo cuatro décadas después (Dennett, Gould and the Controversy over Adaptationism Four Decades Later)	187
<i>Rodrigo Sebastián Braicovich</i>	
Disfunción, desajuste y subdeterminación empírica (Dysfunction, Mismatch, and Empirical Underdetermination)	219
<i>Federico Burdman, Juan R. Loaiza y Francisco Pereira</i>	
Dennett y el materialismo cultural (Dennett and Cultural Materialism)	255
<i>Fernando Broncano</i>	
Simulación y justificación en la era del <i>deep learning</i>: AlphaFold como ejemplar paradigmático de modelos opacos (Simulation and Justification in the Age of Deep Learning: AlphaFold as a Paradigmatic Case of Opaque Models)	288
<i>Erick Manuel Rubio, Leandro Giri, Miguel Fuentes, Elkin Navarro</i>	
Análisis de algunas críticas contemporáneas en torno a la eventual propuesta política de Alasdair MacIntyre (Analysis of Some Contemporary Criticisms of Alasdair Macintyre's Eventual Political Proposal)	315
<i>Alberto Díaz Rivero y Sebastián Buzeta Undurraga</i>	

El reconocimiento identitario en la filosofía de Hannah Arendt: la primacía del espectador y los límites del agente (The Recognition of Identity in Hannah Arendt's Philosophy: The Primacy of the Spectator and the Limits of the Agent)	345
<i>María Camila Sanabria, Jesús María Carrasquilla Ospina, Tatiana Saavedra Flórez</i>	
La interpretación dialéctica de la hermenéutica (The Dialectical Interpretation of Hermeneutics)	373
<i>Luis Eduardo Gama Barbosa</i>	
Interpretation of Space and Understanding of Self: Reading Foucault and Ricoeur in One Frame (Interpretación del espacio y comprensión del yo: una lectura de Foucault y Ricoeur en un mismo marco)	409
<i>Vidya Mary George</i>	
La ausencia de la necesidad en Husserl como concepto explicativo del fenómeno del valor. Una propuesta desde Carl Menger y Ludwig Feuerbach (The Absence of Need in Husserl as an Explanatory Concept of the Phenomenon of Value: A Proposal from Carl Menger and Ludwig Feuerbach)	435
<i>Diego Colomé</i>	
Descartes, un comienzo absoluto. La génesis del pensamiento moderno en la lectura heideggeriana de las <i>Regulae</i> de Descartes (Descartes, an Absolute Beginning: The Genesis of Modern Thought in the Heideggerian Reading of Descartes' <i>Regulae</i>)	475
<i>José Francisco Ramírez Olvera</i>	

- La interrupción en el sistema de la absoluta objetividad de Friedrich Heinrich Jacobi**
(Interruption in Friedrich Heinrich Jacobi's system of absolute objectivity) 511
Federico Vicum

Filosofía en el espacio público

- El liberalismo como religión y su repercusión global. Las interpretaciones críticas de John Gray y John Milbank en perspectiva comparada**
(Liberalism as Religion and Its Global Repercussion: A Comparative Perspective on the Critical Interpretations of John Gray and John Milbank) 551
Luis Aránguiz Kahn

Reseñas

- Koch, K. (2023). *Denken in Zwecken. Bedeutung und Status der Teleologie in der theoretischen Philosophie Kants und Hegels*. Felix Meiner Verlag. 310 pp.** 583
Zaida Olvera
- Aouad, M., Di Vincenzo, S. y Fadlallah, H. (2023). *Averroes' Middle Commentary on Aristotle's Metaphysics: Critical Edition of the Arabic Version, French Translation and English Introduction*. Brill. 632 pp.** 591
Josep Puig
- Gerlach, S. (2023). *Schellings Philosophie der Potenz 1798-1854*. Mohr Siebeck. 300 pp.** 599
Osman Choque-Aliaga

Artículos

<https://doi.org/10.21555/top.v760.3861>

Treinta años de *La peligrosa idea de Darwin*: estudios
en torno a la obra de Daniel Dennett. Introducción
de los editores del número

Thirty Years of *Darwin's Dangerous Idea*: Studies on
the Work of Daniel Dennett. Introduction by the
Issue's Editors

Joaquín Suárez-Ruíz

Universidad Nacional de La Plata

Argentina

jsuarez@fahce.unlp.edu.ar

<https://orcid.org/0000-0002-0299-8893>

Rodrigo López-Orellana

Universidad de Valparaíso

Chile

rodrigo.lopez@uv.cl

<https://orcid.org/0000-0002-3576-0136>

Cómo citar este artículo: Suárez-Ruíz, J. y López-Orellana, R. (2026). Treinta años de *La peligrosa idea de Darwin*: estudios en torno a la obra de Daniel Dennett. Introducción de los editores del número. *Tópicos, Revista de Filosofía*, 76, 11-19. <https://doi.org/10.21555/top.v760.3861>



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

Hay filósofos que construyen sistemas, hay filósofos que los derriban. Daniel Dennett (1942-2024) fue, con notable destreza, ambas cosas a la vez. Su influencia se extendió más allá de los límites disciplinarios: hizo borroso el límite entre una investigación “propriadamente” científica y una “propriadamente” filosófica. *Content and Consciousness* (1969) inauguró una carrera que pasaría por hitos como *Brainstorms* (1978), *Elbow Room* (1984), *The Intentional Stance* (1987), *Consciousness Explained* (1991), *Freedom Evolves* (2003) y *From Bacteria to Bach and Back* (2017), entre muchos otros. Cada uno de estos libros desafió consensos establecidos, provocó polémicas fértiles y, sobre todo, amplió el horizonte de lo que la filosofía podía preguntar y de las herramientas con las que podía responder. Durante más de medio siglo, su obra recorrió los territorios más disputados de la filosofía contemporánea —la consciencia, la intencionalidad, la evolución biológica, el libre albedrío, la religión— sin perder jamás ni el rigor argumentativo ni una vocación contagiosa por hacer de la filosofía una empresa pública, accesible y urgente.

Este número especial se inspira en un libro en particular del filósofo: *Darwin's Dangerous Idea* (1995). Treinta años después de su publicación, es una obra que sigue ocupando un lugar central en las discusiones contemporáneas sobre la mente, la evolución y el alcance explicativo de la ciencia. En este libro ambicioso y provocador, Dennett retomó la teoría de Charles Darwin no solo como un hito biológico, sino como una auténtica “idea peligrosa”, esto es, en tanto en cuanto un principio capaz de transformar radicalmente nuestra comprensión de la vida, la cultura y la mente humana. Esta peligrosidad a la que alude el filósofo no reside en una amenaza externa, sino en su poder corrosivo sobre ciertas intuiciones profundamente arraigadas: la creencia en una mente excepcional, en una consciencia irreductible o en explicaciones últimas que trascienden los mecanismos naturales. Frente a estas posiciones, Dennett propone una visión decididamente naturalista, según la cual la complejidad puede explicarse mediante procesos graduales, acumulativos y ciegos, como los que describe la selección natural. Así, la filosofía de Dennett, especialmente desde *Darwin's Dangerous Idea*, se construyó sobre una convicción que nunca abandonó, a saber, que *la mente humana es un fenómeno natural* y, más específicamente, *el resultado de procesos evolutivos graduales*. En esta clave, explicarla no exige sino

extender con rigor y audacia las herramientas que la ciencia ya tiene a su disposición. No había en esa postura, vale decir, ningún desprecio por la riqueza de la experiencia subjetiva. Había, en cambio, la sospecha, argumentada con paciencia de décadas, de que esa riqueza nos había llevado a postular fantasmas donde había procesos y misterios donde había complejidad aún no explicada.

Dennett falleció en abril del 2024, tiempo a partir del cual las preguntas a las que había dedicado su vida comenzaron a dejar de ser interrogantes discutidos únicamente por filósofos: ¿qué distingue una mente que comprende de un sistema que simula comprensión?, ¿puede emerger la intencionalidad de procesos que no la tienen?, ¿dónde termina la metáfora y empieza la descripción literal cuando hablamos de “máquinas que piensan”? Durante décadas, estas preguntas fueron patrimonio de un grupo reducido de pensadores, científicos cognitivos e ingenieros con vocación especulativa. De pronto, se han vuelto preguntas de interés común: desde legisladores que intentan regular sistemas que no entienden y usuarios que no saben bien con qué están interactuando, hasta técnicos que construyen esas máquinas y dudan, en privado, de lo que han construido. Este desplazamiento de la academia al mundo, de la disputa tecnológica a la urgencia pública, es el contexto en que este número especial aparece, y también, en cierta medida, la razón de su oportunidad. No se trata de que Dennett haya anticipado el presente con precisión profética, sino que las herramientas conceptuales que forjó a lo largo de cincuenta años de trabajo resultan, ahora, de una utilidad que excede ampliamente el dominio para el que fueron construidas.

Vale resaltar que este número no consiste solamente en un ejercicio académico. Es también una respuesta a una forma de hacer filosofía. Dennett creyó siempre que las grandes preguntas sobre la mente y el lugar del ser humano en la naturaleza eran demasiado importantes para quedarse únicamente en seminarios, coloquios y congresos. Escribió para los científicos y para los lectores curiosos, para los que llegaban desde la biología evolutiva y para los que llegaban desde la literatura. Esa vocación de apertura, que algunos quizás confundieron con simplificación, fue en realidad una de las expresiones más consecuentes de su naturalismo: si la mente es un fenómeno del mundo, hablar de ella es hablar del mundo, y esa conversación no tiene por qué ocurrir a puertas cerradas. El fundamento de este número especial es, entonces, una disposición intelectual: aquella que invita a seguir una pregunta hasta donde lleve sin declarar ningún territorio vedado a la explicación,

sin proteger ninguna intuición de la crítica. Esta disposición es más rara de lo que parece, y más difícil de sostener de lo que sus resultados sugieren. Dennett la mantuvo frente a adversarios formidables —p. ej., Chalmers, Nagel, Searle, Fodor— y frente a la incomodidad, a veces genuina, que sus conclusiones producían incluso entre quienes compartían sus premisas. Los trabajos reunidos en este número son herederos de esa disposición, aunque no siempre de sus conclusiones. Algunos defienden sus posiciones con argumentos nuevos; otros las cuestionan con la misma seriedad con que él cuestionó las ajenas.

El primer artículo, de Carolina Scotto, se propone reconstruir y analizar el lugar que ocupa el lenguaje dentro del proyecto filosófico global de Daniel Dennett, cuyo libro *Darwin's Dangerous Idea* constituye el eje central de la indagación. Para comprender la importancia que Dennett le asigna al fenómeno lingüístico, es necesario partir de la valoración que hace de la teoría de la evolución por selección natural. A juicio de Dennett, ninguna otra teoría científica ha contribuido tanto a la filosofía como la de Darwin, precisamente porque habría logrado lo que ninguna otra consiguió, a saber, unificar en un único y coherente marco explicativo dos reinos que la tradición filosófica y científica había mantenido rigurosamente separados: por un lado, el reino de la física, los mecanismos y la materia ciega; por otro lado, el reino del significado, el propósito, la cultura y la mente humana. Esta unificación implica una tesis de enorme alcance, en el sentido tradicional del término. Esto es, todos los logros de la civilización humana (el lenguaje, el arte, la religión, la ética, la ciencia misma) son el resultado del mismo proceso gradual, acumulativo y carente de propósito que dio origen a las formas de vida más simples, sin necesidad de postular saltos cualitativos inexplicables, fines prefigurados, diseño inteligente ni ninguna forma de creación especial. Esta “peligrosa idea de Darwin”, sintetizada en la expresión dennettiana de una “extraña inversión del razonamiento”, según la cual la competencia siempre precede a la comprensión, es la que Scotto se propone rastrear específicamente en el terreno del lenguaje mostrando cómo impacta en cada una de sus dimensiones de análisis. La filósofa enfatiza que las ideas de Dennett sobre el lenguaje no son dispersas, periféricas ni de significación secundaria dentro de su obra, sino que conforman una visión global, coherente y sustantiva que resulta absolutamente indispensable para sostener el objetivo unificador más ambicioso del proyecto dennettiano, hasta el punto de que, si dichas interpretaciones logran integrarse exitosamente en ese marco,

las consecuencias serían de largo alcance: buena parte de los supuestos, teorías y herramientas conceptuales que integran el *mainstream* de la filosofía analítica del lenguaje, así como de diversas corrientes de la lingüística teórica contemporánea, deberían ser sometidos a una profunda revisión crítica o, en algunos casos, directamente abandonados.

En segundo lugar, Antoni Gomila, partiendo de un vínculo personal e intelectual con Dennett que se remonta a un encuentro en Salamanca en 1990, se propone examinar críticamente los capítulos finales de *Darwin's Dangerous Idea* dedicados a la moral y ofrecer, en su lugar, una explicación evolutiva más rigurosa y completa de la psicología moral humana. El punto de partida es una observación paradójica: a pesar de que los capítulos 16 y 17 deberían constituir la culminación del esfuerzo sostenido a lo largo de más de quinientas páginas por mostrar el poder explicativo de la selección natural aplicada a todos los dominios de lo humano, Dennett no llega en ellos a ofrecer una explicación evolutiva propiamente dicha de la moral. Frente a estas limitaciones, Gomila propone una genealogía alternativa de la moralidad articulada en dos fases sucesivas y complementarias que integran, en lugar de contraponer, la evolución biológica y la evolución cultural a través del desarrollo ontogenético individual. El artículo concluye reconociendo en Dennett un modelo ejemplar de filosofía naturalista, científicamente comprometida, culturalmente relevante y alejada de la aridez del análisis conceptual, cuya contribución a reivindicar la importancia de la perspectiva evolutiva para las ciencias sociales y humanidades resulta duradera y valiosa, aun cuando su tratamiento específico de la moral haya quedado inconcluso y necesite ser desarrollado y corregido en la dirección que el propio autor propone.

En tercer lugar, el artículo de Leonardo González Galli examina los análisis filosóficos de Daniel Dennett sobre la teleología en la biología y explora sus implicancias didácticas para la enseñanza de las ciencias naturales. En el plano epistemológico, González Galli reconstruye tres ideas centrales e interrelacionadas de Dennett: la “postura intencional”, que propone tratar a los organismos y sus partes como si fueran sistemas intencionales para predecir y explicar su conducta; la biología como ingeniería invertida, que consiste en analizar los rasgos adaptativos como si fueran artefactos de diseño para inferir las razones que la selección natural “descubrió” durante el proceso evolutivo; y la defensa de un concepto “fuerte” de diseño biológico, según el cual el diseño en la naturaleza no es una mera metáfora sino diseño auténtico, dado que la

selección natural constituye un proceso genuinamente análogo al diseño deliberado humano, aunque sin intencionalidad ni representación mental previa. En el plano didáctico, González Galli parte de constatar que los estudiantes exhiben un sesgo teleológico intuitivo universal y resistente al cambio, y critica el enfoque “eliminacionista” tradicional —que busca suprimir todo pensamiento teleológico por considerarlo científicamente ilegítimo— argumentando que tanto la teleología en la biología científica como el sesgo teleológico en la biología intuitiva son aspectos legítimos, funcionales e ineliminables. Frente a ello, propone una perspectiva alternativa basada en el desarrollo de una “vigilancia metacognitiva” que permita a los estudiantes reconocer, regular y discriminar entre formas de teleología científicamente aceptables e inaceptables, tomando como fundamento conceptual la analogía dennettiana entre selección natural y diseño deliberado, cuyo análisis explícito de alcances y limitaciones en el aula podría constituir una herramienta pedagógica poderosa para mejorar la comprensión de la teoría de la evolución por selección natural.

En cuarto lugar, el artículo de Malena León parte de la definición estándar de la creatividad como la producción de algo que sea simultáneamente novedoso y valioso o eficaz, y cuestiona la concepción tradicional que atribuye los logros creativos a individuos excepcionales dotados de una capacidad casi extraordinaria, conocida como la “idea del genio creativo”. Frente a esta visión, recupera diversas críticas provenientes de los estudios contemporáneos sobre creatividad, que destacan su carácter no necesariamente excepcional, su dependencia del contexto cultural, su dimensión colectiva y la importancia del conocimiento previo del dominio. En este marco, propone profundizar estas críticas a partir de la filosofía naturalista de Daniel Dennett, quien, apoyado en la teoría darwiniana de la evolución, sostiene que la mente humana no es una entidad especial separada del mundo natural ni la única fuente de significado o diseño. Es decir, Dennett cuestiona la idea predarwiniana de que todo diseño requiere un diseñador consciente mostrando que procesos ciegos, como la selección natural, pueden generar estructuras complejas sin intención ni comprensión. A partir de estas tesis, León argumenta que la creatividad debe entenderse como un fenómeno plural, que no se limita a procesos mentales individuales ni requiere necesariamente comprensión consciente, sino que puede surgir de dinámicas acumulativas, mecánicas y también supraindividuales o colectivas. En consecuencia, propone una concepción darwinista de la

creatividad que desmitifica el papel del individuo genial, enfatiza la interacción entre múltiples niveles y procesos y busca alinear el estudio de la creatividad con una visión científica contemporánea, alejándola de interpretaciones románticas o intuitivas.

En quinto lugar, el artículo de Vicente Claramonte y Rodolfo Guarinos Rico ofrece un análisis crítico y detallado de la propuesta filosófica de Daniel Dennett en *Freedom Evolves* (2003). Defienden que la libertad humana puede comprenderse como un fenómeno natural surgido de la evolución biológica y, por tanto, compatible con las leyes causales del universo. Para sostener esta tesis, los autores examinan en profundidad el concepto de “causalidad” tal como se emplea en la ciencia —especialmente en la física— y distinguen diversos grados de posibilidad (lógica, física, biológica e histórica), mostrando cómo la libertad debe situarse en este marco naturalista y no en el metafísico tradicional. Asimismo, cuestionan la clasificación de Dennett entre universos deterministas, indeterministas y nihilistas, argumentando que estos últimos carecen de plausibilidad empírica y deberían subsumirse en alguna de las otras categorías, además de señalar ambigüedades en la distinción entre causalidad lógica y empírica. También revisan críticamente la noción clásica de “causa-efecto” e incorporan aportes de la física moderna (como la teoría del caos, la relatividad y la mecánica cuántica) que problematizan la idea de una precedencia temporal estricta entre causa y efecto, debilitando así las versiones más rígidas del determinismo. En este contexto, defienden una posición compatibilista según la cual el determinismo no elimina la libertad, ya que esta depende de capacidades cognitivas evolucionadas (como la representación de escenarios posibles, el aprendizaje intencional y el control conductual) que permiten a los organismos actuar en función de un rango de posibilidades reales. Finalmente, Claramonte y Guarinos concluyen que la libertad debe entenderse como un rasgo fenotípico adaptativo ligado al conocimiento y a la capacidad de anticipar y modificar el futuro, aunque subrayan la necesidad de refinar conceptualmente las nociones de “causalidad” y “posibilidad” en Dennett para lograr una mayor coherencia filosófica.

En sexto lugar, el artículo de Rodrigo Braicovich analiza la célebre controversia entre Daniel Dennett y Stephen Jay Gould en torno al alcance explicativo del adaptacionismo en la teoría de la evolución. Muestra que, pese a sus desacuerdos, ambos compartían ciertos puntos básicos: la selección natural no explica todos los rasgos, algunas explicaciones

adaptacionistas son válidas y todas son falibles. A partir de ahí, expone la crítica de Gould al “adaptacionismo radical”, que atribuye a la selección natural un poder explicativo total, defendiendo en cambio un enfoque pluralista que incluya otros factores como la deriva genética o las limitaciones estructurales. Aunque Dennett también reconoce abusos del adaptacionismo, mantiene su defensa de versiones moderadas, especialmente en campos como la sociobiología y la psicología evolucionista, mientras que Gould critica duramente estas corrientes por su carácter especulativo, su falta de evidencia empírica y su tendencia a construir relatos plausibles pero infalsables, especialmente en el caso de la psicología evolucionista y su tesis de la modularidad mental. Braicovich argumenta que, pese a décadas de debate, estas corrientes no han superado sus debilidades metodológicas ni teóricas, y que muchas de sus hipótesis (a menudo sesgadas cultural e ideológicamente) han persistido más por su atractivo narrativo y mediático que por su solidez científica. Finalmente, sugiere que el éxito de estas ideas, incluso en su versión “popular”, refleja tanto limitaciones epistemológicas como afinidades con ciertas visiones culturales dominantes, y se plantea la necesidad de un enfoque evolucionista de la mente más riguroso, empíricamente fundamentado y libre de reduccionismos.

En séptimo lugar, Federico Burdman, Juan Loaiza y Francisco Pereira examinan el debate en filosofía de la psiquiatría entre las teorías de la disfunción, que entienden los trastornos mentales como resultado de fallas en los mecanismos psicológicos o biológicos, y las teorías del desajuste, que explican el malestar como consecuencia de la diferencia entre el entorno actual y aquel en el que evolucionaron nuestros mecanismos mentales, sin necesidad de suponer una falla. Focalizándose en la adicción a sustancias, los autores analizan el funcionamiento del sistema de recompensa cerebral y plantean que el desacuerdo central entre ambas posturas radica en cómo se define la “función propia” de dicho sistema: puede concebirse como un mecanismo que rastrea y responde a aumentos de dopamina o, alternativamente, como uno que detecta estímulos con valor biológico como la comida o la interacción social. A partir de la filosofía de Dennett, sostienen que ambas interpretaciones son igualmente compatibles con la historia evolutiva del sistema, lo que genera una subdeterminación teórica que impide decidir de manera objetiva si la adicción constituye una disfunción o un caso de desajuste. En consecuencia, la disputa no puede resolverse apelando únicamente a la evidencia empírica, sino que requiere considerar criterios pragmáticos

y valorativos. Los autores concluyen que este cambio de enfoque no solo evita un falso dilema, sino que permite orientar de manera más fructífera el debate científico y clínico, incorporando explícitamente el papel de los valores en la comprensión y tratamiento de las adicciones.

Finalmente, en la sección *In memoriam*, Fernando Broncano realiza una evaluación filosófica rigurosa y matizada del materialismo de Daniel Dennett, reconstruyendo su posición en el marco de los grandes debates sobre la mente, la evolución y la cultura que dominaron la filosofía anglosajona en el último tercio del siglo XX. El punto de partida es el contexto del predominio filosófico del cognitivism computacional fodoriano, cuyo dualismo encubierto entre *hardware* cerebral y *software* mental Dennett se propuso dismantelar mediante una doble estrategia teórica: por un lado, la heterofenomenología, que escala entre niveles explicativos según la complejidad del sistema; por otro lado, el adaptacionismo darwinista, que unifica la explicación de todo fenómeno vivo o artificial bajo el poder del algoritmo de variación, reproducción y selección natural. En el texto se identifica con precisión los límites de este programa: el adaptacionismo de Dennett tiende a colapsar las tres formas distinguidas por Godfrey-Smith (empírica, explicativa y metodológica) en una versión fuerte que sostiene que la selección natural es la única fuerza evolutiva relevante, lo cual genera controversias tanto en el plano científico como en el filosófico. Frente a esto, autores como Sterelny, Godfrey-Smith y Malafouris proponen enriquecer el marco darwinista incorporando la construcción de nichos, la plasticidad fenotípica, la selección de grupo y, sobre todo, la cultura material como factores activos en la evolución humana, señalando que la relación entre mente y artefactos no es unidireccional (del artefacto hacia la mente, como sostiene Dennett) sino constitutivamente recíproca, conformando una “zona gris” en la que cerebros, cuerpos y cosas se fusionan y constituyen mutuamente. Broncano concluye con un balance que reconoce el valor iluminador y el impulso desmitificador del materialismo dennettiano, pero señala, al mismo tiempo, sus límites estructurales.

Por último, esperamos que este número especial contribuya a difundir la obra de Dennett entre los y las lectoras hispanohablantes.

<https://doi.org/10.21555/top.v760.3371>

La “extraña inversión del razonamiento” sobre el lenguaje en la obra de Dennett

The “Strange Inversion of Reasoning” about Language in Dennett’s Work

Carolina Scotto
Universidad Nacional de Córdoba
Argentina
carolina.scotto@unc.edu.ar
<https://orcid.org/0000-0002-2013-8539>

Recibido: 29 - 01 - 2025.
Aceptado: 08 - 05 - 2025.
Publicado en línea: 20 - 05 - 2026.

Cómo citar este artículo: Scotto, C. (2026). La “extraña inversión del razonamiento” sobre el lenguaje en la obra de Dennett. *Tópicos, Revista de Filosofía*, 76, 21-65. <https://doi.org/10.21555/top.v760.3371>



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

Resumen

En este trabajo me propongo mostrar de qué maneras, según Dennett, la “peligrosa idea” de Darwin alcanza también al lenguaje y a sus propiedades distintivas. Según Dennett, en virtud de una “extraña inversión del razonamiento” atribuible a la obra de Darwin, todo “diseño inteligente” obedece al principio de la *competencia sin comprensión*, según el cual la primera precede y está presupuesta por la segunda, por lo que, en algunos casos, es posible explicarla sin que la comprensión tenga lugar. Como se verá, la reconfiguración que resulta de aquella “inversión del razonamiento” acerca de la relación de los humanos con el lenguaje —cómo emergió, cómo es aprendido, cómo es usado, cuáles son sus efectos cognitivos y culturales— así como del lenguaje mismo como un conjunto de herramientas comunicativas y cognitivas, es de signo opuesto a la que ofrecen muchas concepciones filosóficas y lingüísticas contemporáneas, arraigadas en el principio opuesto: la primacía de la comprensión sobre la competencia.

Palabras clave: extraña inversión del razonamiento; competencia sin comprensión; lenguajes; Darwin; Dennett; evolución del lenguaje; memes; palabras; significados; comunicación.

Abstract

In this paper, I aim to show how, according to Dennett, Darwin’s “dangerous idea” also extends to language and its distinctive properties. According to Dennett, by Darwin’s “strange inversion of reasoning,” all “intelligent design” follows the principle of *competence without comprehension*: the former precedes and is presupposed by the latter, so that, sometimes, it is possible to explain it without comprehension taking place. As will be seen, the reconfiguration that results from that “inversion of reasoning” about the relationship between humans and language—how it emerged, how it is learned, how it is used, what its cognitive and cultural effects are—as well as language itself as a set of communicative and cognitive tools, is contrary to that offered by many contemporary philosophical and linguistic conceptions, rooted in the opposite principle: the primacy of comprehension over competence.

Keywords: strange inversion of reasoning; competence without comprehension; language; Darwin; Dennett; evolution of language; memes; words; meanings; communication.

Introducción: el proyecto dennettiano

Darwin's Dangerous Idea: Evolution and the Meanings of Life (1995) (en adelante: DDI),¹ uno de los libros más representativos de la obra de Daniel Dennett, es un homenaje entusiasta a la obra de Charles Darwin. Según su evaluación, en comparación con otras teorías científicas, la teoría de la evolución por selección natural contribuyó más que cualquier otra a la filosofía (cfr. Dennett, 2009a, p. 10061). El libro es una exposición argumentada de la tesis según la cual la teoría darwiniana, enriquecida por sus desarrollos y extensiones más recientes, habría conseguido unificar en un único marco explicativo “dos reinos” hasta entonces considerados independientes uno del otro: “el reino de la física y los mecanismos” y “el reino del significado y el propósito”, es decir, el de los seres humanos, sus características y sus creaciones. En otras palabras: “De acuerdo con la peligrosa idea de Darwin, todas las *posibles* exploraciones del espacio del diseño están conectadas” (1995, pp. 143-4) y, por lo tanto, “todos los logros de la cultura humana —*el lenguaje*, el arte, la religión, la ética y la ciencia misma— son artefactos (de artefactos de artefactos...) del mismo proceso fundamental que desarrolló las bacterias, los mamíferos, el *Homo Sapiens*. *No hay Creación Especial del lenguaje*, y ni el arte ni la religión tienen literalmente inspiración divina” (1995, p. 144; mi énfasis).

Según Dennett (2009a, p. 10061), la contribución de la teoría darwiniana sería revolucionaria no solo por la ambición de su alcance sino también por sus logros explicativos. Para justificar dicha valoración es preciso mostrar cómo la teoría puede explicar tanto el origen de la diversidad biológica como de la diversidad cultural a partir de precursores más simples o menos inteligentes y mediante procesos graduales, sin “saltos mágicos” ni fines prefigurados. Para ello, el libro pone al descubierto algunos fundamentos indiscutidos de la tradición filosófica, propone nuevas herramientas conceptuales, saca provecho de diferentes hipótesis y resultados provenientes de las más variadas disciplinas científicas y elabora un sinnúmero de conjeturas que, una vez

¹ Me referiré a este título tanto por sus iniciales como por el año de su publicación. A las demás publicaciones del autor me referiré solo por el año. Con excepción de las ediciones en español listadas en las referencias, las restantes traducciones de publicaciones en inglés son mías.

desarrolladas apropiadamente, podrían comportar logros significativos, sobre todo en el terreno más controversial: la teoría de la evolución cultural.² Teniendo el objetivo “de hacer que los pensadores en otras disciplinas tomen en serio el pensamiento evolucionista” (Dennett, 1995, p. 12), es previsible que el libro prescinda de “argumentos filosóficos altamente técnicos”, de interés más acotado. Ahora bien, como señala Zawidzki (2007, p. 154), “detrás del estilo exploratorio y casual de Dennett se encuentra un sistema filosófico profundamente coherente y ambicioso”.

En este trabajo me propongo mostrar de qué maneras la “peligrosa idea de Darwin” alcanza también al lenguaje y a sus propiedades distintivas. El fenómeno lingüístico, en todas sus dimensiones de análisis, suscita una variedad de desafíos científicos y filosóficos. Dennett (1995, p. 21) se interesa especialmente por aquellas características del lenguaje que reciben una nueva luz en virtud de la adopción de la perspectiva darwiniana. Como se verá, la imagen que resulta de la relación de los humanos con el lenguaje —cómo emergió, cómo es aprendido, cómo es usado, cuáles son sus efectos cognitivos y culturales— es de signo opuesto a la que proponen muchas concepciones filosóficas y lingüísticas contemporáneas, arraigadas en distintos presupuestos heredados. Me referiré a estas últimas de manera genérica como la visión o el enfoque tradicional, clásico o canónico sobre el lenguaje.

La importancia del tópico es crucial para el proyecto dennettiano de poner en evidencia y tornar persuasivo el objetivo *unificador* que le asigna a la teoría de Darwin (cfr. Dennett, 1995, p. 21).³ Si sus interpretaciones pueden integrarse satisfactoriamente a dicho objetivo, habrá logrado su cometido más desafiante. Por otro lado, buena parte de los supuestos y las teorías que integran el *mainstream* en la filosofía analítica del lenguaje

² Estas consideraciones también pueden aplicarse a *From Bacteria to Bach and Back* (Dennett, 2017a). Para refinar o complementar el análisis de las ideas expuestas en DDI, además de otros escritos, me valdré de este libro. En algún sentido, se trata “de la misma obra”, aunque corregida, ampliada y actualizada dos décadas después.

³ Dennett se pregunta en qué medida dicho proyecto unificador es o no reduccionista, diferenciando una variedad anodina y correcta (la suya) de otra disparatada o “insaciable” (*greedy*) (cfr. DDI, pp. 80-84). Esta cuestión es importante para comprender algunas polémicas que suscitaron sus propuestas en relación con las ciencias de la cultura, pero no cabe abordarla aquí.

(y en disciplinas conexas), y algunas corrientes de la lingüística teórica, deberían ser revisados o, incluso, abandonados. En efecto, la perspectiva elaborada en DDI “sobre el reino del significado y el propósito” es, en gran medida, ajena a la tradición filosófica a la que Dennett pertenece, en tanto enfoca el fenómeno lingüístico rebalsando el formato canónico altamente idealizado y carente de historia en el que ha sido estudiado en esa tradición.⁴ No es casual que el libro esté estructurado históricamente tanto en lo que respecta a la reconstrucción de las ideas que examina como a los fenómenos mismos estudiados: desde el origen de la vida hasta las creaciones más sofisticadas de la mente humana, y en el caso de Dennett (2017a), incluye algunas reflexiones sobre el futuro incierto en la era del “diseño post-inteligente”.

Me referiré, en primer lugar, a la estructura de las explicaciones darwinianas que dan fundamento al proyecto crítico y positivo de Dennett, sintetizada en la expresión “extraña inversión del razonamiento” (sección 1). Luego, intentaré reconstruir de qué manera la visión que ofrece sobre el lenguaje se ajusta al principio darwiniano/dennettiano de conformidad con aquella “inversión del razonamiento”, según el cual *la competencia precede a la comprensión*. Para ello me referiré: a la importancia del lenguaje para el proyecto dennettiano (sección 2), a la perspectiva memética que da cuenta de las relaciones del lenguaje con la cultura (secciones 3 y 3.1), a los orígenes evolutivos del lenguaje (sección 4), al aprendizaje del lenguaje (sección 5), a la relación del lenguaje con la cognición (o la “inteligencia”) (sección 6), al significado lingüístico (sección 7) y, por último, a la función comunicativa del lenguaje (sección 8). Algunas reflexiones finales recogerán las conclusiones más generales del trabajo.

⁴ Se atribuye a la corriente principal en la tradición analítica haber elaborado *teorías ideales* sobre el significado y la comunicación, esto es, sin considerar los procesos sociales en cuyo marco deberían ser explicados (cfr. Keiser, 2023). En mi opinión, un enfoque filosófico *no ideal del lenguaje* debería incorporar también una amplia perspectiva diacrónica y concebir las supuestas “imperfecciones” o “anomalías” de las lenguas históricas como rasgos de una herramienta heterogénea y multipropósito, no como obstáculos para el éxito de su función representacional. La crítica de Dennett (2024) a “la lingüística cartesiana” y su propósito de “euclidificar” la teoría del lenguaje es, entre otras, una expresión explícita de su crítica a los enfoques ideales. Pero no cabe ahondar aquí en estas interesantes cuestiones metafísicas.

Cabe aclarar que si bien cada una de esas cuestiones merecería un examen crítico *in extenso*, mi propósito es más general: mostrar cómo todas ellas se articulan en una visión global coherente y sustantiva. Podría pensarse, en cambio que, las ideas de Dennett sobre el lenguaje son más bien dispersas, tienen un escaso desarrollo o una significación secundaria. En segundo lugar, no todas las cuestiones serán analizadas con igual detenimiento, tanto en virtud de la dispar atención que reciben en la obra de Dennett como teniendo en cuenta el grado de novedad teórica que conllevan. Por mi parte, he intentado prestar más atención a sus propuestas positivas.

1. La “extraña inversión del razonamiento”

Según explica Dennett, la contribución revolucionaria de Darwin fue posible en virtud de una “extraña inversión del razonamiento”.⁵ Esta consistió en poner “patas para arriba” (*upside down*) el modo tradicional de concebir todas las entidades que exhiben diseño, tanto biológicas como artificiales, incluyendo a las criaturas inteligentes (cfr. Dennett, 1995, 2009a y 2017a). Como lo explica en DDI, la expresión “extraña inversión del razonamiento” pertenece a un crítico de Darwin que, en una publicación anónima de 1868, afirmó que “el propósito esencial de la Teoría” era, “en pocas palabras”, el siguiente: “por medio de una extraña inversión del razonamiento, [Darwin] parece pensar que la Ignorancia Absoluta está plenamente calificada para tomar el lugar de la Sabiduría Absoluta en todos los logros de la habilidad creativa” (citado en Dennett, 1995, p. 65). Dennett hace propio el diagnóstico que sintetiza esta expresión, pero invierte su valoración: en esa “inversión” está, justamente, el gran mérito de Darwin.

La idea de una “Pirámide Cósmica” —que tiene a Dios en la cúspide y al Caos, la materia inerte o la Nada en su base— concebida como un “mapa general de las cosas desde arriba hacia abajo” (Dennett, 1995, p. 64) ilustra, para Dennett, la forma de pensar que ha sido dominante en

⁵ El concepto es introducido en DDI, pero es explorado con detalle en textos posteriores, sobre todo en Dennet (2009a y 2017a). Dennett (2009a) atribuye la misma “extraña inversión” a Alan Turing por haber mostrado que puede haber computación sin un computador humano, es decir, que “era posible diseñar máquinas sin mente [...] pero que podían hacer aritmética perfectamente, siguiendo ‘instrucciones’ que podían ser implementadas mecánicamente” (2017a, pp. 55-56).

la filosofía (y en la tradición humanística) acerca del orden metafísico y explicativo correcto. Dennett ejemplifica esta forma de pensamiento con las ideas de algunos filósofos clásicos. Así, por ejemplo, en el *Ensayo sobre el entendimiento humano* (de 1690), Locke argumenta que “es imposible concebir que cualquier materia simple no cogitativa pudiera producir un Ser inteligente pensante” (IV, X, § 10; 2005, p. 26). La intuición de Locke es que la “comprensión” o el “entendimiento” es una capacidad original y autónoma que *derrama* sus efectos sobre todas las operaciones y los productos de las criaturas que la poseen. Esta perspectiva vincula a Locke, Descartes y Kant con Fodor, Searle y Grice, entre muchos otros filósofos contemporáneos.

La inversión o el “asalto” de Darwin sobre esa Pirámide Cósmica, en particular sobre la idea de que la Mente de Dios ha creado la mente humana y que esta es responsable de los actos creativos y las obras inteligentes, supone el derrumbe del “modo de pensar de la primacía de la Mente” (*the Mind’s first way of thinking*) (DDI, p. 70) o “visión del Cosmos de la primacía de la Mente” (*the Mind-first vision of the Cosmos*) (DDI, p. 204). Según Dennett, esta idea es una expresión de “chauvinismo filogenético” (1995, p. 376) convertido en filosofía. Su cuestionamiento es “profundamente contraintuitivo” (2009a, p. 10061), pero solo roza “un prejuicio mal examinado” (1995, p. 66). Se trata, precisamente, del prejuicio “*Mind-first*”, que deriva de “analizar el ‘producto terminado’ independientemente de su historia” (Dennett, 2018a, p. 145). Ahora bien, “¿[p]or qué las cosas más importantes de todas no podrían haber surgido de cosas no importantes?” (1995, p. 66) o, más precisamente, ¿por qué “fuerzas carentes de propósito y sin mente” (1995, p. 66) no podrían explicar la emergencia de criaturas con mente? DDI es un esfuerzo por persuadir, sobre todo a los filósofos, de que Darwin elaboró una respuesta afirmativa y correcta a esas preguntas.

El principio darwiniano que invierte el razonamiento puede resumirse en la siguiente fórmula: “para hacer una máquina perfecta y bella, no es requisito *conocer* cómo hacerla” o *conocer la teoría que da cuenta de cómo hacerla* (cfr. Dennett, 2009a, p. 10061; mi énfasis). Es decir, el origen de cualquier producto diseñado puede explicarse (total o parcialmente) mediante una ruta “laboriosa y no inteligente” (2009a, p. 10061) que explica la “acumulación de diseño” (1995, pp. 68 y ss.). Este

es, en síntesis, el principio de la *competencia sin comprensión*.⁶ Según este principio, cualquier tipo de criatura o subsistema orgánico y cualquier producto cultural, incluyendo el lenguaje, deben ser conceptualizados como “un efecto emergente de sistemas de competencia que carecen de comprensión” (Dennett, 2017a, p. 75), porque “la competencia viene primero” (2017a, p. 94). En el caso humano, como veremos luego, también la competencia precede a la comprensión, e incluso más: la comprensión no es “su *ingrediente activo*”, sino que “está *compuesta* de competencias” (2017a, p. 94). Generalizando el conocido canon de Morgan de la psicología comparada, Clatterbuck (2018) bautiza a este principio “el canon de Dennett”. Reza así: “deberíamos abstenernos de atribuir comprensión cuando la mera competencia bastaría para explicar la conducta” (Clatterbuck, 2018, p. 129).⁷ O, como afirma Rathkopf (2017, p. 1358) más breve y precisamente, Dennett recomienda “extender el canon de Morgan a la antropología”.

El concepto de “comprensión” (o “entendimiento”) refiere, de manera muy general, al modo humano de “apreciar los principios que explican por qué algo es como es” (Rathkopf, 2017, p. 1357), mientras que “competencia” refiere al conjunto de capacidades y habilidades de los organismos, y también de los microorganismos y los artefactos, que les permite satisfacer alguna función o lograr un cierto fin. Según la teoría darwiniana, la competencia es el fruto de procesos de selección, y ello explica que podamos describirla con provecho predictivo empleando el lenguaje intencional (Dennett, 1998).⁸ En el caso humano, no siempre involucra capacidades cognitivas sofisticadas y “razones” que las justifican, es decir, comprensión. En el caso de otras criaturas

⁶ Este principio se desarrolla en Dennett (2009b) y, sobre todo, en Dennett (2017a), pero está prefigurado en los capítulos 3, 12 y 15 de DDI.

⁷ El canon de Morgan de la psicología comparada recomienda: “En ningún caso podemos interpretar una acción como el resultado del ejercicio de una facultad psíquica superior, si podemos interpretarla como el resultado del ejercicio de una que está más abajo en la escala psicológica” (2007, p. 53).

⁸ Dennett (1998) distingue dos tipos de estrategias predictivas emparentadas: la *actitud de diseño*, aplicable a las funciones biológicas y los artefactos, y la *actitud intencional*, aplicable a los agentes humanos y a criaturas o entidades semejantes en cierto grado (algunos animales y artefactos inteligentes). Adoptamos ambas estrategias sin tomar compromisos acerca de su implementación, pero solo la segunda atribuye propósitos e intenciones bajo presunciones de racionalidad.

con distintos talentos (instintos, respuestas aprendidas mediante condicionamiento o ensayo y error, imitación o instrucción), no hay “mentes entendiendo las razones” que les atribuimos para entender su comportamiento. En estos casos operan “razones de flotación libre”: las razones por las cuales existen todos los productos naturales del diseño, surgidas de procesos selectivos ciegos y graduales. Son “de flotación libre” porque no son razones conscientes de algún agente inteligente, es decir, existen sin ser pensadas o representadas por alguna mente (cfr. Dennett, 2017a, pp. 50-52). Aunque pueden ser atribuidas a agentes competentes en tanto explican lo que hacen, no son razones “para ellos”. Los castores, los pájaros y las abejas, por ejemplo, tienen competencias sofisticadas para hacer diques, nidos y panales, respectivamente, pero no poseen un “saber que” que ponga a su alcance los porqués de sus comportamientos. Es decir, una cosa es la existencia de razones, otra es su representación por parte de un organismo (Rathkoff, 2017). “La apreciación de las razones es un producto más tardío, más avanzado que las razones” (Dennett, 2014b, p. 51). Estas, a su vez, emergieron de procesos causales: antes que los “para qué” (*what for*) vienen los “cómo llega a ser” (*how comes*) (Dennett, 2017a, p. 38).

Frente a muchos intentos de “contener la idea de Darwin dentro de una revolución ‘segura’, aceptable y meramente parcial”, es decir, confinada al dominio de la biología (Dennett, 1995, p. 63), Dennett argumenta que ella extiende sus efectos no solo hacia abajo (*all the way down*), sino también hacia arriba (*all the way up*). Sus efectos “hacia arriba”, como señalamos antes, impactan sobre nuestras mentes inteligentes y sus creaciones: ambas están sometidas al mismo “ácido universal”.⁹ Esta metáfora pretende ilustrar el carácter revolucionario del pensamiento darwiniano, en el sentido de que “corroe casi todos los conceptos tradicionales, y deja a su paso una visión revolucionaria del mundo, con la mayor parte de los más antiguos hitos todavía reconocibles, pero transformados de maneras fundamentales” (Dennett, 1995, p. 63). Otras metáforas aluden a los tipos opuestos de explicación del diseño: una, la tradicional, que apela a “ganchos celestes” (*skyhooks*); la otra, darwiniana, que recurre a “grúas” (*cranes*). La primera apunta a los mecanismos, poderes o agentes no terrenales postulados con el objeto de explicar el

⁹ La expresión “ácido universal” (el título del capítulo 3 de DDI) alude a una fantasía popular infantil acerca de un líquido que podría corroer o “devorar” todos los materiales y destruir el planeta.

diseño y así mantenerlo de algún modo “suspendido en el cielo”: una clase de “fuerza, poder o proceso *Mind-first*” (Dennett, 1995, p. 76). Las grúas, en cambio, realizan una tarea mecánica “elevando” cada pieza y, de ese modo, logran conservar y reproducir diseños previos.

De esta manera, la teoría darwiniana rebalsa el marco de una explicación acerca de los orígenes de la especiación biológica para convertirse en una “teoría general del diseño”, lo que se conoce como “darwinismo universal” (Dawkins, 1983). Ese único “Espacio de Diseño” está gobernado por los mismos principios fundamentales (cfr. Dennett, 1995, p. 135). Su principal mecanismo, la selección natural, es un proceso algorítmico “ciego” —carente de inteligencia— que opera sobre diferentes sustratos y produce cierta clase de resultados, si se dan las condiciones apropiadas (Dennett, 1995, pp. 54-55). Entonces, en virtud de aquella “inversión del razonamiento”, tanto el diseño natural como el cultural pueden explicarse sin recurrir a un “diseñador inteligente”, es decir, invocando procesos ciegos de variación, selección y copia. Así, el orden explicativo heredado, *i. e.* qué fenómenos deberían explicarse primero, cambia radicalmente: sobre la base de dichos procesos, ya sea a través de una secuencia evolutiva filogenética, del desarrollo, del aprendizaje o la innovación, será posible explicar cómo pueden emerger los restantes. Como dijimos, esta “inversión del razonamiento” se aplica también al lenguaje y a su relación con la mente humana.

Según la caracterización predarwiniana, en cambio, la comprensión en el caso humano —si se dan las condiciones óptimas— es un fenómeno ubicuo o generalizado, de modo que solo ella permite explicar la naturaleza de la cognición, la acción y la creación. Incluso más, se da por supuesta ante cualquier muestra de comportamiento o producto inteligente (Dennett, 2017b). Consiste en un tipo de *metacognición consciente* en el siguiente sentido: solo un agente que entiende lo que piensa y por qué, y entiende también otras alternativas y las razones que podrían avalarlas o desacreditarlas, posee verdadera o genuina competencia cognitiva, es decir, *competencia con comprensión*. Dicho agente está situado —para decirlo en el idioma sellarsiano— en el “espacio lógico de las razones”, aquel en el que “se es capaz de justificar lo que uno dice” (cfr. Sellars, 1997, § 36). Este “espacio lógico” está cerrado para las criaturas que no pueden dar o pedir razones.

Ahora bien, aunque existe en muchos casos una estrecha relación entre las capacidades representacionales conscientes y los comportamientos inteligentes y racionales, esa relación es “un hecho reciente de la historia

natural”, no un “principio metafísico” que vale para todos los casos (Dennett, 1995, p. 205). Así, desde una *perspectiva diacrónica*, la inversión consiste en tratar a la mente como un efecto antes que como una causa (cfr. Dennett, 1995, p. 63). Desde una *perspectiva sincrónica*, la genuina comprensión humana también debería concebirse como un producto, pero en este caso de “las actividades de todos los *bits* semimentales y no mentales” que la componen (1995, p. 206). En las secciones siguientes veremos cómo estas dos perspectivas se complementan para dar cuenta de distintas dimensiones del lenguaje.

La teoría darwiniana proporciona, además, las herramientas para identificar no sólo tipos y grados de competencia, sino también tipos y grados de comprensión: en tanto efectos de procesos seleccionistas, ambas capacidades se presentan de maneras y en grados diferentes, dependiendo de una variedad de factores. Así, no todos los comportamientos humanos que consideramos inteligentes —por ejemplo, anticipar, rectificar o implementar un curso de acción— requieren tener representaciones y metarepresentaciones conscientes acerca de esas operaciones y sus contenidos, sus presupuestos e implicaciones, y las razones que justifican las decisiones adoptadas (cfr. Dennett, 2017a, p. 95). Aun cuando la comprensión puede ser útil e incluso inevitable en algunos casos, también ocurre con frecuencia de manera opcional, parcial, superficial, provisional, confusa, autoengañoso o *a posteriori*. Así mismo, puede desempeñar un rol relevante para la economía total del organismo y de su sistema cognitivo, con distintas funciones —por ejemplo, planificar, racionalizar, manipular a otros, etc. (cfr. Dennett, 2017a, pp. 340 y ss.)—, no solo justificar las competencias específicas puestas en juego. Es decir, los humanos no siempre nos representamos las razones que están operando en nuestros modos de actuar, y cuando lo hacemos, estas no siempre explican o justifican nuestro comportamiento, o no lo hacen del todo adecuadamente. Ahora bien, “el hecho de que podamos *poseer* estas razones (inteligente o estúpidamente) nos hace susceptibles de hablar de ellas y hablar con los demás” (Dennett, 2017a, p. 219). Es decir, tenemos la posibilidad de someterlas a escrutinio público, revisarlas o convertirlas en “nuestras” razones, porque el uso del lenguaje nos proporciona las herramientas, tanto mentales como comunicativas, para hacerlo. Y ello introduce una diferencia crítica, como veremos en las secciones 6 y 8.

Ahora bien, la idea no es que la *comprensión* es siempre secundaria, inerte o incluso prescindible, sino, más bien, que deberíamos recurrir

a un modelo híbrido para dar cuenta de nuestra capacidad para crear y apreciar diseños inteligentes, ajustando en cada caso el peso relativo que corresponde asignar a diferentes capacidades: apelando, por una parte, a procesos cognitivos y creativos de alto nivel y, por otro lado, a condiciones de posibilidad basadas en capacidades, destrezas y disposiciones de “bajo nivel”, donde no interviene inteligencia alguna. Es decir, “hay una codependencia entre las capacidades de razonamiento de arriba abajo de nuestras mentes y los talentos de abajo arriba, sin comprensión, de nuestros cerebros animales” (Dennett, 2017a, p. 415). En síntesis, según la perspectiva darwiniana, por una parte, hay “un parentesco entre las más finas producciones del arte y la ciencia y las glorias de la biosfera” (Dennett, 1995, p. 511). Pero, por la otra, los procesos “profundamente darwinianos” que dieron origen a la cultura humana, basados en competencias sin comprensión, “*gradualmente se desdarwinizaron, aumentando la comprensión y la capacidad de organización de arriba-abajo*” (Dennett, 2017a, p. 148; énfasis de Dennett). Retomaré esta cuestión más adelante.

Años después de la publicación de DDI, Dennett reflexionaba así:

He llegado a reconocer en los últimos años que quizás la idea revolucionaria central que nos dio Darwin, su “extraña inversión del razonamiento” [...] es la idea de *competencia sin comprensión*. Tendemos a pensar —especialmente si somos filósofos— que la competencia debe *fluir de* la comprensión, que primero necesitamos comprender *para poder* ser competentes [...]. No, realmente es al revés: nuestra comprensión es el producto de cascadas de competencias para semicomprender, pseudocomprender, no comprender con las que estamos dotados, primero por la selección natural, luego por el aprendizaje y el rediseño cultural, especialmente lingüístico (2013, pp. xi-xii).

Consideremos, entonces, cómo la “inversión del razonamiento” permitiría explicar distintas dimensiones del lenguaje humano.

2. La importancia del lenguaje para un enfoque darwinista

La formación de diferentes lenguajes y de especies distintas y las pruebas de que ambas se han desarrollado siguiendo un proceso gradual, son curiosamente las mismas.

Darwin (2009b, p. 59).

Como afirma Darwin, vistos desde la perspectiva de sus orígenes, los lenguajes son suficientemente parecidos a las especies: emergen y se desarrollan gradualmente por medio de mecanismos similares. Por otro lado, así como la vida, el lenguaje es un fenómeno “suficientemente complicado” como para pretender estudiarlo con provecho buscando identificar condiciones necesarias y suficientes (cfr. Dennett, 1995, pp. 201-202).

Ahora bien, según el enfoque tradicional, para explicar los fenómenos lingüísticos debemos tomar como modelos a las mentes evolucionadas adultas, conscientes de los contenidos de sus pensamientos, produciendo actos con significados intrínsecos. Dicho enfoque no es solo sincrónico sino también homogéneo en un sentido específico: está basado en “nuestra tendencia natural a interpretar *todo* diseño como de arriba-abajo, es decir, en tanto conducido por la representación” (Dennett, 2014b, p. 58). Pero es también homogéneo por privilegiar ciertos rasgos del lenguaje, considerados nucleares, y ciertas funciones, consideradas constitutivas, dejando fuera del foco de atención las que se conciben como periféricas o secundarias. Me referiré a otros rasgos de este enfoque en las secciones que siguen.

Por su parte, siguiendo a Darwin, Dennett está interesado en otros aspectos, tales como la “formación de diferentes lenguajes”, reconstruir cómo adquieren significado los signos lingüísticos y, más ampliamente, cómo distintas entidades adquieren la propiedad de ser intencionales, cómo los niños llegan a aprender su lengua materna, de qué maneras las palabras nos permiten pensar pensamientos cada vez más complejos y contribuyen a transmitir y acumular información cultural; en suma, cómo usamos el lenguaje para diferentes fines sociales e individuales. Para dar cuenta de estos diferentes fenómenos, propone Dennett, se requiere una “inversión del razonamiento”, adoptando los dos

caminos complementarios ya mencionados: el diacrónico gradualista y el sincrónico, pero ahora de “abajo arriba”. La combinación de ambos permitirá explicar mejor que los enfoques clásicos no solo las cuestiones de las que estos se ocuparon, sino también identificar y explicar una variedad más amplia de dimensiones del lenguaje que aquellos desatendieron. Por ello, tanto en DDI como en su obra previa y posterior, el lenguaje ocupa un lugar muy relevante. Las principales referencias al tema están expuestas en *Darwin's Dangerous Idea* y en *From Bacteria to Bach and Back*. Otros escritos previos y posteriores, no recogidos en ellos, también serán consultados.

3. Los memes y la evolución de la cultura

Tampoco [...] la facultad del lenguaje articulado es una objeción irrefutable a la creencia de que el hombre se haya desarrollado procediendo de una forma inferior.

Darwin (2009b, p. 62).

La teoría de la evolución expuesta en *El origen de las especies* (de 1859) se extiende también al *Homo Sapiens*, tal como Darwin lo hizo explícito en *El origen del hombre* (de 1871). No es ocioso destacar el alcance de su teoría: desde entonces, algunos debates (y vestigios culturales) muestran la supervivencia de una visión antidarwiniana sobre nosotros basada en una u otra versión de la tesis de que somos criaturas especiales. En efecto, dicha tesis asumió con frecuencia el supuesto de que la teoría de la evolución no es aplicable a nosotros, o al menos no a nuestras características distintivas. Según este razonamiento, nuestra condición única solo podría ser explicada apelando a factores *sui generis*, porque, de lo contrario, “debe crecer a partir de algo que es *menos*, algo *cuasi*, algo simplemente *como si* más que algo intrínseco” (Dennett, 1995, p. 341). Como advierte Dennett (1995, p. 345), este tipo de explicaciones ha sido predominante en las ciencias de la cultura de todas las épocas. Por su parte, los grandes logros de la cultura humana parecieran aportar la evidencia más contundente de que somos especiales. Ahora bien, para que la cultura se desarrolle y acumule se requiere del lenguaje. Nuevamente, o bien el lenguaje carece de una historia evolutiva, o tuvo que evolucionar a partir de precursores cuasilingüísticos. A su vez, antes de contar con el lenguaje, no podemos presuponer ni inteligencia ni cooperación: “todo esto tuvo que construirse afanosamente” (Dennett,

1995, p. 341). ¿Cómo explicar, entonces, la emergencia y la probable coevolución de estas características que nos hacen especiales?

Sobre los orígenes evolutivos del lenguaje, Dennett ofrece poco más que una reconstrucción general de las hipótesis y los debates sobre el tema, de la que me ocuparé, brevemente, en la siguiente sección. Antes presenta una detallada propuesta sobre la evolución cultural, y en particular sobre la evolución cultural del lenguaje que es, en gran medida, su propia explicación darwiniana acerca de nuestra singularidad. Ambos fenómenos ilustran la “extraña inversión del razonamiento”.

Apoyado en la propuesta de Dawkins de los memes como replicadores culturales, desde su formulación original en *The Selfish Gene* (1976) hasta sus versiones menos entusiastas en *The Extended Phenotype* (1982),¹⁰ Dennett intenta explicar tanto la evolución cultural en general como la emergencia y la propagación del lenguaje en particular. Según Dawkins, un meme es “una unidad de transmisión cultural, o una unidad de *imitación*” (1976, p. 206).¹¹ Lo que se imita o copia es una *unidad memorable distinta* de información o *unidad de información semántica* (cfr. Dennett, 2017a, p. 206), transmitida por medio de diferentes vehículos, incluyendo las palabras (cfr. Dennett, 1995, p. 344). El concepto de “información semántica” refiere a todo aquello que una criatura puede percibir en su entorno que puede hacer una diferencia para ella en términos adaptativos, es decir, “diseño que vale la pena obtener” (Dennett, 2017a, p. 115). Esta noción difiere de la más abstracta de “contenido semántico” de los enfoques clásicos, porque es relativa a la perspectiva, al entorno y a las acciones de una criatura.¹²

¹⁰ Dennett conjetura que la incorrecta identificación, por parte de las “mentes humanistas”, de la memética con la sociobiología humana —simplificadamente, la reducción de los fenómenos socioculturales a fenómenos biológicos, *v. gr.* a predisposiciones genéticas (cf. Dennett, 1995, pp. 481 y ss.)— es responsable de la versión más escéptica de Dawkins acerca de su contribución a las ciencias de la cultura (cfr. Dennett, 1995, pp. 361 y ss.). Es probable que buena parte de la resistencia perdurable a la memética obedezca a una interpretación de ese tipo.

¹¹ El término elegido por Dawkins, suficientemente parecido a “gen” (*gene* en inglés), deriva del griego *mimeme*, que significa “cosa imitada”.

¹² Este concepto de “información semántica” está emparentado con el de *affordance*, que es la posibilidad u oportunidad para la acción suscitada por los objetos del entorno de una criatura, incluidos los artefactos culturales en el caso humano, que son percibidos por ella como directamente significativos.

Otro rasgo definicional de los memes es que son “los elementos más pequeños que se replican a sí mismos con fiabilidad y fecundidad” (Dennett, 1995, p. 344). Un replicador es una unidad de información que puede reproducirse empleando los recursos de algún sustrato material. Ahora bien, “la evolución de los memes no pudo haber comenzado hasta que la evolución de los animales hubiera recorrido el camino que llevó a crear una especie —el *Homo Sapiens*— con cerebros capaces de proporcionar refugio y hábitos de comunicación que facilitasen a los memes medios de transmisión” (Dennett, 1995, p. 345). Ahora bien, aunque los hábitos de comunicación pudieron comenzar a desarrollarse mediante el uso de medios no verbales, fue la emergencia del lenguaje la que hizo posible la acumulación cultural (cfr. Dennett, 1995, p. 347). Es decir, las unidades mínimas de transmisión cultural solo habrían sido posibles por la disponibilidad de mecanismos y herramientas comunicativas, primero preverbales y luego verbales. Dennett introduce esta hipótesis sobre los orígenes comunicativos del lenguaje recién en (2017a). Me referiré a ella en la sección 8.

En tanto entidades semánticas, los memes son tan invisibles como los genes, aunque necesitan “encarnarse” en algún vehículo físico para poder replicarse. Ahora bien, diferentes vehículos y formatos pueden transmitir un mismo meme, por ejemplo, una misma obra dramática (es decir, una misma historia puede ser tanto escrita como filmada). Por otra parte, los memes pueden ser individuados en diferentes niveles de abstracción respecto a la materialidad de sus soportes. De ese modo, su propagación eficiente podría depender de criterios de identidad de grano más o menos fino. Veremos ejemplos de este tipo en algunos memes lingüísticos, *v. gr.*, vehiculizados por palabras. Estos dos vínculos tan diferentes de los memes con sus vehículos evidencian una tipología muy variada.¹³ Ahora bien, aunque el criterio de identidad

Una *affordance* es “un objeto ecológico rico en valor” (Gibson 1979/2015, p. 132). Dennett aplica esta noción a todos los memes, incluidas las palabras.

¹³ Los memes también pueden tener diferentes tamaños y complejidad, *v. gr.* los mitos, las instituciones, los credos, etc. (cfr. Dennett, 2017a, p. 188), lo que hace recomendable distinguir entre memes simples y complejos. Pero no es claro cómo descomponer un meme complejo en sus elementos simples ni cuál podría ser el papel explicativo de estos. Dennett (2017, pp. 221-247) intenta responder a esta y a otras objeciones a su propuesta concediendo muchas críticas e intentando mostrar el grado de compatibilidad de los memes con las herramientas teóricas

semántico puede suscitar dificultades frente a casos tan dispares, sería imprescindible —aunque probablemente insuficiente— para dar cuenta de la transmisión y transformación de la información para las “ciencias de la cultura”. Como se ha dicho, los memes son “suficientemente discretos” (Blackmore, 1999). Ello les permitiría desempeñar papeles causales relevantes en la explicación de los fenómenos culturales, identificando primero cuál información está en juego para formular, luego, predicciones basadas en la detección de patrones comunes.

Dennett sigue a Dawkins al asemejar los memes a los genes. Además de ser invisibles, ambos transportan información —cultural y biológica, respectivamente— y pueden concebirse como unidades de selección, es decir, como las entidades cuyas copias proliferan, más o menos. En consecuencia, principios explicativos comunes se podrían aplicar a ambos, o sea, obedecerían a los mismos procesos de selección natural (cfr. Dennett, 1995, p. 345; 2017a, cap. 11). Pero hay también diferencias entre ellos. La primera es que los memes son transmitidos con menor fidelidad que los genes. En efecto, los cerebros que son sus “hogares temporarios” no solo los replican de manera fidedigna, sino que también los censuran, transforman y alteran de diferentes maneras. Las condiciones ambientales también influyen en el grado y manera en que se replican. En síntesis, los memes mismos tienen criterios de identidad más o menos inestables e indeterminados: aunque algunos pueden ser imitados fielmente, otros solo son reproducidos preservando un suficiente parecido, es decir, son “solo un poco meméticos [*memish*], en un grado u otro [...] sea como fuere que lleguen a la escena” (Dennett, 2014a). Finalmente, a diferencia de los genes, los memes son criaturas de la imagen manifiesta (Sellars, 1992), es decir, del mundo tal como es “para nosotros”, no de la imagen científica (Dennett, 2017a, p. 287).¹⁴ Dado que son objetos intencionales, solo son discernibles desde la actitud intencional. Aún más: conforman una red que da forma a las características de las criaturas culturales que somos, es decir, se presentan a través de nuestras diferentes expresiones fenotípicas. Incluso “nuestros yoes han sido creados a partir de la

de las ciencias de la cultura, así como refinando su evaluación de los beneficios que la perspectiva memética podría agregar a ellas.

¹⁴ El asunto es más complejo porque las palabras se aprenden y emplean muchas veces “de forma inadvertida e inesperada”, por lo que solo serían parte de la imagen manifiesta cuando “nos percatamos y somos capaces de reflexionar sobre ellas” (Dennett, 2017a, p. 267).

interacción de memes” (Dennett, 1995, p. 367). Por lo tanto, adoptar la perspectiva memética no conlleva ninguna reducción biologicista ni de la cultura ni de la evolución cultural.

La idea original de Dawkins incluye otra característica crítica: un rasgo cultural podría evolucionar “simplemente porque es *ventajoso para sí mismo*” (Dawkins, 1976, p. 214). Este es el concepto de “meme egoísta” (análogo al de “gen egoísta”). Esta caracterización busca cuestionar el supuesto de una conexión necesaria entre el poder de replicarse o el *fitness* de un meme, visto desde el “punto de vista del meme” y su contribución a “nuestro *fitness*”, desde nuestra propia perspectiva, según algún estándar (cfr. Dennett, 1995, p. 363). Es evidente que el concepto mismo de un “punto de vista del meme desafía uno de los axiomas centrales de las humanidades” (1995, p. 362), porque colisiona con lo que podríamos llamar “el punto de vista egocéntrico”, según el cual, dadas ciertas condiciones, *nosotros* elegimos (apreciamos, comprendemos, etc.) “lo verdadero, lo bello y lo bueno”, encarnados en ciertos productos culturales, o al menos, lo hacemos en la medida que nuestras capacidades hayan podido desarrollarse apropiadamente. El punto de vista según el cual *nosotros* estamos a cargo de nuestras decisiones y somos los beneficiarios de ellas, no nuestros memes, pareciera el más natural (cfr. Dennett, 1995, pp. 327-328).

Sin embargo, argumenta Dennett, el *desacoplamiento* entre estos “dos puntos de vista” podría explicar de una manera sencilla fenómenos que son ubicuos en la cultura humana. En primer lugar, podría explicar por qué muchos memes proliferan a través de nosotros a pesar de no contribuir a nuestro ajuste con el entorno, “manipulando” incluso nuestra apreciación de su supuesto beneficio para nosotros. Es decir, podría explicar la permanencia y proliferación de memes perniciosos o que carecen de ciertas virtudes apreciadas por sus portadores, sin que debamos recurrir a explicaciones en términos de factores excepcionales o desviaciones de las normas (p. ej., casos de irracionalidad, inmoralidad o similares) bajo presupuestos racionales idealizados (cfr. Dennett, 1995, p. 363). De modo similar, permite explicar por qué algunos memes pueden extinguirse a pesar de ser beneficiosos para sus huéspedes, o por qué pueden ser, al mismo tiempo, benignos para sus huéspedes (p. ej., “orgullo étnico”) pero perjudiciales para otros (“xenofobia”)

(cfr. Dennett, 1995, p. 364, n. 11).¹⁵ A todos nos sucede, incluidas las mentes más lúcidas, que ciertos *memes* “persisten [...] sin ser dirigidos ni apreciados, o que se extienden —como los rumores— a pesar de la general desaprobación de esta expansión por parte de aquellos que contribuyen a propagarlos” (Dennett, 1995, p. 347). ¡Algunos se robustecen a pesar de su comprobada falsedad! También permitiría explicar por qué no “adoptamos” las buenas ideas cuyas credenciales tenemos las capacidades para apreciar o, al revés, por qué cuando lo hacemos, ello puede suceder sin nuestra contribución activa.

Por su parte, dado que “el punto de vista del meme” incorpora las creaciones “no apreciadas por nadie”, no restringe la atención de los estudios culturales a las manifestaciones de la “alta” cultura (cfr. Dennett, 2017a, p. 215). Finalmente, puede dar cuenta de cómo se puede preservar y transmitir la información que portan sus vehículos, aun cuando esta no sea comprendida en absoluto, *v. gr.* el caso de “la instalación de un lenguaje nativo en un infante humano” (2017a, p. 194), caso al que me referiré más adelante. En resumen, “el punto de vista del meme” permite sortear la “perspectiva psicológica tradicional”: el requerimiento de que detrás de cada innovación y transmisión cultural hay ideas o creencias *entendidas* y “*uptake* consciente” por parte de “artífices individuales” (2017a, pp. 212-213). Por lo tanto, “es valiosa para *despsicologizar* la propagación de las innovaciones (buenas y malas)” (2017a, p. 239). De ese modo, no solo evita sobrestimar el rol de la comprensión en la cultura (cfr. Dennett, 2009b, p. 435), sino que contribuye a rectificar el relato de la historia cultural como una acumulación de proezas y éxitos que son el fruto de la labor de inventores lúcidos o creadores geniales en un camino sin tropiezos ni fracasos onerosos (cfr. Dennett, 2017a, p. 309). Aunque es muy discutible que este relato tenga actualmente el mismo asidero en las ciencias de la cultura que el que tuvo en épocas de mayor confianza en la racionalidad humana, sigue operando como un supuesto implícito que impregna el sentido común.

Dennett (2017a, pp. 210-211) sintetiza los rasgos distintivos del concepto de “memes”: (1) son *unidades informacionales* relativamente independientes de sus vehículos, es decir, identificables por su función —

¹⁵ Este contraste entre los memes que hospedamos y nuestras mentes es, en otro sentido, artificioso, porque “la mente humana es ella misma un artefacto creado cuando los memes reestructuran un cerebro humano para hacerlo un mejor hábitat para memes” (Dennett, 1995, p. 365).

almacenar y transmitir información—; (2) que *pueden ser no comprendidas*, es decir, pueden ser receptadas (aprendidas, transmitidas) sin que los huéspedes humanos seamos capaces de representarnos su función, y (3) *son egoístas*, es decir, promueven su propio éxito reproductivo, no el de sus huéspedes, aunque también pueden ser beneficiosos para ellos. Como veremos enseguida, las palabras satisfacen estas tres características.

Como se dijo antes, la perspectiva memética es compatible con el papel del diseño inteligente, sobre todo en la historia cultural reciente. En este sentido, ofrecería un modelo que podría dar cuenta de la evolución cultural pasada en mucha mayor medida que de la reciente (retomaré esta cuestión al final). En todo caso, sería apropiado afirmar que los frutos de la cultura humana son el efecto de “una mezcla de respeto a la tradición, improvisación imprudente y oportunista y una I+D deliberada, intencional y sistemática, interrumpida de vez en cuando por episodios de ‘inspirada’ genialidad” (2017a, p. 371), incluyendo la “comprensión colaborativa” o grupal. La sugerencia que se desprende de esta síntesis es que la evolución cultural debería ser explicada por la intervención de mecanismos y capacidades de tipos muy diferentes (cfr. Sterelny, 2006 y Mesoudi, 2021 para propuestas de ese tipo).

Dennett (2014a; 2017a, pp. 221-247; 2023, cap. 32) hizo varios esfuerzos infructuosos por acercar las diferentes propuestas sobre la evolución cultural, minimizando los desacuerdos entre ellas y explorando su posible complementariedad. En cualquier caso, mi interés está limitado a reconstruir los argumentos con los que Dennett defiende que el examen de la cultura en unidades meméticas agrega una *perspectiva* útil a la explicación de la evolución cultural de conformidad con la “inversión del razonamiento” darwiniana, dando fundamentos a una explicación unificada, aunque no reductiva, del mundo natural y cultural.

Al respecto, Dennett aclaró que la noción de “meme” es un constructo teórico útil, pero no para fundamentar una disciplina científica, la memética, o una teoría científica, la teoría memética de la cultura, sino porque aporta una *perspectiva* fructífera al estudio de los ítems culturales (cfr. Dennett, 1995, pp. 345-346 y 368; 2017a, pp. 210, 216-217 y 237 y ss.). Es decir, los mismos productos que antes eran *solo* descriptos como el fruto de la intervención de procesos deliberados pueden ser caracterizados también como efectos de mecanismos de creación, imitación y transformación relativamente “ciegos”, susceptibles de ser evaluados, desechados o propagados de maneras

demasiado graduales para ser percibidas y que, además, no siempre son beneficiosos para las comunidades que los conservan (Dennett, 2017a, p. 240). Simplificadamente, la perspectiva memética “llena la gran e incómoda brecha entre los instintos genéticamente transmitidos y las invenciones comprendidas, entre animales competentes y diseñadores inteligentes” (Dennett, 2017a, p. 241). Consideremos ahora los memes lingüísticos.

3.1. Las palabras como memes

La supervivencia o preservación de determinadas palabras favorecidas en la lucha por la existencia es selección natural.

Darwin (2009b, pp. 60-61).

Según Dennett (1995, p. 347; 2009b, p. 437; 2017a, p. 206), las palabras “son los mejores ejemplos de memes” o los memes paradigmáticos.¹⁶ Según mi reconstrucción, esta afirmación se fundamenta en los siguientes hechos: (a) son criaturas “opcionales”, no innatas, *i. e.* tienen un origen cultural, e incluso más: son “el medio dominante de innovación y transmisión cultural” (cfr. Dennett, 2017a, p. 208); (b) tienen una “fuerte individualidad”: en tanto entidades discretas y más simples que otros memes, pueden transmitirse con mucha fidelidad, es decir, pertenecen a “la clase que puede ser pronunciada” (2017a, p. 206), una característica que las hace “memorables”,¹⁷ y (c), se transmiten (copian, recuerdan, enseñan, denuncian, etc.) a través de las comunidades lingüísticas. La conjetura es que el lenguaje habría evolucionado hacia formas codificadas de vehiculizar la información semántica porque de ese modo se facilitó su preservación de manera fidedigna, pudiendo evolucionar por replicación diferencial, es decir, por selección natural, como sostuvo

¹⁶ Compuestos de palabras, como las frases hechas y las expresiones idiomáticas, también son unidades informacionales que se replican. Procesos no dirigidos “toman combinaciones replicadas con frecuencia y las solidifican gradualmente en unidades que pueden replicarse por sí solas como unidades combinatorias” (Dennett, 2017a, p. 276). Planer (2024), siguiendo a Dennett, enfatiza el mismo punto: muchas construcciones y estructuras sintácticas que combinan rasgos fonológicos y semánticos también son memes porque se replican con facilidad y fidelidad.

¹⁷ Incluso “son autónomas en ciertos aspectos” (Dennett, 2017a, p. 189), ya que pueden migrar a otras lenguas.

el propio Darwin en los textos de *El origen del hombre* citados *supra*. Sus usos comunicativos desempeñaron un papel crítico en dar su perfil a las palabras.

Ahora bien, en tanto criaturas concretas, las palabras y antes las protopalabras están compuestas por secuencias de fonemas, es decir, tienen un carácter digital. Ello permite que puedan ser percibidas y recordadas mejor. Esa propiedad de sus vehículos es una solución de diseño que favoreció su reproducción fidedigna (más sobre este rasgo en la sección siguiente). Así, cada caso o *token*, cuando es replicado por otros ejemplares, genera un meme, es decir, genera su descendencia. El conjunto de esa descendencia forma una especie: esta es la palabra como tipo o *type* (cfr. Dennett, 2017a, p. 189). Así, en tanto memes, las palabras no son sus *tokens* o instanciaciones concretas, en una u otra lengua, o en una u otra modalidad, hablada o escrita. Tampoco tendrían que concebirse como entidades abstractas, como se sostiene desde los enfoques formales (cfr. Dennett, 2017a, p. 188). Por el contrario, antes que nada, las palabras son “maneras de hacer algo [...], una clase de *modo de comportamiento* que puede ser copiado, transmitido, recordado, enseñado, evitado, denunciado, lucido, ridiculizado, parodiado, censurado, consagrado” (2017a, p. 206). Más aún, una palabra es “una clase mínima de *agente: quiere lograr ser dicha* [...] porque si no, se extinguirá” (2017a, p. 189). Estas caracterizaciones metafóricas intentan facilitar la comprensión de la dinámica propia de las palabras como memes.

Como señalamos acerca de los memes en general, también en el caso de las palabras los mismos contenidos semánticos pueden ser transportados por diferentes vehículos fonológicos y sintácticos, por ejemplo, un mismo texto puede ser traducido a diferentes lenguas. Y al revés: la identidad de algunos *memes* depende de maneras críticas de sus “rasgos de bajo nivel”, por ejemplo, de sus propiedades fonéticas, ortográficas, rítmicas y otras (como con los refranes, los poemas, los *jingles*, las marcas, las “frases hechas”). En tales casos, el vínculo entre forma y contenido los hace más fácilmente memorables, lo que permite que se repliquen más fidedignamente. Se podría argumentar que solo este tipo de expresiones deberían conceptualizarse como memes y que sus diferencias con otro tipo de palabras hacen difícil subsumir bajo un mismo concepto a criaturas informacionales tan diferentes. Sin embargo, si se aprecia la utilidad funcional del concepto general de “meme” como

“unidad de información semántica”, es razonable aceptar que esta pueda encarnarse en vehículos y formatos heterogéneos.

Dennett sigue la sugerencia de Darwin de que “ningún lenguaje ha sido deliberadamente inventado; cada uno ha sido lenta e inconscientemente desarrollado a través de muchos pasos” (Darwin, 2009b, p. 55). Según la perspectiva memética, las palabras carecen de un origen inteligente. De hecho, casi todas “tienen muchos padres [...] exceptuando, por supuesto, las invenciones y mejoras de las nomenclaturas científicas y aquellos esquemas de lenguaje filosófico que se han propuesto” (Darwin, 2009b, p. 379). Por cierto, el lenguaje debió ser inventado, pero no de manera intencionada o voluntaria, sino a través de procesos de “selección inconsciente” (Darwin, 2009a, p. 241), un tópico sobre el cual se generó incluso cierta polémica, contemporánea a la publicación de *La descendencia del hombre*. Dennett propone que la selección de las palabras es atribuible a diferentes mecanismos que llevarían a distinguir entre palabras *sinantrópicas*, criaturas adaptadas más que creadas artificialmente, y palabras *domesticadas*, que son el fruto de la intervención activa de los humanos en su diseño y propagación (cfr. Dennett, 2009b y 2017a). Finalmente, de manera similar a los virus, deberíamos distinguir entre palabras *parásitas*, *comensales* y *mutualistas* (cfr. Dennett, 2009b y 2017a, pp. 192 y ss.). Mientras que las primeras son perjudiciales para sus huéspedes, las *comensales* son neutras y las *mutualistas* los benefician, es decir, mejoran su aptitud sociocultural (cfr. Dennett, 2017a, p. 193, n. 51). Estas diferentes relaciones de las palabras con sus usuarios tendrían que contribuir a explicar diferentes tipos de usos en el discurso.

Las propiedades de los memes lingüísticos difieren de las atribuidas a las palabras y los significados en los enfoques canónicos: tanto de sus propiedades formales en cuanto unidades léxicas, como de sus funciones semánticas en cuanto nombres y predicados, en la medida en que son concebidos como los componentes mínimos de las estructuras oracionales y de las proposiciones que estas expresan. En cambio, las palabras y los compuestos de ellas conceptualizadas como memes poseen estructuras, funciones y grados de complejidad muy dispares entre sí. Tampoco se compadecen con la visión canónica sobre los hablantes, según la cual, por una parte, estos comprenderían plenamente los contenidos semánticos de las palabras que forman parte de sus repertorios léxicos porque, como dijimos, para aprender y replicar un meme no se requiere comprensión semántica plena; y, por otra parte, porque los hablantes

no siempre “deciden” cuáles palabras usar más o menos (o no usar): algunos memes se instalan y reproducen más eficazmente que otros sin nuestra colaboración consciente. Estos dos rasgos ponen en evidencia que “la comprensión humana no es ni necesaria ni suficiente para la fijación de un meme en la cultura” (Dennett, 2017a, p. 211). Volveré sobre esta cuestión en la sección 7.

La caracterización de las palabras como memes introduce, pues, una perspectiva diferente. Como veremos enseguida, nos permite comprender su trayectoria evolutivo-cultural como unidades con “historias de descendencia con modificación” (2017a, p. 207), así como la trayectoria de su aprendizaje, ambas dimensiones diacrónicas virtualmente ausentes en los enfoques filosóficos formales. Me referiré ahora, brevemente, a ambas cuestiones.

4. El origen del lenguaje

En “Chomsky contra Darwin: cuatro episodios”, Dennett (1995) reconstruye la génesis de la hipótesis chomskiana de una competencia lingüística innata: una estructura de reglas innatamente especificadas. La controversia con Chomsky se había iniciado en *Behavioral and Brain Sciences* (Dennett, 1980). Allí Dennett había propuesto que, si la hipótesis chomskiana fuera correcta, debía buscarse en la biología evolutiva una explicación de su carácter innato, dado que “lo que es innato debe haber sido ‘aprendido’ en la historia evolutiva” (1980, p. 19). Sin embargo, “el órgano del lenguaje, pensaba Chomsky, *no* era una adaptación sino [...] un misterio” (Dennett, 1995, p. 389) o un “don inexplicable” (p. 390). Dennett (2017a, pp. 276 y ss.) retoma el debate sobre la postulación de un Dispositivo de Adquisición del Lenguaje innato por parte de Chomsky y vuelve a referirse a “su terca resistencia” (Dennett, 2017a, p. 277) a explicar dicha competencia mediante la selección natural, convirtiéndola casi en un “gancho celeste” —una “mutación azarosa” o una “coincidencia cósmica”— antes que en una “grúa”.

La versión “minimalista” de esa hipótesis, defendida en Hauser *et al.* (2002), que limita la capacidad innata a una operación lógica de recursión o “ensamble” (*merge*) —un mecanismo para anidar expresiones en otras expresiones—, según Dennett, en el mejor de los casos, no sería más que una “elegante idealización matemática” de una primera adaptación para la gramática, seguida de muchas otras en un proceso evolutivo gradual (cfr. 2017a, p. 281). Sugiere, incluso, que la gramática podría ser el efecto de un proceso de “abajo arriba” intermedio, ni heredado genéticamente

ni conseguido mediante un tipo de aprendizaje profundo e inconsciente de búsqueda de patrones (cfr. 2017a, pp. 275-276). Desde *Darwin's Dangerous Idea* hasta *From Bacteria to Bach and Back*, el interés de Dennett por el enfoque chomskiano disminuyó notablemente en proporción al suscitado por la proliferación de hipótesis sobre los orígenes evolutivos del lenguaje en las últimas décadas.

El apartado más extenso de este capítulo, “El problema del huevo y la gallina”, está dedicado al examen de “una lista tentativa de las funciones a las que han servido eventualmente todos los lenguajes” (2017a, p. 250). Sin embargo, solo la primera es claramente una función: su “utilidad comunicativa”. Abarca la gran variedad de acciones que podemos realizar mediante el uso de las señales verbales: ordenar, preguntar, pedir, informar, entretener, fingir, etc. Los restantes son los rasgos distintivos de las señales, en la modalidad vocal que fue la predominante, sin los cuales dicha función, incluyendo la *transmisión fiel de la información*, no habría podido ser satisfecha. Ellos son la *productividad* o el poder de producir infinitos significados diferentes con un repertorio limitado de ítems léxicos; la *digitalidad* o la capacidad para corregir el ruido que pudiera acompañar a la producción o a la comprensión de la señal, una propiedad crítica que permite preservar la información que transportan los fonemas que componen la señal; y la *referencia desplazada* o la capacidad para referir a lo que no está presente, no existe o ya ocurrió. Por último, dado que el lenguaje debe poder ser *fácilmente adquirido*, ello habría favorecido la modalidad vocal frente a otras modalidades. En virtud de las propiedades antes mencionadas, el lenguaje vocal habría podido satisfacer la función de ser *comunicativamente útil*.¹⁸

Su conjetura más fundamental es que el lenguaje se habría apoyado en “un instinto para cooperar” con coespecíficos (2017a, p. 251), una capacidad que está presente en otras criaturas sociales, en combinación con “un hábito instintivo para la imitación” (p. 251) o “imitación primitiva” (Dennett, 2018a, p. 149). Ambas capacidades habrían creado una presión selectiva para cooperar cada vez mejor. Aquí la hipótesis memética podría aportar una solución al “problema del huevo y la gallina”: ¿qué fue primero, las capacidades cognitivas o el lenguaje?

¹⁸ Dos aspectos muy relevantes diferencian su perspectiva de otras más conocidas acerca de los rasgos distintivos de todo lenguaje: la importancia que asigna a la digitalidad como propiedad de los vehículos lingüísticos, por un lado, y la primacía de la función comunicativa, por el otro.

Su conjetura es que unos memes más toscos, un protolenguaje de expresiones breves quizás primero gestual y luego vocal, precedió a las primeras palabras. Estas habrían colonizado gradualmente nuestros cerebros antes de que fuéramos capaces de comprender su utilidad, lo que supone que algunos talentos y disposiciones innatas deberían estar presentes para que puedan emerger. Luego, los sonidos, la morfología, los contenidos semánticos y la gramática habrían atravesado procesos selectivos graduales que fueron optimizando las herramientas con las que intercambiar información para distintos fines (cfr. Dennett, 2017a, p. 254; más en la sección 8).

En este capítulo, Dennett pasa revista a diferentes hipótesis que intentan explicar las motivaciones y los mecanismos que habrían permitido desarrollar las propiedades críticas antes mencionadas bajo la guía de la misma conjetura darwiniana: la intervención de procesos graduales de selección inconscientes o sin una dirección inteligente. Sin embargo, pareciera que las diferentes hipótesis examinadas acerca de los orígenes evolutivos del lenguaje fueran evaluadas por su mayor o menor contribución a fortalecer su propia perspectiva memética, que tiene un alcance más general.

5. El aprendizaje del lenguaje

[...] un niño adquiriendo el lenguaje es como un pájaro construyendo un nido; es una cuestión “instintiva”.

Dennett (1997, p. 223, nota).

Dennett conjetura sobre el origen de las palabras y sobre las capacidades cognitivas involucradas en su uso apoyándose en la posible trayectoria que seguiría el aprendizaje gradual de la lengua materna por parte de los niños. En esa etapa del desarrollo, desde las primeras vocalizaciones aleatorias hasta la imitación defectuosa de los sonidos de las primeras palabras como el fruto de su exposición repetida (cfr. Dennett, 1996, pp. 147 y ss.; 2017a, pp. 190 y ss.), los sonidos son reproducidos como “microhábitos vocales”, sin que tengan significado para sus usuarios. Así, “la fonología viene primero”, antes que la semántica y la sintaxis (cfr. Dennett, 2017a, p. 193), que también se adquieren gradualmente y “casi sin instrucción” (2017a, p. 195).

Las “primeras palabras” serían aprendidas como “garabatos” — serían comparables a los primeros trazos en el proceso de aprendizaje

de la escritura— porque el niño no conoce todavía sus significados (cfr. Dennett, 1993, p. 546). Esas palabras aún no comprendidas suscitan gradualmente asociaciones con las circunstancias experimentadas al oírlas, lo que permitiría “etiquetar” eventos u objetos. El uso reiterado de tales etiquetas haría posible que lleguemos a ser capaces de “contemplar las representaciones” o “etiquetas internas”, lo que nos convertiría “en *entendedores* de los objetos que hemos creado” (cfr. Dennett, 1997, p. 229; 1996, pp. 150-151). Es decir, esas etiquetas verbales, al ser desacopladas de sus instanciaciones sonoras, generan “rótulos internos” o conceptos (cfr. Dennett, 1997, p. 229). En síntesis, el aprendizaje va de las palabras escuchadas y repetidas a los conceptos comprendidos, en el sentido de que las primeras son los antepasados de los segundos (cfr. Dennett, 1996, p. 151; más sobre este proceso y sus efectos cognitivos en la sección siguiente).

Las primeras palabras se transforman de a poco en palabras especializadas (nombres, adjetivos, verbos, etc.) cuando empieza a ser necesaria cierta estructura para expresar pensamientos complejos: la gramática y la sintaxis (cfr. Dennett, 1997 y 2017a). Sean cuales fueren las capacidades o constreñimientos innatos (sintácticos o semánticos) que hacen posible ese aprendizaje, los niños adquieren su lengua materna como un hábito, sin construir “teoría” alguna, es decir, por medio de un proceso “de abajo arriba” que les permite desarrollar *competencia sin comprensión* (2017a, p. 197).¹⁹ “El lento comienzo y la subsecuente mejora de la comprensión (compartida) es un *efecto* de la práctica lingüística, no un prerrequisito, tanto ontogenética como filogenéticamente” (Dennett, 2018a, p. 149).

Esta manera de caracterizar el logro de la competencia lingüística da cuenta de un fenómeno ubicuo en el hablante adulto, manifiesto, por ejemplo, en el hecho de que “entender una palabra *no* es lo mismo que haber adquirido una definición de ella” (Dennett, 2017a, p. 195) o, al revés, en la imposibilidad de establecer “cuánto necesitamos

¹⁹ Dennett disiente de Quine (1968) porque este asemeja las tareas del lingüista de campo o “traductor radical” frente a un lenguaje extraño a las del niño que intenta aprender su lengua materna. Sin embargo, el traductor hace su tarea como un teórico, es decir, formula y testea hipótesis, recoge evidencias, etc. El niño, en cambio, balbucea, tantea y avanza mediante “un proceso inconsciente de ensayo y error masivo”, logrando “la misma proeza epistemológica que el adulto, pero sin el beneficio de una teoría” (Dennett, 2017a, pp. 196-197).

conocer para entender nuestros propios conceptos” (Dennett, 2009a, p. 10063), tanto porque la competencia no siempre asciende al nivel de la comprensión consciente como porque la mayor parte de las palabras no tienen significados únicos, estables y determinados. Nuevamente, la competencia no está subordinada a la comprensión.

6. El lenguaje como herramienta mental

Si se quiere sostener que ciertas facultades, como la conciencia de sí mismo, la abstracción, son peculiares al hombre, es fácil también que sean resultados accesorios de otras facultades intelectuales muy adelantadas, que, a su vez, sean principalmente el resultado del uso continuo de un lenguaje que haya alcanzado un alto grado de desarrollo.

Darwin (2009b, p. 105).

Las palabras también pueden ser definidas como *herramientas mentales* (*thinking tools*) diseñadas por la evolución natural-cultural (cfr. Dennett, 2017a, p. 292; DDI; 1993; 1996; 1997; 2009b y 2017a): mediante ellas podemos producir y testear modelos acerca del entorno antes de actuar, así como pensar acerca de los propios pensamientos, involucrarnos en soliloquios, hacer autocríticas, mentir, etc. En primer lugar, estas herramientas funcionan como “muletas cognitivas” (Dennett, 1993, p. 545) que son el fruto de “nuestra costumbre de *descargar* al máximo posible nuestras tareas cognitivas en el propio entorno [...] en el que un montón de dispositivos periféricos que construimos pueden almacenar, procesar y volver a representar nuestros significados” (Dennett, 1996, pp. 134-135). Estas descargas consisten en “colocar deliberadamente marcas en el entorno” (p. 135). No es una estrategia exclusivamente humana, aunque solo nosotros convertimos esas marcas en “objetos que se pueden manipular, rastrear, mover, atesorar, alinear, estudiar” (p. 143). ¿Cuáles son sus efectos sobre nuestras mentes?

El capítulo 13 de DDI comienza con un apartado dedicado a “El rol del lenguaje en la inteligencia” en el que contrasta la gran tradición filosófica, que concibe la diferencia entre nosotros y los demás animales como “absoluta y metafísica” (Dennett, 1995, p. 370), con la perspectiva basada en las evidencias de la etología, la biología evolutiva y las ciencias cognitivas, según la cual, “aunque somos diferentes [...] solo hay una enorme diferencia de grado” (1995, p. 370). Esa enorme diferencia podría

explicarse por la confluencia de dos factores: (a) nacemos con cerebros diferentes, cuyos rasgos han evolucionado a lo largo de seis millones de años, y (b) esos rasgos permitieron incrementar nuestros poderes para elaborar y compartir nuevos diseños mediante la transmisión cultural. Ambos están entrelazados por el lenguaje. Por lo tanto, la diferencia que nos distingue es que somos la especie más inteligente y “somos también la única especie con lenguaje” (1995, p. 371), es decir, contamos con “un medio *extra* para preservar y comunicar el diseño: la cultura” (1995, p. 338). Como veremos, los procesos que conectan la inteligencia, el lenguaje y la cultura se explican por la misma “extraña inversión del razonamiento” darwiniana.

Aunque diferentes especies tienen rudimentos de cultura y algunas clases de “lenguaje”, es decir, sistemas de señales que satisfacen algunas de las funciones de ciertas porciones de nuestro lenguaje, el lenguaje humano es único en cuanto a sus propiedades para codificar información (como vimos en las secciones anteriores), pero también porque amplifica los poderes mentales de sus usuarios. Así como nuestras mentes son el fruto de la restructuración de nuestros cerebros por los memes, esos mismos memes se convierten en *herramientas de pensamiento* porque “se mueven desde afuera hacia adentro”, es decir, siguen una “ruta social” que va del uso público de las “etiquetas” o señales verbales, a través de procesos de familiarización, hasta crear gradualmente prácticas significativas que, finalmente, se internalizan en las mentes de los hablantes (cfr. Dennett, 2000, p. 21). Como se puede ver, el lenguaje es una herramienta mental con un impacto cognitivo-individual porque antes es una herramienta pública con un uso sociocognitivo y comunicativo: las palabras, como dijimos antes, son “modos de comportamiento”.

Así, la evolución cultural que creó las palabras “añadió novedades en el entorno (en forma de miles de *affordances* gibsonianas) que enriquecieron las ontologías de los seres humanos y, a su vez, generaron más presiones selectivas que favorecieron las adaptaciones (herramientas de pensamiento) que permitían controlar y rastrear todas estas nuevas oportunidades” (Dennett, 2017a, p. 412). Esta “tecnología” está compuesta por “una nueva clase de objetos para contemplar [...] que podrían ser revisados en cualquier orden y a cualquier velocidad” (1995, p. 380). Como vimos en la sección anterior, cuando esos nuevos “objetos” se convierten en conceptos, pueden ser aplicados productivamente para pensar sobre una variedad potencialmente infinita de nuevas cuestiones. Antes que ninguna otra cosa, son objetos

públicos que usamos para comunicarnos, que también nos permiten beneficiarnos “de los trabajos cognitivos de otros”, lo que amplifica aún más nuestros poderes mentales individuales (1995, p. 381).²⁰

Dennett sugiere que la identificación de las variadas opciones de diseño de los cerebros de distinto tipo de criaturas permitiría comprender mejor el papel crítico del lenguaje en el caso del cerebro humano. Según su hipótesis, hay diferentes tipos de mentes y de criaturas: *darwinianas*, *skinnerianas*, *popperianas* y *gregorianas*. Las criaturas darwinianas se adaptan por medio de mutaciones genéticas y selección natural; las criaturas skinnerianas se adaptan, además, por medio del aprendizaje por la vía del ensayo y el error; las criaturas popperianas también pueden evaluar internamente entre distintas alternativas antes de actuar: acopian información acerca del entorno externo y pueden considerar, con distintos mecanismos, cursos alternativos de acción antes tomar una u otra decisión. Finalmente, las criaturas gregorianas son las únicas que pueden albergar “entornos internos” ricamente informados por porciones diseñadas del entorno cultural, porque disponen, entre otras herramientas, de las palabras.²¹ Las criaturas gregorianas son capaces de manipular herramientas en el doble sentido que dicha capacidad es tanto un signo de inteligencia como que ellas les confieren aún más inteligencia. En tanto herramientas mentales, las palabras proporcionan un entorno interno que permite pensar mejor.

Ciertas capacidades innatas, incluso conceptuales, podrían apoyar la incorporación gradual del repertorio léxico de una criatura gregoriana (cfr. Dennett, 1995, pp. 378-379). Como advierte Dennett, Darwin entrevió correctamente que el lenguaje era un prerequisite para tener no solo pensamientos abstractos y pensamientos acerca de pensamientos, ejerciendo control sobre ellos, sino:

[...] que el uso continuo y el perfeccionamiento de esta facultad han debido influir a su vez en la inteligencia, permitiéndole y facilitándole el enlace de una serie más extensa de ideas. Nadie puede emitir una sucesión

²⁰ También los mapas, los cálculos, los microscopios, las estadísticas son herramientas para pensar más y mejor. Pero, a diferencia de las palabras, son diseños inteligentes.

²¹ Dennett homenajea al psicólogo británico Richard Gregory (1923-2010) por haber subrayado la importancia de la información como herramienta mental que confiere inteligencia potencial a su portador.

prolongada y compleja de pensamientos sin el auxilio de palabras, sean habladas o en silencio, como tampoco hacer un extenso cálculo sin el uso de figuras o el álgebra (Darwin, 2009b, p. 57).

Así, una vez provistos de lenguaje, nuestros cerebros difieren de los “cerebros desnudos de los animales” (Dennett, 1995, p. 381) porque pueden pensar pensamientos más complejos. En efecto, las palabras nos permiten anticipar, imaginar, planear cursos posibles de acción, en suma, un poder para conocer que no tiene comparación con los poderes cognitivos de otras criaturas. Pero también difieren de los “cerebros desnudos de los animales” porque los nuestros son “mentes enculturadas” por esas herramientas mentales: las palabras son memes opcionales que transportan información cultural acumulada que nos interesa incorporar.

En cuanto al carácter y alcance de esta tipología mental, es importante señalar, en primer lugar, que las diferencias entre uno y otro tipo de mentes son el resultado de transformaciones graduales. En segundo lugar, los mecanismos que podrían explicar el pasaje de unos a otros tipos mentales se explican por procesos seleccionistas, es decir, no dependen de una mutación azarosa única.²² En tercer lugar, que la comprensión solo es posible porque nuestras mentes gregorianas disponen (entre otras) de unas especiales herramientas mentales para su uso deliberado no solo por tratarse de una tecnología cuyo diseño hace posible una explosión de pensamientos novedosos y complejos, sino porque en virtud de la información semántica que ellas transportan nos permiten usufructuar el conocimiento culturalmente acumulado (cfr. Dennett, 1995, p. 381; 1996, p. 101). Finalmente, esta tipología mental identifica los hitos cognitivos más destacados que las diferencian (cfr. Dennett, 1995, p. 373; 1996, p. 83; 2018a, p. 146), no tipos de mentes homogéneas y excluyentes entre sí.²³ Nosotros, criaturas gregorianas, somos también criaturas popperianas, skinnerianas y darwinianas.

²² Chomsky propone un mecanismo de ese tipo para dar cuenta de la recursión, la propiedad crítica que habría hecho posible la emergencia del lenguaje. Pero, según Dennett: “La idea de que una mutación azarosa puede transformar una especie de una vez no es una historia ni remotamente creíble” (2017a, p. 280).

²³ Esta tipología ha suscitado varios cuestionamientos: ¿son esas *todas* las clases de mentes o solo sirven para destacar algunos hitos evolutivos mayores?

Esta forma de entender el rol tanto cognitivo-individual como cognitivo-social del lenguaje difiere de otras variantes propuestas en la literatura sobre las relaciones constitutivas entre lenguaje y pensamiento. Me refiero a los enfoques ya referidos según los cuales los componentes conceptuales de las etiquetas léxicas y las reglas sintácticas de las estructuras oracionales canónicas en las que participan se correlacionan con la naturaleza proposicional y la función representacional de los pensamientos expresados por ellas.²⁴ Los rasgos del lenguaje involucrados en la constitución de esas estructuras de pensamiento preexisten y se diferencian de los rasgos de las lenguas históricas; por lo tanto, no son aprendidos ni transportan información cultural que pueda transformar las mentes de quienes los aprenden.

7. El significado de las palabras

En el capítulo 14 de DDI, “La evolución de los significados”, Dennett señala acertadamente que ningún tópico ha recibido más atención en la filosofía que el significado en sus “variadas manifestaciones” (1995, p. 401). En la filosofía del lenguaje de la tradición analítica, donde ha sido objeto de un “escrutinio microscópico”, pueden identificarse, entre otros, los siguientes supuestos comunes, encadenados entre sí: (a) el significado es un fenómeno “dependiente del lenguaje”; (b) “las palabras [...] obtienen sus significados de nosotros” (Dennett, 1995, p. 404), de nuestras mentes; y (c) el significado es un rasgo *sui generis*, independiente de la función biológica y no evolutivo, por lo que no admite gradualidad, variación ni indeterminación. Consideremos brevemente cada uno de ellos.

Muchos filósofos asumen que el significado, *stricto sensu*, es una propiedad de ciertos ítems lingüísticos y solo derivadamente de otras entidades. O, en su defecto, asumen que hay una clase de significado — el significado *real* — que es “lenguajependiente”, por lo que debemos

Cfr. las propuestas de Godfrey-Smith (2018) y Clatterbuck (2018) sobre otros tipos de criaturas que harían el pasaje hacia las criaturas gregorianas aún más gradual.

²⁴ Una concepción de ese estilo está presupuesta por la lingüística chomskiana y es elaborada por la hipótesis del *Lenguaje del Pensamiento* de Fodor. Para un análisis de las diferencias entre las propuestas de Fodor y Chomsky y la de Dennett sobre las relaciones entre pensamiento y lenguaje, cfr. Miguens (2022).

explicar primero cómo las expresiones lingüísticas tienen significados para examinar luego, sobre la base de ese modelo, cómo otro tipo de entidades podrían tener una propiedad similar o análoga. Este supuesto ha generado una “tradición linguacéntrica” (cfr. Dennett, 2017b). Ahora bien, se suele asumir, además, que el significado lingüístico es una variedad de un fenómeno más fundamental, restringido a las mentes humanas, que es la intencionalidad: la propiedad de “ser acerca de algo”. Esta propiedad emparenta a los significados con los contenidos de los estados mentales intencionales: aquellos derivan de estos, su fuente original.

Esta idea nos lleva al segundo supuesto: “las palabras [...] obtienen sus significados de nosotros” (Dennett, 1995, p. 404). Dennett ridiculiza la versión extrema de esa idea, expresada por el personaje de Lewis Carroll (2015), Humpty Dumpty: “Cuando yo utilizo una palabra [...] significa precisamente solo lo que yo quiero que signifique, ni más ni menos” (p. 179). Según esta versión, los significados no solo son transferidos a las palabras por el hablante, quien primero los comprende y luego los vierte en las palabras que profiere, sino que, por esta razón, tiene una autoridad final sobre su contenido. En algunas variantes de esta idea, los hablantes deben ser capaces de pensamientos de orden superior acerca de los contenidos semánticos de sus pensamientos de primer orden para poder emplear comunicativamente el lenguaje que los expresa (veremos esto en la sección 8). Ahora bien, los significados de las emisiones lingüísticas *no dependen siempre ni enteramente* de las actitudes y los contenidos mentales del hablante, es decir, la “intuición” según la cual poseemos autoridad semántico-pragmática plena sobre lo que decimos y hacemos con el lenguaje se asienta en una concepción individualista e intelectualista cuestionable. La noción de “meme” que sustituye a la noción idealizada de “significado” intenta corregir esta distorsión.

Si concebimos, en cambio, que la propiedad de ser intencionales es compartida por entidades muy diferentes —además de las palabras y los estados mentales, también las representaciones gráficas, las piezas musicales, los artefactos, las representaciones perceptuales en distintas especies, es decir, no se trata de una propiedad sólo lingüística—, es natural reconocer que tampoco se trata de una propiedad homogénea. Y, a menos que asignemos intencionalidad original a algunas de ellas —típicamente a ciertas entidades mentales—, el otro camino disponible es indagar cómo pueden haber derivado, en virtud de cierto tipo de

procesos, de fenómenos con menos o ninguna intencionalidad. Es decir, o bien asumimos el principio “*Mind-first*” guiados por una intuición *a priori* (separamos *nuestras mentes* del mundo y de las explicaciones naturales) o bien intentamos explicar cómo emerge, se desarrolla y se manifiesta esta propiedad en tales variedades. De esa manera, la “extraña inversión del razonamiento” alienta una “visión *Mind-last*” (2017a, p. 58). Bajo esta perspectiva, carece de justificación ordenar los distintos tipos de fenómenos intencionales según una distinción esencial entre “genuina” intencionalidad y significado *stricto sensu*, por una parte, y sus manifestaciones meramente derivadas o “como si”, por la otra.

Los primeros dos supuestos impidieron adoptar una perspectiva evolutiva acerca de cómo emergieron los significados y, más ampliamente, una perspectiva diacrónica acerca del lenguaje. Es claro que “los filósofos se han mostrado poco dispuestos a aceptar la hipótesis de que el pensamiento evolucionista pueda arrojar luz sobre los problemas de cómo es que estas palabras, y sus fuentes y sus destinos en las mentes o cerebros humanos, poseen significado” (Dennett, 1995, p. 402).²⁵ En otras palabras, asumieron una visión sincrónica sobre los fenómenos semánticos en combinación con una perspectiva sobreintelectualizada del significado. Ejemplos de tales puntos de vista se advierten en las ideas de Fodor, Searle, Putnam, Kripke, entre otros conspicuos representantes del denominado “Grupo de Trabajo de la Actitud Proposicional”. Dennett les atribuye la defensa de una u otra variante de la tesis de la intencionalidad intrínseca y original de los contenidos mentales, que luego heredan los significados lingüísticos. Los argumentos a favor de una concepción de los significados derivados y dependientes de la interpretación son expuestos detalladamente en el capítulo 8 de *La actitud intencional* (Dennett, 1998).

Dennett reivindica en distintos lugares las lecciones de Quine contra “el mito de museo”: la concepción de los significados como entidades, y lamenta que “la mayoría complaciente” en la filosofía analítica del lenguaje haya desoído o no tomado en serio sus argumentos sobre la “indeterminación del significado” (1995, pp. 402-403; 2018c). Todas las palabras pueden vehiculizar contenidos potencialmente indeterminados —incluyendo (*pace* Putnam) nuestros términos de clase natural en escenarios de tierra gemela— (cfr. Dennett, 1995, pp. 409-410). Así,

²⁵ Dennett (1995, pp. 402-403) destaca las ideas naturalistas de Quine, Sellars, Dewey y sobre todo Millikan como excepciones al canon.

los “significados determinados”, que suelen tomarse como la piedra de toque en la argumentación filosófica, deberían concebirse más bien como “artefactos filosófico-antropológicos”, es decir, idealizaciones elaboradas para los propósitos de la argumentación.²⁶ Aunque puedan servir como buenos puntos de partida para la reflexión filosófica, no deberían desempeñar el papel de “la última palabra” (Dennett, 1995, p. 60). Sin que Dennett les dedique una atención parecida, otras dos propiedades de los significados de la mayor parte de las palabras corrientes recomiendan más claramente aún la adopción de una perspectiva diacrónica: no son estables, es decir, sufren cambios, muchas veces imperceptibles para los hablantes, y no son únicos, en el sentido de que una misma palabra transporta con frecuencia significados diferentes, algunos emparentados entre sí de maneras muy intrincadas.

Ahora bien, aunque Dennett vincula los contenidos semánticos de las emisiones lingüísticas con otros contenidos expresados en formatos y vehículos no lingüísticos, no se interesa por el papel de las señales expresivas o comunicativas en distintas modalidades (gestual, facial, corporal) que podrían haber apuntalado y complementado las primeras emisiones protoverbales, es decir, que formarían parte de su trayectoria evolutiva, pero que también desempeñan un papel crítico en la comunicación adulta. Ellas también evidencian que el significado no es solo lingüístico y que diferentes señales pueden desempeñar, de maneras eficientes, variadas funciones en las acciones comunicativas, sin la intermediación de procesos de comprensión autoconscientes.²⁷ Además, esta es una evidente implicación de la noción de “meme”, ya que pone de relieve el parentesco del lenguaje con otras herramientas informacionales que pueden ser explicadas, en este nivel fundamental, del mismo modo. Es probable que esta “laguna” se deba a su interés primario por dar cuenta de las singulares ventajas adaptativas que proporcionó el desarrollo del lenguaje verbal, con sus propiedades y funciones especializadas y su enorme impacto cognitivo y cultural en la historia evolutiva del *Homo Sapiens*.

²⁶ Los filósofos, además, intercambian intuiciones sobre esos conceptos sin advertir que, cuando son “obsoletos” o “reliquias de un previo *Zeitgeist*”, “fossilizan la vieja ciencia” (Dennett, 2018b, p. 57).

²⁷ Cf. Scotto (2022 y 2024).

8. La comunicación lingüística

El significado de las palabras, lo que está detrás de ellas, no me preocupa en la comunicación lingüística normal. Las palabras fluyen, y se lleva a cabo la transición de las palabras a las acciones y de las acciones a las palabras. Nadie piensa, al hacer cálculos aritméticos, en si los hace “reflexivamente” o “a la manera de un papagayo” [...].

Wittgenstein (1997, § 603).

El examen de la función comunicativa del lenguaje está prácticamente ausente en DDI.²⁸ En Dennett (2017a), en cambio, ocupa un lugar relevante, especialmente en el capítulo 13, donde una sección está dedicada a “Las razones de flotación libre de la comunicación humana”. Allí se examina el modelo griceano, que concibe a la comunicación como un proceso de razonamiento articulado, con sus órdenes anidados de estados intencionales en el hablante y el oyente. Según Grice, para que un hablante H signifique algo emitiendo *x*, debe *intentar* (a) que *x* produzca cierta respuesta *r* en una audiencia *A*, (b) que *A reconozca* su intención (la de *H*), y (c) que dicho reconocimiento sea al menos parte de la razón de *A* para *producir su respuesta r*. Distintos autores han señalado que un modelo así no parece adecuado para dar cuenta de las conversaciones ordinarias, en particular, del grado de complejidad cognitiva requerido por parte de quienes participan en ellas. También lo hace Dennett (2017a, p. 292), para quien Grice ofrece un modelo “idealizado” de la comunicación, en el sentido de que incorpora las “aplicaciones sofisticadas” de todas las herramientas de la comunicación ordinaria que podrían explicar su funcionamiento óptimo.²⁹ Pero las condiciones y finalidades bajo las cuales ocurre la conversación presentan muchas variantes. Incluso los

²⁸ Únicamente hay una breve referencia en DDI a la teoría griceana, en la que solo señala su compromiso con un enfoque del significado “lenguajedependiente” (cfr. DDI, p. 402).

²⁹ Diferentes filósofos han argumentado que la *idealización* griceana de la comunicación conlleva muchos compromisos objetables: presupone que las interacciones conversacionales ordinarias *deben* ser cooperativas, veraces y sinceras, lo que implica que buena parte de ellas son desviaciones de la norma (cfr. Keiser, 2023).

hablantes difieren mucho entre sí en sus habilidades comunicativas. En cualquier caso, no necesitan *comprender* las razones que justifican su empleo más o menos competente de una gran variedad de herramientas.

El “acierto” de Grice residiría, en cambio, en haber ofrecido una reconstrucción, mediante ingeniería inversa, de las complejidades de diseño involucradas en el funcionamiento exitoso de las interacciones comunicativas. De ese modo, habría mostrado que los comunicadores humanos “tienen la *competencia para explotar* estos rasgos” (Dennett, 2017a, p. 292) y que, en consecuencia, “[L]as sucesivas capas de cognición de Grice quizá no representan con exactitud los rasgos inmediatos de una representación, pero sí que especifican una competencia” (2017a, p. 395). En síntesis, la teoría griceana no debiera tomarse como una teoría de la actuación comunicativa, sino, quizás, de la *competencia ideal*, como tal no representada en la mente de los hablantes. En otras palabras, la conversación humana funciona con gran eficiencia porque los actos de habla están gobernados por “razones griceanas de flotación libre” (2017a, p. 395).

Una interesante elaboración de esta interpretación puede encontrarse en Planer (2023, p. 1291): “Los agentes (*e. g.*, niños pequeños, adultos en conversaciones rápidas, posiblemente grandes simios) pueden involucrarse en una comunicación plenamente griceana sin albergar estructuras cognitivas griceanas en sus cabezas”. Las herramientas de la atribución intencional propuestas por Dennett (1991b y 1998), incluyendo presupuestos de racionalidad y coherencia entre las creencias y otros estados para la atribución de patrones intencionales, pueden aplicarse también a las interacciones comunicativas ordinarias como simplificaciones de la psicología *folk*, es decir, solo justificadas por su poder predictivo. En tal sentido, siguiendo a Planer (y a Dennett, 1991b), cuando conversamos no ejemplificamos “estructuras cognitivas griceanas” sino solo “patrones reales griceanos”.

Si interpretamos la explicación griceana según el modelo de la comprensión como requisito de la competencia, es decir, atribuyendo en exceso razones conscientes para explicar las prácticas comunicativas, pondríamos las cosas en un orden incorrecto. Como afirma Wittgenstein en el fragmento antes citado, la mayor parte de los usos competentes del lenguaje no están acompañados por un conocimiento reflexivo de lo que cada interlocutor comprende acerca de lo que el otro está queriendo

decir o hacer.³⁰ Incluso los más diestros “conversadores” pueden no ser capaces de analizar cómo y por qué las variadas herramientas comunicativas disponibles operan eficazmente en las conversaciones corrientes de las que participan. Como sugiere Dennett, ni los más astutos manipuladores ni los más ingenuos novatos “tienen por qué comprender las razones por las cuales sus recursos comunicativos cotidianos ofrecen todas las opciones que ofrecen” (2017a, p. 292). Es decir, “pueden desarrollar agudas sensibilidades a los riesgos y peligros y oportunidades en la comunicación que ellos mismos no pueden analizar” (2017a, pp. 292-293).

Planer (2024) desarrolla esta idea, apoyado en la perspectiva memética de la evolución cultural. Muestra cómo podría explicarse la emergencia de los primeros mapeos entre formas fonológicas y significados antes de que la sintaxis sea necesaria para proporcionar una estructura a nuevos mapeos conforme las interacciones comunicativas adquieren creciente complejidad. Según su propuesta, el sistema lingüístico emerge gradualmente a partir de tanteos comunicativos cada vez más exitosos. Es decir, deberíamos invertir el enfoque de la comunicación como un fenómeno derivado de la previa posesión de un sistema lingüístico. Pero también deberíamos remover el supuesto de que nuestra competencia comunicativa adulta requiere autoconsciencia semántica. Consideremos esta cuestión ahora.

Dennett (2017a, pp. 335-370) examina la conjetura según la cual la evolución de la comunicación tendría que ser explicada no ya sobre la base de la conciencia, sino sobre la necesidad que surge, entre organismos en competencia, de manipular estratégicamente a otros mediante comportamientos de “engaño guiado” que tienen la función de ocultar o confundir a otros acerca de los propios propósitos. Ello habría creado un “espacio privado” para los pensamientos que no se quieren revelar a otros, y que pueden complejizarse a medida que evoluciona el lenguaje como una herramienta más versátil para la comunicación. Compartir información reflexiva y cooperativamente, entonces, presupone la

³⁰ En una formulación “patas para arriba”, Wittgenstein sostiene que los niños aprenden a hablar —y, en general, a seguir reglas— sin entender, es decir, por medio de mecanismos como la imitación, la ejercitación y el adiestramiento, hasta que logran una *competencia* “ciega”. En ese mismo proceso, también llegan a ser capaces de comprender, en el sentido de Dennett (cfr. Wittgenstein, 2003, §§ 208-211 y 219).

participación en interacciones comunicativas que puedan crear, gradualmente, un espacio para los propios pensamientos e intenciones, es decir, para la emergencia de la conciencia. Según esta “inversión del razonamiento”, “ser uno mismo” derivaría de la capacidad adquirida para decirles a otros (o abstenerse de hacerlo) “cómo es ser uno mismo” (Dennett, 2017a, p. 344). Esa es la idea del “yo” como un “centro de gravedad narrativa” (Dennett, 1991a).

La autoconsciencia de cada individuo como una “ilusión del usuario” emerge así de una previa “comunidad de ilusiones de usuarios” (Dennett, 2017a, p. 342) o, lo que es igual, la comunicación es la innovación que hace que un organismo requiera autoconsciencia. Dennett sugiere, incluso, que esta se apoya en el previo desarrollo de la interacción diádica o de segunda persona a través de la atención conjunta a las *affordances* a las que prestar atención: “Si no tuviéramos que ser capaces de hablar con los demás sobre nuestros pensamientos y proyectos actuales, sobre nuestros recuerdos de cómo eran las cosas, etc., nuestros cerebros no gastarían tanto tiempo, energía y materia gris en un boletín revisado de nuestras actividades actuales, que es precisamente lo que es nuestro flujo de conciencia” (Dennett, 2017a, pp. 344-345). En síntesis, tanto la cooperación como la autoconsciencia serían efectos de la evolución de prácticas comunicativas originalmente manipuladoras y no reflexivas.

Reflexiones finales

[...] esa actividad, basada en el lenguaje, conocida como filosofía.

Dennett (2020, p. 3).

El origen del hombre ahora está demostrado. —La metafísica debe florecer. —Quien entienda a los babuinos haría más por la metafísica que Locke.

Darwin (1987, D26, M84).

Según Zawizdki (2007), la obra de Dennett se caracteriza por su “visión sinóptica”, en el sentido de que entiende la filosofía como el esfuerzo por *reconciliar*, en alguna medida, la imagen manifiesta con la imagen científica. En este aspecto, Dennett sigue a Sellars (1992, p. 15). En su caso, esa reconciliación es *revisionista* respecto a los conceptos

de la imagen manifiesta puesto que, si bien los fenómenos admitidos por ella son reales, “no son como pensábamos que eran”. En un sentido similar, Schliesser (2019) caracteriza a su enfoque metodológico como una “filosofía sintética”,³¹ porque busca producir una visión coherente e integrada de los sistemas complejos, recogiendo los resultados de las ciencias especiales pero apoyándose en una teoría científica general, también en el espíritu de otra clásica formulación sellarsiana: “entender cómo las cosas, en el más amplio sentido posible del término, se relacionan entre sí [*hang together*] unas con otras, en el más amplio sentido del término” (Sellars, 1992, § 1). Las “cosas” que integran la ontología de la imagen manifiesta tienen que “relacionarse entre sí” con aquellas más o menos desconocidas de la imagen científica. Una tarea que es para filósofos. Como afirma Dennett, esa “es la mejor definición de la filosofía que he encontrado” (2023, cap. 33). En ese marco meta-filosófico, el optimismo de Darwin sobre la filosofía (“La metafísica debe florecer...”) claramente inspiró su obra.

Dennett resume el aporte que intenta hacer en su último libro, y que puede extenderse también a DDI, así:

Mi propósito global en este libro es que la perspectiva evolutiva en general y la perspectiva memética por lo que respecta a la cultura transformen muchos de los aparentemente eternos rompecabezas de la vida, como el significado y la conciencia, en caminos inaccesibles para quienes nunca miran más allá de la imagen manifiesta con la que crecieron y de las disciplinas en las que se han formado (2017a, pp. 242-243).

Justamente por trascender el marco estrecho de una u otra disciplina y, en particular, de aquellas más apegadas a las intuiciones que apoyan la imagen manifiesta —a eso refiere la mordaz definición de la filosofía como una “actividad basada en el lenguaje”—, la “filosofía sintética” tiene que ser evaluada por el potencial heurístico de las conjeturas que puede producir y por su capacidad para apoyar o reorientar las explicaciones ya disponibles en distintas disciplinas con vistas a lograr una visión coherente y científicamente informada. En suma, en tanto

³¹ Aunque se refiere a Dennett (2017a), aplica también a DDI.

sea capaz de relacionar de manera apropiada la imagen manifiesta con la imagen científica.

He intentado mostrar que, para Dennett, esa tarea solo puede ser encarada poniendo “patas para arriba” algunas intuiciones basadas en la imagen manifiesta que alimentaron la especulación filosófica, más precisamente, reconceptualizando las piezas de la estructura “*Mind-first*” y sus funciones. Ello requiere admitir *razones sin razonadores y competencia sin comprensión*. Desde esa “perspectiva invertida” es posible recorrer el camino evolutivo que condujo hacia el diseño inteligente para, finalmente, llegar a “poner nuestras mentes al revés” (Dennett, 2017a, pp. 4-6) haciendo emerger una “visión *mind-last*” (Dennett, 2017a, p. 58). Por su importancia crítica, ello también requiere que la “inversión del razonamiento” pueda explicar cada uno de los aspectos y dimensiones del lenguaje, no solo como herramienta mental y como medio de comunicación sino, sobre esos dos fundamentos, como condición de posibilidad de la cultura acumulativa. De ese modo, el lenguaje recibe una explicación evolutiva que no solo da cuenta de sus orígenes graduales desde “abajo hacia arriba”, sino que también corrige el modo idealizado de entender sus usos y funciones insertándolos en el territorio diacrónico de la evolución natural-cultural. Así, sería posible cerrar la brecha que abrió Descartes al establecer una conexión constitutiva entre el diseño inteligente, la competencia lingüística y la singularidad humana en un marco explicativo predarwiniano. Ese es el objetivo ambicioso de *Darwin’s Dangerous Idea*: universalizar el darwinismo.

En este trabajo solo identifiqué los aspectos más salientes (varios de ellos muy polémicos) que convierten a ese ambicioso objetivo en un programa desafiante. Hay que señalar, también, que se trata de un programa que tiene pocos cultores en la filosofía, quizás porque todavía es evaluado como “peligroso” para el canon. Después de todo, como advierte el propio Dennett, “la filosofía ha sido el principal hogar académico de la metarepresentación durante milenios (2017a, p. 300). O, como también afirma Dennett, quizás se deba a otro rasgo que ha sido distintivo de la filosofía: “la voluntad de poner fin a la curiosidad prematuramente” (2014b, p. 59; para una interesante conjetura que sintetiza las dos anteriores, cfr. Zawidzki, 2007, p. 174). Expresado más sintéticamente, la suya es una batalla contra el error de considerar que las razones que nos parecen evidentes son “la última palabra”.

En sus primeros escritos, Dennett exhibió un optimismo por momentos exultante sobre las promesas de los diseños inteligentes humanos —su propia versión de la excepcionalidad humana—: sobre la ciencia, el arte y todas las sofisticadas manifestaciones de la cultura y, en particular, sobre las potencialmente ilimitadas capacidades de los dispositivos inteligentes creados por las mentes humanas. En su último libro (Dennett, 2017a), en cambio, se advierte una actitud diferente. Los procesos de *desdarwinización de la cultura* que, por una parte, explican la creación de los diseños inteligentes artificiales paradójicamente, por otra parte, amenazan con poner fin a esos mismos procesos que les dieron origen en virtud de una gradual enajenación de nuestra autonomía, o lo que es igual, de una inusitada expansión de la *competencia sin comprensión*. Dennett se refiere a esta nueva época como la era del *diseño postinteligente* justamente porque “los algoritmos han comenzado a reemplazar a las personas como generadores de novedad cultural” (Rathkopf, 2017, p. 1355). Se trata de nuestra sumisión creciente a los “nuevos oráculos computacionales” (Dennett, 2018c, p. 106), a los procesos ciegos de tráfico de información (que muchas veces es insidiosa, falaz o tóxica), asociada a una alarmante “abdicación de nuestras responsabilidades cognitivas” (Dennett, 2017a, p. 405). La pregunta desafiante sería ahora: ¿seremos capaces de revertir el actual proceso de *redarwinización de la cultura* “invirtiendo” el tipo y la magnitud de nuestras interacciones con esos diseños inteligentes? Una pregunta difícil que quedará para otra ocasión.

Referencias

- Blackmore, S. J. (1999). *The Meme Machine*. OUP.
- Carroll, L. (2015). *Los libros de Alicia*. E. Stilman (trad.). Ediciones de la Flor.
- Clatterbuck, H. (2018). Dennett’s Canon and Major Cognitive Transitions. *Teorema*, 37(3), 129-143.
- Darwin, C. (1987). *Charles Darwin’s Notebooks 1836-1844*. P. H. Barrett, P. J. Gautrey, S. Herbert, D. Kohn y S. Smith (eds.). Cambridge University Press.
- Darwin, C. (1992). *El origen de las especies*. A. de Zulueta (trad.). Planeta.
- Darwin, C. (2009a). *The Correspondence of Charles Darwin. Volume 20*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139815000>
- Darwin, C. (2009b). *The Descent of Man, and Selection in Relation to Sex*. CUP. <https://doi.org/10.1017/CBO9780511703836>

- Dawkins, R. (1976). *The Selfish Gene*. OUP.
- Dawkins, R. (1982). *The Extended Phenotype*. OUP.
- Dawkins, R. (1983). Universal Darwinism. En D. S. Bendall (ed.), *Evolution from Molecules to Men* (pp. 403-425). CUP.
- Dennett, D. C. (1980). Passing the Buck to Biology. *Behavioral and Brain Sciences*, 3, 19. <https://doi.org/10.1017/S0140525X00001552>
- Dennett, D. C. (1991a). *Consciousness Explained*. Little, Brown and Co.
- Dennett, D. C. (1991b). Real Patterns. *Journal of Philosophy*, 88(1), 27-51. <https://doi.org/10.2307/2027085>
- Dennett, D. C. (1993). Learning and Labelling (Comments on Clark and Karmiloff-Smith, “The Cognizer’s Innards”). *Mind & Language*, 8(4), 540-547. <https://doi.org/10.1111/j.1468-0017.1993.tb00302.x>
- Dennett, D. C. (1995). *Darwin’s Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster.
- Dennett, D. C. (1996). *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfeld & Nicolson.
- Dennett, D. C. (1997). How to do Other Things with Words. *Royal Institute of Philosophy Supplement*, 42, 219-245. <https://doi.org/10.1017/S1358246100010262>
- Dennett, D. C. (1998). *La actitud intencional*. G. Ventureira (trad.). Gedisa.
- Dennett, D. C. (2000). Making Tools for Thinking. En D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 17-29). OUP. <https://doi.org/10.1093/oso/9780195141146.003.0002>
- Dennett, D. C. (2009a). Darwin’s “Strange Inversion of Reasoning”. *PNAS*, 106(suppl. 1), 10061-10065. <https://doi.org/10.1073/pnas.0904433106>
- Dennett, D. C. (2009b). The Cultural Evolution of Words and Other Thinking Tools. *Cold Spring Harbor Symposia on Quantitative Biology*, 74, 435-441. <https://doi.org/10.1101/sqb.2009.74.008>
- Dennett, D. C. (2013). Foreword. En D. Ryder, J. Kingsbury y K. Williford (eds.), *Millikan and Her Critics* (pp. ix-xii). John Wiley & Sons.
- Dennett, D. C. (2014a, 2 de agosto). *Perspectives on Cultural Evolution*. International Cognition & Culture Institute. <https://cognitionandculture.net/webinars/sfi-cultural-evolution-workshop/perspectives-on-cultural-evolution-by-daniel-c-dennett/>
- Dennett, D. C. (2014b). The Evolution of Reasons. En B. Bashour, B. y H. D. Muller (eds.), *Contemporary Philosophical Naturalism and Its Implications* (pp. 47-62). Routledge.
- Dennett, D. C. (2017a). *From Bacteria to Bach and Back: The Evolution of Mind*. W. W. Norton & Co.

- Dennett, D. C. (2017b). *Haig's 'Strange Inversion of Reasoning'* [Manuscrito]. Center for Cognitive Studies, Tufts University.
- Dennett, D. C. (2018a). Reflections on Clatterbuck. *Teorema*, 37(3), 145-150.
- Dennett, D. C. (2018b). Reflections on Tadeusz Zadwidzki. En B. Huebner (ed.), *The Philosophy of Daniel Dennett* (pp. 57-61). OUP.
- Dennett, D. C. (2018c). Resumen de *From Bacteria to Bach and Back: The Evolution of Minds*. *Teorema*, 37(3), 103-106.
- Dennett, D. C. (2020). Comment on "Can Memes Explain the Birth of Comprehension?". *Avant*, 11(2), 1-4. <https://doi.org/10.26913/avant.2020.02.13>
- Dennett, D. C. (2023). *I've Been Thinking*. Norton & Co.
- Dennett, D. C. (2024). Commentary on Mark Richard, *Meanings as Species*. *Inquiry*, 67(2), 626-632. <https://doi.org/10.1080/0020174X.2022.2039852>
- Gibson, J. J. (2015). *The Ecological Approach of Visual Perception*. Psychology Press. <https://doi.org/10.4324/9781315740218>
- Godfrey-Smith, P. (2018). Between Towers and Trees in Cognitive Evolution. En B. Huebner (ed.), *The Philosophy of Daniel Dennett* (pp. 225-249). OUP. <https://doi.org/10.1093/oso/9780199367511.003.0015>
- Hauser, M. D., Chomsky, N. y Fitch, T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298(5598), 1569-1579. <https://doi.org/10.1126/science.298.5598.1569>
- Keiser, J. (2023). *Non-Ideal Foundations of Language*. Routledge. <https://doi.org/10.4324/9781003188537>
- Locke, J. (2005). *Ensayo sobre el entendimiento humano*. E. O'Gorman (trad.). FCE.
- Mesoudi, A. (2021). Blind and Incremental or Directed and Disruptive? On the Nature of Novel Variation in Human Cultural Evolution. *American Philosophical Quarterly*, 58(1), 7-20. <https://doi.org/10.2307/48600682>
- Miguens, S. (2022). Animal Brains and the Work of Words: Daniel Dennett on Natural Language and the Human Mind. *Topoi*, 41(3), 599-607. <https://doi.org/10.1007/s11245-021-09745-2>
- Morgan, C. L. (2007). *An Introduction to Comparative Psychology*. Kessinger Publishing.
- Planer, R. J. (2023). Going Dennettian about Gricean Communication. *Philosophical Psychology*, 1271-1294. <https://doi.org/10.1080/09515089.2023.2204887>
- Planer, R. J. (2024). Memetics and the Parallel Architecture. *Topics in Cognitive Science*, 17, 898-908. <https://doi.org/10.1111/tops.12735>

- Rathkopf, C. (2017). Mental Evolution: A Review of Daniel Dennett's *From Bacteria to Bach and Back*. *Biology & Philosophy*, 32, 1355-1368. <https://doi.org/10.1007/s10539-017-9586-y>
- Quine, W. O. (1968). *Palabra y objeto*. Manuel Sacristán (trad.). Editorial Labor.
- Scotto, C. (2022). A Pragmatics-First Approach to Faces. *Topoi*, 41(3), 641-657. <https://doi.org/10.1007/s11245-022-09821-1>
- Scotto, C. (2024). Signos liminares en el corazón de la comunicación: un enfoque interactivo. En I. Audisio y F. G. Rodríguez (eds.), *Pensamiento cuerpo a cuerpo. Intersubjetividad, cuerpo y lenguaje* (pp. 101-144). Ed. Universidad Nacional de Córdoba.
- Sellars, W. (1992). Philosophy and the Scientific Image of Man. En *Science, Perception and Reality*, Routledge & Kegan Paul.
- Sellars, W. (1997). *Empiricism and the Philosophy of Mind*. En R. Brandom (ed.), Harvard University Press.
- Schliesser, E. (2019). Synthetic Philosophy. *Biology & Philosophy*, 34. <https://doi.org/10.1007/s10539-019-9673-3>
- Sterelny K. (2006). Memes Revisited. *British Journal of Philosophy of Science*, 57, 145-165. <https://doi.org/10.1093/bjps/axi157>
- Zawidzki, T. W. (2007). *Dennett*. Oneworld Pub.
- Wittgenstein, L. (1997). *Observaciones sobre la filosofía de la psicología*. 2. L. F. Segura (trad.). UNAM.
- Wittgenstein, L. (2003). *Investigaciones filosóficas*. A. García Suárez y C. U. Moulines (trads.). UNAM.

<https://doi.org/10.21555/top.v760.3239>

Por el camino de Dennett: una genealogía de la moralidad

On Dennett's Path: A Genealogy of Morality

Antoni Gomila

Universitat de les Illes Balears
España

toni.gomila@uib.cat

<https://orcid.org/0000-0001-7325-961X>

Recibido: 03 - 09 - 2024.

Aceptado: 26 - 05 - 2025.

Publicado en línea: 20 - 05 - 2026.

Cómo citar este artículo: Gomila, A. (2026). Por el camino de Dennett: una genealogía de la moralidad. *Tópicos, Revista de Filosofía*, 76, 67-85. <https://doi.org/10.21555/top.v760.3239>



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

Resumen

En este artículo discuto los dos capítulos finales de *Darwin's Dangerous Idea* dedicados a la evolución de la moralidad y ofrezco una versión resumida de la genealogía de la moral que he desarrollado en trabajos recientes, como respuesta en parte a las limitaciones señaladas en la propuesta de Dennett. Respecto al capítulo 16 del libro, argumento que se puede ir más allá de su planteamiento y ofrecer una explicación evolutiva del origen y evolución de la moralidad. Respecto al capítulo 17, señalo la inespecificidad moral de su propuesta de un "manual de primeros auxilios morales", lo cual lleva a la necesaria caracterización del ámbito de la moralidad. En la tercera sección se argumenta que la genealogía de la moral que he propuesto permite avanzar en la línea pretendida por Dennet al ofrecer una explicación evolucionista de la moralidad. Esta propuesta no separa la evolución biológica y la evolución cultural como dos procesos alternativos, al modo de Dennet y de la memética, sino que los integra a través del proceso de desarrollo individual, por el cual cada nuevo individuo biológico humano se constituye culturalmente.

Palabras clave: moralidad; evolución cultural; Dennett; genealogía de la moral; memética; motivación moral; moralización; perspectiva de segunda persona.

Abstract

In this paper I discuss the final two chapters of *Darwin's Dangerous Idea* devoted to the evolution of morality and offer a summarized version of the genealogy of morality that I have developed in recent work, in part as a response to the shortcomings identified in Dennett's proposal. Regarding chapter 16 of Dennet's book, I argue that one can go beyond his approach and offer an evolutionary explanation of the origin and evolution of morality. Regarding chapter 17, I point out the moral non-specificity of his proposal for a "moral first aid manual", which leads to the necessary characterization of the scope of morality. In the third section it is argued that this genealogy of morality offers a way to fulfill Dennett's basic goal of an evolutionary account of morality. This proposal does not create a gap between biological evolution and cultural evolution as two alternative processes, as Dennett and memetics assumed, but rather integrates them through the process of individual development, by which each new human biological individual is culturally constituted.

Keywords: morality; cultural evolution; Dennett; genealogy of morals; memetics; moral motivation; moralization; second-person perspective.

1. Introducción¹

Coincidió con Dennett en diversas ocasiones. Recuerdo especialmente la primera, en Salamanca, en 1990, en el contexto de uno de los primeros seminarios de filosofía de la mente y ciencia cognitiva, que organizó Fernando Broncano, por la cercanía e interés que mostró con nuestro grupo de jóvenes balbuceantes en inglés. Dennett atendió a todas las presentaciones con interés y respondió a todas las cuestiones que se le plantearon, además de interactuar afable y amistosamente en las instancias informales del encuentro. Mi trabajo se titulaba “Interpretation in Biology and Psychology” (Gomila, 1991), una crítica del instrumentalismo de su “posición intencional” mediante la comparación con el modo en que se atribuyen funciones en biología. En ese momento Dennett estaba acabando la que considero su obra más brillante, *Consciousness Explained* (Dennett, 1992), pero siempre me gustó pensar que mi trabajo tuvo algo que ver con su interés subsiguiente en Darwin y el poder explicativo de la teoría de la selección natural, un movimiento de transición muy significativo en su trayectoria intelectual, iniciada en el marco de referencia del conductismo lógico de Ryle. En cualquier caso, considero a Dennett un modelo ejemplar de cómo hacer filosofía en estos tiempos, desde el naturalismo y la continuidad con la ciencia en diálogo con la cultura en general. Se trata de una filosofía “impura” y desacomplejada, comprometida con la elaboración de los conceptos necesarios para articular nuestra experiencia, y con la pretensión de influir en la elaboración de los conceptos científicos.

La celebración del vigesimoquinto aniversario de la publicación de *Darwin's Dangerous Idea* ofrece una oportunidad excelente para repensar la contribución de Dennett en ese libro pero especialmente para desarrollar su perspectiva. En este artículo quisiera centrarme en la cuestión de la evolución de nuestra psicología moral. Los capítulos finales (16 y 17) de *Darwin's Dangerous Idea* están dedicados a la moralidad y suponen la culminación del esfuerzo desarrollado por Dennett a lo largo de las más de quinientas páginas del libro. El capítulo 18 funciona

¹ Este trabajo se inserta en el proyecto “Autonomy as address”, financiado por el Gobierno de la Comunitat València dentro de la convocatoria “Prometeu” (CIPROM/2023/55). . Agradezco a los dos revisores de este trabajo para *Tópicos, Revista de Filosofía* sus comentarios y sugerencias para mejorar el texto.

como un epílogo en el que se resumen las ideas principales del libro y se anticipan las conclusiones ateas que desarrollará Dennet posteriormente (en relación con la religión, sobre todo). Aquí discutiré los contenidos de estos dos capítulos acerca de la moralidad y de la ética y ofreceré una versión resumida de la explicación evolutiva de la moralidad que, en diversas colaboraciones, he desarrollado en estos últimos años (Pérez y Gomila, 2021, pp. 141-154; Isern-Mas y Gomila, 2022a y 2022b), y que considero que Dennett hubiera visto con simpatía como un modo sólido de desarrollar el programa de una explicación naturalista de la moralidad, tal como él pretendía.

En realidad, lo que llama la atención de estos capítulos es la ausencia de una explicación evolutiva de la moralidad por parte de Dennett, a pesar de que el planteamiento inicial del capítulo 16 remite a tal objetivo. En realidad, ese capítulo consiste principalmente en un crítica a la sociobiología para sugerir, más que detallar, la idea de que nuestra psicología moral es fruto de la evolución cultural, que puede ir en contra de la evolución biológica. Y el capítulo 17 está basado en sus Tanner Lectures de 1988, donde articuló su noción de un “manual de primeros auxilios morales” (Dennett, 1988), propuesta que tiene que ver con la dimensión normativa de la ética. La “peligrosa idea” de Darwin, en relación con la moralidad, para Dennett, supone que la moralidad es el resultado de la evolución, pero de la evolución cultural, que consiste en un proceso mucho más rápido de explorar el espacio de diseño y en encontrar algún “buen truco” que funcione, que no requiere de la herencia genética, y que puede ir contra los intereses genéticos del individuo. Sin embargo, pienso que la caracterización de este “buen truco” deja mucho que desear, del mismo modo que los argumentos que le llevan a esa noción de “manual de primeros auxilios morales” tienen que ver con el razonamiento humano en general más que con el razonamiento y el juicio moral en particular. Ciertamente el tema de la moralidad no aparece en *From Bacteria to Bach and Back* (Dennett, 2017), que puede considerarse la última versión del pensamiento de Dennett sobre la evolución de la mente, lo que puede interpretarse como un signo de que no siguió trabajando sobre este tema.

Nuestro planteamiento, en cambio, parte de una caracterización del ámbito de la moralidad a partir del sentido del deber: de la motivación moral como elemento distintivo. En efecto, lo propio de la moralidad, a diferencia del razonamiento y el juicio, es que el juicio moral es motivador: impulsa a la acción. Si bien no garantiza el cumplimiento

del deber, implica una forma interior de penalización en tal caso, bien la vergüenza o la culpa, según los casos. Nuestra caracterización de la moralidad es por tanto estructural, no de contenido —dado que cualquier cuestión puede ser moralizada, cuestión sobre la que volveremos—. Y permite plantear la cuestión de su evolución de una forma muy clara y operativa. Por ello, creo que constituye un modo riguroso de dar cuenta de la moralidad humana de un modo naturalista, pero no reduccionista, tal como Dennett pretendía.

Así, comenzaremos por revisar el debate de la sociobiología y la posición de Dennett al respecto. En esa sección, mostraremos que su crítica a la sociobiología se combina con una actitud aprobatoria con respecto a la psicología evolucionista, que resulta inconsistente con su posición global sobre Darwin y la evolución cultural, pero que sirve para plantear la necesidad de una explicación evolutiva del origen y evolución de la moralidad que integre, en lugar de escindir, el ámbito biológico con el cultural. La segunda sección señalará la inespecificidad moral de la propuesta de Dennett de un “manual de primeros auxilios morales”, lo cual nos llevará a la necesaria caracterización del ámbito de la moralidad. La tercera sección intentará combinar los resultados de las dos secciones anteriores en una explicación evolucionista de la moralidad, tal como la caracterizamos, que no separa evolución biológica y evolución cultural como dos procesos alternativos, sino que los integra a través de la importancia del desarrollo individual, por el cual cada nuevo individuo biológico humano se constituye culturalmente. De este modo, nuestra propuesta puede considerarse como un modo de superar las limitaciones de lo que Dennett llegó a ofrecer en *Darwin's Dangerous Idea*.

2. “On the origin of morality”

El título del capítulo 16 de *Darwin's Dangerous Idea* plantea la cuestión del origen de la moralidad. Se trata de un capítulo extenso, de cuarenta páginas. A pesar del anuncio y de la extensión, Dennett no ofrece en ella una explicación evolutiva de la moralidad. El núcleo del capítulo es su crítica a la explicación evolutiva de la moralidad ofrecida por la sociobiología, por reduccionista, pero no se nos ofrece una explicación alternativa. Sin embargo, el tenor del capítulo parece sugerir que tal explicación alternativa, no reduccionista, es deseable y posible. Y se sugiere que tal explicación debería seguir la línea del planteamiento de la psicología evolucionista, pero con matices.

Antes de la crítica a la sociobiología, Dennett recoge los planteamientos de Hobbes y de Nietzsche. El capítulo abre con una extensa cita del *Leviathan* donde se plantea la analogía de la sociedad con el organismo humano. Para Dennett, Hobbes es el primero que se plantea explicar el origen de la moralidad y la sociedad. En el “estado de naturaleza” hobbesiano ya existen los seres humanos, pero se rigen por relaciones de fuerza y dominación. No existen aún los conceptos normativos: ni justicia, ni bien, ni maldad, ni ley. La vida humana, en tal situación, es solitaria, pobre, brutal y breve, regida por el miedo y la violencia. Hasta que aparece la civilización, la vida en sociedad, gracias a un contrato por el que los individuos, por su propio interés, renuncian a seguir sus deseos y aceptan una autoridad absoluta con poder para limitarlos. Esta autoridad absoluta permite la cooperación al establecer normas y disponer del poder para hacerlas respetar. Dennett menciona los nombres de Rousseau, Locke y Rawls como versiones posteriores, más sofisticadas, de esta “reconstrucción racional” que inaugura la explicación contractualista del origen de la sociedad, y distingue acertadamente el proyecto de una explicación del origen de la civilización del proyecto normativo de la justificación de normas de justicia, que Hobbes funde en uno solo —a diferencia de Rawls, por ejemplo, interesado únicamente en la cuestión normativa—. Pero a Dennett le interesa destacar del planteamiento de Hobbes su afinidad con el evolucionismo: Hobbes articula un punto de partida presocial, donde ya existen agentes capaces de actuar para promover su interés, y un punto de llegada social “evolutivamente estable”, un equilibrio evolutivo. Por eso lo considera un antecedente de su planteamiento adaptacionista, donde las soluciones que funcionan pueden ser descubiertas por la vía de la evolución cultural, si bien Dennett se cuida de añadir que los beneficios actuales de una novedad evolutiva de tal calibre no constituyen por sí mismos una explicación evolutiva. La cuestión radica en explicar el modo en que tales agentes autointeresados pasan a actuar con base en normas en primer lugar, y de qué modo se evitan los aprovechados, los tramposos. Este énfasis en la capacidad de decisión de los agentes humanos es lo que le lleva a rechazar los modelos basados en la teoría de juegos que pretenden ver la sociedad como un organismo compacto, al modo de Hobbes, por reduccionistas. Dennett enfatiza la capacidad de decisión y acción de los agentes humanos, que no está circunscrita por los intereses genéticos. Este aspecto va a

constituir su crítica principal a la sociobiología y su reivindicación de la evolución cultural y la memética.

En la segunda sección, Dennett comenta el proyecto de Nietzsche de una genealogía de la moral, ahora sí bajo la influencia, si bien indirecta, de Darwin. En realidad, sostiene Dennett, Nietzsche reacciona al darwinismo social de Spencer, tan simplista que pretende consagrar y legitimar como normas sociales las fuerzas que Hobbes atribuía al estado de naturaleza. El eje de la genealogía de la moral que propone Nietzsche, por el contrario, consiste en que se da cuenta de que para contratar, es decir, para que el contrato social pueda tener lugar, hacen falta agentes capaces de proyectarse hacia el futuro, de comprometerse, y esta capacidad también debe formar parte del *explanans*. Del mismo modo, los agentes involucrados en el contrato deben ser capaces de reconocer a los demás con quienes interactúan y de recordar interacciones pasadas: llevar la cuenta de a quien le debemos y quien nos debe, así como identificar a quien no cumple. Por ello, Nietzsche propone dos fases en el origen de la moral: en la primera habrían aparecido agentes con tales capacidades de memoria y acción, con una vida social primaria, aún premoral, pero (a diferencia de Hobbes) ya fuera del estado de naturaleza; en la segunda se habría producido la famosa “inversión de los valores” nietzscheana: lo que era bueno en esas formas de vida social premoral (regida por relaciones de dominancia basadas en la fuerza y la ley del más fuerte, que consagra una aristocracia de señores de la guerra) pasó a ser condenado y se promovió la cooperación y la confianza.

Estas secciones parece que van a anticipar la propuesta dennettiana del origen de la moral. En lugar de eso, lo que nos encontramos en las otras dos secciones del capítulo es una extensa crítica de la sociobiología por su reduccionismo. En estas secciones, Dennett critica la visión desarrollada en los capítulos anteriores sobre los memes y la evolución cultural como un proceso independiente y autónomo de la evolución biológica. La mayor parte de la crítica sigue el liderazgo de Philip Kitcher (1993). La sociobiología (Wilson, 1975 y 1978) fue un intento de explicar la conducta social humana mediante los mismos recursos teóricos que la conducta social de los insectos sociales, básicamente por el altruismo de parentesco (genético). Obviamente, tales recursos no permiten dar cuenta de normas morales o del sentimiento de obligación moral, ni de la estructura de la acción humana en general. Además, la sociobiología constituye un ejemplo indiscutible de falacia naturalista, al tratar de derivar el deber ser moral del ser biológico. En sentido positivo, Dennett

simplemente hace referencia a los trabajos de Trivers sobre la evolución de la inversión parental, la selección de parentesco y la reciprocidad, y a los de Axelrod (1984) sobre la estrategia “tit-for-tat” para explicar la evolución de la cooperación. Pero como señaló Darwin mismo en *Descent of Man* (Darwin, 1872a), la conducta de la madre que se sacrifica por una cría no puede considerarse aún dentro de la esfera del *moral sense*, puesto que tal conducta no está guiada por un juicio moral ni por sentido del deber. En este sentido, Darwin sigue resultando una referencia inexcusable para no confundir la evolución de la cooperación o del altruismo (presente en multitud de especies) con la evolución de la moralidad (característica exclusivamente humana). Darwin convierte la explicación de la evolución de la moralidad en la de la emergencia del *moral sense*, la capacidad de juzgar moralmente y sentirse motivado a actuar según tal juicio, siguiendo la idea tradicional de la moralidad humana en la versión específica de Hume.

Más allá, en la dirección de una explicación de la moralidad, Dennett se limita a considerar “what limitations we are born with and what variations there are among us that might have ethical relevance” (1995, p. 481). Pero el desarrollo de este planteamiento es escaso. La línea argumental asume la clásica dicotomía entre biología y cultura y se centra en señalar la dificultad central a la que se enfrenta una explicación biológica de alguna de estas capacidades: que siempre es posible una explicación alternativa en términos culturales —lo cual no excluye, insiste, que alguna explicación biológica pueda ser correcta—. Del mismo modo, de la universalidad de una conducta no puede inferirse que sea genética o instintiva. Pero a nivel propositivo nos encontramos únicamente con una opinión favorable a la psicología evolucionista (Barkow *et al.*, 1993; Pinker, 1997), que propone la selección de módulos de capacidades cognitivas específicas, a los que llaman “algoritmos darwinianos”. Desde esta perspectiva se propone como innata la capacidad de detectar tramposos, por ejemplo. Para este enfoque, la mente humana sería como una navaja suiza compuesta de mecanismos de propósito específico, seleccionados evolutivamente para dar una respuesta adaptativa a problemas de coordinación social.

¿Qué aspecto podría tener en este marco una genealogía de la moral? Si bien Dennett, como ya hemos dicho, no llega a abordar la cuestión, la respuesta debería ser que la psicología moral humana resultaría de una serie de «algoritmos darwinianos», un conjunto de módulos seleccionados por sus consecuencias positivas para afrontar

los problemas de coordinación que tuvieron que afrontar nuestros ancestros. En otros trabajos he argumentado contra un enfoque de este tipo (Gomila, 2014), si bien de modo general, básicamente porque pasa por alto el papel central del desarrollo infantil en la constitución de la mente humana. Dicho de otro modo: la ontogenia no recapitula la filogenia. En primer lugar, porque el contexto en que tuvo lugar la evolución de los homíninos no fue fijo e idéntico en el tiempo; de hecho, la innovación cultural fue modificando el nicho ecológico, y como el desarrollo cognitivo es sensible al contexto, es erróneo pensar que se dio en algún momento del pasado del género *Homo*, un fenotipo cognitivo único y universal. Además, dada la evolución del propio nicho ecológico humano por la intervención de nuestra especie —mediante el uso de herramientas, por ejemplo—, el punto de partida del desarrollo de cada generación va cambiando progresivamente. Esta modificación del proceso de desarrollo ontogenético humano ha sido llamada por Tomasello “efecto trinquete”, para referirse a su naturaleza acumulativa (Tomasello, 1999).

Sin embargo, una de las teorías más influyentes actualmente sobre la psicología moral humana, la de los “fundamentos de la moral” (Haidt y Joseph, 2007), adopta este planteamiento innatista y modularista propio de la psicología evolucionista. Según esta propuesta, venimos equipados de serie con cinco módulos, cada uno de los cuales sería la solución adaptativa a un problema que tuvieron que resolver nuestros ancestros: un módulo para el daño/cuidado, que permite proteger y cuidar a los parientes indefensos y vulnerables; otro relativo a la justicia y la reciprocidad, para permitir la cooperación con los no parientes; un tercero sobre la afinidad y lealtad con el propio grupo, que permitiría la cooperación grupal; un cuarto módulo sobre la autoridad y el respeto, que permite negociar las jerarquías sociales; y un quinto módulo sobre la pureza y la santidad, cuya función sería proteger de los microbios y parásitos. Pero estos módulos iniciales no generan por sí mismos juicios morales, sino que tienen que estructurarse en función de las normas sociales efectivas existentes a través de un proceso de aprendizaje. Y una vez estructurados, funcionan dando lugar a intuiciones basadas en respuestas emocionales, no a razonamientos. No es este el lugar para una discusión en profundidad de esta teoría (para una crítica detallada, cfr. Suhler y Churchland, 2011), pero es importante observar dos cosas: que la evidencia que aporta Haidt tiene que ver con la segunda parte, con el aprendizaje social y la dimensión intuitiva y emocional de los juicios

morales, mientras que para la primera —el innatismo y la modularidad— la evidencia sigue faltando. De hecho, la evidencia disponible apunta en el sentido contrario: que el juicio moral no es modular, sino que es sensible a cualquier aspecto de la situación, por menor o irrelevante que pueda parecer, y novedoso o hipotético (Christensen y Gomila, 2012). En definitiva, la teoría de los fundamentos de la moralidad no da cuenta del origen de las normas sociales, que deben existir ya para que puedan ser aprendidas, ni del sentimiento genuino de obligación moral, que la teoría no contempla. Lo que hace, en realidad, es tratar de simplificar la diversidad moral humana identificando cinco áreas temáticas y enfatizando que los juicios morales son intuitivos, no resultado de largos razonamientos. Pero en realidad puede haber normas morales sobre cualquier cosa, como demuestra el fenómeno de la moralización (Rozin, 1999): fumar se convirtió en moralmente reprochable cuando antes no lo era, la homosexualidad ha dejado de serlo, etc.

Creo que Dennett sería sensible a estas consideraciones, a pesar de la simpatía que manifestó por la psicología evolucionista al enfatizar la evolución cultural y la memética, y la plasticidad humana para desarrollar las conductas más diversas. Por nuestra parte, suscribimos su crítica a la sociobiología pero, como trataremos de mostrar, falta en Dennett una caracterización adecuada de lo que debe ser explicado evolutivamente: la aparición del *moral sense*, del juicio moral motivador de la conducta. Los contenidos de la moral van a ser naturalmente diversos, dependientes del contexto cultural e histórico, pero la cuestión básica es qué convierte las normas en morales en primer lugar. Pero esta caracterización estructural de la moralidad es compatible con una concepción de la evolución cultural como parte de la evolución biológica de nuestra especie: la estrategia adaptativa seguida por nuestro linaje fue precisamente la plasticidad fenotípica que resulta de una mente que necesita tiempo para estabilizarse y se constituye en interacción, con el nicho y con los demás. Y la clave para esta integración es el desarrollo ontogenético. La evolución biológica y la cultural no se contraponen como procesos diferentes e incompatibles, como asume Dennett, sino que se integran a través del proceso de desarrollo mental de cada individuo.

3. “Redesigning morality”

En el siguiente capítulo, Dennett trata de ofrecer una ética substantiva de inspiración evolucionista, que llama una “ética naturalizada” —en

alusión al “Epistemology Naturalized” de Quine (1969). En ese capítulo 17 trata de responder a la siguiente pregunta: “¿Qué implicaciones tiene para la ética el hecho de que somos buscadores heurísticos finitos, y con presión temporal, de verdades éticas?” (1995, p. 493). La respuesta a esta pregunta, no obstante, va a tener poco de específicamente moral, a pesar de que se plantee como una cuestión de ética normativa. El argumento que desarrolla Dennett, basado en sus Tanner Lectures (Dennett, 1988), que denomina “Manual de primeros auxilios morales”, consiste en explorar las consecuencias normativas de las limitaciones de la cognición humana en general.

Esto es así porque su adversario dialéctico inicial en este caso es el utilitarismo. El utilitarismo afirma que lo correcto es la alternativa que genera mayor felicidad o placer global. Esta posición se enfrenta al problema de calcular la utilidad, puesto que no es algo sencillo e inmediato. Dennett ilustra su argumento con una analogía. La máxima utilitarista “actúa de modo que maximices la felicidad, o el placer” es como la recomendación “compra barato y vende caro”: no sirve en la práctica para decidir qué comprar y cuándo. Dennett se basa en la crítica de Herbert Simon (1969) a los modelos abstractos que necesitan explorar todo el espacio del problema para poder establecer el razonamiento óptimo. Pero tales cálculos suponen tiempo, esfuerzo y capacidades ilimitadas, elementos todos ellos fuera de nuestro alcance. Como Simon, Dennett propone renunciar a la pretensión de encontrar la opción óptima y conformarse con la que sea satisfactoria. En la práctica, supone substituir la pretensión de disponer de algoritmos —procesos efectivos que garantizan alcanzar la solución óptima— y reconocer que contamos únicamente con heurísticos —procesos rápidos y frugales en términos de procesamiento de información—, que nos ofrecen soluciones satisfactorias. La cuestión subsiguiente de cómo esto afecta al propio proyecto normativo de la ética, no obstante, no se afronta en realidad.

En cualquier caso, Dennett va más allá del rechazo al utilitarismo. Según su razonamiento, lo mismo ocurre con el enfoque kantiano, que dispone actuar según máximas universalizables: ¿cómo decidir si una máxima es universalizable o no?, ¿cómo decidir qué máxima es la adecuada en un contexto determinado si el número de máximas universalizables podría ser infinito? El modo de disimular el problema sería mediante cláusulas *ceteris paribus* e idealizaciones semejantes. El problema de fondo, para ambas teorías éticas que conciben el razonamiento moral como un proceso de deducción normativa, tiene

que ver con el problema de la explosión computacional exponencial que generan.

La solución que propone Dennett, ese “manual de primeros auxilios morales”, sin embargo, se parece poco al modo en que se producen los juicios morales: por intuición, generalmente basada en una respuesta emocional, y como ya señalamos, siendo muy sensibles a los mínimos detalles de la situación. Dennett, en cambio, considera el problema del juicio moral como un problema de las limitaciones del diseño de nuestras mentes como resultado de su origen evolutivo, pero al hacerlo se desliza del plano normativo al psicológico, a cómo implementar nuestra capacidad de encontrar razones para nuestros juicios. Y se limita al ámbito del razonamiento prudencial, de la ponderación de razones a favor y en contra, para el que necesariamente vamos a disponer de un tiempo limitado e información insuficiente para alcanzar juicios garantizadamente correctos. En esa medida, sus consideraciones no son específicamente morales: Dennett aplica al ámbito de la moralidad restricciones propias de la cognición humana en general. De este modo, pasa por alto la especificidad del modo en que se produce el juicio moral —ese sentimiento de obligación que predispone a la acción, así como el modo en que el juicio refleja las normas morales que el agente tiene en consideración—. Estos dos aspectos son los que una genealogía de la moral debe ser capaz de explicar. Y son los que hemos tratado de aclarar en nuestra propuesta.

4. Una genealogía de la moral alternativa

Nuestra explicación de la evolución de la moralidad humana asume como elemento preliminar la perspectiva de segunda persona de la atribución psicológica: busca en la estructura de la interacción con los demás los aspectos que van a permitir la emergencia de la moralidad en el caso humano. El punto de partida, por tanto, es premoral y se caracteriza por la existencia de agentes capaces de atribuirse estados mentales como modo de organizar sus interacciones, incluyendo las reacciones emocionales. En el libro *Social Cognition and the Second Person in Human Interaction* (Pérez y Gomila, 2021) caracterizamos con detalle la naturaleza de este tipo de atribuciones, que explotan la dimensión expresiva y corporal de la intencionalidad, y en diversos trabajos hemos argumentado que esta forma de atribución está presente en animales no humanos (Pérez-Manrique y Gomila, 2022) y en niños pequeños (Barone y Gomila, 2020). El punto de partida de nuestra genealogía, por

tanto, consiste en una situación donde agentes no morales interactúan y reaccionan emocionalmente a sus interacciones a partir de atribuciones psicológicas básicas que toman como base la dimensión expresiva, corporal, de tales estados, de modo que pueden ser reconocidos directamente.

El *explanandum*, por su parte, como hemos visto en la discusión de Dennett, consiste en explicar el origen y la evolución de nuestra psicología moral, tanto de la motivación moral, del sentirse moralmente obligado, como de la sensibilidad a las demandas de los demás, en base a juicios que implican normas y valores. En otros términos, una explicación evolutiva que se pretenda satisfactoria debe dar cuenta de nuestra psicología moral completa, no de un aspecto particular —la empatía o la prosocialidad, por ejemplo—. Ciertamente puede resultar contenciosa la caracterización que proponemos, pero a su favor debe tenerse en cuenta que es un aspecto central de la comprensión de la moralidad (de Aristóteles a Kant, por lo menos). Ser moral consiste en ser sensible a las razones de orden moral y ello abarca tanto poder juzgar con base en normas morales, más allá de las propias inclinaciones y preferencias, y que tales juicios promuevan la disposición a actuar con base en tales juicios de un modo genuino, lo que se manifiesta en el resultado si no se actúa como se considera moralmente correcto: una sanción en forma de emoción, de culpa o vergüenza. Esta dimensión está ausente de la acción prudencial o estratégica, donde lo que prima es el interés propio. En este sentido, nuestra propuesta va más allá de las explicaciones que confunden la evolución de la moralidad con la evolución del altruismo. La conducta altruista, entendida como aquella en que un agente actúa contra su interés para beneficiar a otro, puede explicarse mediante la teoría de la selección de parentesco, o de la reciprocidad, directa e indirecta, pero no constituye un caso de conducta moral, tal como ya Darwin afirmó en *Descent of Man* en relación con el ejemplo de una hembra que se sacrifica para salvar a sus crías, tal como hemos señalado *supra*.

Por su parte, el *explanans* de una capacidad de esta complejidad no puede ser directo e inmediato: la explicación evolutiva de la aparición de esta estructura agentiva requiere de al menos dos fases. En la primera, explicamos cómo las reacciones emocionales que median nuestras interacciones llegan a depender de normas o valores propios del grupo. En la segunda, explicamos el sentimiento de obligación moral como resultado de la internalización de sanciones inicialmente externas,

sociales. Conjuntamente explicamos la estructura psicológica de la moralidad como universal y derivada de la manera en que interactuamos con los demás, y su contenido como resultado de la interiorización de las normas del grupo en donde se produce el desarrollo.

Con respecto a la primera fase, el núcleo de la propuesta parte de reconocer que nuestras reacciones emocionales constituyen elementos centrales para la regulación y la coordinación social. De nuevo, Darwin es referencia obligada en este punto: en *The Expression of Emotions in Animals and Man* (Darwin, 1872b) argumenta que la expresión de las emociones no es algo secundario y opcional. Al contrario: la función de las emociones es la coordinación y para ello tienen que ser expresadas, para poder ser reconocidas; correlativamente, cada especie es capaz de reconocer fácilmente esa expresión. Además, las emociones con que nos relacionamos con los demás implican un momento de valoración (generalmente denominado *appraisal* en la teorías de las emociones). No se trata aún de una valoración normativa, basada en una norma, pero puede llegar a convertirse en norma si es reconocida y aceptada por los demás, hasta el punto de generalizarse al grupo y convertirse en la base de las expectativas recíprocas de interacción. La idea que se propone, desde la perspectiva de segunda persona, es que las “actitudes reactivas” (Strawson, 1962) que median nuestras interacciones no tienen garantizada su validez inmediatamente. Podríamos recurrir al ejemplo de Darwall (2006) en su caracterización de esta dimensión intrínsecamente interpersonal de la moralidad: el de mi enfado si alguien me pisa; esta respuesta emocional, que implica una valoración de que pisar está mal, se plantea como una demanda de reconocimiento que debe ser validada por la otra parte (en este caso, el pisador). Su reacción puede ser distinta: podría rechazar mi demanda. Por ejemplo, si patino y me caigo, a un espectador cercano podría parecerle divertido y reírse. Esa risa no me haría ninguna gracia y puedo dirigir una demanda parecida hacia quien se burla de mí, sin que tenga ningún efecto, sin que sea reconocida.

¿De qué va a depender que sea reconocida y, por tanto, validada? Proponemos dos aspectos que contribuyen a esta transición: en primer lugar, el intercambio de posiciones entre los agentes que interactúan repetidamente; en segundo lugar, la perspectiva del observador (*bystander*), que puede intervenir para avalar la demanda o para denegar su validez. Cuando se considera la perspectiva de la dinámica de un grupo con esta estructura emergen como “atractores”, como puntos de equilibrio para la coordinación del grupo, ciertas valoraciones aceptadas

por el grupo (por supuesto, pueden emerger también expectativas relativas a diádas o subgrupos). Tales valoraciones equivalen a normas sociales entendidas como expectativas compartidas (Bicchieri, 2006 y 2016). Estas normas no son leyes, no son reglas explícitas, no forman parte de un código público. Son expectativas que regulan la conducta al permitir anticipar las reacciones validadas socialmente, basadas en formas de control social implícitas. No es preciso dedicar mucho esfuerzo a argumentar la mejora de la aptitud biológica de un grupo de este estilo: mayor coordinación, mayores oportunidades de supervivencia.²

En la segunda fase, estas normas sociales se convierten en morales al dejar de ser prudenciales y condicionales respecto de las circunstancias del contexto e internalizarse; con ello, obtienen su poder motivador a través del sentimiento de obligación. Dicho de otro modo, caracterizamos el ámbito moral por su fuerza motivacional incondicional, frente a las normas sociales que son contextuales y por tanto condicionales. Es este carácter dual de tales normas, que se experimentan como externas y objetivas tanto como subjetivas (Stanford, 2018), lo que caracteriza el dominio de la moral —no su contenido, que puede ser variado y diverso—.

¿Cómo pudo tener lugar esta transición? Según nuestra propuesta, el sentimiento de obligación moral emergió al producirse una sutil transformación de un cierto tipo de emociones, aquellas que suponen un *appraisal* del propio agente, bien por uno mismo o por el agente con el que se actúa. Estas emociones —orgullo, autoestima, envidia, agradecimiento, culpa, vergüenza, resentimiento...— se denominan a veces “autoconscientes” porque su dimensión valorativa se refiere al propio agente. Como ya observara Darwin (1872b), estas emociones se elicitán por lo que uno piensa que los demás piensan de uno. En esa medida, la valoración tiene también una dimensión de recompensa

² Hemos ilustrado esta dinámica con el caso de la confianza, que es una actitud compleja que encierra componentes cognitivos y emocionales, y que ilustra muy claramente la dimensión valorativa implícita en la confianza cuando alguien la defrauda. Lo interesante es darse cuenta de que las relaciones de confianza entre agentes individuales gracias a una experiencia previa positiva acaban repercutiendo en el nivel de confianza de un individuo hacia un desconocido, de modo que se incrementa muy significativamente el grado de cooperación dentro del grupo en general. Cfr. Acedo-Carmona y Gomila (2014 y 2019).

o castigo interiorizado: el agradecimiento que percibo me resulta gratificante; el resentimiento, en cambio, me hace sentir culpable, que es una forma de malestar. Además, tales emociones tienen una dimensión motivacional: tu resentimiento por lo que te hice (o omití hacer), me motiva a tratar de reparar la relación (Sznycer *et al.*, 2016).

Es importante en este punto mostrar la ventaja evolutiva de un grupo social con normas morales además de las sociales, frente a un grupo únicamente con normas sociales. Es decir, normas cuya fuerza normativa se ha interiorizado a través del sentimiento de obligación moral, que lleva aparejada también una sanción emocional en forma de vergüenza o culpa en caso de transgresión. La respuesta consiste en reconocer la función de compromiso que supone sentir una obligación (Frank, 1988), que previene cursos de acción miopes. En grupos sociales pequeños, la cooperación puede sostenerse gracias a vínculos afectivos (del tipo de la empatía), pero en grupos más grandes, donde la cooperación depende de la coordinación entre individuos no emparentados, un mecanismo que permitiera un grado de compromiso semejante resultó adaptativo. Y, recíprocamente, disponer de un sistema para extender el cumplimiento normativo incondicional permitió a su vez grupos más y más grandes.

En resumen, la propuesta parte de agentes capaces de interactuar emocionalmente, con un repertorio emocional limitado y estructurado motivacionalmente. Explica, por un lado, la emergencia de normas sociales y, por otro lado, la transformación de un cierto tipo de emociones especializadas en las interacciones intersubjetivas. El resultado de ambos cambios es la emergencia de normas morales como un subconjunto de las normas sociales, aquellas que se vuelven incondicionales al asociarse a un sentimiento de obligación. Este planteamiento permite dar cuenta, además, de la diversidad de normas morales y de los cambios que pueden experimentar, a veces en el curso de una generación, proceso que se denomina “moralización” (Rozin, 1999). Rozin se centra en la moralización del consumo de tabaco, en el sentido de constatar cómo lo que inicialmente se planteaba en términos de conveniencia pasó a plantearse en términos morales. Pero cambios semejantes se observan en procesos de toma de decisión política, con la satanización de adversarios (Bernstein y Gomila, 2024).

A modo de resumen, cabría decir que tomamos de Dennett el reto de ofrecer una explicación evolutiva de la moralidad humana, y que hemos tratado de resolverlo superando las limitaciones que señalamos en los dos capítulos finales de *Darwin's Dangerous Idea*: una caracterización

insuficiente de nuestra psicología moral, y un mayor énfasis en la evolución cultural como proceso desconectado de la evolución biológica. Hemos argumentado que el núcleo de nuestra psicología moral no es ningún contenido, ninguna norma acerca de algún ámbito específico de la vida humana, sino la estructura motivacional específica de los juicios morales. Esta estructura psicológica es la que heredamos evolutivamente y que en el curso de la ontogénesis se llena de contenido, de las normas especificadas de cada tiempo y lugar, de cada sociedad humana. Compartimos el rechazo de Dennett de las teorías morales puramente normativistas, que conciben el juicio moral como un proceso algorítmico, pero su alternativa en términos de heurísticas pierde de vista la especificidad moral.

A pesar de todo, y como conclusión, me gustaría insistir en la idea de que Dennett nos proporcionó un modelo de cómo hacer una filosofía naturalista que resultara científica y culturalmente relevante. Y en relación con Darwin y la teoría de la evolución por selección natural, contribuyó a reivindicar la importancia de la perspectiva evolutiva para las ciencias sociales. Y lo hizo siempre con ingenio, con una habilidad especial para el uso de analogías, y con un estilo ameno y entretenido, alejado de la aridez del análisis conceptual. Un camino a seguir.

Referencias

- Acedo-Carmona, C. y Gomila, A. (2014). Personal Trust Increases Cooperation beyond General Trust. *PLoS One*, 9(8), e105559. <https://doi.org/10.1371/journal.pone.0105559>
- Acedo-Carmona, C. y Gomila, A. (2019). Personal Trust Extends Cooperation beyond Trustees: A Mexican Study. *International Journal of Psychology*, 54(5), 687-704. <https://doi.org/10.1002/ijop.12500>
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.
- Barkow, J., Tooby, J. y Cosmides, L. (1993). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press. <https://doi.org/10.1093/oso/9780195060232.001.0001>
- Barone P. y Gomila A. (2020). Infants' Performance in the Indirect False Belief Tasks: A Second-Person Interpretation. *WIRE Cognitive Science*, 12(3), e1551. <https://doi.org/10.1002/wcs.1551>
- Bernstein, A. y Gomila, A. (2024). The Truth in Social Media. *Topoi*, 44, 127-138. <https://doi.org/10.1007/s11245-024-10039-6>

- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616037>
- Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190622046.001.0001>
- Christensen, J. F. y Gomila, A. (2012). Moral Dilemmas in Cognitive Neuroscience: A Principled Review. *Neuroscience and Biobehavioral Reviews*, 36, 1249-1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008>
- Darwall, S. (2006). *The Second-Person Standpoint*. Harvard University Press.
- Darwin, C. (1872a). *The Descent of Man, and Selection in Relation to Sex*. John Murray.
- Darwin, C. (1872b). *The Expression of Emotions in Man and Animals*. John Murray.
- Dennett, D. C. (1988). The Moral First Aid Manual. En S. M. McMurrin (ed.), *Tanner Lectures on Human Values*. VIII (pp. 120-147). University of Utah Press.
- Dennett, D. C. (1992). *Consciousness Explained*. Little, Brown & Co.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster.
- Dennett, D. C. (2017). *From Bacteria to Bach and Back*. Penguin.
- Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. Norton.
- Gomila, A. (1991). Interpretation in Biology and Psychology. *Epistemología*, 14, 275-292.
- Gomila, A. (2014). Sobre la relación entre ontogenia y filogenia: una crítica de la Psicología Evolucionista. En S. Español (ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 261-283). Paidós.
- Haidt, J. y Joseph, C. (2007). The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. En P. Carruthers, S. Laurence y S. Stich (eds.), *The Innate Mind*. 3 (pp. 367-392). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195332834.003.0019>
- Isern-Mas, C. y Gomila, A. (2022a). A Second-Personal Approach to the Evolution of Moral Motivation and the Feeling of Obligation. *Biological Theory*, 17(3), 199-209. <https://doi.org/10.1007/s13752-022-00397-y>

- Isern-Mas, C. y Gomila, A. (2022b). Love, Friendship, and Moral Motivation. *Journal of Theoretical and Philosophical Psychology*, 42(2), 93-107. <https://doi.org/10.1037/teo0000166>
- Kitcher, P. (1993). The Evolution of Human Altruism. *Journal of Philosophy*, 90, 497-516. <https://doi.org/10.2307/2941024>
- Pérez, D. y Gomila, A. (2021). *Social Cognition and the Second Person in Human Interaction*. Routledge. <https://doi.org/10.4324/9781003133155>
- Pérez-Manrique, A. y Gomila, A. (2022). Emotional Contagion in Non-Human Animals: A Review. *WIRE Cognitive Science*, 13(1), e1560. <https://doi.org/10.1002/wcs.1560>
- Pinker, S. (1997). *How the Mind Works*. Norton.
- Quine, W. V. O. (1969). Epistemology Naturalized. En W. V. O. Quine, *Ontological Relativity and Other Essays* (pp. 69-90). Columbia University Press. <https://doi.org/10.7312/quin92204-004>
- Rozin, P. (1999). The Process of Moralization. *Psychological Science*, 10(3), 218-221. <https://doi.org/10.1111/1467-9280.00139>
- Simon, H. (1969). *The Sciences of the Artificial*. MIT Press.
- Stanford, P. K. (2018). The Difference between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation. *Behavioral and Brain Sciences*, 41, e95. <https://doi.org/10.1017/S0140525X17001911>
- Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 1-25.
- Suhler, C. y Churchland, P. (2011). Can Innate, Modular “Foundations” Explain Morality? Challenges for Haidt’s Moral Foundations Theory. *Journal of Cognitive Neuroscience*, 23(9), 2103-2116. <https://doi.org/10.1162/jocn.2011.21637>
- Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S. y Halperin, E. (2016). Shame Closely Tracks the Threat of Devaluation by Others, Even Across Cultures. *Proceedings of the National Academy of Sciences*, 113(10), 2625-2630. <https://doi.org/10.1073/pnas.1514699113>
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press. <https://doi.org/10.4159/9780674044371>
- Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Harvard University Press.
- Wilson, E. O. (1978). *On Human Nature*. Harvard University Press.

<https://doi.org/10.21555/top.v760.3379>

¿Hay diseño en la naturaleza? Daniel Dennett y el problema de la teleología: cuestiones epistemológicas y didácticas

Is There Design in Nature? Daniel Dennett and the Problem of Teleology: Epistemological and Didactic Issues

Leonardo González Galli
Universidad de Buenos Aires
Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas
Argentina
leomgalli@gmail.com
<https://orcid.org/0000-0002-0713-164X>

Recibido: 02 - 02 - 2025.

Aceptado: 05 - 04 - 2025.

Publicado en línea: 01 - 06 - 2026.

Cómo citar este artículo: González Galli, L. (2026). ¿Hay diseño en la naturaleza? Daniel Dennett y el problema de la teleología: cuestiones epistemológicas y didácticas. *Tópicos, Revista de Filosofía*, 76, 87-118. <https://doi.org/10.21555/top.v760.3379>



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

Resumen

En este artículo se discuten algunos análisis (asociados a las nociones de “postura intencional”, “ingeniería reversa” y “diseño”) de Daniel Dennett sobre el problema de la teleología en la biología. A partir de esa discusión se propone un posible modo en que dichas nociones se articulan en una perspectiva general sobre el rol y legitimidad de la teleología en la biología. Por último, se discuten algunas implicancias didácticas de dichos análisis en relación con el problema de la teleología en la enseñanza de la biología evolutiva. Más específicamente, se argumenta que los análisis de Dennett constituyen un adecuado fundamento epistemológico para un enfoque didáctico centrado en fomentar una “vigilancia metacognitiva” de los estudiantes sobre sus intuiciones teleológicas.

Palabras clave: Daniel Dennett; teleología; diseño; postura intencional; postura de diseño; evolución; enseñanza; metáforas; obstáculos epistemológicos; metacognición.

Abstract

This paper discusses some of Daniel Dennett’s analyses (associated with the notions of “intentional stance,” “reverse engineering,” and “design”) of the problem of teleology in biology. On the basis of this discussion, a possible way in which these notions can be articulated within a general perspective on the role and legitimacy of teleology in biology is proposed. Finally, some didactic implications of such analyses are discussed in relation to the problem of teleology in the teaching of evolutionary biology. More specifically, it is argued that Dennett’s analyses constitute an adequate epistemological foundation for a didactic approach focused on fostering students’ “metacognitive vigilance” over their teleological intuitions.

Keywords: Daniel Dennett; teleology; design; intentional stance; design stance; evolution; teaching; metaphor; epistemological obstacles; metacognition.

1. Introducción

Aunque, por la agenda de temas que abordó, Daniel Dennett no podría considerarse un típico filósofo de la biología, sino más bien de las ciencias cognitivas, ha producido algunos de los más lúcidos análisis filosóficos del darwinismo, especialmente en su libro *La peligrosa idea de Darwin* (Dennett, 1995). Fue un pensador audaz y original cuyos análisis arrojaron luz y promovieron productivos debates en diversas áreas de la filosofía, las ciencias biológicas y las ciencias cognitivas. En este trabajo reseñaré algunos de sus análisis vinculados especialmente con el problema de la teleología en la biología (Allen *et al.*, 1998; Gayon *et al.*, 2023), esto es, el debate sobre la naturaleza y legitimidad de las nociones y explicaciones teleológicas, especialmente en el ámbito de la biología. Como discutiré más adelante, este problema epistemológico tiene un correlato educacional, a saber, la cuestión de cómo lidiar en la enseñanza de las ciencias naturales, especialmente de la biología, con las intuiciones teleológicas de los estudiantes (González Galli *et al.*, 2020; Kelemen, 2012).

El principal objetivo de este artículo es reseñar los diferentes análisis de Dennett vinculados con el problema de la teleología, elaborar una síntesis integradora de ellos, y argumentar que dichos análisis no solo constituyen un valioso aporte a la filosofía de la biología sino también un fundamento epistemológico especialmente adecuado para repensar el problema didáctico mencionado. La obra de Dennett es extensa, y para este trabajo me basaré especialmente en tres libros: *La actitud intencional* (Dennett, 1991), *La peligrosa idea de Darwin* (Dennett, 1995), y *Bombas de intuición y otras herramientas de pensamiento* (Dennett, 2015).

Dado que se trata de un trabajo de carácter filosófico y didáctico (teórico), la metodología empleada es principalmente conceptual y reconstructiva. Consiste en la revisión, selección y recontextualización de marcos conceptuales provenientes de la filosofía de la biología y de la didáctica de las ciencias, con el objetivo de identificar puntos de convergencia entre ambos campos que permitan repensar un problema específico de la enseñanza. En particular, he realizado una reconstrucción interpretativa del enfoque de Daniel Dennett sobre la teleología en la biología; destaco aquellos aspectos que considero relevantes para el ámbito educativo. Posteriormente, he vinculado ese marco filosófico con resultados y análisis provenientes de investigaciones en didáctica a fin

de establecer nexos argumentativos (que se especifican en el desarrollo del texto) entre las perspectivas filosóficas y los desafíos pedagógicos. La metodología utilizada se basa en el análisis crítico de textos y en la articulación argumentada de enfoques teóricos desde una lógica interna, con un criterio de coherencia conceptual y relevancia educativa.

2. Cuestiones epistemológicas

En rigor, Dennett no abordó el problema de la teleología como tal, como sí lo han hecho numerosos filósofos de la ciencia (cfr. Hempel, 1959; Nagel, 1961; Wright, 1976) y de la biología (cfr. Diéguez, 2012; Godfrey-Smith, 2022; Ruse, 2023; Sober, 2018) sino que, más bien, se aproximó al tema desde sus intentos por analizar cómo la teoría de la evolución o, más precisamente, la mirada darwinista del mundo orgánico (la “peligrosa idea de Darwin”, el “enfoque adaptacionista”)¹ contribuye a comprender el origen y naturaleza de la mente humana. A continuación, reseñaré tres ideas clave de Dennett, estrechamente relacionadas, sobre este asunto: lo que definió como la “postura intencional”,² su interpretación de la biología como ingeniería (y la idea de “ingeniería reversa”), y su defensa de una noción “fuerte” de diseño biológico.

2.1 La postura intencional

Dennett (1991, pp. 26 y ss.; 1995, pp. 371 y ss.; 2009; 2015, pp. 65 y ss.) parte de señalar que existen diferentes modos de entender e intentar predecir el comportamiento de cierta entidad bajo estudio. Una primera estrategia consiste en adoptar la “postura física” (PF), de acuerdo con la cual se trata de caracterizar detalladamente el estado del sistema analizado para determinar las “condiciones iniciales” y luego aplicar las leyes de la física para predecir el comportamiento del sistema en cuestión. Sin embargo, no es difícil imaginar que en muchas ocasiones

¹ El término “adaptacionismo” tiene varias acepciones (cfr. Lewens, 2007), pero aquí se refiere al enfoque en biología evolucionista que metodológicamente se centra en la teoría de la selección natural para basar la investigación en la hipótesis de que los rasgos analizados son adaptativos, es decir, son producto de la selección natural. En este sentido, no supone asumir que de hecho todos los rasgos son adaptativos ni que la adaptación sea perfecta.

² Dennett utiliza el vocablo inglés “stance”, que en algunas ocasiones es traducido como “actitud” y en otras como “postura”. Aquí utilizaré el segundo término.

esta estrategia no será viable, por ejemplo, porque no disponemos de toda la información pertinente para describir con el suficiente detalle el estado inicial del sistema (intente el lector anticipar la conducta de su gato o de su computadora personal aplicando esta estrategia). Así, aunque en principio esta postura podría adoptarse para cualquier caso, en muchas ocasiones no sería lo más práctico, o sería directamente imposible hacerlo. Frente a estas limitaciones, podemos intentar un segundo enfoque adoptando la “postura de diseño” (PD), que consiste en suponer que el sistema está diseñado para cierto fin y que se comportará tal como está diseñado para hacerlo. Este enfoque puede funcionar muy bien cuando analizamos objetos de diseño, artefactos tales como un despertador o la mencionada computadora personal. La gran ventaja de esta estrategia es que no necesitamos conocer los detalles físicos. Sin embargo, como contraparte, es una estrategia más arriesgada (Dennett, 2015, p. 66) que la “postura física”, ya que descansa en dos supuestos que podrían no cumplirse: que el sistema está diseñado para cierto fin (podría estar diseñado para otro fin que ignoramos o para ninguno) y que se comportará según dicho diseño (podría estar descompuesto). Queda aún una tercera estrategia, que es una variante de la anterior, y que consiste en adoptar la “postura intencional” (PI) (Dennett, 2015) de acuerdo con la cual “la cosa diseñada se trata como un tipo de agente, con creencias y deseos, y suficiente racionalidad para hacer lo que debería dadas esas creencias y deseos” (p. 66).

Dennett argumenta que, en la práctica, en ciertas ocasiones nos veremos obligados a adoptar la PI. Como ejemplo, discute cómo predecir qué hará una computadora que juega ajedrez. Para predecir la próxima jugada, uno podría adoptar la PF y calcular el flujo de electrones que resultaría de pulsar cierta tecla, pero de más está decir que hacer eso sería imposible en la práctica. Uno podría entonces recurrir a la PD, lo que implicaría obtener el código fuente de la computadora y simular “a mano” los millones de pequeños pasos que la máquina dará en su búsqueda de la mejor jugada posible. Nuevamente, esto sería inviable en la práctica. En este caso, entonces, uno estaría obligado a recurrir a la PI, lo que supondría (1) hacer una lista de todas las jugadas legales posibles; (2) clasificar las posibles jugadas, de la mejor (la que se condice con el deseo o intención de ganar la partida) a la peor (la que conspira contra la intención de ganar), y, finalmente, (3) predecir que la computadora hará la mejor jugada disponible (Dennett, 2015, pp. 66-67).

La principal pregunta para los fines de este trabajo es a qué entidades es conveniente o legítimo aplicar cada una de las tres posturas. Empezaré por caracterizar las opiniones más ortodoxas y conservadoras en relación con este asunto, opiniones que, como veremos, Dennett no comparte.

Tal como mencionamos, la PF es en principio aplicable a cualquier entidad, ya que todas son sistemas físicos sometidos a las leyes físicas, pero en muchos casos se trata de una opción inviable en la práctica. Con respecto a las otras dos posturas la cuestión es más compleja, y en ambos casos podemos distinguir instancias de aplicación no problemáticas y problemáticas. Comenzando por los casos no problemáticos: en principio, la aplicación no problemática de la PD sería aquella que tiene por objeto los artefactos de origen humano (objetos producto del diseño humano deliberado).³ Estos artefactos exhibirían, además, una intencionalidad derivada o secundaria, pues la intencionalidad intrínseca o primaria es una cualidad exclusiva del diseñador humano —esta distinción es sostenida por muchos autores, por ejemplo, Searle (1983)—. Por su parte, la PI solo sería aplicable de modo no problemático a organismos con agencialidad o, simplemente, agentes. Por “agentes”⁴ en este contexto entendemos entidades intencionales y racionales, capaces de definir fines y ordenar su conducta racionalmente para tender a la consecución de esos fines. Dada esta definición, el único caso no problemático de aplicación serían los sujetos humanos, únicos que, como ya mencioné, dispondrían de intencionalidad primaria o intrínseca.

Pasemos ahora a los casos de aplicación problemáticos. En relación con la PD, ¿qué diríamos sobre los artefactos construidos (¿diseñados?) por organismos no humanos?⁵ Es muy conocido el hecho de que muchos animales no humanos utilizan herramientas. Algunos incluso no se limitan a utilizar objetos tal como los encuentran, sino que modifican dichos objetos (de un modo aparentemente intencional) de forma que

³ O a todas las entidades existentes si uno es un creacionista, pero en este trabajo me restringiré a análisis compatibles con la perspectiva científica.

⁴ Hay diferentes conceptos de “agencia”, y no todos incluyen el componente intencional aquí mencionado; cfr., por ejemplo, Schlosser (2019) y Sultan *et al.* (2022).

⁵ La idea de “artefacto construido por no humanos” es de por sí problemática, dado que, por lo general, se asume que, por definición, los artefactos son creaciones humanas. Por ejemplo, en un reciente glosario de filosofía de la técnica se define “artefacto” como “*entidades producidas por humanos que poseen una función práctica en la cultura material*” (Crelier, 2022, p. 62).

los transforman en herramientas (Sanz *et al.*, 2013; Shumaker *et al.*, 2011). Sin embargo, tal vez al lector le llame la atención que hable de artefactos construidos por organismos no humanos. Más allá de que no es clara la distinción entre “herramienta” y “artefacto”, la segunda idea se asocia intuitivamente a una mayor complejidad (partes que interactúan, mecanismos de regulación por retroalimentación, etc.). Pero, ¿construyen algunos animales no humanos auténticos artefactos? En principio, tenderíamos a pensar que una ramita modificada por un chimpancé para “pescar” termitas podría considerarse una herramienta, pero no un artefacto. Pero pensemos en los nidos construidos por algunas hormigas y termitas. Estos nidos tienen una estructura (¿diseño?) compleja que, por ejemplo, incluye túneles y aberturas dispuestos de modos muy precisos que contribuyen a mantener los valores de humedad y temperatura finamente regulados (Yang *et al.*, 2022). Pensemos también en numerosos dispositivos con partes móviles que habitualmente calificamos de “trampas”, desde aquellas construidas con seda por arañas hasta las complejas flores de ciertas orquídeas cuyos movimientos aseguran que el insecto que las visita buscando néctar quede cargado con el polen de la flor (en algunos casos, previa captura y confinamiento temporal del insecto) (cfr., por ejemplo, Gerlach, 2011). En tanto artefactos (de factura no humana), estos dispositivos solo exhibirían una intencionalidad derivada o secundaria. ¿Podemos aplicar la PD en estos casos? El problema con estos ejemplos es que —a diferencia de lo que sucede en el caso de los artefactos humanos— no disponemos inequívocamente de una intencionalidad intrínseca a la cual remitirnos. Desde ya, podríamos atribuir intencionalidad al chimpancé, incluso podríamos hacer lo mismo con las hormigas, aunque en este último caso esa atribución sería muy polémica, pero difícilmente nos atrevamos a atribuir intencionalidad a una orquídea —o a los genes que guían (pero no determinan) su desarrollo—. Aunque en la actualidad se tiende a considerar que la conciencia y la intencionalidad no son rasgos exclusivamente humanos, y que no pueden definirse fácilmente en términos de todo o nada, siempre es polémica su atribución a animales no humanos (Godfrey-Smith, 2017 y 2021). Por otro lado, ¿por qué excluir de la categoría de “artefacto” a los órganos que son parte de un organismo pluricelular o, incluso, a las complejísimas organelas de una célula eucariota (pensemos en el flagelo bacteriano, uno de los ejemplos favoritos de presunta “complejidad irreductible” de los partidarios

de la “teoría del diseño inteligente”)?⁶ Como vemos, la aplicación de la PD a artefactos no humanos resulta extremadamente compleja y problemática.

Con respecto a la PI tenemos el problema de si puede aplicarse, como intencionalidad primaria o intrínseca, a organismos no humanos y, como intencionalidad secundaria o derivada, a artefactos contruidos por organismos no humanos. Con respecto a lo primero, es muy probable que, como ya señalé, algunos animales no humanos presenten algún grado y tipo de conciencia e intencionalidad, pero siempre es problemática la atribución de intenciones a organismos no humanos (básicamente, porque, por el momento, es entre difícil e imposible abordar empíricamente la cuestión) y, más aún, la eventual determinación de *cuáles* serían las intenciones de esos animales. Podemos aceptar que los chimpancés tienen algún tipo y grado de conciencia e intencionalidad, pero eso no es lo mismo que demostrar que cuando modifican una ramita que luego utilizan para “pescar” termitas tenían la intención de construir una herramienta. Estos son solo algunos de los arduos problemas que complican el campo de la etología cognitiva (Proust y Vauclair, 2003). Con respecto a lo segundo, sería problemático atribuir intencionalidad derivada a la “caña” que el chimpancé utiliza para “pescar” termitas por el hecho ya mencionado de que resulta problemático atribuir intencionalidad intrínseca al chimpancé.

Estas consideraciones sobre aplicaciones no problemáticas y problemáticas de las PD y PI suponen, como ya advertí, un enfoque muy conservador del asunto. En este sentido, y ahí reside en parte lo innovador de su enfoque, Dennett argumentaría que sería legítimo y sobre todo extremadamente práctico y, más aún, obligado analizar a los organismos, a sus partes y a los artefactos por ellos contruidos desde PD y PI, aun en caso de que no pudiéramos resolver los problemas

⁶ Aclaremos, en este punto, que la denominada “teoría del diseño inteligente” no es más que una perspectiva religiosa que intenta presentarse a sí misma, espuriamente, como científica (Kampourakis, 2013). En tanto perspectiva religiosa, no tendrá ningún lugar en los análisis que aquí ofrezco. También es importante dejar en claro que la eventual aceptación de una noción amplia de “diseño” que incluya los rasgos adaptativos de los organismos (tal es la propuesta de Dennett que discutiré más adelante) no supone ningún apoyo a la “teoría del diseño inteligente”: por el contrario, la idea de Dennett es que el gran mérito de Darwin es haber mostrado que puede haber diseño *sin* diseñador.

conceptuales planteados en los párrafos previos. Por ejemplo, en el caso de los artefactos (sean de origen humano o no), cabe suponer que se comportarán como si tuvieran creencias, deseos y racionalidad. Por ejemplo, un termostato se comportará (al menos mientras no se descomponga) como si deseara mantener la temperatura constante y como si comprendiera qué debe hacer para cumplir con tal fin. Del mismo modo, la flor de la orquídea se comportará como si quisiera atrapar al insecto polinizador y cargarlo con polen y el corazón como si quisiera asegurar la circulación sanguínea.

El “como si” en estos enunciados sugiere que entendemos que, estrictamente hablando, el termostato, la flor o el corazón no desean ni comprenden nada: por eso hablamos de una intencionalidad derivada o de segundo orden, esto es, que solo está presente porque una entidad auténticamente intencional (el diseñador humano) la puso en él. Pero esta distinción nos lleva de nuevo a los problemas anteriores: como, salvo para los casos no problemáticos, no tenemos certeza de si estamos o no en presencia de una intencionalidad intrínseca, entonces tampoco tendremos certeza sobre en qué casos atribuir intencionalidad derivada. Dennett (1991, pp. 254-285; 2015, pp. 115-148) cuestiona estas distinciones, para él lo más importante es si las diferentes posturas funcionan, en el sentido de permitirnos predecir con cierta confianza la conducta del sistema analizado. En este sentido, no es muy relevante poder distinguir si un organismo (o una parte de un organismo, o un artefacto) tiene o no intencionalidad intrínseca o derivada. Así, por ejemplo, podemos predecir que una planta expandirá su sistema radicular frente a una sequía, como si quisiera evitar la deshidratación y supiera qué debe hacer para alcanzar tal fin. Y podemos hacer esta predicción sin necesidad alguna de distinguir si la planta tiene o no alguna forma de intencionalidad, sea esta intrínseca o derivada. Dennett (1991, p. 37) sostiene que la PI no solo es útil cuando queremos predecir el comportamiento de un sistema biológico, sino que se trata de un recurso *indispensable* que, en rigor, nunca podríamos reemplazar por la PF.

La razón por la que la PD y la PI funcionan en el caso de los organismos (y sus partes y artefactos) es que estos se comportan como si fueran agentes porque son producto de la selección natural. Estos sistemas biológicos pueden no ser racionales en sí mismos, pero se comportan como si lo fueran porque encarnan las razones, “razones

de flotación libre” (*free-floating rationales*),⁷ que la selección descubrió durante el proceso evolutivo. La PI también es imprescindible cuando queremos reconstruir la historia evolutiva de estos sistemas, ya que — al menos desde el enfoque adaptacionista adoptado y defendido por Dennett— en tal caso nos veremos obligados a identificar las razones que la selección descubrió y que explican por qué cierta variante de un rasgo prevaleció sobre sus alternativas (Dennett, 1995, pp. 232-233; 2015, pp. 171-172).

En síntesis, para Dennett, las PD y PI son útiles y, de hecho, indispensables⁸ para analizar los sistemas biológicos, sean estos organismos completos o partes de ellos. Además, eliminando la distinción entre lo intrínseco o auténtico y lo derivado, tanto para el caso de la intencionalidad como para el diseño, Dennett (2015, p. 64) sostiene que es intencional todo sistema cuya conducta se puede predecir con éxito adoptando la PI. Esta última consideración es importante porque, cuando analicemos las implicancias educativas de estas discusiones, habilitará a aquellos que no están dispuestos a abandonar esas distinciones a adoptar una postura instrumentalista según la cual,

⁷ Dennett (2015, pp. 171-172) utiliza esta curiosa expresión para referirse a las razones que “descubre” la selección natural, es decir, las razones por las que cierta variante fue seleccionada, sin que nadie se representara esas razones. Por ejemplo, la razón por la cual el color blanco fue seleccionado en las poblaciones ancestrales de zorros árticos es que dicho color confiere un buen camuflaje a los zorros, lo que les permite cazar mejor y evitar mejor a sus depredadores, todo lo cual resulta en un mayor *fitness*. Esa es la razón por la cual se seleccionó el color blanco, pero se trata de una razón que ni los zorros ni la selección ni la naturaleza, ni nadie, se representó hasta que alguien esbozó una explicación darwiniana de ese rasgo: se trata de una “razón de flotación libre”.

⁸ Entiendo que, para Dennett, esta necesidad se da tanto en el plano psicológico como epistemológico. En relación con lo primero, simplemente no podemos concebir hipótesis adaptacionistas sin recurrir a la postura intencional y de diseño. Por lo tanto, en la medida en que al menos algunos rasgos orgánicos sean producto de la selección, la comprensión de los seres vivos demandará el recurso a dichas posturas. En relación con lo segundo, no es claro cómo podría exigirse en el plano epistemológico un enfoque que esté más allá de las capacidades cognitivas humanas. Pero, por otro lado, Dennett no ve ningún problema epistemológico en relación con adoptar estas posturas. Se trata, para Dennett, de una estrategia claramente fructífera, y rechaza los cuestionamientos que estos enfoques han recibido desde el positivismo lógico, por lo que, simplemente, no ve razones para renunciar a ellos.

en biología, se trata a los sistemas biológicos *como si* fueran sistemas intencionales con independencia de si consideramos que realmente lo sean. Adelantemos en este punto que el propio Dennett (1991, pp. 72 y ss.) no dice tener una perspectiva epistemológica instrumentalista sino más bien realista moderada.

2.2 La biología como ingeniería

Desde un enfoque adaptacionista, Dennett argumenta que en nuestro intento de explicar el origen de los rasgos orgánicos nos vemos obligados a analizar al organismo y sus partes como si fueran artefactos, es decir, como si fueran objetos de diseño. La expresión “ingeniería reversa” (Dennett también la llama “hermenéutica artefactual”) se refiere a la estrategia utilizada por los ingenieros de una empresa tecnológica cuando adquieren un nuevo producto lanzado por una empresa competidora y lo desarman y examinan, preguntándose por qué fue construido de ese modo, asumiendo siempre que deben existir buenas razones por las cuales el fabricante competidor lo hizo de ese modo y no de otro (cfr. Dennett, 1995, pp. 299 y ss.). La PI es imprescindible para la ingeniería reversa (cfr. Dennett, 1995, pp. 376-377), ya que el análisis del artefacto analizado requiere una guía conceptual que permita inferir las razones de su diseño, y para tal fin conviene asumir que ese objeto se comportará como un sistema intencional.

En la ingeniería “normal”, el proceso de concepción y construcción de artefactos se mueve desde las razones y fines hacia el diseño, mientras que en la ingeniería reversa se parte del objeto de diseño para inferir las razones y fines del diseñador. Vemos claramente aquí una analogía que atraviesa todos los análisis de Dennett sobre estos asuntos: la analogía entre el proceso de diseño deliberado (intencional y racional), el proceso de “investigación y desarrollo” (*research and development*) o ingeniería y sus productos, los artefactos o diseños (en tanto producto), por un lado, y la selección natural o adaptación (como proceso) y sus productos, los rasgos adaptativos y los organismos, por el otro.⁹

Comencé este apartado aclarando que la propuesta de Dennett se relaciona con el llamado enfoque adaptacionista. Esto se debe a que la mencionada analogía entre organismos (y sus partes) y artefactos,

⁹ Para un análisis crítico de la analogía entre organismos y artefactos ver Lewens (2000, 2002 y 2004).

así como entre los procesos que les dan origen (diseño deliberado y selección natural, respectivamente), solo tiene sentido y está justificada cuando asumimos que los rasgos analizados son producto de la selección natural y, por lo tanto, que son adaptativos. Conviene aclarar aquí que esto no supone afirmar que todos los rasgos son adaptativos, idea que cualquier persona mínimamente instruida en biología evolucionista considerará un error. Dennett no cree eso, pero sí ha defendido con vehemencia (especialmente frente a las críticas de autores como Gould y Lewontin, 1979) la legitimidad y utilidad del enfoque adaptacionista, esto es, la perspectiva de investigación en Biología que prioriza, como guía para la indagación, aquellas hipótesis que suponen que los rasgos analizados son adaptativos (Dennett, 1991, pp. 211-253; 1995, pp. 371-428). Conviene insistir en que dicha hipótesis se adopta con fines metodológicos: la investigación dentro del paradigma adaptacionista podría eventualmente arrojar como resultado que, de hecho, cierto rasgo *no es adaptativo*.

2.3 Auténtico diseño biológico

Con frecuencia se habla en Biología del “diseño” de tal o cual rasgo. Sin embargo, en general se asume que dichas referencias a tan problemático concepto son la expresión de una metáfora o analogía (Reynolds, 2022), a la que se recurre para facilitar el discurso, y que, hablando con propiedad, no existe tal cosa como el diseño biológico (cfr., por ejemplo, Mahner y Bunge, 2000). Lo mismo puede decirse de todas las referencias a otras nociones con connotaciones teleológicas, incluyendo la propia idea de “selección”. Dennett evita esa solución fácil pero ilusoria en relación con el uso de estas nociones en biología. Digo “ilusoria” porque decir que las expresiones y conceptos teleológicos en biología son la mera expresión de una metáfora con fines comunicativos deja sin explicar varias cuestiones, entre ellas por qué, de hecho, se recurre a esa metáfora con mucha mayor frecuencia en biología que en otras ciencias (Ruse, 2023). Esa respuesta tampoco nos dice nada sobre los roles de esa metáfora (¿son solo comunicativos o también heurísticos o explicativos?) ni sobre por qué no se la ha abandonado, dado su carácter epistemológicamente problemático. Frente a esta postura habitual, que sintoniza con el mandato de la filosofía de la ciencia clásica (positivismo lógico y concepción heredada) de evitar toda referencia a las causas finales en las explicaciones científicas (Allen *et al.*, 1998), Dennett, sin ser explícito sobre cómo respondería las preguntas que recién planteé,

afirma que en el mundo orgánico hay propósitos y diseño, no “como si” sino propiamente hablando:

La biosfera está completamente saturada de diseño, de propósito, de razones. Lo que llamo la postura de diseño predice y explica características presentes a lo largo del mundo viviente mediante las mismas suposiciones que funcionan tan bien cuando aplicamos ingeniería inversa a artefactos hechos por diseñadores humanos (un poco) inteligentes. La evolución por selección natural es un conjunto de procesos que “encuentran” y “rastrear” razones para que las cosas se configuren de una manera más que de otra (Dennett, 2015, p. 163, énfasis de Dennett).

Por supuesto, hay diferencias entre los procesos mediante los cuales los humanos encontramos razones y aquellos que operan en la evolución. La principal de esas diferencias —como se adelantó *supra* en una nota al pie— es que, en el caso del diseño humano, las razones son primero representadas en las mentes de los diseñadores, cosa que no sucede en la evolución.¹⁰ Las razones encontradas por la selección son representadas por primera vez (si es que alguna vez llegan a serlo) cuando algún biólogo o bióloga hace ingeniería reversa sobre algún rasgo adaptativo y de ese modo descubre y se representa las razones (de “flotación libre”) que la selección descubrió antes.

Intentaré a continuación sintetizar los distintos conceptos reseñados en los párrafos previos, relacionándolos con el problema de la teleología. Admito que esta síntesis es de mi cosecha, y que el propio Dennett no la ha expresado con estos términos.

Si nos preguntamos por qué en la biología persisten términos y nociones teleológicas, la respuesta que podemos inferir de los análisis de Dennett sería que la selección natural constituye una teoría central en

¹⁰ Es interesante señalar que, a pesar de esta diferencia, y contra lo que suele creerse, el proceso cultural de diseño tecnológico no es totalmente intencional. Por el contrario, hay aspectos del cambio cultural (incluido el tecnológico) que obedecen a “causas descentralizadas”, procesos que no están bajo control consciente e intencional de los individuos y que, incluso, pueden ser parcialmente análogos al proceso de selección natural (Hanisch y Eirdosh, 2023; Micheli *et al.*, 2019).

biología, que explica específicamente el carácter adaptativo de muchos rasgos biológicos, y que dicha teoría propone que las adaptaciones¹¹ son consecuencia de un proceso natural fuertemente análogo al proceso de diseño deliberado. Es esa analogía lo que está en la base de todo el enfoque de Dennett. Así, para Dennett, el proceso de I y D (investigación y desarrollo) humano, el proceso de diseño “verdadero”, y la selección natural, son procesos análogos. Y también son necesariamente análogos los productos de dichos procesos: los artefactos, los objetos diseñados, por un lado, y las adaptaciones biológicas y —en gran medida— los organismos completos, por el otro. Como consecuencia de esta analogía, para comprender e investigar el mundo orgánico, y especialmente los rasgos adaptativos, nos vemos obligados a adoptar la PD y, muchas veces, su variante, la PI. En el fondo, lo que hace que todo esto sea adecuado, legítimo y, de hecho, inevitable es la fuerte analogía entre ambos procesos. Dennett señala, además, y provocativamente, que los límites entre lo “auténtico” o “intrínseco” (sea la intencionalidad, sea el diseño) y lo “como sí”, lo aparente, lo secundario, lo derivado, son difusos y, en definitiva, arbitrarios o ilusorios, por lo que acaba por sugerir que mejor haríamos en reconocer que el diseño biológico no es diseño “como si” sino diseño sin más, y que todo sistema cuya conducta se pueda predecir con la PI es, de hecho, un auténtico sistema intencional. Así, concluimos que hay dos procesos naturales capaces de producir diseño: la actividad tecnológica humana y la selección natural.

¹¹ Recordemos aquí que “adaptación” tiene muchos significados. En la actual biología evolucionista, “adaptación” tiene principalmente dos significados técnicos: como sustantivo significa “proceso de evolución por selección natural” (“cierta población de plantas se adaptó a la escasez de agua reduciendo la superficie foliar”); como adjetivo significa “rasgo producto de la selección natural” o, lo que es lo mismo, “rasgo que incrementa el *fitness* de sus poseedores” (“la versión de las hojas de menor superficie es una adaptación”) (Futuyma y Kirkpatrick, 2018). Cuando decimos que la teoría de la selección natural explica la adaptación no nos referimos a estos significados, ya que, en caso de hacer eso, incurriríamos en la famosa tautología tantas veces denunciada. Cuando decimos que la teoría de la selección natural explica la adaptación estamos haciendo referencia a otro significado de “adaptación”: la adecuación de los organismos a sus condiciones de vida. Ese es el fenómeno que Darwin quería explicar y que los naturalistas que lo precedieron explicaron apelando a Dios (Ginnobili, 2010). Actualmente existe un amplio consenso en que la teoría de la selección natural explica satisfactoriamente el fenómeno en cuestión.

En relación con el concepto de “diseño”, dice Dennett:

Algo tiene que concederse. O bien definimos “diseño” como el producto de un Diseñador Inteligente, como el diseñador de poemas o coches, por ejemplo, o reconocemos que puede haber diseño —genuino diseño— sin un Diseñador Inteligente. La tradición y la etimología parecerían favorecer la primera opción [...]. Sostengo que la ciencia (también) ha descubierto que el diseño sin diseñador (sin un diseñador con una mente, previsión e intención) no sólo es posible, sino que está en todas partes. El diseño por evolución es un proceso real y bien entendido. Difiere de manera interesante del diseño por ingeniero, pero también es sumamente parecido y puede darle una lección de humildad al “ingenio” (2015, p. 165).

Así, diseño, finalidad e intencionalidad son rasgos de los sistemas biológicos en virtud de que estos son (en gran medida) producto de la selección. En lo que sigue exploraré algunas implicancias didácticas de estos análisis.

3. Cuestiones didácticas

3.1. El problema de la teleología en la enseñanza de la biología

Comenzaré esta sección caracterizando lo que denominaré “el problema de la teleología en la enseñanza de la biología”. Este problema es consecuencia de que las personas tenemos una tendencia a producir explicaciones intuitivas de los fenómenos biológicos basadas en nociones tales como “fin”, “meta”, “propósito”, “necesidad”, “función” y otras relacionadas (Coley y Tanner, 2012; Kelemen, 2012). Lo que estas nociones tienen en común es un supuesto teleológico o finalista general. Así, las personas tendemos a creer que en la naturaleza todo existe y sucede para alcanzar ciertos fines predeterminados (sobrevivir, reproducirse, etc.). Probablemente, el origen de este sesgo teleológico obedezca a una combinación de factores biológicos y culturales. A favor de la hipótesis de que este sesgo tiene alguna base biológica se aduce que su presencia se evidencia en todas las culturas y que se expresa en los primeros meses de vida (Atran *et al.*, 2001; Kelemen, 2012; Schachner

et al., 2017). Por supuesto, factores culturales tales como la cosmovisión religiosa (en la que todo tiene una razón de ser derivada de las intenciones divinas) o la cultura tecno-capitalista (que nos acostumbra a pensar en términos utilitarios o instrumentales) probablemente refuercen el sesgo teleológico (Rottmann *et al.*, 2017). Como fuere, innumerables estudios muestran que estudiantes de países muy diversos exhiben este sesgo a la hora de razonar sobre los organismos (Brock, *et al.*, 2024; González Galli y Meinardi, 2015; González Galli *et al.*, 2020; Kampourakis, 2020).

¿Por qué esto constituye un problema didáctico? Por dos razones. En primer lugar, porque durante mucho tiempo prevaleció en el ámbito educativo el supuesto de que toda forma de teleología es científicamente ilegítima. Este supuesto epistemológico proviene de la filosofía de la ciencia clásica (positivismo lógico y concepción heredada), que cuestionó todo recurso a las causas finales por considerarlo metafísicamente sospechoso (Allen *et al.*, 1998). Al adoptar semejante supuesto epistemológico, inevitablemente, el sesgo teleológico de los estudiantes se convirtió en un problema central para la enseñanza de las ciencias, en especial de la biología. En segundo lugar, porque el sesgo teleológico hace que las personas tiendan a producir explicaciones intuitivas sobre los fenómenos biológicos que con frecuencia resultan contradictorias con los modelos científicos (González Galli *et al.*, 2025). Así, por ejemplo, cuando buscamos que los estudiantes produzcan explicaciones de tipo mecanicista sobre cómo y por qué el páncreas libera insulina tras la ingesta, muchos considerarán que la mera observación de que el páncreas hace eso *para* regular la concentración de la glucosa en la sangre, o de que el organismo *necesita* de esa acción, basta para explicar ese hecho. En el caso de la enseñanza de la teoría de la evolución por selección natural, la idea de que las mutaciones se producen *para* adaptarse al ambiente, o en respuesta a ciertas *necesidades* impuestas por ese ambiente, es muy común. También son comunes las explicaciones intuitivas basadas en el supuesto de que el proceso evolutivo como un todo está orientado en cierta dirección predeterminada (por ejemplo, tendiendo al origen de los humanos).

Este tipo de explicaciones teleológicas resultan ser una “comodidad de la mente” para los estudiantes (en realidad, para las personas en general), es decir, es el tipo de explicaciones que tienden a producir

espontáneamente, de un modo rápido y fácil.¹² Además, como dijimos, se trata de modos de pensar que son altamente resistentes al cambio mediante la instrucción.

3.2 Dos perspectivas contrapuestas frente al problema didáctico

Frente al problema caracterizado en la sección anterior, muchos autores (y profesores) han asumido que la teleología es científicamente ilegítima y que, por lo tanto, el principal objetivo educativo sería que, mediante un “cambio conceptual”,¹³ los estudiantes abandonaran el pensamiento teleológico, o bien, que sean capaces de inhibirlo totalmente, al menos siempre que la situación demande una explicación científica. Denominaré esta perspectiva “eliminacionista”, considerando que su objetivo último es eliminar la teleología de la enseñanza de la biología. En otros textos hemos argumentado *in extenso* que dicho enfoque se basa en dos supuestos cuestionables (González Galli, 2019). El primero, epistemológico y ya mencionado, es aquel según el cual la teleología no cumple ningún rol legítimo en la biología. El segundo, didáctico-psicológico, supone que la mencionada eliminación-inhibición es posible y deseable.

Contra el primer supuesto, numerosos autores (Caponi, 2020 y 2023; Lennox, 1993; Ruse, 2023; Short, 2002, entre otros) han argumentado, desde diferentes marcos teóricos y estrategias analíticas, que, de hecho, ciertas formas de teleología son una parte intrínseca, legítima y no eliminable de la biología y, más específicamente, de la teoría de la evolución por selección natural. Sugiero que los análisis de Dennett glosados en las secciones previas son parte de ese universo de enfoques que convergen en una naturalización de la teleología en la biología (Sober, 2018). Contra el segundo supuesto, los análisis de psicología cognitiva (Kelemen, 2012) sugieren que el sesgo teleológico es una parte intrínseca, funcional y no eliminable de la “biología intuitiva” de las personas, por lo que su eliminación o inhibición total no sería posible ni deseable.

¹² Dada esta caracterización, podríamos decir que este sesgo teleológico es parte del “sistema 1”, en la terminología del modelo de la mente dual de Kahneman (2019; cfr. también Frankish, 2010 y González Galli *et al.*, 2025).

¹³ Para una discusión de las limitaciones de este modelo de aprendizaje, cfr. Krajcik y Shin (2023).

En relación con la no eliminabilidad del sesgo teleológico intuitivo, alguien podría plantear que aun siendo dicho sesgo un aspecto constitutivo de la biología intuitiva, podría ser deseable y posible eliminar todo sesgo teleológico de la biología científica. De hecho —podría agregarse en la misma línea argumentativa—, uno de los aspectos del avance científico consiste en su progresivo alejamiento de los modos intuitivos de comprender el mundo. Sin embargo, contra la pretensión de eliminar la teleología de la biología científica cabe decir que quienes han sostenido dicha postura no han tenido ningún éxito. Por un lado, los filósofos (ceranos al positivismo lógico y la concepción heredada, como Carl Hempel y Larry Wright) que han hecho intentos formales de operar tal eliminación, especialmente en relación con la noción de “función”, han fracasado en un doble sentido: dichos análisis han sido cuestionados en el propio ámbito de la filosofía de la biología y, además, no han tenido ningún impacto en las prácticas discursivas y analíticas de los biólogos reales. Por otro lado, otros autores —entre ellos Dennett, como mostramos— han argumentado que no hay ninguna necesidad de tal eliminación. Recordemos que el objetivo de eliminar la teleología de la biología es deudor de ciertos supuestos epistemológicos impuestos por la primera filosofía de la ciencia que tomaba la física clásica como ciencia modelo. Entre esos supuestos se encuentra el ideal nomológico deductivo de explicación y el supuesto de que toda noción de “causa final” tiene necesariamente connotaciones metafísicas incompatibles con la ciencia (intencionalidad, entidades sobrenaturales). Rechazados estos supuestos epistemológicos (como hacen Dennett, Ruse, Caponi y otros), desaparece la necesidad de eliminar la teleología de la biología científica.

A partir de estas críticas a la perspectiva eliminacionista, y basándonos en el concepto didáctico de “obstáculo epistemológico” (Camilloni, 2001; González Galli *et al.*, 2022) y en el de “metacognición”¹⁴ (Pérez y González Galli, 2020; McGuire, 2023), hemos elaborado un enfoque alternativo en el que el principal objetivo es fomentar en los estudiantes el desarrollo de una “vigilancia metacognitiva” sobre el sesgo

¹⁴ Dado que en una nota anterior he mencionado que el sesgo teleológico puede considerarse parte del “sistema 1”, resulta interesante señalar aquí que la relación entre la metacognición y los modelos de la mente dual es compleja y controvertida, sin que haya consensos al respecto (cfr. Ackerman y Thompson, 2017; Arango Muñoz, 2011; Houdé, 2019).

teleológico (González Galli *et al.*, 2020). Dicha vigilancia supone conocer en qué consiste el pensamiento teleológico, ser capaz de reconocer sus múltiples expresiones (desde las más explícitas y evidentes hasta las más veladas), y ser capaz de regular el uso de esa forma de pensar de acuerdo con el contexto y los objetivos de la tarea cognitiva emprendida.

Como todo objetivo metacognitivo, el desarrollo de esa vigilancia constituye un desafío de gran magnitud (Pérez y González Galli, 2020). Entre otras complicaciones, tender a dicho objetivo supone ofrecer a los estudiantes algunos criterios o “ideas clave” (metaconocimiento) que les resulten útiles para decidir si cierta forma específica de teleología resulta o no aceptable en cada contexto (Cupo *et al.*, 2024; Hartelt y Martens, 2024). Por ejemplo, los estudiantes deberían ser capaces de comprender que la afirmación de que “las bacterias mutan para resistir el antibiótico” es teleológica y científicamente errónea por ser inconsistente con los modelos científicos aceptados (las mutaciones son aleatorias), mientras que la afirmación de que “el pelaje tupido de los zorros árticos fue seleccionado porque sirve para protegerse del frío” también es teleológica, pero es científicamente correcta por ser consistente con los modelos científicos. Como vemos, y tal como ha señalado Kampourakis (2020), no es la teleología *per se* el problema. Además, enseñar biología desde este enfoque implicaría enseñar a los estudiantes sobre la naturaleza de la teleología en la biología. Esto supone trabajar en las clases cuestiones tales como qué es la teleología, cómo aparece en diversas teorías de la biología, por qué biólogos y biólogas recurren a razonamientos y expresiones teleológicas, y demás cuestiones metateóricas.

En otros trabajos (González Galli, 2025) hemos argumentado que la biología evolutiva (como, tal vez, toda la ciencia) implica modos de ver el mundo fuertemente contraintuitivos. Paradójicamente, los análisis previos implican que la teleología, en sentido general, es un aspecto común a la biología intuitiva y la científica. Sin embargo, de allí no se sigue que, después de todo, la biología científica no sea contraintuitiva: lo es. Esto se debe a que la biología intuitiva implica una forma de “teleología promiscua” (Kelemen, 2012) no regulada, mientras que la biología científica implica ciertas formas específicas de teleología cuya correcta aplicación requiere regular finamente las intuiciones teleológicas. Para decirlo de otro modo, las intuiciones teleológicas del profano persisten en el experto, pero en este último caso dichas intuiciones se ven redescritas y reinterpretadas en el contexto de

un nuevo marco teórico (la teoría de la selección natural) (Pozo, 2014). En síntesis, la biología (más específicamente la biología evolucionista darwiniana) es contraintuitiva pero no por ser teleológica, sino por implicar una forma más restringida y regulada de teleología.

Sugiero que análisis como los de Dennett (para un análisis análogo en relación con la propuesta de Ruse, cfr. González Galli, 2025) pueden ser de utilidad en relación con los propósitos didácticos enunciados. En este sentido, podríamos, por ejemplo, analizar con los y las estudiantes en qué sentido los biólogos hacen “ingeniería reversa”. Por supuesto, algunas ideas de Dennett podrían resultar excesivamente abstractas y confusas para los estudiantes; tal sería el caso, tal vez, de la noción de “razones de flotación libre”. A continuación, exploraré con más detalle en qué podría consistir un enfoque de enseñanza de la biología basado en los análisis de Dennett.

3.3 Una versión dennettiana de la perspectiva metacognitiva

Retomando los análisis de Dennett, sugeriré aquí que las ideas de que hay diseño auténtico en la naturaleza y de que, de acuerdo con lo que sabemos al día de hoy, existen dos procesos que lo generan (el diseño deliberado humano y la selección natural), podrían constituir uno de los fundamentos conceptuales de una enseñanza de la biología que asuma el reconocimiento de la centralidad y legitimidad de la teleología en la biología y que, al mismo tiempo, contribuya a mejorar el aprendizaje en relación con el problema de la teleología.

Tal como mencioné antes, es común que los biólogos hablen de diseño. Consideremos los siguientes ejemplos, todos tomados de reconocidos libros de texto de nivel universitario: “Los coágulos no son estructuras permanentes. Al contrario, están diseñados para disolverse cuando se restaura la integridad estructural de las zonas dañadas” (Stryer, 1995, p. 257; mi traducción), y: “Los ribosomas eucarióticos y procarióticos son muy similares en diseño y función” (Alberts *et al.*, 1994, p. 231, mi traducción). Tomemos nota en este punto de que, con independencia de qué tan correctas o convenientes consideremos este tipo de expresiones, es un hecho que los estudiantes las encontrarán con frecuencia, no solo en el discurso de divulgación sino incluso, como vemos, en textos técnicos.

La constatación del uso de estas expresiones teleológicas ha dado lugar a un complejo debate entre quienes consideran que se trata solo de un recurso retórico (eventualmente eliminable) y quienes, como

Dennett y Ruse, consideran que se trata más bien de la expresión de un rasgo epistemológico de la biología (no eliminable). No puedo ni es mi objetivo caracterizar aquí ese debate. Solo señalaré que, entre otros problemas, la primera postura no logra dar cuenta de por qué esas expresiones persisten en la biología, incluso en textos técnicos dirigidos a estudiantes universitarios como lo citados.

Una postura intermedia consiste en reconocer que los razonamientos y expresiones teleológicos no cumplen solo un rol comunicativo sino también heurístico (en tal caso, aún serían, al menos en teoría, eliminables). La idea es que pensar sobre los seres vivos *como si* (esta expresión supone que, en realidad, estamos tratando con una metáfora o analogía) fueran objetos de diseño potencia la exploración científica del mundo orgánico al sugerir preguntas e hipótesis que, de otro modo, no vendrían fácilmente a la mente. Esta postura puede sostenerse (al menos asumiendo ciertos supuestos sobre la naturaleza de las teorías científicas) aun negando que en su forma más técnica la teoría implique alguna forma de teleología, y es en ese sentido que la considero intermedia entre la perspectiva que afirma que la teleología *es parte de* la teoría y aquella que solo reconoce un rol comunicativo para ese modo de razonar.

Volviendo al problema didáctico, el punto es que aun adoptando esta perspectiva intermedia se impone la necesidad de enseñar a los estudiantes cómo lidiar con las ambigüedades derivadas de ese uso de las nociones teleológicas: ¿cómo deberían los estudiantes entender expresiones que hablan de diseño al tiempo que se afirma que, en realidad, no hay diseño? Un enfoque posible se basa en proponer a los estudiantes pensar que se trata de una metáfora o analogía (Ruse, 2023). Al adoptar este enfoque (González Galli, 2016) podemos beneficiarnos de la vasta literatura existente sobre cómo enseñar ciencias con metáforas y analogías (Clement, 1993; Else *et al.*, 2008; González Galli, 2023; González Labra, 1997; Linares e Izquierdo Aymerich, 2006; Oliva-Martínez y Aragón-Méndez, 2009).

Las analogías y metáforas (no discutiré aquí la distinción, poco clara y sobre la que no hay consenso, entre ambas nociones) implican poner en relación dos campos semánticos, dos entidades —en principio diferentes— con sus respectivos universos de significados asociados (Hills, 2024; Reynolds, 2022). Muchas veces, especialmente cuando estos recursos se utilizan con fines didácticos, se distingue un “campo fuente” y un “campo blanco” (McComas, 2020). El primero sería una

entidad o sistema que nos resulta familiar, mientras que el segundo sería una entidad menos conocida y cuya comprensión buscamos facilitar recurriendo a la analogía. En la literatura sobre enseñanza con metáforas y analogías, una de las principales sugerencias que se encuentra consiste en proponer a los estudiantes analizar explícitamente los “alcances y limitaciones” de la analogía utilizada (Galagovsky y Adúriz-Bravo, 2001). Esto supone identificar aquellos componentes del “campo fuente” que tienen un elemento correspondiente en el “campo blanco” (los “alcances” de la analogía), así como aquellos componentes para los cuales no existe tal correlato (las “limitaciones” de la analogía). Se asume en general que este trabajo analítico ayuda a comprender en qué consisten ambos campos, especialmente el “blanco” que, en el contexto de la enseñanza de las ciencias naturales, será un modelo científico. Pero, más específicamente, este tipo de análisis previene que los estudiantes establezcan correspondencias espurias, extrapolaciones incorrectas, entre ambos campos. Si, por el contrario, utilizamos las analogías de un modo más bien espontáneo e improvisado (como usualmente hacemos en el discurso cotidiano, pero con frecuencia también en contextos educativos), sin detenernos a realizar estos análisis detallados y explícitos, el riesgo de que se establezcan tales relaciones indeseables es mayor.

Así, si decidiéramos introducir en las clases de evolución la analogía entre el diseño deliberado humano y la selección natural, deberíamos, entre otros criterios didácticos, diseñar actividades para que los estudiantes, con la guía del docente, analicen cuidadosamente los “alcances” y “limitaciones” de la analogía en cuestión (González Galli, 2016). Y el análisis de Dennett nos da las pistas principales para tal análisis. El rasgo más importante compartido por el proceso de selección y el proceso de diseño deliberado es que ambos producen objetos de diseño. Por otro lado, la principal diferencia es que solo en un caso (el diseño deliberado) hay intencionalidad y representación mental del objeto de diseño.

Si no aceptáramos la propuesta de Dennett de que en ambos casos tenemos “auténtico diseño” podríamos decir algo ligeramente diferente, señalando que la semejanza principal es que ambos procesos producen entidades que exhiben una organización funcional, esto es, sistemas cuyas partes, procesos y propiedades en general permiten que esas entidades (y sus partes) cumplan eficazmente funciones específicas: volar, nadar, hablar, capturar presas, hacer fotosíntesis, etc. Pero no parece que evitar

la palabra “diseño”¹⁵ cambie nada en términos conceptuales, y al evitarla perderíamos la oportunidad de enseñar a los estudiantes qué significa exactamente ese término cuando aparece en el contexto del discurso biológico como, de hecho, y tal como mostré antes, sucede con los libros de texto de biología.

Por su parte, la principal limitación de la analogía entre el diseño humano y la selección sería que en el primer caso —pero no en el de la selección— las razones que explican ese diseño o funcionalidad fueron previamente representadas en la mente de un diseñador intencional y consciente. Por el contrario, en el caso de la selección, es el proceso ciego de variación y reproducción diferencial lo que explica el diseño resultante.

Así, tras haber enseñado una versión básica de la teoría de la selección natural, se puede proponer a los estudiantes completar una tabla con una columna dedicada a la selección y otra dedicada al diseño tecnológico, de modo que en cada fila se compare algún aspecto específico en ambos procesos. Por ejemplo, se les puede pedir que especifiquen, para ambos casos, cuál es el rol del azar, qué factores limitan la producción de diseño (resultando en un diseño “subóptimo”), qué procesos generan la variación, qué mecanismos de herencia operan, etc. Se puede dar a los estudiantes algún texto que narre la historia de alguna innovación tecnológica para que dispongan de una fuente detallada del “campo fuente” que servirá de insumo para el análisis comparativo. Como resultará evidente, no es este un análisis trivial, por lo que qué aspectos comparar dependerá del nivel educativo en el que se esté trabajando. Un

¹⁵ Es probable que la reticencia a utilizar el término “diseño” derive, al menos parcialmente, del temor a dar la impresión de que se está concediendo algo a la “teoría del diseño inteligente”, una perspectiva religiosa que, engañosamente, y tal como ya señalé, se presenta a sí misma como científica. Esto ha llevado a algunos autores a acuñar neologismos que difícilmente contribuyen a aclarar el panorama conceptual. Así, por ejemplo, Richard Dawkins (1998) se ha referido a los organismos como “objetos diseñoides”: “Los objetos diseñoides son organismos vivos y sus productos. [...] parecen responder a un designio, tanto que mucha gente [...] piensa que han sido diseñados. Quienes así lo creen se equivocan, pero no en su convicción de que los objetos diseñoides no pueden ser resultado del puro azar. Los objetos diseñoides no son accidentales. En realidad, han sido contruidos por un magnífico proceso no aleatorio que crea una ilusión casi perfecta de designio” (Dawkins, 1998, p. 15). Desde ya, ese proceso es la selección natural.

modo de “andamiar” el proceso es ofrecer a los estudiantes una versión de la tabla en la que una de las columnas esté completa, de modo que sean ellos quienes deban pensar en cuál sería el factor equivalente en el otro campo.

En relación con la implementación de esta propuesta didáctica, cabe señalar que, aunque el enfoque epistemológico general (basado en la obra de Dennett) y el objetivo didáctico general (el desarrollo de una “vigilancia metacognitiva”) serían transversales para todo nivel educativo y todo contexto comunicativo, las posibles estrategias específicas deberán ajustarse a cada contexto. Por ejemplo, pretender que un sujeto “vigile” las expresiones teleológicas analizando su consistencia con la teoría de la selección natural solo tendrá sentido tras la instrucción formal en relación con dicho contenido. Así, actividades dirigidas a fomentar esa destreza solo podrían implementarse a partir de los últimos años de la educación secundaria. Dicho esto, sugiero que *algún* trabajo didáctico podría llevarse cabo en todo contexto de educación formal y que, además, cualquier comunicador en cualquier contexto podrá afinar su discurso ejerciendo él mismo una “vigilancia metacognitiva” sobre su propio uso de nociones y expresiones teleológicas.

Por último, una enseñanza basada en este marco teórico supondría que al enseñar la teoría de la selección natural deberíamos enseñar, además, explícitamente, ciertos contenidos metacientíficos (el rol de la teleología en la biología, el rol de las metáforas en la ciencia, etc.) y metaconceptuales (en qué casos la teleología es aceptable o no). Además del potencial de mejorar el aprendizaje de un contenido específico (la teoría de la selección natural, en este caso) ambos objetivos contribuirían a algunas de las grandes metas de la enseñanza en ciencias, a saber, enseñar sobre la *naturaleza de la ciencia* (McComas, 2020) y fomentar la metacognición y el aprendizaje autorregulado (Jorba y Casellas, 1997; Swartz *et al.*, 2017).

4. Conclusiones

En este trabajo reseñé algunos aspectos de la filosofía de Daniel Dennett vinculados con el problema de la teleología. Aunque espero haber sido fiel a las ideas de este autor, lo ofrecido es en rigor mi interpretación y síntesis de dichos desarrollos. Así, intenté mostrar (sección 2) que las ideas de Dennett en relación con lo que implica la teoría de la selección natural para nuestra comprensión del mundo orgánico y la mente humana suponen una mirada original y coherente en relación

con varios problemas relacionados con la teleología. Las posturas de diseño e intencional (sección 2.1), la biología como ingeniería reversa (sección 2.2) y una idea “fuerte” de diseño biológico (sección 2.3) son distintas caras del supuesto fundamental de que la selección natural es un proceso significativamente análogo al de diseño humano deliberado, perspectiva que se extiende por igual a los productos de ambos procesos.

Luego, abordé cuestiones educativas (sección 3), presenté lo que denominé “el problema de la teleología en la enseñanza de la biología” (sección 3.1) y caractericé dos perspectivas en relación con dicho problema (sección 3.2). En relación con esta última cuestión, apartándome de las perspectivas tradicionales “eliminacionistas”, propuse una alternativa general que apunta a fomentar una “vigilancia metacognitiva” de los estudiantes sobre sus intuiciones teleológicas. Luego desarrollé una versión más específica de esta segunda perspectiva basada en la idea dennettiana de que existe auténtico diseño en la naturaleza y en el supuesto de que dicho enfoque se basa en una analogía entre los procesos de diseño humano y selección (sección 3.3). Por falta de espacio, me limité a ofrecer algunas sugerencias generales en relación con cómo llevar a la práctica esta propuesta didáctica.

¿Aprenderían mejor los estudiantes biología evolutiva —y, en particular, la teoría de la selección natural— si en vez de buscar la eliminación o inhibición de las intuiciones teleológicas fomentáramos una vigilancia metacognitiva basada en el enfoque aquí propuesto?, ¿daría mejores resultados una enseñanza basada en la noción de que lo que encontramos en la biología es auténtico diseño? En el actual estado de la investigación, no puedo afirmar que dicho enfoque vaya a producir necesariamente mejores resultados en términos de aprendizaje. Sin embargo, sí nos consta que el enfoque tradicional no ha dado buenos resultados, por lo que necesitamos transitar nuevos caminos en la enseñanza de estos contenidos. Espero con este trabajo haber ofrecido algunas pistas conceptuales que sean de utilidad para explorar algunos de esos nuevos caminos posibles, caminos que también habremos de explorar empíricamente.

Por último, creo que los problemas aquí analizados ponen en evidencia que los límites entre la biología, la filosofía de la biología y la didáctica de la biología son menos nítidos de lo que habitualmente se asume. Esto es consistente con la propuesta de algunos filósofos que sugieren que, además de los clásicos contextos de descubrimiento y justificación, deben considerarse los de aplicación y enseñanza (Echeverría, 1998). Esto

es, la enseñanza de la biología sería parte de la propia biología. En este sentido, el problema didáctico que he abordado ha dado lugar a nuevas preguntas en relación con la teleología. La inadecuación para los fines didácticos de ciertos enfoques epistemológicos sobre el problema de la teleología ha vuelto la mirada hacia la filosofía de la biología en busca de otros enfoques más útiles. No podría afirmar que el hecho de que enfoques como el de Dennett resulten más prometedores desde el punto de vista didáctico pueda considerarse un argumento a su favor en la arena epistemológica. Sin embargo, me permito finalizar este escrito sugiriendo que una naturalización de la teleología como la que resulta del análisis de Dennett parece al mismo tiempo una mejor solución al problema en la filosofía de la biología (principalmente porque da cuenta de las prácticas de los biólogos reales) y una guía útil en relación con el problema didáctico. Considerando que ciencia, filosofía de la ciencia y didáctica de la ciencia son solo enfoques perspectivas sobre diferentes aristas de esa única y compleja empresa que llamamos “ciencia”, el enfoque de Dennett parece ofrecer soluciones satisfactorias en todos los planos.

Bibliografía

- Ackerman, R. y Thompson, V. A. (2017). Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences*, 21, 607-617. <https://doi.org/10.1016/j.tics.2017.05.004>
- Alberts, B., Bary, D., Lewis, J., Raff, M., Roberts, K. y Watson, J. (1994). *Molecular Biology of the Cell*. Garland Publishing Inc.
- Allen, C., Bekoff, M. y Lauder, G. (1998). *Nature's Purposes: Analyses of Function and Design in Biology*. MIT Press. <https://doi.org/10.1080/713659339>
- Arango-Muñoz, S. (2011). Two Levels of Metacognition. *Philosophia*, 39, 71-82. <https://doi.org/10.1007/s11406-010-9279-0>
- Atran, S., Medin, D., Lynch, E., Vapnarsky, V., Ucan Ek, E. y Sousa, P. (2001). Folkbiology Doesn't Come from Folkpsychology: Evidence from Yukatek Maya in Cross-Cultural Perspective. *Journal of Cognition and Culture*, 1(1), 3-42. <https://doi.org/10.1163/156853701300063561>
- Brock, R., Tsourakis, N. y Kampourakis, K. (2024). Using Text Mining to Identify Teleological Explanations in Physics and Biology Textbooks: An Exploratory Study. *Science & Education*, 34, 2167-2188. <https://doi.org/10.1007/s11191-024-00513-3>

- Caponi, G. (2020). The Darwinian Naturalization of Teleology. En L. Zaterka y L. Baravalle (eds.), *Life and Evolution: Latin American Essays on the History and Philosophy of Biology* (pp. 121-142). Springer. https://doi.org/10.1007/978-3-030-39589-6_8
- Caponi, G. (2023). Function, Adaptation, and Design in Biology. En J. Gayon, A. de Ricqlès y A. Dussault (eds.), *Functions: From Organisms to Artefacts* (pp. 115-134). Springer. https://doi.org/10.1007/978-3-031-31271-7_9
- Clement, J. (1993). Using Bridging Analogies and Anchoring Intuitions to Deal with Student's Preconception in Physics. *Journal of Research in Science Teaching*, 30(10), 1241-1257. <https://doi.org/10.1002/tea.3660301007>
- Coley, J. D. y Tanner, K. D. (2012). Common Origins of Diverse Misconceptions: Cognitive Principles and the Development of Biology Thinking. *CBE Life Sciences Education*, 11(3), 209-215. <https://doi.org/10.1187/cbe.12-06-0074>
- Cupo, B., González Galli, L. y Soto, I. (2024, 2-6 de septiembre). *Avances en la caracterización de la regulación metacognitiva en el aprendizaje del modelo de evolución por selección natural*. [Conferencia]. 17th International History, Philosophy and Science Teaching Conference, Buenos Aires, Argentina.
- Crelier, A. (2022). Artefacto. En D. Parente, A. Berti y C. Celis (eds.), *Glosario de filosofía de la técnica* (pp. 62-66). La Cebra.
- Dawkins, R. (1998). *Escalando el monte improbable*. J. Ros (trad.). Tusquets.
- De Camilloni, A. R. W. (2001). *Los obstáculos epistemológicos en la enseñanza*. A. Bixio (trad.). Gedisa.
- Dennett, D. (1991). *La actitud intencional*. D. Zadunaisky (trad.). Gedisa.
- Dennett, D. (1995). *La peligrosa idea de Darwin*. C. Pérez Blanco-Morales(trad.).Galaxia Gutemberg.
- Dennett, D. (2009). Intentional Systems Theory. En A. Beckermann, B. McLaughlin y S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind* (pp. 339-350). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199262618.003.0020>
- Dennett, D. (2015). *Bombas de intuición y otras herramientas de pensamiento*. L. Leucona (trad.). Fondo de Cultura Económica.
- Diéguez, A. (2012). *La vida bajo escrutinio. Una introducción a la filosofía de la biología*. Buridán.
- Hanisch, S. y Eirdosh, D. (2023). Teaching for the Interdisciplinary Understanding of Evolutionary Concepts. En A. du Crest, M. Valković,

- A. Ariew, H. Desmond, P. Huneman y T. Reydon (eds.), *Evolutionary Thinking Across Disciplines: Problems and Perspectives in Generalized Darwinism*. Springer. https://doi.org/10.1007/978-3-031-33358-3_8
- Else, M., Clement, J. y Rea-Ramirez, M. (2008). Using Analogies in Science Teaching and Curriculum Design: Some Guidelines. En J. Clement. y M. Rea-Ramirez (Eds.), *Model Based Learning and Instruction in Science* (pp. 215-232). Springer. https://doi.org/10.1007/978-1-4020-6494-4_12
- Frankish, K. (2010). Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5(10), 914-926. <https://doi.org/10.1111/j.1747-9991.2010.00330.x>
- Futuyma, D. y Kirkpatrick, M. (2018). *Evolution*. Oxford University Press.
- Galagovsky, L. y Adúriz-Bravo, A. (2001). Modelos y analogías en la enseñanza de las ciencias naturales. El concepto de modelo didáctico analógico. *Enseñanza de las Ciencias: Revista de Investigación y Experiencias Didácticas*, 19(2), 231-242. <https://doi.org/10.5565/rev/ensciencias.4000>
- Gayon, J., de Riquès, A. y Dussault, A. (2023). *Functions: From Organisms to Artefacts*. Springer. <https://doi.org/10.1007/978-3-031-31271-7>
- Gerlach, G. (2011). The Genus *Coryanthes*: A Paradygm in Ecology. *Lankesteriana: International Journal on Orchidology*, 11(3), 253-264. <https://doi.org/10.15517/lank.v11i3.18280>
- Ginnobili, S. (2010). *La teoría de la selección natural. Una exploración metacientífica*. Universidad Nacional de Quilmes Editorial.
- Godfrey-Smith, P. (2017). *Otras mentes. El pulpo, el mar, y los orígenes profundos de la consciencia*. J. Ros i Aragonés (trad.). Taurus.
- Godfrey-Smith, P. (2021). *Metazoos. La evolución de la vida y el nacimiento de la consciencia*. J. Ros i Aragonés (trad.). Shackleton Books.
- Godfrey-Smith, P. (2022). *Filosofía de la biología*. B. Villa Pacheco (trad.). Bauplan.
- González Galli, L. (2016). El problema de la teleología y la metáfora del diseño en biología: cuestiones epistemológicas e implicancias didácticas. *Tecné, Episteme y Didaxis*, 40, 149-173. <https://doi.org/10.17227/01203916.6151>
- González Galli, L. (2019). Permitido decir “para”: el problema de la teleología en la enseñanza de la biología. *Revista Científica*, 34(1), 49-62. <https://doi.org/10.14483/23448350.13710>
- González Galli, L. (2023). Analogías y enseñanza de la genética y la biología evolucionista. *Enseñanza de las Ciencias. Revista de Investigación y Experiencias Didácticas*, 41(1), 63-78. <https://doi.org/10.5565/rev/ensciencias.5615>

- González Galli, L. (2025). La evolución de un evolucionista: Michael Ruse y la teleología. *Metatheoria. Revista de Filosofía e Historia de La Ciencia*, 14(1), 47-66. <https://doi.org/10.48160/18532330me14.355>
- González Galli, L. y Meinardi, E. (2015). Obstáculos para el aprendizaje del modelo de evolución por selección natural en estudiantes de escuela secundaria de Argentina. *Ciencia y Educação*, 21(1), 101-122. <https://doi.org/10.1590/1516-731320150010007>
- González Galli, L., Pérez, G., Cupo, B. y Alegre, C. (2022). Revisión y revalorización del concepto de obstáculo epistemológico para la enseñanza de las Ciencias Naturales. *Ciência & Educação*, 28(13). <https://doi.org/10.1590/1516-731320220040>
- González Galli, L., Pérez, G., Cupo, B., Suárez, J., Dop, S. y Alegre, C. (2025). Un mundo extraño forjado por nuestra mejor ciencia. Aspectos contraintuitivos de la teoría de la evolución. En A. Gómez Galindo y G. Peñaloza (eds.), *Enseñar y aprender evolución. Perspectivas latinoamericanas* (pp. 13-50). Lito-Grapo.
- González Galli, L., Pérez, G. y Gómez Galindo, A. (2020). The Self-Regulation of Teleological Thinking in Natural Selection Learning. *Evolution: Education & Outreach*, 13(6). <https://doi.org/10.1186/s12052-020-00120-0>
- González Labra, M. J. (1997). *Aprendizaje por analogía. Análisis del proceso de inferencia analógica para la adquisición de nuevos conocimientos*. Trotta.
- Gould, S. y Lewontin, R. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London B*, 205(1161), 581-598. <https://doi.org/10.1098/rspb.1979.0086>
- Hartelt, T. y Martens, H. (2024). Self-Regulatory and Metacognitive Instruction Regarding Student Conceptions: Influence on Students' Self-Efficacy and Cognitive Load. *Frontiers in Psychology*, 22(15). <https://doi.org/10.3389/fpsyg.2024.1450947>
- Hempel, C. (1959). The Logic of Functional Analysis. En C. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (pp. 297-330). Free Press.
- Hills, D. (2024). Metaphor. En E. Zalta y U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2024/entries/metaphor/>
- Houdé, O. (2019). *3-System Theory of the Cognitive Brain: A Post-Piagetian Approach to Cognitive Development*. Routledge. <https://doi.org/10.4324/9781315115535>

- Jorba, J. y Casellas, E. (1997). *La regulación y la autorregulación de los aprendizajes*. Síntesis.
- Kahneman, D. (2019). *Pensar rápido, pensar despacio*. J. Chamorro Mielke (trad.). Debate.
- Kampourakis, K. (2013). Philosophy of Biology and Biology Education: An Introduction. En K. Kampourakis. (ed.), *The Philosophy of Biology: A Companion for Educators* (pp. 1-30). Springer. <https://doi.org/10.1007/978-94-007-6537-5>
- Kampourakis, K. (2020). Students' "Teleological Misconceptions" in Evolution Education: Why the Underlying Design Stance, Not Teleology per se, Is the Problem. *Evolution Education & Outreach*, 13(1). <https://doi.org/10.1186/s12052-019-0116-z>
- Kelemen, D. (2012). Teleological Minds: How Natural Intuitions about Agency and Purpose Influence Learning about Evolution. En K. Rosengren, S. Brem, E. Evans y G. Sinatra (eds.), *Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution* (pp. 66-92). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730421.003.0004>
- Krajcik, J. y Shin, N. (2023). Student Conceptions, Conceptual Change, and Learning Progressions. En N. G. Lederman, D. L. Zeidler y J. S. Lederman (eds.), *Handbook of Research on Science Education* (pp. 121-157). Routledge. <https://doi.org/10.4324/9780367855758-7>
- Lennox, J. (1993). Darwin Was a Teleologist. *Biology and Philosophy*, 8(4), 409-421. <https://doi.org/10.1007/BF00857687>
- Lewens, T. (2000). Function Talk and the Artefact Model. *Studies in History and Philosophy of Biology and Biomedical Sciences*, 31(1), 5-111. [https://doi.org/10.1016/S1369-8486\(99\)00040-0](https://doi.org/10.1016/S1369-8486(99)00040-0)
- Lewens, T. (2002). Adaptationism and Engineering. *Biology and Philosophy*, 17(1), 1-31. <https://doi.org/10.1023/A:1012915007444>
- Lewens, T. (2004). *Organisms and Artifacts. Design in Nature Elsewhere*. MIT Press. <https://doi.org/10.7551/mitpress/5172.001.0001>
- Lewens, T. (2007). Adaptation. En D. Hull y M. Ruse (eds.), *The Cambridge Companion to The Philosophy of Biology* (pp. 1-12). Cambridge University Press.
- Linares, R. e Izquierdo Aymerich, M. (2006). El rescate de la princesa encerrada en lo más alto de la más alta torre. Un episodio para aprender sobre analogías, símiles y metáforas. *El Hombre y la Máquina*, 27, 24-37.
- Mahner, M. y Bunge, M. (2000). *Fundamentos de biofilosofía*. Siglo XXI.

- McComas, W. (2020). *Nature of Science in Science Instruction: Rationales and Strategies*. Springer.
- McGuire, S. (2023). *Teach Students How to Learn Strategies You Can Incorporate into Any Course to Improve Student Metacognition, Study Skills, and Motivation*. Routledge.
- Micheli, P., Wilner, S. J., Bhatti, S. H., Mura, M. y Beverland, M. B. (2019). Doing Design Thinking: Conceptual Review, Synthesis, and Research Agenda. *Journal of Product Innovation Management*, 36(2), 124-148. <https://doi.org/10.5465/AMBPP.2018.16071abstract>
- Nagel, E. (1961). *The Structure of Science*. Harcourt.
- Oliva-Martínez, J. y Aragón-Méndez, M. (2009). Contribución del aprendizaje con analogías al pensamiento modelizador de los alumnos en ciencias: marco teórico. *Enseñanza de las Ciencias: Revista de Investigación y Experiencias Didácticas*, 27(2), 195-208. <https://doi.org/10.5565/rev/ensciencias.3731>
- Pérez, G. y González Galli, L. (2020). Una posible definición de metacognición para la enseñanza de las ciencias. *Investigações em Ensino de Ciências*, 25(1), 384-404. <https://doi.org/10.22600/1518-8795.ienci2020v25n1p384>
- Pozo, J. (2014). *Psicología del aprendizaje humano. Adquisición de conocimiento y cambio conceptual*. Morata.
- Proust, J. y Vauclair, J. (2003). Cognición animal. En O. Houdé, D. Kayser, O. Koenig, J. Proust y F. Rastier (eds.), *Diccionario de ciencias cognitivas. Neurociencia. Psicología. Inteligencia artificial. Lingüística. Filosofía* (pp. 56-60). C. Molinari Marotto (trad.). Amorrortu.
- Reynolds, A. (2022). *Understanding Metaphors in the Life Sciences*. Cambridge University Press. <https://doi.org/10.1017/9781108938778>
- Rottmann, J., Zhu, L., Wang, W., Schillaci, R. S., Clark, K. J. y Kelemen, D. (2017). Cultural Influences on the Teleological Stance: Evidence from China. *Religion Brain & Behavior*, 7(1), 17-26. <https://doi.org/10.1080/2153599X.2015.1118402>
- Ruse, M. (2023). Teleología: ¿ayer, hoy y mañana? *Metatheoria*, 13(2), 125-142. <https://doi.org/10.48160/18532330me13.334>
- Sanz, C. M., Call, J. y Boesch, C. (2013). *Tool Use in Animals: Cognition and Ecology*. Cambridge University Press.
- Schachner, A., Zhu, L., Li, J. y Kelemen, D. (2017). Is the Bias for Function-Based Explanations Culturally Universal? Children From China Endorse Teleological Explanations of Natural

- Phenomena. *Journal of Experimental Child Psychology*, 157, 29-48.
<https://doi.org/10.1016/j.jecp.2016.12.006>
- Schlosser, M. (2019). Agency. En E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2019/entries/agency/>
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139173452>
- Short, T. (2002). Darwin's Concept of Final Cause: Neither New nor Trivial. *Biology and Philosophy*, 17(3), 322-340.
<https://doi.org/10.1023/A:1020173708395>
- Shumaker, R., Walkup, K., Beck, B. y Burghardt, G. (2011). *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. John Hopkins University Press.
- Sober, E. (2018). *Philosophy of Biology*. Routledge.
- Stryer, L. (1995). *Bioquímica*. Reverté.
- Sultan, S., Moczek, A. y Walsh, D. (2022). Bridging the Explanatory Gaps: What Can We Learn from a Biological Agency Perspective? *Bioessays*, 44(1), e2100185. <https://doi.org/10.1002/bies.202100185>
- Swartz, R., Costa, A., Beyer, B., Reagan, R. y Kallick, B. (2017). *El aprendizaje basado en el pensamiento. Cómo desarrollar en los alumnos las competencias del siglo XXI*. A. Belén Flete (trad.). SM.
- Wright, L. (1976). *Teleological Explanations*. University of California Press.
<https://doi.org/10.1525/9780520333697>
- Yang, G., Zhou, W., Qu, W., Yao, W., Zhu, P. y Xu, J. (2022). A Review of Ant Nests and Their Implications for Architecture. *Buildings*, 12(12), 2225. <https://doi.org/10.3390/buildings12122225>

<https://doi.org/10.21555/top.v760.3265>

Cuando el “ácido universal” darwiniano entra en contacto con la creatividad humana

When Darwin’s “Universal Acid” Comes into Contact with Human Creativity

Malena León

Universidad Nacional de Córdoba
Argentina

malena.leon@unc.edu.ar

<https://orcid.org/0000-0002-9520-2267>

Recibido: 27 - 09 - 2024.

Aceptado: 30 - 01 - 2025.

Publicado en línea: 01 - 06 - 2026.

Cómo citar este artículo: León, M. (2026). Cuando el “ácido universal” darwiniano entra en contacto con la creatividad humana. *Tópicos, Revista de Filosofía*, 76, 119-158. <https://doi.org/10.21555/top.v760.3265>



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

Resumen

Dennett ha sostenido que una recepción cabal de la teoría evolutiva darwiniana debe modificar nuestra concepción de los fenómenos mentales, dentro de los cuales se ha incluido tradicionalmente la creatividad humana. En este artículo sistematizo y desarrollo algunos lineamientos que debería adoptar un enfoque darwinista de la creatividad humana siguiendo las ideas de Dennett. Defiendo que el enfoque darwinista dennettiano de la creatividad es de carácter pluralista —dado que incorpora procesos de distinto tipo y que operan en distintos niveles—, incluye explicaciones que tienen en cuenta el carácter acumulativo del diseño y contempla que los procesos creativos son, en gran medida, colaborativos y que no necesariamente involucran comprensión. Así, es un enfoque que sostiene que la mente humana, actuando de manera individual, no detenta la potencia que le asignaron los enfoques predarwinianos de la creatividad.

Palabras clave: Daniel Dennett; creatividad; darwinismo; evolución cultural; diseño sin diseñador; competencia sin comprensión; genio creativo; nivel comunitario; ácido universal; procesos de abajo arriba; intencionalidad derivada.

Abstract

Dennett has argued that a thorough reception of Darwinian evolutionary theory must modify our conception of mental phenomena, which traditionally includes human creativity. In this article, I aim to systematize and develop some guidelines that a Darwinian approach to human creativity should adopt, following Dennett's ideas. I argue that Dennett's Darwinian approach to creativity is pluralistic in nature—since it incorporates processes of different kinds which operate at different levels—it includes explanations that account for the cumulative nature of design, and it considers that creative processes are largely collaborative and do not necessarily involve understanding. Thus, it is an approach that maintains that the human mind, acting individually, does not possess the power assigned to it by pre-Darwinian approaches to creativity.

Keywords: Daniel Dennett; creativity; Darwinism; cultural evolution; design without a designer; competence without comprehension; creative genius; community level; universal acid; bottom-up processes; derived intentionality.

Y si la evolución sin mente podía explicar los asombrosamente ingeniosos artefactos de la biosfera, ¿cómo podrían los productos de nuestras propias mentes “reales” estar exentos de una explicación evolutiva? La idea de Darwin, por lo tanto, también amenazaba con extenderse hacia lo más alto, disolviendo la ilusión de nuestra propia autoría, nuestra propia chispa divina de creatividad y comprensión.

Dennett (1995b, p. 63; mi traducción).

Introducción¹

De acuerdo con la definición estándar de la creatividad (Barron, 1955; Simonton, 1999; Runco y Jaeger, 2012; Kaufman y Glăveanu, 2019), un producto creativo debe exhibir dos propiedades para ser considerado como tal. Por una parte, debe ser original o novedoso. Por la otra, debe ser valioso en algún sentido específico: ser adecuado o funcional o útil; debe resolver un problema. Otra manera de caracterizar este segundo componente de la definición estándar es afirmando que un producto creativo debe ser eficaz. En suma, cuando le otorgamos a un producto el “título honorífico” (Briskman, 2009, p. 17) de “creativo” reconocemos que no solo se trata de algo nuevo, sino también de algo bueno.

En nuestras concepciones clásicas y tradicionales sobre las creaciones culturales subyace la idea de que determinados logros creativos extraordinarios son posibles porque los han originado individuos excepcionales, ya sea en virtud de su inspiración e intuición, ya sea en virtud de su raciocinio prodigioso. Podemos identificarla como “la idea del genio creativo”. En la literatura especializada en creatividad

¹ Este artículo es, de algún modo, el resultado de una investigación doctoral dirigida por la Dra. Carolina Scotto y codirigida por la Dra. Laura Danón. Los resultados obtenidos no hubieran sido posibles sin su orientación atenta, sus correcciones agudas y, en general, su ayuda, siempre generosa e inteligente. Asimismo, los comentarios y aportes del Dr. Santiago Ginnobili, el Dr. Sergio Barberis y el Dr. Ariel Roffé fueron sumamente valiosos para el desarrollo de esta investigación.

se pueden encontrar diversas críticas a dicha idea. Se ha sostenido, por poner algunos ejemplos, que la creatividad no tiene por qué ser concebida como un atributo excepcional (Boden, 2002), que lo que se considera valioso depende del contexto histórico y cultural (Amabile, 1983),² que para que una persona sea creativa tiene que interiorizarse con el dominio antes de poder hacer su contribución (Csikszentmihalyi, 2014), que existen formas de crear colectivamente —pensando en procesos en los que distintos individuos se involucran simultáneamente en una tarea creativa— (Glăveanu, 2011 y 2015), etc.. En este trabajo mostraré que la filosofía de Dennett provee razones adicionales para oponerse a la idea del genio creativo. Contar con una concepción de este tipo resulta valioso para avanzar con uno de los objetivos que tienen los estudios académicos sobre la creatividad: liberarse del halo romántico que encubre nuestras creencias en torno al fenómeno en cuestión y adoptar una visión de este compatible con algunas de las teorías científicas contemporáneas y los compromisos ontológicos que las acompañan (Boden, 2002; Simonton, 1999; Kronfeldner, 2009).

Durante por lo menos el último medio siglo, el filósofo estadounidense Daniel Dennett (1990a, 1990b, 1991, 1995a, 1995b, 1998, 2001, 2002 y 2017) ha realizado importantes contribuciones al proyecto filosófico naturalista de explicar el mundo que nos rodea de un modo concordante con lo que proponen las mejores teorías científicas contemporáneas. Su perspectiva es naturalista en, al menos, dos sentidos fundamentales: uno ontológico y otro epistemológico. Dennett es un naturalista en un sentido ontológico porque adopta la tesis según la cual se debe incluir en la ontología solo entidades naturales, en contraposición con hipotéticas entidades sobrenaturales o no naturales. A su vez, es un naturalista epistemológico porque nos ofrece un modo de entender la filosofía como una disciplina no apriorística que ha de trabajar de manera cooperativa con las ciencias particulares.

En el corazón de sus desarrollos filosóficos se encuentra una recepción cabal de la teoría darwiniana de la evolución de las especies, lo que constituye otra marca de su naturalismo. Desde su visión, nuestras

² Amabile (1983) demostró que los test psicométricos que pretendían medir la creatividad contenían a menudo una evaluación subjetiva implícita. Según Amabile, un producto será entendido como creativo solo en la medida en que lo determinen los expertos en la materia, cuyo juicio está condicionado cultural e históricamente.

explicaciones filosóficas acerca de fenómenos como la mente humana no han cumplido la exigencia de ser permeadas por este “ácido universal” que todo lo corroe (Dennett, 1995b). Dennett (1995b) introduce esta fantasía popular que refiere a la idea de un ácido tan corrosivo que atravesaría cualquier material y que, por ende, presentaría un dilema a la hora de decidir dónde guardarlo. De acuerdo con Dennett, la teoría de Darwin es semejante a un ácido universal en un sentido específico: el de transformar todos los conceptos tradicionales, dejando tras su paso una concepción del mundo completamente revolucionada. Como se puede ver en el fragmento citado como epígrafe, Dennett ha señalado que la creatividad humana no es una excepción: nuestras concepciones y teorías en torno a este fenómeno también se modifican de maneras cruciales cuando se desarrollan las consecuencias de adoptar las tesis de la teoría darwiniana. Sin embargo, Dennett no se ha dedicado a desarrollar en profundidad las consecuencias de adoptar una visión darwiniana para concebir este fenómeno en particular. Este es el paso teórico que querría dar aquí.

Este artículo está motivado por la idea de que algunas nociones filosóficas fundamentales sobre la mente humana y la evolución propuestas por Dennett constituyen un marco prometedor e innovador para concebir la creatividad humana. Mi objetivo es sistematizar, clarificar y enriquecer dos lineamientos centrales que debería adoptar un enfoque darwinista de la creatividad humana siguiendo el marco teórico filosófico planteado por Dennett. Según uno de estos lineamientos, la creatividad no es un fenómeno exclusiva ni primariamente mental. Según el otro, los procesos creativos son de diverso tipo, no necesariamente involucran comprensión y pueden operar en un nivel supraindividual.

En lo que respecta al marco teórico desarrollado por Dennett, en primera instancia me centraré en dos tesis defendidas por él y en ciertos compromisos filosóficos que las acompañan. Una de ellas versa sobre la mente humana y afirma que los estados mentales exhiben *intencionalidad derivada* (Dennett, 1995b y 1998). La otra, que trata sobre los procesos de creación, sostiene que no es necesario que exista un diseñador para producir un diseño (Dennett, 1995b). Se trata de dos tesis basadas en nociones que, según Dennett, estarían implicadas por la teoría de la evolución de las especies propuesta por Darwin (1964) y que se contraponen a una visión predarwiniana del mundo.

Con el propósito de ofrecer una presentación un tanto simplificada, pero que puede resultar esclarecedora, del cambio de concepción que

conlleven estas dos tesis que, de acuerdo con Dennett, se siguen de la teoría darwiniana, recuperaré dos ideas que son propias de la visión predarwiniana del mundo. Por una parte, de acuerdo con esta visión, la mente es un fenómeno excepcional, por lo que no parece ser parte del mundo natural. Antes bien, es la fuente última de los significados de todos los objetos culturales. Por otra parte, una de sus características distintivas es que es capaz de producir diseño: de ordenar la materia desordenada de modo tal que adquiriera una función. De hecho, pareciera que el único modo de producir diseño depende de la acción intencional de la mente.

Dennett ha señalado que Locke (1847) sintetizó estas tesis predarwinianas como ningún otro pensador. Bajo la forma de un argumento deductivo, Locke presentó la siguiente idea, que, a su parecer, todos damos por sentado: la mente y el diseño solo pueden venir de otra mente, por lo que en el origen de todo lo que hay debe haber una mente (cfr. Dennett, 1995b, pp. 26-28). Siguiendo los planteamientos de Dennett podemos vincular las dos tesis predarwinianas con una concepción tradicional de la comprensión de la siguiente manera. Por un lado, la capacidad de comprender parece ser el ingrediente extra que solo poseen las mentes y que no se encuentra en el mundo natural. Por otro lado, la capacidad de comprender es lo que permitiría a la mente producir diseños. Dennett (1995b y 2017) sugiere que, si bien Darwin brindó buenos motivos para abandonar estas ideas, estas siguen presentes, de maneras más o menos explícitas, en nuestras concepciones de sentido común y también en diferentes teorías filosóficas y científicas vinculadas a la mente humana.

Las tesis dennettianas que tomaré como punto de partida de este artículo pueden ser vistas como las dos caras de una misma moneda: una que versa sobre la mente y la otra sobre los procesos creativos. En primer lugar, Dennett ha defendido que la mente humana forma parte del mundo natural. Ha reconocido que los estados mentales tienen una característica bastante especial: pueden *ser acerca de* otra cosa. Esta característica ha recibido en filosofía de la mente el nombre técnico de “intencionalidad”.³ Ahora bien, la *intencionalidad* también es portada por fenómenos no mentales (pensemos, por ejemplo, en una frase, una señal

³ A lo largo de este artículo emplearé las bastardillas para señalar el uso técnico del término “intencional”, es decir, su acepción filosófica como la propiedad de *ser acerca de algo*. En cambio, cuando el término “intencional”

de tránsito, un dibujo). A diferencia de una gran cantidad de filósofos de la mente, Dennett (1998) ha sostenido que los estados mentales exhiben *intencionalidad* del mismo modo en el que lo hacen los fenómenos no mentales: en otros términos, que debería cuestionarse la distinción, aparentemente intuitiva, entre una forma de *intencionalidad* original —la de la mente humana— y la que es derivada —la de sus productos—. Así, con respecto a ese aspecto en particular, no habría motivos de peso para considerar que la mente constituye una excepción al mundo natural. En segundo lugar, la teoría de Darwin provee una alternativa a la explicación mentalista sobre el origen del diseño. En ese sentido, se puede entender que dicha teoría conduce a entender que la comprensión no es un ingrediente indispensable para producir cualquier tipo de diseño; antes bien, un mecanismo ciego y sin objetivos puede dar como resultado productos de gran complejidad y ajuste con su entorno. En suma, se trata de una visión según la cual no hay nada esencialmente misterioso en los procesos creativos humanos de producción de diseños. Antes bien, un proceso mecánico y sin mente, como el de la selección natural, puede dar lugar a diseños sumamente complejos.

En lo que sigue, voy a argumentar que estas dos tesis proporcionan un marco prometedor para abordar el fenómeno de la creatividad. El mismo Dennett también pareció entenderlo así, dado que ha señalado que es preciso modificar nuestra concepción de la creatividad en virtud de las implicancias de la teoría de Darwin (Dennett, 1995b, p. 63; 2001, pp. 16-17, 21 y 23; 2017, pp. 255, 286-287 y 290). Sin embargo, como señalé anteriormente, el investigador no desarrolló en profundidad qué implicaría exactamente este cambio de concepción. Ahora bien, algunas sugerencias dispersas, junto con otros elementos de su propuesta teórica general, permiten delimitar en qué consistiría un marco darwinista dennettiano para explicar la creatividad. Concretamente, defenderé que se trata de un marco pluralista —dado que incorpora procesos que operan en distintos niveles—, que incluye explicaciones que tienen en cuenta el carácter acumulativo del diseño y contempla que los procesos creativos son, en gran medida, colaborativos y no necesariamente involucran comprensión. Es decir, se trata de un enfoque que sostiene que la mente humana actuando de manera individual no detenta la potencia creativa que le asignaron los enfoques predarwinianos.

aparezca sin bastardillas, deberá entenderse en su sentido coloquial, como sinónimo de “deliberado”.

Las primeras dos secciones de este artículo estarán destinadas a una reconstrucción informada y selectiva de algunas tesis dennettianas y los compromisos filosóficos que las acompañan, que considero relevantes para la construcción de un enfoque darwinista para explicar la creatividad humana. En el primer apartado, expondré los compromisos filosóficos de Dennett (1978, 1995b y 1998) sobre la mente y la *intencionalidad*. En el segundo apartado, presentaré las ideas dennettianas según las cuales no todo diseño tiene diseñador y no toda competencia necesita estar guiada por la comprensión de lo que se está haciendo, ni de las razones para ello (Dennett, 1995b y 2017). En el tercer apartado, argumentaré que los compromisos desarrollados en los dos apartados previos conducen a objetar la idea de que los estados mentales de los creadores son un elemento crucial a la hora de explicar todo proceso de creación humana. Finalmente, en el cuarto apartado expondré qué elementos formarían parte de una concepción darwinista dennettiana de la creatividad; para lo cual retomaré las escasas sugerencias que Dennett (2001, 2017) ha realizado con respecto a cómo deberíamos modificar nuestra concepción de la creatividad y las pondré en vínculo con otros conceptos dennettianos sobre la mente y la evolución.

1. Dennett versus la excepcionalidad de los estados mentales intencionales

De acuerdo con Dennett (1971 y 1998), el contenido de los estados mentales se especifica mediante la adopción de una estrategia interpretativa que consiste en tratar a un sistema “como si tuviera creencias, objetivos y capacidades que se relacionan de forma sistemática” (Ross, 2000, p. 3; mi traducción). Esta estrategia interpretativa recibe el nombre de enfoque o *actitud intencional* y es uno de los conceptos nucleares de su teoría de los sistemas intencionales, presentada principalmente en las colecciones de ensayos *Brainstorms* (1978) y *La actitud intencional* (1998).

Para explicar mejor a qué hace referencia la *actitud intencional*, es relevante apuntar que Dennett propone distinguir entre tres enfoques o actitudes (*stances*):⁴ la actitud física, la del diseño y la *intencional*. Se trata

⁴ Como señala un revisor de este artículo, quizá la mejor forma de traducir el término *stance* en este contexto sería “perspectiva”. No obstante, las traducciones al español de los escritos de Daniel Dennett han optado generalmente por la palabra “actitud”. La traducción de Marc Figueras en Dennett (2017) constituye, en cierto modo, una excepción al introducir los

de estrategias de razonamiento que nos permiten hacer predicciones acerca de los sistemas que se pretende interpretar. De acuerdo con la teoría, adoptar la actitud física consiste en tratar al sistema cuyo “comportamiento” se quiere predecir en función de su constitución física y utilizar los conocimientos —científicos y/o *folk*— que tenemos acerca de cómo se comportan los objetos macroscópicos para realizar las predicciones correspondientes. Por su parte, la actitud del diseño no requiere tener conocimiento de la constitución física del sistema en cuestión. Antes bien, las interpretaciones se apoyan en la suposición de que las partes del sistema cuyo “comportamiento” se quiere predecir cumplen funciones que obedecen a un diseño satisfactorio. En otros términos, se trata de una actitud interpretativa que consiste en tratar al sistema en cuestión como una entidad diseñada. Por último, la actitud intencional es una estrategia que consiste en tratar al sistema cuyo comportamiento se quiere predecir como un agente racional con creencias, deseos y otros estados mentales dotados de *intencionalidad*. En términos de Dennett:

[...] primero, se decide tratar al objeto cuyo funcionamiento hay que predecir como un agente racional; luego se deduce qué creencias debería tener ese agente, dada su posición en el mundo y su objetivo. Más tarde se deduce qué deseos tendría que tener, siguiendo las mismas consideraciones, y por fin se predice que ese agente racional actuará para conseguir sus metas a la luz de sus creencias (1998, pp. 28-29).

Así, de acuerdo con esta propuesta, tener una creencia *P* no es nada más ni nada menos que ser un sistema *intencional* cuyo comportamiento resulta mejor explicado y predicho si se le atribuye dicha creencia.

Esta concepción de la *intencionalidad* implica que esta no depende de las propiedades intrínsecas del sistema nervioso de los seres humanos, sino de sus propiedades funcionales. Según esta teoría, los estados mentales serían simplemente aquellos estados que atribuimos a un agente con el objetivo de explicar su conducta. Esto implica que no hay algo más profundo ni misterioso más allá de los patrones que

términos “enfoque o actitud” como sinónimos funcionales. En este trabajo, privilegiaré la traducción más habitual (“actitud”), aun cuando ello implique sacrificar cierta precisión conceptual.

encontramos para explicar los comportamientos. De este modo, no hay diferencia posible entre un sistema que se comporta *aparentando* tener creencias y deseos y uno que tiene *verdaderamente* creencias y deseos, dado que tener creencias y deseos no es otra cosa que comportarse *como si* se los tuviera. En suma, la teoría sostiene que la postulación de creencias y deseos depende de la adopción de esta perspectiva interpretativa, que recibe el nombre de “actitud intencional”.⁵

De acuerdo con la teoría en cuestión, cuando se atribuye un estado mental, no hay *detrás* un hecho profundo que determine *realmente* de qué trata. Existen hechos que determinan los contenidos de los estados mentales, pero son múltiples y, además, se entiende que los contenidos de tales estados pueden ser, en algún punto de la interpretación, “simplemente indeterminados” (Dennett, 1998, p. 276).

Ahora bien, hay diferentes tipos de fenómenos que exhiben *intencionalidad*. Podemos mencionar, entre los más distintivos, a las cogniciones, las percepciones y las acciones. Pero, por otra parte, la *intencionalidad* también es una propiedad que podemos hallar en las afirmaciones, los símbolos públicos y artefactos como mapas o programas de computadora. Es decir, la *intencionalidad* es exhibida tanto por fenómenos mentales como por fenómenos no mentales. Según un enfoque clásico, el modo en el que los fenómenos mentales exhiben *intencionalidad* es diferente al modo en el que lo hacen los no mentales (Brentano, 2012). Mientras los primeros tendrían *intencionalidad* de manera original e intrínseca, los segundos tendrían *intencionalidad*

⁵ Dicha dependencia de la interpretación confiere un componente *antirrealista* a su teoría de los estados mentales. Ahora bien, Dennett ha señalado repetidas veces que su teoría no debe ser interpretada como una forma de instrumentalismo (cfr. Dennett, 1991 y 1998; Ross, 2000). De acuerdo con la postura de Dennett, las creencias, deseos e intenciones son algo que habita objetivamente el mundo en el que vivimos, como también lo hacen otras abstracciones. Al mismo tiempo, el éxito o fracaso de una atribución *intencional* es un hecho perfectamente objetivo. Efectivamente, esta estrategia interpretativa, que nos permite guiarnos en el mundo asumiendo que las personas tienen creencias y deseos, no es optativa, sino “una función de la manera en la cual la selección natural ha diseñado nuestros cerebros y nuestro sistema nervioso” (Ross, 2000, pp. 3-4; mi traducción). En rigor, en un artículo que intenta precisar estos matices acerca de los compromisos ontológicos de su propuesta, Dennett (1991) ha afirmado que su teoría podría ser clasificada como una variante de “realismo moderado”.

solo de modo derivado, heredándola de otro fenómeno que ya tenía originariamente esa propiedad. De este modo, se trazaría una distinción entre un tipo de *intencionalidad* que es considerada original y otra que es derivada.

Cuando se presenta esta distinción filosófica entre *intencionalidad* original y derivada, la mayoría de las personas la encuentra bastante intuitiva. Además, la distinción es respaldada por una gran cantidad de destacados filósofos de la mente, como Fodor, Searle, Dretske, Burge y Kripke (cfr. Dennett, 1998, pp. 254-284). Uno de los principales defensores de esta distinción es John Searle (1982, 1983 y 2004). En sus términos:

En lo concerniente a la mente también debemos distinguir entre intencionalidad original o intrínseca, por una parte, e intencionalidad derivada, por otra. Por ejemplo, en la cabeza tengo información sobre la manera de llegar a San José. Tengo un conjunto de creencias verdaderas acerca del camino a esa localidad. Esa información y esas creencias presentes en mí son ejemplos de intencionalidad original o intrínseca. El mapa frente a mí también contiene información sobre el modo de llegar a San José, así como ciertos símbolos y expresiones que se refieren a, versan sobre o representan ciudades, autopistas y cosas por el estilo. Pero si el mapa contiene intencionalidad en forma de información, referencialidad y representaciones, lo hace en un sentido derivado de la intencionalidad original de cartógrafos y usuarios. Intrínsecamente, el mapa es solo una lámina de fibra de celulosa con manchas de tinta. Cualquiera sea su intencionalidad, le es impuesta por la intencionalidad original de los seres humanos (Searle, 2004, p. 19).

Aunque en la teoría de Searle estén solapadas, Haugeland (1990) señala que en realidad es posible establecer, no uno, sino dos pares de distinciones que guardan vínculos entre sí: una entre la *intencionalidad* original y la derivada y otra entre la *intencionalidad* intrínseca y la relativa al observador.

Respecto a la primera distinción, podríamos formular la tesis de la *intencionalidad* original del siguiente modo: hay fenómenos *intencionales*

que lo son de modo original, a diferencia de otros que lo son de modo derivado. Como señalé, Searle sostiene que mientras que los fenómenos mentales exhiben *intencionalidad* original, los fenómenos intencionales no mentales —tales como un texto escrito, un mapa o un acto de habla— derivan su *intencionalidad* de la que es propia de los fenómenos mentales. Como expondré a continuación, Dennett sostiene, en cambio, que toda *intencionalidad* es derivada.

Por otra parte, se puede establecer también la distinción entre la *intencionalidad* intrínseca y la relativa a un observador. La diferencia que hay entre una propiedad que es relativa a un observador y una que es independiente de un observador es, por ejemplo, la diferencia que hay entre la masa y la composición química de una moneda (propiedades independientes del observador) y su valor monetario (propiedad dependiente del observador). Por otra parte, cuando una propiedad de un objeto es intrínseca, nos referimos a que es propia del objeto, como contrapuesta a relacional. Así, la distinción entre *intencionalidad* intrínseca y relativa a un observador se refiere a aquello en virtud de lo cual un fenómeno tiene *intencionalidad*. De acuerdo con Searle, las estructuras o estados mentales tienen intencionalidad intrínsecamente, como una consecuencia causal de las propiedades físicoquímicas del cerebro: así como otros tejidos tienen la propiedad de fotosintetizar, el cerebro tiene la propiedad de representar. Es decir que, según su teoría, “nuestros cerebros son capaces de causar fenómenos mentales con contenido semántico o *intencionalidad*” (Searle, 1982, p. 57; mi traducción). De acuerdo con esto, podemos decir que, según la tesis de la *intencionalidad* intrínseca, los fenómenos con contenido semántico exhiben dicha *intencionalidad* de manera intrínseca. En las antípodas de la posición intrínsecalista encontramos teorías como la de Dennett (1998), que sostiene que la *intencionalidad* es siempre relativa al observador.⁶

Si bien Searle es el representante más emblemático de la tesis de que, a diferencia de los fenómenos *intencionales* no mentales, los estados mentales exhiben *intencionalidad* intrínseca y original, esta postura es compartida por una gran cantidad de filósofos de la mente. Podríamos decir que, en este aspecto, todos los que la defienden comparten

⁶ Dennett no es el único en rebatir la distinción entre *intencionalidad* intrínseca y derivada. También se oponen a ella importantes filósofos, como Patricia y Paul Churchland, Davidson, Millikan, Rorty, Stalnaker, entre otros (cfr. Dennett, 1998, p. 261).

una perspectiva cartesiana de lo mental. Al emplear la expresión “perspectiva cartesiana” no me refiero a tesis o ideas expuestas por el propio Descartes, sino, como es usual en la literatura sobre estos temas, a un modelo del que pueden derivarse o inferirse otras tesis que resultan compatibles con él. En los términos de Haugeland, los cartesianos son los filósofos de la mente que sostienen que la *intencionalidad* original es la “provincia exclusiva de los estados mentales con contenido” (1990, p. 132; mi traducción).

Ahora bien, la concepción de la *intencionalidad* defendida por Daniel Dennett (1990a, 1990b, 1991, 1998 y 2002) pone en jaque ambas distinciones, en la medida en que sostiene que ningún fenómeno exhibe *intencionalidad* intrínseca ni original. Según Dennett, toda *intencionalidad* es derivada y relativa al observador.

Por lo general, las tesis de la *intencionalidad* original y la de la *intencionalidad* intrínseca “vienen juntas”, tanto para quienes las defienden como para quienes las niegan. De hecho, en los argumentos que elabora para rebatir estas distinciones, Dennett las reconstruye como si fueran una sola. Por su parte, cuando Searle hace referencia a la intencionalidad de los fenómenos mentales, también emplea los términos “intrínseca” y “original” de modo indistinto.

Tanto la perspectiva cartesiana como la dennettiana entienden que los fenómenos *intencionales* no mentales exhiben *intencionalidad* derivada y relativa al observador. La diferencia entre las posturas radica en que, para Dennett, la *intencionalidad* de los fenómenos mentales también posee estas características. Por eso, de acuerdo con su perspectiva, las distinciones no tienen fundamento: no existe algo así como la *intencionalidad* original e intrínseca. El principal argumento esgrimido por Dennett para rebatir estas distinciones es una especie de legítima inversión de la carga de la prueba. Dennett considera que en realidad no tenemos más motivos para sostener las distinciones en cuestión que nuestras arraigadas, pero no fundamentadas, creencias de sentido común acerca de la naturaleza de nuestra mente (cfr. Dennett, 1990b, p. 106).

En los distintos textos en los que aborda el tema, Dennett (1990b, 1995a, 1995b y 1998) recurre a un experimento mental o “bomba de intuición”⁷ que tiene como objeto mostrar que no tenemos buenas razones

⁷ Para un desarrollo sobre las “bombas de intuición” como un recurso específico y valioso para la filosofía, cfr. Dennett (2013).

para sostener que los seres humanos somos diferentes, en los aspectos relevantes, a robots a los que nos negamos a atribuir *intencionalidad* original. En el experimento mental, se imagina que alguien tiene el deseo de ver cómo es el mundo en el año 2401. Las tecnologías disponibles ofrecerían la posibilidad de congelar y mantener en estado comatoso a esa persona para que sea despertada en ese año. Se debería construir un robot encargado de conservar y cuidar el cuerpo. El robot debería ser resistente y autosustentable energéticamente, por lo que debería contar con dispositivos de transformación y aprovechamiento de la energía. También se debería proveer al robot de un medio de traslado que le permita encontrar nuevos lugares con provisiones energéticas y escapar de los peligros. Dado que la persona estaría incapacitada para tomar decisiones, el dispositivo debería contar con cierta capacidad de autocontrol, de manera tal que sea capaz de derivar metas subsidiarias vinculadas al objetivo final de preservar la vida de la persona, en función de los distintos contextos. El resultado del diseño sería un robot con cierto grado de autonomía que, mientras persigue el objetivo final de preservar el cuerpo que aloja, se verá involucrado en distintos proyectos (muchos de los cuales pueden terminar resultando contraproducentes para la persona que mandó construirlo). Sin embargo, nadie discutiría que este robot solo tendría *intencionalidad* derivada.

La conclusión a la que pretende llegar esta bomba de intuición es la siguiente: si entendemos que el robot solo posee *intencionalidad derivada*, lo mismo también deberíamos pensar de nuestra propia *intencionalidad*. Esto se justifica porque el experimento mental elaborado por Dennett es solo una versión con variaciones de la propuesta de Dawkins (1976), según la cual los diferentes organismos vivos de nuestro planeta (incluyendo a los seres humanos) son máquinas de supervivencia que fueron diseñadas para perpetuar la existencia de sus genes “egoístas”.⁸

Esto no implica que Dennett sostenga que nuestros estados mentales están al servicio de nuestros genes. Por el contrario, el desarrollo ontogenético, que implica la introducción a un lenguaje natural y a

⁸ Si bien Dennett desarrolla este argumento para asentar una posición con respecto al funcionamiento *actual* de la mente humana (como bien señaló uno de los revisores de este artículo), la apelación a los *orígenes* de esta resulta pertinente en la medida en que sus contrincantes estarían de acuerdo en atribuir *intencionalidad* derivada y relativa al observador a un artefacto en virtud de su origen, como muestra el experimento mental de la cápsula-robot.

una cultura, hace de los seres humanos sistemas muy diferentes a una máquina de preservación de genes. Dicho desarrollo los convierte en locus de autocontrol y autodeterminación, en virtud de una integración funcional con el ambiente. Así, aunque se reconozca que el origen de la *intencionalidad* humana guarda vínculo con los “intereses” de sus genes, esta ha adquirido autonomía. Ahora bien, si aceptamos esta manera de entender nuestra propia *intencionalidad*, lo mismo debería valer para un potencial robot que exhibiera todas estas características.⁹ En definitiva, lo que importa es que no parece haber buenos argumentos para sostener que hay una diferencia de carácter ontológico que distinga el modo en el que nuestros estados mentales son acerca de otra cosa, del modo en el que lo pueden ser fenómenos no mentales.

La idea de fondo que sustenta tanto este experimento mental como la posición de Dennett respecto de este tema es que si nos tomáramos en serio la teoría de la evolución darwiniana, deberíamos asumir que las mentes de los seres humanos no son más que el resultado de un proceso histórico que las fue moldeando de modo tal que sus mecanismos de entrada y salida de información fueron adquiriendo cada vez mayor complejidad, sin existir, por ende, un límite o instancia precisa a partir de los cuales se pueda afirmar algo como “aquí es donde milagrosamente aparece la mente”. En síntesis, para Dennett, el complejo equipamiento cognitivo de los humanos es el resultado de un proceso de diseño llevado a cabo por la evolución natural y cultural.¹⁰ Teniendo en cuenta

⁹ El corazón de esta cuestión no radica en la discusión de si algún robot efectivamente exhibe características funcionales de adecuación al entorno lo suficientemente complejas o parecidas a las humanas como para afirmar que poseen una *intencionalidad* similar (la respuesta a esto probablemente sea negativa). Antes bien, la discusión es si eso es imposible por una cuestión de principios. En las antípodas a la de Dennett encontramos posiciones como la de Searle (1983), para quien resultan irrelevantes las capacidades que pueda llegar a exhibir un robot: de cualquier manera, habrá una diferencia de naturaleza entre el tipo de *intencionalidad* que les atribuimos y la *intencionalidad* exhibida por las mentes humanas.

¹⁰ Si bien en este experimento mental —que recurre provocadoramente a la idea de Dawkins (1976) de los “genes egoístas”— el énfasis principal parece estar puesto en la evolución biológica, Dennett (1995a, 1995b, 2006 y 2017) argumenta en distintas oportunidades que la evolución cultural es igual de ineludible a la hora de explicar el origen de las mentes humanas.

esto, debemos sostener que la *intencionalidad* de los fenómenos mentales es derivada y relativa al observador.

He reconstruido las principales tesis de la teoría de lo mental de Dennett, a saber: (a) que poseer estados mentales no es nada más ni nada menos que ser un sistema intencional cuyo comportamiento resulta mejor explicado y predicho si se le atribuye dichos estados, y (b) que toda la *intencionalidad* —incluyendo la de las mentes humanas— es derivada y relativa al observador. En suma, se trata de una perspectiva en la cual no hay un fundamento que permita establecer una diferencia de carácter metafísico u ontológico entre las mentes humanas y el resto del mundo natural. En palabras de Clark, la *actitud intencional* dennettiana implica que “humanos, perros, insectos e incluso meros termostatos son todos capaces de creer y desear, fundamentalmente, en el mismo sentido teórico” (Clark, 2002, p. 187; mi traducción). Esta concepción sobre la mente humana y la *intencionalidad* constituye una parte importante del enfoque dennettiano desde el cual pretendo reconstruir y profundizar un abordaje darwinista para explicar la creatividad. La otra parte importante la constituyen las ideas dennettianas sobre la creación de diseños. A esto me abocaré en el siguiente apartado.

2. Diseños sin diseñador y competencia sin comprensión

De acuerdo con Sober (2004), lo que caracteriza a los llamados “argumentos del diseño” es que encuentran objetivos deliberados en la naturaleza. Concretamente, los argumentos del diseño biológico parten de observaciones sobre los organismos que habitan el planeta Tierra, señalan que exhiben un alto grado de ajuste con respecto a su entorno (dado que las características que les permiten sobrevivir y reproducirse son, en gran medida, complejas y delicadas) y concluyen que existe un diseñador que los ha creado.

Sober se pregunta si es legítimo deducir la existencia de un diseñador inteligente a partir de las características que se observan en un objeto. Responde que se puede construir un argumento deductivo válido. Una manera de formular dicho argumento es la siguiente:

- (1) Los elementos diseñados fueron causados por un diseñador.
- (2) Los organismos biológicos son elementos diseñados.

Entonces:

- (3) Los organismos biológicos fueron causados por un diseñador (Dios).

En esta reconstrucción se entiende que un diseñador es alguien inteligente que actúa de manera intencional en la producción del diseño en cuestión. Dada la forma lógica de este argumento, si las premisas son verdaderas, la conclusión es verdadera.

Ahora bien, en contra de una interpretación bastante extendida, según la cual la teoría de la evolución de Darwin permite rebatir la segunda premisa, que afirma que los organismos son elementos diseñados,¹¹ Dennett (1995b) ha sostenido que dicha teoría no rechaza esta premisa; en cambio, la da por sentada. En sus términos: “la clave para comprender la contribución de Darwin es la *concesión* de esa premisa del argumento del diseño” (Dennett, 1995b, p. 101). Más bien, la teoría de Darwin constituye, para Dennett, una objeción a la primera premisa, dado que explica cómo es posible que haya diseños sin diseñador: es posible gracias a la lenta y persistente acción de la selección natural.

Seguindo a Darwin, Dennett señala que tratar a los organismos vivos como entidades diseñadas o diseños no supone un empleo metafórico del término “diseño”. Podemos sostener que, para Dennett, si resulta provechoso adoptar la *actitud del diseño* frente a un sistema para predecir el “comportamiento” de sus partes, dicho sistema *es un diseño*, como vimos al caracterizar la actitud del diseño en el apartado anterior. Los organismos biológicos cumplen esta condición. Como han constatado los naturalistas de siglos atrás y lo sigue haciendo la biología funcional, las características de los organismos vivos son como son por determinadas “razones”. Se podría objetar que, a diferencia de los diseños artificiales, en el caso de los productos evolutivos, esos motivos no fueron previamente representados en la mente de nadie. Con respecto a ello, Dennett (1990a) defiende que esto no hace ninguna diferencia. Desde su perspectiva, las “razones” surgieron antes de que surgieran las mentes capaces de identificarlas como tales (Dennett, 1995a y 2017). Para dar cuenta de este tipo de “razones”, Dennett acuña el concepto “razones de flotación libre” (*free-floating rationales*), que hace referencia, precisamente, a las razones por las cuales un sistema tiene determinadas

¹¹ Para un análisis profundo de esta cuestión, cfr. Ginnobili (2022).

características que le permiten funcionar adecuadamente o estar bien adaptado a su entorno, aunque no han sido representadas en ninguna mente (Dennett, 1998, pp. 211-253).

La idea de que la actitud del diseño puede ser aplicada tanto a artefactos como a organismos biológicos ha sido criticada desde distintos frentes.¹² En el marco de esta controversia, Dennett ha replicado a las objeciones de sus críticos y ha defendido que, en el campo de la biología funcional, aplicar razonamientos que suponen optimalidad (es decir, adoptar la actitud del diseño) suele conducir a resultados exitosos. A la vez, la teoría de la evolución darwiniana provee buenos fundamentos acerca de por qué esta actitud interpretativa funciona cuando se aplica a organismos vivos. En términos de Dennett: “la adopción de la perspectiva de la ingeniería en la biología no es solo ocasionalmente útil, no es solo una opción valiosa, sino el organizador obligatorio de todo el pensamiento darwiniano, y la fuente principal de su poder” (1995b, p. 187; mi traducción). Por todo esto es que resulta sensato sostener que “Darwin no acabó con la teleología, la naturalizó” (Dennett, 2017, p. 57).

He argumentado que, de acuerdo con Dennett, los organismos vivos son, estrictamente hablando, entidades diseñadas. Y que, por ende, la teoría de Darwin ha mostrado que es posible que existan diseños sin diseñador. Así, sería preciso señalar las diferencias entre el proceso evolutivo llevado a cabo por la acción de la selección natural y lo que habitualmente se entiende como la acción intencional de un diseñador. El motivo de esta diferencia es que la evolución por selección natural es, en última instancia, un proceso “ciego” y sin mente. Dennett sintetiza esta cuestión sosteniendo que se trata de un proceso algorítmico. Si bien esta caracterización ha resultado controvertida,¹³ lo relevante para

¹² Es célebre una objeción proveniente del campo de la biología planteada por Gould y Lewontin (1979), quienes han desarrollado un fuerte ataque a lo que denominan el pensamiento “panadaptacionista”. Dicho pensamiento estaría comprometido con la adopción de un único mecanismo explicativo (el cual supone la adopción de la actitud de diseño), que sería aplicable a todas las características de un organismo. De esta manera, dicho enfoque asumiría que todas las características de un organismo son adaptativas (mejoran su *fitness*) y procedería, según este supuesto, a preguntarse por las “razones de ser” de estas. Se puede ver la respuesta de Dennett a esta objeción en Dennett (1995b y 1998).

¹³ Dennett (1995b) ha sostenido que la selección natural es un proceso algorítmico dado que está constituido por un conjunto de instrucciones o reglas

el presente artículo es que, con ella, se pretende enfatizar que se trata de un proceso automático y recursivo: “primero construye una x , luego modifica los descendientes de x , luego modifica esas modificaciones, y luego modifica las modificaciones de las modificaciones” (Dennett, 1995b, p. 62; mi traducción). Así, si bien no hay un plan o una dirección predeterminada, la recursividad permite que se produzca un proceso de acumulación que da como resultado complejos y elaborados diseños. La “idea peligrosa de Darwin” es que el nivel algorítmico es el mejor nivel para explicar los ajustes de los seres vivos a su entorno. “Es difícil creer que algo tan mecánico y carente de mente como un algoritmo pueda producir cosas tan maravillosas” (Dennett, 1995b, p. 59).

Nótese que esta forma de definir el mecanismo de la selección natural es neutral con respecto al dominio. Es decir, la selección no necesariamente tiene que actuar sobre genes o sobre entidades biológicas. Por el contrario, siempre que se observa variación, reproducción diferencial y herencia, se entiende que está actuando la selección natural.¹⁴

En suma, la selección natural es un proceso que puede dar como resultado elementos con un alto grado de complejidad en su diseño,

ordenadas y finitas que permiten solucionar un problema, realizar un cómputo, o llevar a cabo otras tareas o actividades. La afirmación de que la selección natural es un proceso algorítmico ha sido discutida. Por ejemplo, Miłkowski (2009) argumenta que, dependiendo del concepto de “algoritmo” que se tome, se trata de una afirmación o bien falsa, o bien trivial. Para la presente investigación no resulta relevante si es adecuado sostener que la selección natural es un proceso algorítmico. En cambio, resultan relevantes las características de dicho proceso por las cuales Dennett sostiene que es algorítmico, a saber, el que se trate de un proceso ciego que garantiza determinados resultados en virtud de su recursividad.

¹⁴ Algunos investigadores, como Dawkins (1986) o Dennett (1995b y 2017), han sostenido que el mismo proceso de selección natural puede actuar en otros dominios, como la cultura, y que, por ende, se entiende *literalmente* que la selección natural es el proceso que actúa en la reproducción diferencial de ítems culturales. En cambio, otras teorías que aplican procesos de la misma naturaleza para el caso de la cultura hablan de una *analogía* con la selección natural. Así, utilizan la expresión “selección cultural” (Mesoudi, 2011) o alternativas que hacen referencia al tipo de proceso en términos más generales —por ejemplo, Campbell (1960) utiliza la expresión “variación ciega y retención selectiva” para referir a tales procesos—.

y que, sin embargo, es mecánico, no tiene objetivos ni una dirección determinada y es neutral con respecto al dominio. Por todo ello, Dennett sostiene que Darwin ha demostrado que puede haber diseño sin diseñador. A su parecer (Dennett, 1995b, 2001 y 2017), este componente de la teoría de Darwin ha propiciado “una extraña inversión del razonamiento”. A continuación, presentaré en qué consiste, de acuerdo con Dennett, la inversión del razonamiento darwiniano.

Para presentar las implicancias del pensamiento darwiniano, Dennett recurre a una cita de Robert Mackenzie Beverley, un crítico contemporáneo de Darwin. Al parecer, este fervoroso opositor al evolucionismo habría señalado como un gran defecto de la teoría que, según ella, “el artífice es la ignorancia absoluta [...] [;] para hacer una máquina perfecta y bella no es necesario saber cómo hacerla”, y que “Darwin parece creer que la Ignorancia Absoluta está plenamente capacitada para sustituir a la sabiduría absoluta en los logros de la habilidad creadora” (Beverley, 1868, citado por Dennett, 2017. p 59). Dennett considera que estas frases expresan de manera inmejorable los móviles que motivaban a Beverley a oponerse a la teoría darwiniana: dicha teoría implica que puede haber entidades diseñadas sin que exista una inteligencia que las haya creado. Desde su perspectiva, Beverley estaba legítimamente preocupado. Así, empleando sus mismos términos, Dennett señala que Darwin mostró que para hacer una máquina perfecta y bella no es necesario saber cómo hacerla (cfr. Dennett, 2017, p. 59). En esto consiste la inversión darwiniana del razonamiento. (Como se puede ver, es, de alguna manera, otra forma de presentar la “peligrosa idea de Darwin” según la cual el nivel algorítmico es el mejor a la hora de explicar los ajustes de los seres vivos a su entorno).

En *De las bacterias a Bach*, Dennett (2017) presenta la inversión del razonamiento de Darwin junto con otra idea que comparte una semejanza estructural con ella: la inversión del razonamiento de Turing.¹⁵ Como se

¹⁵ Como señala Schliesser (2018), Dennett (2017) refiere a ciertas implicancias de las teorías de Darwin y Turing, como “inversiones del razonamiento”. Sin embargo, en ningún momento explica qué significa esta expresión. Schliesser (2018) sostiene que, en ambos casos, las expresiones cubren dos conceptos relacionados por un parecido de familia. Dicho parecido de familia consiste en tener tres pasos cruciales. Estos pasos se podrían representar del siguiente modo: “(1) uno puede pensar que alguna propiedad/estructura/causa básica X es requerida para la explicación de las propiedades observadas Y,

sabe, Alan Turing es el creador de la llamada “máquina de Turing”: un modelo matemático computacional que define una máquina abstracta que manipula símbolos en una cinta de acuerdo con una tabla de reglas. Esta es la idea que permitió construir las computadoras modernas, por lo que Turing es considerado el padre de la informática. Ahora bien, antes de la aparición de las máquinas computadoras, ya existían personas computadoras, que efectuaban cálculos científicos y de ingeniería. Si bien estas personas conocían la aritmética y sabían lo que estaban haciendo, lo que Turing señaló es que ese conocimiento no era necesario para efectuar los cálculos. Turing mostró que es posible diseñar máquinas absolutamente ignorantes que —siguiendo instrucciones programadas de manera mecánica— realicen cálculos aritméticos perfectamente. Así, Turing mostró que “para ser una máquina de calcular perfecta y bella no es necesario saber aritmética” (Dennett, 2017, p. 60).

Ahora bien, ambas inversiones del razonamiento tienen en común el supuesto de que puede haber competencia sin comprensión:

Lo que Darwin y Turing concibieron fue la versión más radical de esta idea: toda la comprensión y la genialidad de este mundo surge, en último término, de competencias sin comprensión, que se van combinando a lo largo del tiempo y forman sistemas cada vez más competentes (y, por ende, también más capaces de comprender) (Dennett, 2017, p. 62).

Ya no resulta válido inferir que si hay competencia, entonces hay comprensión, que es algo que se suele dar por sentado,¹⁶ sino que se entiende que, en última instancia, la competencia ha dado lugar a la

(2) donde X exhibe (al menos) algo de inteligencia/mentalidad/intencionalidad y las Y típicamente no lo hacen, (3) pero, en realidad, las propiedades Y son necesarias para explicar X” (Schliesser, 2018, p. 3; mi traducción). Así, tanto la inversión del razonamiento de Turing como la inversión del razonamiento de Darwin se podrían reconstruir mediante un ajuste al esquema propuesto anteriormente.

¹⁶ Dennett (2017) también sostiene que la idea de que detrás de la competencia hay comprensión nos resulta tan intuitiva porque está presente en el *Umwelt* que la evolución ha diseñado como parte de nuestro equipamiento cognitivo, es decir, la adopción de la actitud intencional como una opción básica para lidiar con otros agentes.

comprensión. Así, “todo ser vivo —desde las bacterias hasta Bach— persiste y se reproduce a fuerza de trucos inteligentes” (De Sousa, 2018, p. 113; mi traducción), pero, en un comienzo, todos estos trucos se realizaban sin comprensión de ellos.

¿Qué está entendiendo Dennett por “comprensión”? Como se sabe, para Dennett, todos los seres vivos están diseñados de tal forma que tanto sus partes como sus comportamientos presentan rasgos de diseño: razones por las cuales son como son, que les permiten desempeñar exitosamente determinadas funciones. Ahora bien, existe un pequeño grupo dentro de estos seres vivos que es capaz de representarse dichas razones. Así, restringiéndonos a este contexto, la comprensión puede ser entendida como la capacidad de los organismos para representar (o incorporar con inteligencia de alguna otra forma) las motivaciones de sus diseños para la supervivencia (cfr. Dennett, 2017, p. 88). De este modo, la comprensión permite aplicar los aprendizajes obtenidos por la experiencia a nuevos temas y elementos (cfr. Dennett, 2017, p. 99). En contra de la idea tradicional de “comprensión”, Dennett ha argumentado que la capacidad de comprender debe ser entendida, como ocurre con la gran mayoría de los productos evolutivos, como una cuestión de grados y no “de todo o nada” (Dennett, 2017, p. 95).

En términos de Dennett (2017), de acuerdo con las inversiones del razonamiento llevadas a cabo por Darwin y Turing, se “derroca el punto de vista pre-darwiniano de la creación como un asunto de ‘al principio, la mente’ y lo sustituye por una perspectiva de ‘al final, la mente’ acerca de nuestra propia evolución, como diseñadores inteligentes que somos” (Dennett, 2017, p. 63). Esto quiere decir que las ideas de Darwin y Turing han servido para mostrar que no hay por qué asumir que la comprensión es una condición necesaria para la competencia. Antes bien, la comprensión es el resultado de una gran cantidad de competencia acumulada. A partir de lo expuesto podemos sostener que, según el naturalismo evolucionista de Dennett, la mente y la comprensión no son la fuente de todo diseño, sino productos recientes de mucho diseño acumulado.

En este apartado he reconstruido algunas nociones básicas del darwinismo dennettiano, fundamentalmente las ideas, con fuertes vínculos entre sí, de que es posible que existan diseños sin diseñador y que algunas competencias complejas no precisan comprensión. Por ejemplo, no necesariamente requieren comprensión ni la competencia involucrada en el diseño de algo nuevo, complejo y funcional (inversión

del razonamiento de Darwin) ni la competencia involucrada en la realización de operaciones aritméticas (inversión del razonamiento de Turing). De acuerdo con Dennett, estas ideas se derivan de la teoría de la evolución de Darwin y también de los aportes de Turing, pero van a contramano de nuestras asunciones de sentido común e, incluso, de algunos supuestos que están presentes en teorías científicas y filosóficas.

Si estas ideas del darwinismo dennettiano resultan controvertidas es porque tenemos incorporado de forma acrítica un supuesto como el siguiente: si observamos un producto que ostenta un alto grado de complejidad en su diseño, asumimos que el proceso que llevó a su producción tuvo determinadas características. Específicamente, asumimos que se trata de un proceso llevado a cabo por un agente intencional, actuando de modo inteligente o exhibiendo comprensión en las diferentes etapas de su concepción y ejecución. De hecho, ese es el supuesto del que se sirve el argumento del diseño.

He presentado el argumento del diseño biológico. Sin embargo, considero que también se podría construir un argumento similar para el ámbito de la cultura modificando la segunda premisa del siguiente modo:

- (1) Los elementos diseñados fueron causados por un diseñador.
- (2) Los ítems culturales (artefactos, teorías, costumbres, etc.) son elementos diseñados.

Entonces:

- (3) Los ítems culturales fueron causados por un diseñador.

Como señalé anteriormente, según Dennett, el pensamiento darwiniano rebate la primera premisa del argumento del diseño biológico, que está presente también en esta versión culturalista. Por ende, esta versión del argumento también se encontraría en problemas.

Ahora bien, aunque el argumento del diseño ha sido cuestionado para el caso del mundo natural, sigue operando, en gran medida, para el ámbito de la cultura. Así, cuando nos encontramos frente a un producto cultural que consideramos novedoso y con un complejo diseño que lo hace valioso o útil, solemos inferir que el proceso que lo generó estuvo

guiado por acciones que fueron llevadas a cabo de manera intencional por un sujeto inteligente. Entiendo que la perspectiva dennettiana ofrece un enfoque que permite rebatir esta visión desde un ángulo novedoso. Profundizaré esta idea en los siguientes apartados.

3. Estados mentales y creatividad

En el primer apartado he expuesto la concepción de la *intencionalidad* de Dennett, según la cual no hay una diferencia de naturaleza entre el carácter *intencional* de los fenómenos mentales y no mentales. De acuerdo con dicha concepción, la mente no puede ser considerada excepcional, al menos en ese sentido específico. En el segundo apartado he expuesto los argumentos de Dennett de acuerdo con los cuales es posible y esperable que algunos procesos que poseen características como las de ser mecánicos, no exhibir comprensión ni estar mediados por objetivos pueden dar como resultado diseños complejos. En otros términos, no todo diseño tiene un diseñador. Esto también implica que no toda competencia requiere comprensión. Considero que si juntamos estos elementos obtenemos un marco propicio para sostener que ni los procesos mentales ni las intenciones y la comprensión tienen porqué ser considerados como elementos centrales en todos los procesos creativos.

En contraposición con la perspectiva dennettiana que vengo exponiendo, desde una concepción predarwiniana o cartesiana¹⁷ de lo mental resultaría sensata la idea de que los estados mentales del creador son cruciales para la explicación de los procesos creativos. Según una concepción predarwiniana o cartesiana, siempre que hay diseño hay una mente detrás, dado que solo las mentes pueden propiciar determinado ordenamiento de la materia orientándola a fines. Esto sería así porque solo la mente, a diferencia del resto de los fenómenos de la naturaleza, es considerada capaz de exhibir comprensión. En ese sentido, estos puntos de vista tradicionales confieren un carácter excepcional a lo mental (otros sistemas, como un algoritmo computacional, pueden *aparentar* comprender, pero no pueden comprender *realmente*). A su vez, esta idea se sustenta, en última instancia, en la tesis de que la *intencionalidad* exhibida por las mentes es de una naturaleza diferente: es intrínseca y original.

¹⁷ Si bien los términos “predarwiniano” y “cartesiano” no son equivalentes, guardan entre sí importantes puntos de contacto; lo que pretendo argumentar en este párrafo aplica a ambos términos por igual.

Ahora bien, una vez que la concepción dennettiana de la *intencionalidad* y de la creación de diseños nos obliga a tirar por la borda estas nociones predarwinianas y cartesianas, debemos objetar la idea de que los estados mentales del creador juegan siempre un rol decisivo en los procesos creativos.

Este razonamiento debería afectar fuertemente a distintas teorías de la creatividad que, de formas más o menos explícitas, se comprometen con la tesis de que los estados mentales de los creadores desempeñan un rol explicativo central en la génesis de los procesos creativos. El caso en el que esto ocurre de modo más explícito es la concepción intencionalista de Gaut (2010; Gaut y Livingston, 2003). Como señalé en la introducción, de acuerdo con la definición estándar de la creatividad, un producto creativo debe ser original y valioso. De acuerdo con Gaut, estos requisitos no son suficientes para que un resultado sea considerado creativo. En cambio, es necesario incluir de alguna manera la dimensión de la agencia intencional involucrada en el proceso que llevó al resultado. De acuerdo con su visión, además de original y valiosa, una idea o artefacto creativo debe ser el resultado de un proceso llevado a cabo por un agente que actúa intencionalmente.

Gaut (2010) argumenta que, si no agregamos el tercer requisito, tendremos que considerar como procesos creativos algunos que claramente no merecen tal atributo. Es decir, tendremos un problema de sobreatribución. Propone que pensemos, por ejemplo, en un diamante. Se trata de un producto que puede ser original en su forma y es sin dudas valioso. Gaut sostiene que, de acuerdo con la definición estándar de la creatividad, el movimiento de las placas tectónicas que da como resultado a ese diamante debería ser considerado creativo y que eso no tiene sentido.

Considero que al argumento de Gaut se puede responder señalando que la reconstrucción de la definición estándar que supone es inadecuada. Cuando dicha definición sostiene que un producto creativo, además de ser original, debe ser “valioso”, se utiliza ese término para intentar abarcar los distintos sentidos según los cuales una nueva contribución puede resultar interesante en el dominio en el que se inserta.¹⁸ El tipo

¹⁸ Es por ello que, como bien señala un revisor de este artículo, el término empleado por Runcó y Jaeger (2012), “eficacia”, resulta más adecuado para referir a la segunda dimensión capturada por la definición estándar de la creatividad, en comparación con “valor”. Sin embargo, he decidido conservar

de valor adquirirá una forma diferente en virtud de que se trate de una teoría científica, de una contribución artística o de un artefacto tecnológico (Simonton, 1999). En ninguno de estos sentidos se podría decir que un diamante es valioso, más allá de que tenga un enorme valor comercial.

Ahora bien, más allá del argumento de Gaut en particular, muchas teorías de la creatividad parten de la idea de que lo que hay que explicar es cómo determinados individuos excepcionales fueron atravesados por el rayo de la iluminación para llevar a cabo descubrimientos extraordinarios. Por ello, los casos y ejemplos que se consideran son historias sobre un químico que se queda dormido frente al fuego, un físico a quien le cae una manzana en la cabeza, un poeta que tiene un sueño demasiado vívido en el que lo visitan las musas, en lugar de relatar la historia de necesidades, accidentes y acumulación gradual de diseño que hay detrás de cualquier producto cultural. En clara contraposición con estos enfoques tradicionales, Dennett nos brinda buenas razones para considerar que los estados mentales de un individuo (los deseos, las intenciones, las creencias) no necesariamente son un componente central al que debemos atender para considerar al resultado creativo como tal. Una vez que se deja de lado la suposición de que lo son, es posible modificar nuestra concepción del siguiente modo: incluyendo también a los procesos de acumulación transgeneracional dentro del conjunto de procesos de la creatividad humana que debería considerar cualquier explicación completa sobre el fenómeno.

En suma, en este apartado he defendido que una consecuencia de adoptar una concepción dennettiana de la mente y el diseño es que ya no se considera que los estados mentales desempeñan siempre un rol decisivo en los procesos creativos. Desde una visión predarwiniana o cartesiana, cabe esperar que los estados mentales, considerados tanto excepcionales como la fuente de todo diseño, ocupen un rol central, como componentes indispensables, en toda explicación de los procesos creativos. Desde una visión dennettiana, este papel es rebatible principalmente por dos motivos. Por una parte, porque para Dennett, toda *intencionalidad* es derivada y, por ende, los estados mentales ya no poseen un rasgo distintivo —la intencionalidad original— ausente

en general la formulación de “originalidad y valor” porque es la que prevalece en la literatura sobre creatividad y porque es la que retoma Gaut para elaborar el argumento que aquí pretendo contradecir.

en el resto del mundo natural. Luego, se desdibujan los motivos para pensar que, si algo novedoso y valioso empieza a existir en el mundo, los estados mentales *intencionales* deben desempeñar un rol ineludible e irremplazable en su génesis. Por la otra, porque, de acuerdo con Dennett, existen procesos no mentales (que, por ende, no podrían exhibir comprensión) que pueden dar resultados originales y novedosos.

Hasta el momento he llegado a una conclusión que se puede considerar negativa o crítica: el marco darwinista dennettiano conduce a rechazar la idea de que los estados mentales de los creadores son un elemento crucial a la hora de explicar todos los procesos de creación humana. En el siguiente apartado intentaré formular de manera propositiva qué elementos formarían parte de una concepción dennettiana de la creatividad.

4. ¿Cómo entender la creatividad desde una perspectiva darwinista dennettiana?

En un artículo cuyo título se podría traducir como “En la estela de Darwin, ¿dónde estoy?”, Dennett (2001) sostuvo que la teoría darwiniana debería haber cambiado nuestro modo de concebir la creatividad. Así, señala que habría que elaborar una concepción darwiniana de la creatividad. En sus términos, se trataría de una concepción según la cual “todas las obras del genio humano pueden ser entendidas, a fin de cuentas, como productos generados mecánicamente por una cascada de algoritmos de generación y testeo” (Dennett, 2001, p. 14; mi traducción). Sin embargo, como señalé anteriormente, Dennett no profundiza en el cambio de concepción que supone este enfoque.¹⁹ Para responder a la pregunta de qué ocurre cuando el “ácido universal” darwiniano entra en contacto con la creatividad, a lo largo de este apartado me apoyaré en las escasas y aisladas consideraciones que Dennett ha realizado con

¹⁹ A decir verdad, el artículo explora las consecuencias que tendría una concepción darwiniana de la creatividad (descrita en los términos recién expuestos), centrándose en sus implicancias en torno a la disolución de la noción de “autoría”, lo que, a su vez, pondría en jaque nuestra concepción del “yo”. Además, explora las consecuencias de adoptar una concepción darwiniana a la hora de evaluar si un algoritmo como el de un programa informático puede desempeñarse creativamente. Con respecto a esto, señala que no tenemos buenos motivos para sostener que un programa informático no puede ser considerado creativo.

respecto a la creatividad y las analizaré a la luz de otros elementos de su pensamiento filosófico.

Dos nociones dennettianas resultan centrales para elaborar este enfoque. Se trata de las nociones metafóricas de “grúa” y “gancho celeste”, postuladas con el fin de diferenciar aquellas explicaciones que son, al menos en algunos aspectos, consecuentes con una perspectiva evolutiva darwiniana, de otras explicaciones que no lo son. Para referir a las segundas, Dennett apela a un término folklórico en inglés: *skyhook*, que podríamos traducir como “gancho celeste”. En el campo de la aeronáutica, dicho término se refiere a un invento imaginario que permitiría mantener suspendido un objeto en el cielo (Dennett, 1995b).²⁰

En el contexto de los debates sobre las explicaciones evolucionistas de la mente, la sugerencia de Dennett es que la apelación a una suerte de gancho celeste para explicar algún fenómeno complejo implica renunciar al intento de elaborar una explicación naturalista. Así, cualquier intento serio de explicar un fenómeno natural debería recurrir, en cambio, a las “grúas”. Dennett introduce este segundo concepto metafórico, dándole un significado opuesto al anterior, para referir a las explicaciones que consiguen dar cuenta de fenómenos de alta complejidad (y que en ese sentido pueden entenderse como “elevados”) apelando exclusivamente a los mecanismos no inteligentes de los procesos de la evolución natural. Las grúas apelan a mecanismos evolutivos en los que la acumulación de los procesos de diseño da cuenta de ciertos resultados más complejos, constituyendo, en definitiva, formas de “aumentar” localmente el poder del proceso de selección natural. En suma, por medio de estas metáforas, Dennett ha propuesto abogar por un tipo de abordaje que permita explicar una variedad de fenómenos altamente complejos sin presuponer un principio explicativo superior en el origen.

Empleando el contraste entre estos dos conceptos, la creatividad humana podría entenderse como una grúa y no como un gancho

²⁰ Dennett señala que el origen del término puede tener que ver con la idea del *deus ex machina* de los antiguos dramaturgos griegos: un dios que interviene sobre el final de una tragedia griega y que resuelve de forma sobrenatural un problema que parecía sin salida. Así, el término en aquel contexto sugiere que emplear un “gancho celeste” es introducir un elemento externo para resolver un problema de la historia, que, a su vez, colisiona con la propia lógica interna de esta. Se trata entonces de una especie de solución externa y, de algún modo, superior.

celeste. En contraposición con este enfoque darwiniano, una perspectiva cartesiana de la creatividad asumiría que “siempre se necesita una cosa inteligente, grande y sofisticada para crear una de menor valor” (Dennett, 2001, p. 14; mi traducción).

En Dennett (2017) se hace referencia, bajo un nuevo nombre, a estas dos formas de aproximarse al fenómeno de la creatividad. Ahí Dennett contrapone una teoría de la creación *por burbujeo* (darwiniana) y una teoría de la creación *por goteo* (*trickle-down*) (cartesiana). Como las metáforas sugieren, mientras que la creación por burbujeo es, en última instancia, un fenómeno que se puede explicar “de abajo arriba”, la creación por goteo opera “de arriba abajo”. Se trata de nuevos términos que hacen referencia a la misma idea que venimos examinando.

En suma, un enfoque darwiniano de la creatividad es aquel que es compatible con la idea de que todos los diseños son, en último término,²¹ el resultado de una actividad mecánica no inteligente y sin objetivos. Para entender los alcances de esta afirmación, la siguiente cita puede resultar iluminadora. De acuerdo con Dennett:

Si los mecanismos darwinianos pueden explicar la existencia de una alondra, en todo su esplendor, pueden seguramente explicar la existencia de una oda para un ruiseñor, también. Un poema es una cosa maravillosa,

²¹ Hay cierta ambigüedad en este señalamiento. Cuando Dennett afirma que todos los diseños son, en última instancia, producto de una actividad mecánica y sin objetivos, ¿está señalando, sencillamente, que el *origen* último de todos los fenómenos vinculados a la vida en la Tierra se debe a la acción de la selección natural? Si así fuera, se trataría de una tesis nada controvertida para cualquiera que crea que la teoría de la evolución darwiniana es, en lo fundamental, correcta. De hecho, podría pensarse que la evolución darwiniana dio origen a nuestras mentes, que son una grúa completamente inteligente, que procede de arriba abajo, con un grado de comprensión altísimo (ya veremos que Dennett no niega que exista un modo humano inteligente de diseñar, distinto al darwiniano). Según este razonamiento, el enfoque darwiniano no entraría en conflicto con una concepción de la creatividad según la cual los procesos creativos estarían guiados por planes inteligentes y serían, en última instancia, explicados por las habilidades cognitivas individuales. Sin embargo, Dennett parece extraer otras implicancias de la tesis del origen darwiniano de la mente humana: implicancias que deberían alterar nuestra comprensión sobre el modo humano de diseñar. Ahondaré en este punto a continuación.

pero no claramente más maravillosa que una alondra viva, cantando (2001, p. 15; mi traducción).

Con esta frase, se nos invita a reflexionar sobre el hecho de que no hay nada esencialmente misterioso en los diseños producidos por los seres humanos. Antes bien, un proceso ciego y sin objetivos, como el producido por la selección natural, puede dar lugar a diseños incluso más complejos y maravillosos. Esto debería modificar la manera en la que solemos concebir la creación humana. No sería necesario recurrir a un gancho celeste cuando una torre de grúas montadas en cadena puede producir el mismo resultado.

Ahora bien, es menester señalar que la propuesta de Dennett, según la cual la teoría de Darwin debería hacernos “invertir” nuestra forma de concebir la creación, no implica un completo abandono de la idea de “diseño inteligente humano” en su acepción tradicional. Así, la concepción dennettiana de la mente humana reconoce que existe una diferencia entre el modo darwiniano de diseñar y el modo inteligente de diseñar. El diseño inteligente comienza con un problema u objetivo y trabaja “de arriba abajo”. Por su parte, el diseño por selección natural, o evolución, es un proceso que carece de objetivos y procede “de abajo arriba”, conservando lo no perjudicial, de manera miope y sin dirección. De este modo, mientras que el diseño por selección natural implica competencia sin comprensión y razones de flotación libre, el diseño inteligente tiende a involucrar algún tipo de comprensión y representación de las razones por las cuales el producto diseñado es del modo que es.

Ahora bien, si en última instancia la perspectiva dennettiana de la mente humana no niega que exista un modo de diseñar inteligente, que procede “de arriba abajo” y que involucra comprensión, ¿por qué habría de pensarse que dicha perspectiva implica un cambio en la concepción de la creatividad? En otros términos, ¿qué implicancias tendría concebir al modo de crear humano, incluso cuando involucra comprensión, como una grúa? Considero que concebir la capacidad creativa humana inteligente como una grúa tiene implicancias de dos tipos: por una parte, con respecto al origen que debe atribuírsele y, por la otra, con respecto a cuán ubicua debe entenderse que es (es decir, con respecto a su alcance).

Con respecto a su origen, Dennett propone que la cultura humana y el diseño inteligente son, por supuesto, el resultado del proceso de evolución por selección natural. Ahora bien, eso no significa que

necesariamente sean completamente darwinianos en su naturaleza. En sus términos:

Mi proposición es que la cultura humana se inició como un fenómeno profundamente darwiniano, con competencias sin comprensión que generaban diversas estructuras valiosas de modo parecido a como las termitas producen sus castillos, y luego, poco a poco, se desdarwinizó, aumentando su comprensión y su capacidad de organización de arriba abajo, aumentando su eficiencia para buscar por el espacio de diseño (Dennett, 2017, p. 138).

Para exponer esta idea, Dennett recurre a la noción de “espacios darwinianos”, propuesta por Godfrey-Smith (2009). Dicha noción permite pensar que el hecho de que un fenómeno sea darwiniano o no no es una cuestión “de todo o nada”.²² Así, por ejemplo, un proceso darwiniano “paradigmático” es uno en el que la variación es completamente aleatoria, la fidelidad de las copias que se producen es bastante alta y no se encuentra control “de arriba abajo” ni comprensión alguna. Un proceso de este tipo es el que ha dado lugar, por ejemplo, a la forma como las termitas construyen castillos. Según Dennett, los orígenes del lenguaje humano también habrían tenido lugar gracias a un proceso de este tipo, debido a la selección natural operando sobre las palabras. Ahora bien, a lo largo de miles de años, la cultura humana fue permitiendo incorporar búsquedas más direccionadas, mayor control “de arriba abajo” y una mayor cantidad de comprensión (cfr.

²² Para defender su propuesta, Godfrey-Smith (2009) elabora representaciones tridimensionales que permitirían visualizar dónde se ubica un proceso que produce cambio. La representación consiste en un cubo en el que una de las aristas es el lugar que ocupan los procesos paradigmáticamente darwinianos y la arista que se encuentra más lejos de ella, los procesos que no son darwinianos. Cada una de las tres dimensiones representa una variable, por ejemplo, la fidelidad de las copias que se producen en el proceso en cuestión. En la arista darwiniana encontramos los procesos que tienen copias de muy alta fidelidad, mientras que en el sentido opuesto se encuentran aquellos en los que el “ruido” del copiado es muy grande. Estas representaciones permiten incluir tres de estas variables por vez y ubicar, según esos criterios, a distintos procesos que pueden ser puramente darwinianos, cuasidarwinianos, protodarwinianos, o no ser darwinianos en absoluto.

Dennett, 2017, p. 254). Así, en la actualidad existirían formas del diseño inteligente que proceden de un modo completamente diferente al modo en el que lo hace el proceso ciego y sin objetivos de la evolución por selección natural.

Si esto fuera lo único que puede decir la perspectiva dennettiana de la mente sobre el modo inteligente de diseñar, dicha perspectiva no implicaría un cambio tan radical con respecto a nuestra comprensión clásica de la creatividad. Sin embargo, la perspectiva dennettiana permite, además, ubicar esta forma de crear diseño “de arriba abajo” como una más entre otras y sostener que, de hecho, es un fenómeno mucho más restringido de lo que creemos. En palabras de Dennett:

El diseño inteligente de arriba a abajo funciona [...]. Antes de Darwin se consideraba la única manera de lograr un diseño; el diseño sin un diseñador inteligente se juzgaba imposible. Pero resulta que este diseño de arriba abajo, en realidad, *es el responsable de mucho menos diseño en el mundo de lo que se suele creer*, y para algunos de los “logros de la habilidad creadora”, repitiendo una vez más las palabras de Beverley, este enfoque ha resultado inútil (2017, p. 77; mi énfasis).

Así, el diseño inteligente “de arriba abajo” es solamente uno de los modos de producir diseño, entre otros. Además del diseño inteligente “de arriba abajo”, otro de los mecanismos centrales en la producción de diseños presente en la cultura humana lo constituye la evolución cultural (Dennett, 1991^a y 2017).²³

El de la evolución cultural es un campo científico creciente que intenta proveer una explicación naturalista y cuantitativa del cambio cultural, tanto en especies humanas como no humanas (Mesoudi, 2011). Si bien es un campo compuesto por teorías que guardan diferencias entre sí, en general sostienen que los cambios en la cultura (es decir, en las creencias, los conocimientos, la tecnología, las lenguas, las instituciones

²³ En esta enumeración de diferentes modos de producir diseños también se podría agregar a la selección natural de genes que opera en la evolución biológica, de la que he hablado anteriormente. Sin embargo, no se incluye en este párrafo porque este está dedicado específicamente a la producción de diseños en la cultura humana.

sociales, etc.) se pueden explicar por principios darwinianos (Acerbi y Mesoudi, 2015).

En términos generales, todas estas teorías caracterizan a la cultura como información que se transmite por aprendizaje social (Mesoudi, 2011; Richerson y Boyd, 2005; Sperber, 1996). Es decir, se trata de información con la que los organismos cuentan para adaptarse a su medio, pero que no ha sido heredada biológicamente (genética o epigenéticamente), ni se ha adquirido por aprendizaje individual. Así, es un campo de estudio que pretende explicar parte del comportamiento humano que no se podía explicar apelando a diferencias genéticas ni ambientales (Richerson y Boyd, 2005).

Las unidades informacionales reciben distintos nombres de acuerdo con las teorías: variantes culturales, ítems culturales, rasgos culturales, memes. Este último es el término acuñado por la teoría memética, fundada por Dawkins (1976) y defendida por el mismo Dennett (1995a, 1995b y 2017), entre otros. De acuerdo con la memética, la cultura evoluciona por selección natural de memes. Si bien ha sido objeto de diferentes críticas, Dennett (2017) realiza una consistente defensa de la memética, sosteniendo que es altamente compatible con la gran mayoría de teorías de la evolución cultural.

Las diferentes teorías de evolución cultural explican cómo esta información se produce, se transmite, se expande o desaparece. Pero también entienden que dicha información es el producto de los procesos evolutivo-culturales (Sterelny, 2006; Dennett, 1995b y 2017; Mesoudi, 2011). En otros términos, sin la acumulación transgeneracional, esos cuerpos de información no existirían: se trata de resultados que no podrían haber sido inventados por una persona. Antes bien, son descubrimientos, algunas veces buscados intencionalmente, otras veces más azarosos, que las diferentes comunidades a lo largo de la historia han considerado importantes y han decidido conservar. Sin la lenta pero persistente acción de la preservación y la acumulación de diseño que ocurre a lo largo de las generaciones, estos ítems culturales no serían posibles.

Así, pareciera que las teorías de evolución cultural y las teorías de la creatividad podrían establecer vínculos virtuosos entre sí. De acuerdo con esta visión, los procesos presentes en la evolución cultural son relevantes para entender el surgimiento de productos creativos.

Sin embargo, de hecho las teorías de la creatividad y de la evolución cultural entran en contacto solo excepcionalmente (Fogarty *et al.*, 2015).²⁴

Recapitulando, en lo que respecta a las creaciones humanas, podemos tomar en consideración al menos dos tipos de procesos de diseño: el de la evolución cultural y el del diseño inteligente (el hecho de que se puedan distinguir analíticamente e identificar casos más paradigmáticos de cada tipo de proceso no significa que no operen de manera combinada e híbrida). Estos dos tipos de procesos de diseño actúan a distintos niveles. Así, una explicación que apele a un proceso de diseño inteligente, de “arriba abajo”, tomará como punto de partida, presumiblemente, el *nivel personal* (Dennett, 2002) y las habilidades cognitivas individuales. En cambio, una explicación que apele a procesos de evolución cultural hará referencia al nivel sociocultural o supraindividual y tendrá en cuenta habilidades cognitivas y no cognitivas. En términos de biología evolutiva, deberemos decir que tendrá en cuenta el nivel de las comunidades o poblaciones (Sober, 2006; Baravalle, 2017). De este modo, podemos sostener que un enfoque dennettiano darwiniano de la creatividad es de carácter pluralista en la medida en que apele a procesos que operan en diferentes niveles y son de diverso tipo.

De esta manera, la perspectiva dennettiana permite sostener que existen diferentes maneras de diseñar, algunas más darwinianas que otras. El diseño humano inteligente, que involucra comprensión, es el resultado de incontables herramientas cognitivas construidas por la acción de la selección natural y cultural y, por otra parte, es un fenómeno mucho más restringido de lo que solemos asumir.

Alguien podría pensar que esta concepción pluralista no tiene muchas implicancias para una teoría de la creatividad humana, dado que esta debería encargarse, precisamente, de los procesos de diseño inteligente “de arriba abajo”, por más restringido que sea su alcance. Sin embargo, la perspectiva de Dennett sobre la mente humana y los diseños expuesta en los apartados anteriores —según la cual no existen diferencias entre el modo en el que exhiben *intencionalidad* los fenómenos mentales y los no mentales y según la cual un diseño lo es con independencia de que las razones de su estructura hayan sido o

²⁴ En León (2023) se puede encontrar un desarrollo más elaborado de cómo se podría entender la idea de que los procesos de evolución cultural pueden desempeñar un papel crucial en algunos procesos de creatividad humana, incluyendo algunos casos de análisis que pueden resultar ilustrativos.

no representadas en la mente de alguien— brinda un marco desde el cual se puede objetar la idea de que algunas formas de producir diseño son más relevantes para la creación humana que otras. Así, considero que, desde una perspectiva dennettiana, un abordaje de la creatividad humana debería encargarse de las diversas formas de producir diseño presentes en la cultura humana. Esto implicaría prestar atención a un espacio hasta ahora inexplorado: el lugar que los procesos evolutivo-culturales han desempeñado en la creación humana.

Conclusiones

En este artículo he analizado qué ocurre cuando el “ácido universal” darwiniano entra en contacto con la creatividad humana. En otros términos, he desarrollado algunos lineamientos para la elaboración de un enfoque darwinista dennettiano de la creatividad humana.

En los primeros dos apartados he presentado algunas tesis y argumentos elaborados por Dennett con respecto a la mente humana y la producción de diseños, que consideré que tendrían efectos en la concepción de la creatividad. Concretamente, en el primer apartado presenté la concepción dennettiana de la mente humana y de la *intencionalidad*. Sostuve que, de acuerdo con Dennett, no existe una distinción entre la *intencionalidad* original e intrínseca (asignada a los estados mentales) y la *intencionalidad* derivada y relativa al observador. Antes bien, toda *intencionalidad* es derivada y relativa al observador. En el segundo apartado señalé que, de acuerdo con Dennett, una de las implicancias de la teoría de la evolución de Darwin es que esta ha demostrado que no todo diseño tiene un diseñador. En sus términos, no es necesario saber cómo hacer una máquina perfecta y bella para poder crearla. Se trata de una competencia que no necesariamente requiere comprensión de aquello que se crea ni de las razones por las que funciona.

En el tercer apartado he argumentado que las tesis y argumentos presentados en los primeros dos conducen a objetar la idea de que los estados mentales necesariamente desempeñan papeles decisivos en los procesos creativos. Sostuve que esta consideración debe tener efectos en las teorías de la creatividad porque en general en dichas teorías se confiere un rol explicativo central a los estados mentales de los creadores.

Finalmente, me he propuesto sistematizar y desarrollar el marco darwinista dennettiano en términos más propositivos. Con ese objetivo, en el cuarto apartado recuperé las sugerencias dispersas que Dennett ha

realizado con respecto a cómo debería ser una concepción darwiniana de la creatividad. Poner estos señalamientos en relación con otras tesis y nociones filosóficas de Dennett me permitió argumentar que, de acuerdo con este enfoque, los procesos de creación en la cultura humana son de diversos tipos. Algunos de ellos se pueden identificar con lo que tradicionalmente se ha entendido como diseño inteligente: procesos que comienzan con un problema u objetivo y trabajan “de arriba abajo”, con intervención de la comprensión. Ahora bien, desde este abordaje, este tipo de procesos ocupan en la cultura humana un lugar mucho más restringido de lo que se suele pensar: tienen un alcance más limitado del que se les ha atribuido tradicionalmente. A su vez, el enfoque darwinista dennettiano nos permite considerar procesos de creación que carecen de objetivos y proceden “de abajo arriba”, conservando lo que no es perjudicial, sin una dirección predeterminada. Se trataría de procesos darwinianos que son estudiados por la evolución cultural. Así, desde la perspectiva que he analizado se entendería que los procesos postulados por las teorías de la evolución cultural deben ser tenidos en cuenta a la hora de estudiar la creatividad humana. Mientras que los procesos de diseño inteligente operan en un nivel personal o individual, los analizados por la evolución cultural lo hacen en el nivel supraindividual o comunitario.

En suma, he defendido que el marco darwinista dennettiano de la creatividad es pluralista —dado que incorpora procesos de diferente tipo y que operan en distintos niveles—, que incluye explicaciones que tienen en cuenta el carácter acumulativo del diseño y contempla que los procesos creativos son, en gran medida, colaborativos y no necesariamente involucran un alto grado de comprensión. Considero que este enfoque permitiría cuestionar la preeminencia teórica que se le ha dado a las habilidades cognitivas individuales a la hora de explicar el carácter original y funcional de los logros creativos.

He señalado que alguien podría considerar que los argumentos que he elaborado no deberían tener efectos en las teorías de la creatividad, en la medida en que dichas teorías deberían ocuparse precisamente de aquellos procesos en los que la mente humana procede mediante objetivos y “de arriba abajo”, exhibiendo comprensión del proceso que lleva a cabo. He argumentado que la concepción dennettiana de la mente y del diseño permite objetar la idea de que algunas formas de producir diseño son más relevantes para la creación humana que otras. Antes bien, una concepción darwinista dennettiana de la creatividad humana

debe tener en cuenta tanto el diseño inteligente como los procesos acumulativos transgeneracionales postulados por la evolución cultural. Espero que los desarrollos aquí discutidos puedan ser considerados un ejemplo adicional del poder transformador que Dennett atribuye a la teoría evolutiva darwiniana, concebida como un “ácido universal” que reconfigura nuestra aproximación a los fenómenos que aborda.²⁵

Referencias

- Acerbi, A. y Mesoudi, A. (2015). If We Are All Cultural Darwinians What's the Fuss About? Clarifying Recent Disagreements in the Field of Cultural Evolution. *Biology and Philosophy*, 30(4), 481-503. <https://doi.org/10.1007/s10539-015-9490-2>
- Amabile, T. (1983). *The Social Psychology of Creativity*. Springer. <https://doi.org/10.1007/978-1-4612-5533-8>
- Baravalle, L. (2017). El papel del pensamiento poblacional en la teoría de la doble herencia. *Scientiae Studia*, 15(2), 283-305. <https://doi.org/10.11606/51678-31662017000200005>
- Barron, F. (1955). The Disposition toward Originality. *Journal of Abnormal and Social Psychology*, 51, 478-485. <https://doi.org/10.1037/h0048073>
- Boden, M. (2002). *The Creative Mind: Myths and Mechanisms*. Routledge.
- Brentano, F. (2012). *Psychology from an Empirical Standpoint*. A. C. Rancurello, D. B. Terrell y L. L. McAlister (trads.). Routledge. <https://doi.org/10.4324/9780203202883>
- Briskman, L. (2009). Creative Product and Creative Process in Science and Art. En M. Krausz, D. Dutton y K. Bardsley (eds.), *The Idea of Creativity* (pp. 17-42). Brill. <https://doi.org/10.1163/ej.9789004174443.i-348.12>
- Campbell, D. T. (1960). Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. *Psychological review*, 67(6), 380-400. <https://doi.org/10.1037/h0040373>

²⁵ Este trabajo fue realizado en el marco de proyectos de investigación financiados por los siguientes organismos: el Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (proyecto PIP 11220200103107CO); la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba, Argentina (proyecto CONSOLIDAR 33620230100966CB); la Agencia Nacional de Promoción Científica y Tecnológica de Argentina (proyecto PICT-2020-01653), y el Ministerio de Ciencia e Innovación de España (proyecto PID2021-128835NB-I00).

- Csikszentmihalyi, M. (2014). *The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi*. Springer. <https://doi.org/10.1007/978-94-017-9085-7>
- Darwin, C. (1964). *The Origin of Species*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9xp5>
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- De Sousa, R. (2018). [Reseña de *From Bacteria to Bach and Back: The Evolution of Minds*, de D. C. Dennett]. *Evolutionary Studies in Imaginative Culture*, 2(1), 113-116. <https://doi.org/10.26613/esic.2.1.80>
- Dennett, D. C. (1990a). The Interpretation of Texts, People and Other Artifacts. *Philosophy and Phenomenological Research*, 50, 177-194. <https://doi.org/10.2307/2108038>
- Dennett, D. C. (1990b). The Myth of Original Intentionality. En K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale y K. V. Wilkes (eds.) *Modelling the Mind* (pp. 43-62). Clarendon Press. <https://doi.org/10.1093/oso/9780198249733.003.0004>
- Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, 88(1), 27-51. <https://doi.org/10.2307/2027085>
- Dennett, D. C. (1995a). *La conciencia explicada*. S. Balari Ravera (trad.). Paidós.
- Dennett, D. C. (1995b). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Penguin Books.
- Dennett, D. C. (1998). *La actitud intencional*. D. Zadunaisky (trad.). Gedisa.
- Dennett, D. C. (2001). In Darwin's Wake, Where Am I? *Proceedings and Addresses of the American Philosophical Association*, 75(2), 11-30. <https://doi.org/10.2307/3218710>
- Dennett, D. C. (2002). *Content and Consciousness*. Routledge. <https://doi.org/10.4324/9780203005729>
- Dennett, D. C. (2006). From Typo to Thinko: When Evolution Graduated to Semantic Norms. En S. Levinson (ed.) *Evolution and Culture: A Fyssen Foundation Symposium* (pp. 133-145). The MIT Press.
- Dennett, D. C. (2013). *Intuition Pumps and Other Tools for Thinking*. W. W. Norton & Company.
- Dennett, D. C. (2017). *De las bacterias a Bach: la evolución de la mente*. M. Figueras (trad.). Pasado & Presente.
- Fogarty, L., Creanza, N., Feldman, M. W. (2015). Cultural Evolutionary Perspectives on Creativity and Human Innovation. *Trends in Ecology & Evolution*, 30(12), 736-754. <https://doi.org/10.1016/j.tree.2015.10.004>

- Gaut, B. y Livingston, P. (2003). *The Creation of Art: New Essays in Philosophical Aesthetics*. Cambridge University Press. <https://doi.org/10.1111/j.1747-9991.2010.00351.x>
- Gaut, B. (2010). The Philosophy of Creativity. *Philosophy Compass*, 5(12), 1034-1046.
- Ginnobili, S. (2022). Darwinian Functional Biology. *Theoria. An International Journal for Theory, History and Foundations of Science*, 37(2), 233-255. <https://doi.org/10.1387/theoria.22645>
- Glăveanu, V. (2011). How Are We Creative Together? Comparing Sociocognitive and Sociocultural Answers. *Theory and Psychology*, 21, 473-492. <https://doi.org/10.1177/0959354310372152>
- Glăveanu, V. (2015). Creativity as a Sociocultural Act. *The Journal of Creative Behavior*, 49(3), 165-180. <https://doi.org/10.1002/jocb.94>
- Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection*. Oxford University Press.
- Kaufman, J. C. y Glăveanu, V. (2019). A Review of Creative Theories: What Questions Are We Trying to Answer? En J. Kaufman y R. Sternberg (eds.), *The Cambridge Handbook of Creativity* (pp. 27-43). Cambridge University Press. <https://doi.org/10.1017/9781316979839.004>
- Kronfeldner, M. E. (2009). Creativity Naturalized. *The Philosophical Quarterly*, 59(237), 577-592. <https://doi.org/10.1111/j.1467-9213.2009.637.x>
- León, M. (2023). Is Cultural Selection Creative?. En J. M. Viejo y M. Sanjuán (eds.), *Life and Mind: New Directions in the Philosophy of Biology and Cognitive Sciences* (pp. 133-164). Springer. https://doi.org/10.1007/978-3-031-30304-3_7
- Locke, J. (1847). *An Essay Concerning Human Understanding*. Kay & Troutman.
- Miłkowski, M. (2009). Is Evolution Algorithmic? *Minds and Machines*, 19(4), 465-475. <https://doi.org/10.1007/s11023-009-9170-6>
- Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226520452.001.0001>
- Richerson, P. y Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226712130.001.0001>
- Ross, D. (2000). Introduction: The Dennettian Stance. En D. Ross, A. Brook y D. Thompson (eds.), *Dennett's Philosophy:*

- A Comprehensive Assessment* (pp. 1-26). The MIT Press.
<https://doi.org/10.7551/mitpress/2335.003.0003>
- Runco, M. A. y Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24(1), 92-96.
<https://doi.org/10.1080/10400419.2012.650092>
- Schliesser, E. (2018). A Genealogy of Modernity and Dennett's Strange Inversion of Reasoning. *Teorema: Revista Internacional de Filosofía*, 37(3), 171-180.
- Searle, J. (1982). The Myth of the Computer. *The New York Review of Books*, 29(7), 3-6.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139173452>
- Searle, J. (2004). *La mente: una breve introducción*. H. Pons (trad.). Grupo Editorial Norma.
- Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press.
- Sober, E. (2004). *The Design Argument*. Blackwell.
- Sober, E. (2006). Evolution, Population Thinking, and Essentialism. En E. Sober (ed.), *Conceptual Issues in Evolutionary Biology* (pp. 329-359). The MIT Press.
- Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Blackwell.
- Sterelny, K. (2006). Memes Revisited. *British Journal for the Philosophy Science*, 57(1), 145-165. <https://doi.org/10.1093/bjps/axi157>

<https://doi.org/10.21555/top.v760.3296>

Human Free Will and Determinism: Critical Considerations on Daniel Dennett's Philosophy and the Dangerous Idea of Freedom in Evolutionary Epistemology

El libre albedrío humano y el determinismo:
reflexiones críticas sobre la filosofía de Daniel Dennett
y la peligrosa idea de la libertad en la epistemología
evolutiva

Vicente Claramonte Sanz

Universitat de València

España

vicente.claramonte@uv.es

<https://orcid.org/0000-0002-0477-6995>

Rodolfo Guarinos Rico

Universitat de València

España

rodolguarinos@gmail.com

<https://orcid.org/0000-0003-2616-2713>

Recibido: 28 - 11 - 2024.

Aceptado: 01 - 08 - 2025.

Publicado en línea: 01 - 06 - 2026.

Cómo citar este artículo: Claramonte Sanz, V. & Guarinos Rico, R. (2026). Human Free Will and Determinism: Critical Considerations on Daniel Dennett's Philosophy and the Dangerous Idea of Freedom in Evolutionary Epistemology. *Tópicos, Revista de Filosofía*, 76, 159-185. <https://doi.org/10.21555/top.v760.3296>



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

Abstract

This paper proposes some critical considerations on Daniel Dennett's philosophy, particularly on his notion of human free will in relation to determinism. To this effect, it analyses the four types of possibility studied by Dennett and their relationship with causality and chance in the context of the philosophy of biology. In addition, it also critically evaluates the distinction made by Dennett between deterministic, indeterministic, and nihilistic universes, with special attention to the treatment conferred in physics to the idea of causality in classical, relativistic, and quantum mechanics. Finally, it explains how the compatibility between human free will and deterministic natural laws defended by Dennett could be developed and explained in evolutionary epistemology.

Keywords: free will; determinism; indeterminism; causality; philosophy of biology; evolutionary epistemology; Daniel Dennett's philosophy; philosophy of science; materialism; libertarianism.

Resumen

Este artículo propone algunas reflexiones críticas sobre la filosofía de Daniel Dennett, en particular sobre su concepción del libre albedrío humano en relación con el determinismo. Para ello, analiza los cuatro tipos de posibilidad estudiados por Dennett y su relación con la causalidad y el azar en el contexto de la filosofía de la biología. Además, evalúa críticamente la distinción que establece Dennett entre universos deterministas, indeterministas y nihilistas, prestando especial atención al tratamiento que la física hace de la idea de "causalidad" en la mecánica clásica, la relativista y la cuántica. Por último, explica cómo la compatibilidad entre el libre albedrío humano y las leyes naturales deterministas defendida por Dennett podría desarrollarse y explicarse en la epistemología evolutiva.

Palabras clave: libre albedrío; determinismo; indeterminismo; causalidad; filosofía de la biología; epistemología evolutiva; filosofía de Daniel Dennett; filosofía de la ciencia; materialismo; libertarismo.

1. Introduction

In *Freedom Evolves*, Dennett (2003) examines the relationship between human freedom and natural causation on the basis of a notion of freedom grounded in an evolutionary epistemology—that is, one not understood as a metaphysical issue, but rather as the evolutionary outcome of a biological process. Such an analysis, in our view, requires an examination of the concept of causality as it is generally employed in the sciences, and particularly in physics, which in turn entails elucidating the ideas of cause and effect and the relation between them.

For the purposes of his study of the relationship between freedom and causation, Dennett proposes a distinction among deterministic, indeterministic, and nihilistic universes—a distinction upon which we wish to offer several critical observations. These consist primarily in questioning the factual existence of nihilistic universes and in discussing their logical viability. We shall therefore advance arguments to suggest that nihilistic universes ought rather to be regarded as subsumed under either the deterministic or the indeterministic category.

Finally, we shall show that Dennett's compatibilist position regarding human freedom and natural causation rests on the conception of freedom as a natural, biologically grounded phenomenon—one that should be understood primarily in biological rather than metaphysical terms. On this view, freedom emerges as an evolved phenotypic trait conferring adaptive advantage through the guidance of intentional learning oriented towards the range of real possibilities available to an organism. Freedom thus becomes a matter of cognition, whereby the future delineated by both genetics and environment can be represented and modified through a representation of the present and of the various possible futures.

2. Casual and causal: on the real and the biologically possible

As a propaedeutic for his analysis of the idea of freedom within the scope of an evolutionary epistemology, Dennett (1999, p. 161) examines the concept of possibility based on a typology that distinguishes at least four different types or degrees: logical, physical, biological, and historical. Logical possibility requires only the absence of contradiction: an organism cannot simultaneously be alive and dead. Secondly, a

physically possible phenomenon or process, like any living being, would only appear if its existence does not violate any known or yet-to-be-discovered physical or chemical law of nature: in this sense, an organism faster than light would be impossible. Biological possibility entails, in addition to not violating any known or yet-to-be-discovered physical or chemical law of nature, an organism capable of evolving from a genotype constituted by any combination of the four nucleotides—adenine, cytosine, guanine, and thymine—that make up the DNA alphabet, and also being viable for survival and reproduction within a given ecological environment: a flying elephant in and of itself would not transgress any physical or chemical law, but it would only be biologically possible if the evolution of a genotype capable of implementing a Bauplan or body development pattern suited to the requirements of weight, mass, gravity, strength, aerodynamics, etc., necessary for flight in a specific ecosystem were to occur, while also ensuring a minimal level of fitness or biological reproductive efficacy. Finally, historical possibility requires synchrony between an evolutionarily viable genotype and an ecosystem potentially compatible with the expression and development of such viability: the superorder of the Dinosauria was undoubtedly the dominant taxon on the planet until the impact of an asteroid created the Chicxulub crater in the Yucatán Peninsula around sixty-five million years ago at the end of the Cretaceous-Tertiary period, but in the current biosphere, they would be biologically-historically inviable, just as the class Mammalia could only diversify and attain its predominant position in biological systematics after—and not before—the collision of the uranolite with Earth. Furthermore, the philosophical perspective that denies any alternative possibility to reality, that is, the idea that only the real is possible, is often referred to as actualism, and Dennett himself (1999, p. 165) argues against it, defending compatibilism, according to which the notions of free will and determinism are perfectly compatible, and holding a simultaneous belief in both does not entail any logical inconsistency.

Once this quadruple distinction between types or degrees of possibility has been established and exemplified, Dennett (1999, pp. 176 et seq.) points out certain complexities or difficulties regarding biological possibility, based on the relationship between genotype and phenotype, upon which we would like to highlight certain nuances that we consider relevant, especially regarding the certainly complex issues of DNA reading and the viability of the phenotype.

Regarding the former, difficulties arise when considering that DNA sequences cannot specify any organismic traits without the action of readers. When a new organism is generated, the first basic effect of this action consists in the production of proteins from amino acids, and among the countless possible types of proteins, only those dependent on the read DNA sequence will perfect themselves, subsequently interacting with each other through a process that begins with the division of the zygote into two cells—each of which obviously has its own duplicated copy of the entire DNA sequence under reading—these in turn dividing into two more, and so on, until the embryo is developed. Ultimately, without a reader capable of deciphering the genomic recipe, it is impossible for the DNA to be specified in phenotype, a circumstance that makes the plot of the film *Jurassic Park* purely fantastical, as the DNA of *Velociraptor mongoliensis*, even if complete, would contribute at most half of the necessary genetic information and could not model a viable phenotype without the other half provided by the egg cell. Thus, the causal mechanism through which biological possibility is realised in phenotype depends simultaneously on the DNA sequence as well as the reader.

Regarding the latter, that is, the viability of the phenotype of a given species, it seems to be an immediate consequence of the historical possibility described in the previous paragraph, as it depends on the environmental setting and the ecosystem in which the organism lives and is adapted: it would hardly be viable for any species without a breathable atmosphere and one relatively protected from solar radiation, for instance. Moreover, considering that the environmental and ecological surroundings change in a slow but continuous manner, one can understand why there are so many cases of taxa whose extinction can only be comprehensively understood and explained in macroevolutionary terms. Thus, chance would influence causality arising from the interaction between the DNA sequence and its required reader, for if, due to bio-historical circumstances external to the biological possibility itself, the genetic mechanism specifying the phenotype ceases to be operational in the environment to which the organism had previously been adapted, viability becomes ineffective in itself, and sooner or later the species will inevitably become extinct.

In the last third of the 20th century, Jacques Monod (1910-1976) had already outlined this relationship between chance and causality in terms of randomness and necessity in *Biology* (Monod, 1993). From

his perspective, the notion of bio-historical possibility proposed by Dennett also underlies, in one way or another. Natural selection, the unrelenting and decisive judge between adaptation and extinction, would play a role, albeit with certain differences, analogous to necessity or causality, while mutation would operate equivalently to chance through a dual stochastic variable: one of random changes that alter the nucleotide sequence of DNA, and the other concerning the greater or lesser adaptive fit or misfit to the given ecological environment within the existing temporal span—here lies the confluence with bio-historical possibility—until the ultimate effect of positive or negative selection of the mutated trait takes place. All of this, finally, within the framework of a materialist conception of the universe, long ago formulated by Democritus of Abdera (ca. 460–ca. 370 BCE), whose *arkhé* or organising principle of all natural entities was also postulated in terms of chance and necessity.

Additionally, Dennett (2004, p. 92) links his concept of causality to a series of factors that are hierarchically organised in an approximate order of precedence, such as causal necessity, causal sufficiency, independence, temporal priority, and a miscellany of further criteria that are not expressly listed. We understand that these notions of causal necessity and sufficiency proposed by Dennett vividly and respectively echo those of necessary and sufficient conditions used in logic, a circumstance which warrants pointing out certain details in order to eliminate or at least reduce any potential confusion in meaning.

By themselves, necessary conditions do not express the existence of a causal relationship between the antecedent and consequent of an implication: thus, for example, drinking water is an essential requirement for life, but it is evident that ingesting water does not cause life. Sufficient conditions can express a causal relationship, but not always or in every case. In the statement, “If you take poison, you die,” there exists a causal relationship between the antecedent “taking cyanide” and the consequent “dying,” but this is not the case in the statement, “If Pedro comes, we will go to the cinema,” where the antecedent “Pedro coming” is a sufficient condition for the consequent “going to the cinema.” However, it seems clear that the cause of going to the cinema does not strictly reside in Pedro’s arrival, but rather in the prior will or agreement of those who wish to share the cinematic experience upon Pedro’s arrival.

Thus, sufficient conditions would only point to a causal relationship when they express a singular factor, but not when a co-cause is involved

that is not entirely subsumed within the antecedent—that is, when the causation of the consequent is multifactorial, and the antecedent does not express all the required factors for its production. Given all the above, only those conditions that are simultaneously necessary and sufficient could be interpreted as causal. Therefore, an approximate definition of cause could be proposed, understood as the result of the set of necessary and sufficient conditions for the production of an event, such that conditions $c_1, c_2, c_3 \dots c_n$ would produce effect E if and only if $c_1, c_2, c_3 \dots c_n$ are both necessary and sufficient conditions, in such a way that if these conditions occur, the effect will occur, and if they do not occur, the effect will not take place. Both the set of necessary and sufficient conditions, on the one hand, and the effect, on the other, are terms on either side of an equivalence equation, and this, obviously, necessitates the transition from the implicator to the co-implicator or equivalator.

However, in his analysis, Dennett does not include anything resembling a “simultaneous causal necessity and sufficiency” in his prioritised list of factors. In our opinion, this circumstance warrants a revision in pursuit of a preferable logical consistency. Lastly, among the factors associated with the concept of causality included in the Dennettian list cited above, we wish to specifically address the issue of temporal priority. As Dennett observes in reference to this, “a reliable way to distinguish causes from effects is to note that causes occur earlier” (2004, p. 93). In physics, this assertion appears entirely congruent with the causal determinism characteristic of classical mechanics. According to its principles, rooted in the foundational work of Isaac Newton (1643–1727), the cause precedes the effect in time. That is, every event or effect EEE is assumed to be produced or caused by a prior event c_1 or prior events $c_1, c_2, c_3 \dots c_n$. Consequently, with complete and exact information about the current state of the universe, it would be possible to accurately predict any future event EEE .

The most extreme formulation of causal determinism was proposed in the *Philosophical Essay on Probabilities* by the Marquis Pierre-Simon Laplace (1749–1827). Its preface includes a famous passage presenting the thesis of the so-called “Laplace’s demon,” according to which “we ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that will follow it” (Laplace, 1985, p. 25). This extrapolated the causal determinism of classical mechanics to all phenomena and processes, enabling this deterministic causality to be formulated mathematically in terms of the laws of nature.

The seemingly invulnerable solidity of this thesis began to erode when Henri Poincaré (1854–1912) conceived the possibility of chaos, understood as a systematic sequence of effects exhibiting a strong dependence on initial conditions. This hypothesis was later confirmed and expanded by chaos theory. According to this theory, slight variations in the initial conditions of certain types of complex systems and nonlinear dynamic systems—such as vehicular traffic, fluid dynamics, or meteorology—even when strictly deterministic, can result in future effects so markedly different that they render Laplacian predictability entirely impossible, particularly in the medium to long term. Thus, chaos theory establishes the unpredictability of certain physical systems typical of classical mechanics, due to their stochastic or chaotic causal behaviour.

In the framework of relativistic mechanics, however, the temporal precedence between cause and effect inherent in causal determinism, and defended by Dennett as an intrinsic factor of the notion of causality, faces even greater challenges. First, in the special theory of relativity formulated by Albert Einstein (1879–1975) in 1905, although the previous idea of causality remains broadly intact, the absence of an absolute, observer-independent time and the introduction of relative time render the classical mechanical concept of simultaneity inapplicable—particularly for events unrelated causally, where synchrony now depends on the observer. Consequently, the very notion of temporal priority or precedence between cause and effect appears unstable. In other words, within the framework of the special-relativistic principle of causality, a cause only temporally precedes its corresponding effect in the case of inertial observation, that is, observation free of any force. The practical impossibility of a truly inertial observer—neutral to all forces (gravity can only be reduced exactly to zero at infinite distance, a condition that cannot factually exist in the universe)—was precisely the anomaly that led Einstein to reformulate special relativity.

Thus, in the general theory of relativity, the notion of causality is generalised—it becomes redundant to observe this. Temporal precedence between cause and effect is no longer required: it suffices for the effect to belong to the future light cone of its respective cause. The quasi-classical notion of causality employed by Dennett now appears to suffer from greater inconsistency. It is, therefore, unsurprising that authors such as Patrick Suppes (1922–2014) developed a probabilistic theory of causality (Suppes, 1970). However, despite the laudable effort

of its author and the incorporation of temporal considerations, the notion of positive statistical relevance is “insufficient to convey the intuitive idea of production, fundamental to any causal description” (Álvarez Toledo, 1999, p. 28).

Finally, in quantum mechanics, the notion of causality seems almost entirely decoupled from the Dennettian requirement of temporal precedence between cause and effect. This is particularly true since Werner Heisenberg (1901–1976) introduced the uncertainty principle in 1927. This principle states that it is impossible to simultaneously and precisely determine the values of certain pairs of physical variables, such as the position and momentum of a particle. Consequently, greater certainty in determining a particle’s position is inversely proportional to the certainty of its momentum or linear motion and, therefore, its velocity and mass. Within this conceptual framework of particle physics, how could the temporal precedence of the relationship between cause and effect be established? If Heisenberg’s uncertainty principle is valid — and it appears to be — there would be no conceptual obstacle to accepting “that the cause may be prior to or simultaneous with the effect” (Álvarez Toledo, 1999, p. 26). As Álvarez Toledo (1999, p. 26) illustrates, when an aeroplane takes off, the movement of the air above and below the wings causes its elevation. However, this movement is not prior to the elevation but simultaneous.

Recapitulating the considerations outlined in the preceding paragraphs, it can be understood that Dennett seeks to clarify the everyday and technical-philosophical notions of possibility, necessity, and causality. He begins with an exploration of possibility through a quadruple filter—logical, physical, biological, and historical—through which he attempts to conceptualise the idea of free will from a naturalistic, materialist, and Darwinian perspective. This approach is directed not so much towards an individual understanding as towards a phylogenetic one. His meticulous analysis of the connections between freedom and causality, as developed in *Freedom Evolves*, undoubtedly aims to demonstrate the compatibility between the deterministic laws of nature and the existence of a human being endowed with free will. However, his concept of freedom is articulated exclusively within the framework of biological-evolutionary coordinates, without veering into metaphysical terrain beyond what is strictly necessary.

That said, in our view, this analysis does not seem to distinguish particularly clearly or explicitly between the notion of logical or

philosophical causality—logically possible worlds—and the concept of causality as used in science—physically possible worlds. Furthermore, his conception of causality in physics aligns closely with classical mechanics but could be said to correspond far less to relativistic mechanics and almost not at all to quantum mechanics. Nonetheless, we do not believe this represents a personal philosophical difficulty for Dennett. Instead, it stems from the pressing need within philosophy to revisit the concepts of cause and effect and the relationship between them. This is especially true in light of the current state of scientific knowledge, particularly regarding the issue of temporal precedence between cause and effect, which is often treated as though it were a logical axiom or a universal physical constant.

3. Materialism, determinism, and indeterminism

In his materialist analysis of the meaning of determinism and indeterminism and their distinction, Dennett adopts, as a premise, the simplifying allegory advanced by Willard Quine (1908–2000) under the label of “Democritean universes” (Quine, 1974). These are imaginary elemental cosmoses consisting of a handful of atoms moving through space, though these atoms are not understood in the contemporary sense, with a complex quantum structure, but rather in the pre-Socratic sense as homologous and indivisible units of matter. Dennett defines determinism in line with the notion proposed by Peter van Inwagen (1942–), which states, in the latter’s literal terms, that “at any given moment, there is exactly one possible future” (1983, p. 3), aligning this with the aforementioned model characteristic of Laplace’s demon. He further introduces a reference to the laws of physics to conceptualise determinism and distinguish it from indeterminism (Dennett, 2004, p. 45): a universe is deterministic when there exist transition rules, constituted by physical laws, which precisely determine the subsequent state that will follow any particular prior state. Conversely, if there is even the slightest deviation or uncertainty, the universe would be indeterministic.

Shortly thereafter, Dennett identifies a *tertium genus* within the typology of configurations that emerge from his elaborate thought experiment concerning these Democritean universes, derived from Quine’s proposal: “those that are nothing but rubbish” (Dennett, 2004, p. 51). He labels these nihilistic universes, as they exhibit no enduring regularity between the past, present, and future states of matter. Thus,

we have three categories of universes: deterministic, indeterministic, and nihilistic.

We understand that Dennett's examination of this issue, while indeed consistent with his philosophical aim in *Freedom Evolves* to demonstrate that a universe governed by deterministic laws can be compatible with freedom conceived not in a metaphysical but in a biological sense, is insufficient—or at least presents certain deficiencies—which we aim to outline and develop in detail below.

Firstly, Dennett does not distinguish, as he does in the case of possibility discussed above, between the three types of determinism commonly acknowledged, namely: logical, epistemic, and causal (Swartz, 2004). From the perspective of logical determinism, the truth value (true/false) of every proposition is timeless, such that a true proposition is and will always be true, while a false proposition is and will always be false. The philosophical background here, already raised by Aristotle (384–322 BCE) in *On Interpretation* and later developed within Scholastic logic, primarily through the Aristotelianism of Robert Grosseteste (1175–1253) and his *De veritate propositionis* (López Cuétara, 2005), centres on the so-called contingent futures: simple declarative propositions about events that may or may not occur in the future. With regard to such propositions, the task is to decide which truth values may be ascribed to them—true/false, necessarily true/necessarily false, possibly true/possibly false, or even an absence of truth/falsity, indeterminacy, etc. (Ostalé & Pradier, 2010, p. 30).

The connection between this issue and Dennett's analysis of the notion of freedom is closer than it might initially appear, as one of the key arguments for fatalism—the doctrine that everything occurs by inescapable predetermination or fate, rendering human action powerless to alter outcomes—derives precisely from the attribution of truth values to contingent futures. This belief leads the fatalistic view to reduce individual freedom to a strictly subjective experience, ultimately devoid of any substantive ontological grounding. Epistemic determinism, in turn, not only attributes truth/falsity values to propositions about the future but also asserts that such propositions are known prior to the events they describe. Consequently, if any future event is known in advance, that event must inevitably occur. Similarly, *mutatis mutandis*, to self-fulfilling prophecies—understood as propositions about the future (analogous to Grosseteste's contingent futures) that, once formulated, themselves

cause their materialisation—epistemic determinism transforms these into a peculiar but unique duality of cause-effect and conversely.

Finally, causal determinism holds that all phenomena and processes are no more than the final outcome of certain initial conditions governed by the laws of nature. Carl Hempel (1905–1997) illustrated this idea of deterministic causality with his famous example of the broken radiator (Hempel, 1942): the malfunction is the result of the radiator's structure—filled with water but lacking antifreeze (initial conditions)—subject to the action of laws concerning the tensile strength of iron, the expansion of frozen water, and the thermodynamic circumstance of the temperature dropping below freezing for hours (laws of nature). In the case of human behaviour, this is the crux of the matter: strong causal determinism would assert that the same principle applies.

To summarise, we are faced with a triad of arguments, each leading via distinct routes to the same conclusion: human action is not free but determined, and therefore, the notion of free will is merely a pre-Kantian metaphysical delusion without any plausible ontological correlate in any region of spacetime. Hence, it is often said that the acceptance of determinism, in any form, is wholly incompatible with the existence of free will. If the states of the universe have been mechanically predictable for thousands of years and presumably will remain so for thousands more, how can one argue that *Homo sapiens* has emerged within it, determined by rigid and inexorable evolutionary laws, while simultaneously claiming that it possesses free will to self-determine?

Although Dennett's work only examines causal determinism, in our view, this is precisely the central thesis he seeks to challenge in *Freedom Evolves*, arguing instead that human freedom is compatible with the deterministic physical, chemical, and biological laws that serve as the *explanans* for its evolutionary origins. Additionally, there is another minimum common denominator across the three aforementioned deterministic arguments that we believe is crucial for elucidating the critical issue at hand, particularly concerning causal determinism, which is presumed to be intrinsically and inevitably tied to classical mechanics. This pertains to the relationship between determinism and time (Hoefer, 2024).

The three deterministic arguments presented are temporally symmetric. If, as discussed earlier, every effect EEE is caused by a prior condition c_1 or by prior conditions $c_1, c_2, c_3 \dots c_n$, thereby implicitly assuming that the preceding states of the universe determine all

subsequent states by pointing univocally to a single future, we may also assume that all subsequent states similarly determine—inasmuch as they are determined by—all preceding states. Since a univocal temporal relationship (forward-backward or before-after) between states has been established, subsequent states retroactively point to a single past. In a chain whose links are understood as a succession of ordinal numbers, link c_1 fixes link c_2 as inexorably as link c_2 fixes c_1 , neither allowing for alternatives. In the words of Pierre Curie (1859–1906), “When certain causes produce certain effects, the symmetry elements of the causes must be found in their effects” (1894, p. 401), which suggests that effects retroactively determine their causes. Hence the exploration of hypotheses and thought experiments concerning retro causality.

The difficulty of grasping the bidirectional temporal symmetry (precedent-subsequent) implicit in determinism—and seemingly well-aligned with classical mechanics—lies in our psychological propensity to accept uncritically the assumption that physical causality operates solely in the past-present-future direction, but not vice versa. This may be due to the influence of the concept of the arrow of time, coined by Arthur Eddington (1882–1944) to denote a temporal direction within a four-dimensional relativistic universe (Eddington, 1928). This concept inherently entails consequences of irreversibility and asymmetry: however, its demonstration is accepted—albeit with caveats—in cosmology, radiation, or thermodynamics (the second law), but not in classical mechanics. Nevertheless, the current state of scientific knowledge in physics does not independently support the belief that the past possesses a fixed quality absent from the present and future. Consequently, in relation to Dennett’s analysis of the idea of freedom, the past exercises no ontological dominion over human action and behaviour that the present and future lack. This conclusion removes obstacles to arguing for the compatibility of determinism with human free will, as Dennett himself contends, albeit through different logical paths.

Secondly, we argue that the triad of deterministic, indeterministic, and nihilistic universes, defined in the preceding paragraphs in strict accordance with Dennett’s treatment, is not entirely consistent with his own global analysis regarding the compatibility or incompatibility between human free will and the action of deterministic laws of nature, at least when considered as three disjoint sets or entirely separate and disconnected concepts. A separate issue, however, would be to discuss whether there are intersections between them or the potential condition

of subsets among them, if the matter is approached from set theory. By contrast, the dichotomy of determinism/indeterminism, in light of the historical-philosophical tradition that seemingly consists of terms that are incompatible or mutually exclusive, appears, at first glance, easier to accept without further objection, as it fits neatly within the bivalent true/false parameters of classical or Aristotelian standard logic. Perhaps for this very reason, the vast amount of literature devoted to the subject tends to treat determinism and indeterminism not only as disjoint sets, whose intersection, precisely because of the disjunction they embody, is an empty set, but, even more so, as what could be called opposing conceptual entities, such as “alive” and “dead” or “single” and “married,” since, to the best of our knowledge, there is no technical-mathematical concept of opposite sets, though there is one for opposite numbers. This association of the analysis of determinism and indeterminism with the bivalence of true or false inherent to classical logic has likely played a significant role in intensifying and perpetuating the debate between proponents and detractors of each view. This is because, firstly, a demonstrative argument for the former seems to entail, consequently, the refutatory argument for the latter, and vice versa; and secondly, it appears entirely irrational or a mere paralogism to even attempt to propose the hypothesis of their compatibility, since the next thing that comes to mind would be the asterisk symbol as a graphical representation of the contradiction involved in asserting the simultaneous validity of two opposites, and, applying the aforementioned examples, accepting without reservation the truth of statements such as “The living are dead” or “The single are married.”

Moreover, another element often underlies this debate that can also foster confusion, and against which, according to our best knowledge and understanding, Dennett does not explicitly or sufficiently guard in his analysis: this involves addressing the issue without consciously realising the shift from the logical to the causal interpretation of the concepts of determinism and indeterminism. One should recall the warning formulated by David Hume (1711–1776) regarding the inadmissibility of confusing propositions whose truth is a relation of ideas with those whose truth concerns matters of fact (Hume, 1994, pp. 47–48): Hume’s famous “fork,” by which every proposition is either necessary or contingent, but never both at the same time. The cumulative effect of both elements— understanding the determinism-indeterminism dichotomy through the true/false bivalence of classical

logic and failing to distinguish between their logical and causal senses—increases the level of confusion in the philosophical debate to the point where it often becomes an inextricable chaos of meaning. Obviously, logical determinism entails a type of logical causality, just as causal determinism entails empirical causality—it would be strikingly redundant to call it “causal causality,” and the expression would end up being meaningless due to its sheer tautological nature—and we have just invoked the gnoseological blunder of treating propositions about relations of ideas the same as those concerning matters of fact. This is why, in the previous section, we drew attention to Dennett’s insufficient distinction when studying both types of causality.

Thus, returning now to the triad of deterministic-indeterministic-nihilistic kosmos, we understand that, in proposing or defining the latter—those universes where no regularity exists between the past, present, and future states of matter—Dennett inadvertently confuses logical causality with empirical causality. In set theory, a category can be defined as any grouping of material or abstract entities, which also includes a universe whose dynamics lack any regularity, and this would lead to attributing it, in essence, a logical causality once we accept for it a non-regular, and even ultimately non-causal, causality, just as, due to the comprehensive nature of set theory, we accept a set of non-existent sets or an empty set of elements and affirm that the empty set is a set in itself and not a non-set or one that cannot be encompassed by the theory or mathematics in general. But, in physics, could we find a nihilistic universe in Dennett’s sense, lacking any regularity to cause the transitions between its successive states? With the current state of scientific knowledge, this seems impossible, and Dennett certainly does not cite any empirical example, in the sense of being observable or identifiable by any method in any spatiotemporal region of the universe. It should be noted that in logic and set theory, the concept of a universe of discourse is common, in which all kinds of entities can be included, even causality aberrations—those that lack any regularity—but it is hardly admissible to extrapolate its properties to an empirical or physical universe merely because of the confusing coincidence of the term “universe.” Consequently, at the logical level of analysis of this issue, a nihilistic *tertium genus* alongside the deterministic and indeterministic universes would be admissible, but at the strictly physical level, in our view, *tertium non datur*, and for this reason, we believe that Dennett’s assertion of nihilistic universes unintentionally falls into a fallacy of equivocation,

which should be avoided. This is a circumstance not entirely mitigated by the explicit admission of Heisenberg's indeterminacy relation in physics for quantum mechanics, as its basic epistemological consequence is to shift from absolutely precise knowledge—at least theoretically—to knowledge based on probabilities, which, although probabilistic, remains valid and relevant.

However, the latter is by no means equivalent to knowledge articulated from the total absence of any regularity, if that were possible. And this seemingly highly plausible impossibility would not arise merely from the fact that “there are absolute limits to our ability to predict the future” (Dennett, 2004, p. 50), but primarily because the inability to process and systematise any information according to stable categories would render any cognitive endeavour a failure. In fact, in an empirically considered nihilistic universe, life, let alone human life, would not even be possible: total absence of regularity would imply the non-existence of gravity, electromagnetism, solar cycles, the translational and rotational movement of the stars, and so on, *ad infinitum*. And without human life, considering Nietzsche's certifiable declaration of the death of God, there would also be no knowledge. A nihilistic universe, in any case, could only be admitted at an exclusively logical level of analysis and as a subset of indeterministic universes, characterised by the total absence of regularities and causality due to the extreme nature of its properties.

Thirdly and finally, the considerations established in the preceding paragraphs lead us to contemplate the viability of plausibly discussing the possibility that there may be an intersection between the set of deterministic and indeterministic universes as proposed by Dennett's analysis. Where would we, for example, classify the previously (briefly) mentioned theory of chaos, which forms the theoretical basis for the study of complex or dynamic systems? In fact, the epistemology of specialist determinism unambiguously posits the existence of deterministic chaotic systems, essentially characterised by two main attributes (Hoefer, 2024): a radical short-term divergence in the evolution of their successive states from identical or nearly identical initial conditions, such that the movement underlying the system's behaviour exhibits extreme instability in its trajectory—authors such as Batterman (1993) focus solely on this requirement—and consequently, a lack of predictability or algorithmic computability, since in the long term they effectively emulate genuine random or stochastic processes. Furthermore, these epistemologists

also acknowledge a varied typology of chaotic dynamic systems, such as discrete, continuous, two-dimensional, three-dimensional, higher-dimensional, particle-based, fluid-based, and so on. Lastly, they debate fervently for and against the qualification of certain phenomena studied in physics as chaotic deterministic systems, such as models of escape to infinity and spatial invasion, particle collision phenomena, classical mechanics' systems of supertasks and hypertasks, and the static black holes of Karl Schwarzschild (1873–1916), the dynamic black holes of Roy Kerr (1934–), or singularities in quantum mechanics.

Now, based on the current state of scientific knowledge, the empirical existence of systems like those investigated by chaos theory and others of a similar nature can only be explained within the framework of the distinction between deterministic and indeterministic worlds, as established by Dennett, through two plausible hypotheses. First, the laws of nature governing the behaviour of such systems are genuinely, and not merely apparently, random or stochastic, in which case they would fall within the set of indeterministic universes; or, second, these systems possess a genuine physical idiosyncrasy of a chaotic nature despite being governed by deterministic laws, the rules of transition between states of matter referred to earlier by Dennett. This is why Suppes (1993) asserts that chaotic deterministic examples such as those listed above can perfectly be classified as deterministic systems fully aligned with classical mechanics, as well as processes that are simply and unequivocally indeterministic. Similarly, Van Fraassen (1941–) sought to establish the epistemic foundations for explaining physical phenomena without presupposing that they are caused by laws that cause or explain their behaviour (Van Fraassen, 1989).

However, leaving aside genuine randomness and returning to the second hypothesis, which returns us to Dennett's dichotomous distinction, given their formal properties and phenomenology, systems characterised by deterministic chaos or chaotic determinism—a phrase that is, in fact, close to paradoxical, as its meaning almost equates to indeterminable determinism, indeterminate determinism, or even indeterministic determinism—only serve to confirm our earlier assertions in favour of seriously considering the possibility of admitting an intersection between both sets, and at the same time, opposing the presumption of their strictly disjoint nature. This corollary, curiously, helps to sustain Dennett's thesis regarding the compatibility between human free will and the action of deterministic laws in nature.

4. Darwinism and the libertarian belief

It is important to highlight that, for our author, not only are free will and determinism compatible, but the issue of freedom, in its essence, has little to do with the metaphysical debates between libertarians and determinists of various kinds. This is because his compatibilist position rests on the idea that freedom is a natural biological phenomenon which, therefore, must be understood and studied as part of the science of biology, rather than being treated as a metaphysical issue. The same applies to the metaphysical debates surrounding phenomena such as avoidability, the existence of real and not merely illusory possibilities, and so on. In this sense, he makes the following statement of intent: “I will try to show that this traditional problem of free will is, in spite of its pedigree, a distractor, a puzzle of no real importance that draws our attention away from some neighboring concerns that truly matter” (Dennett, 2003, p. 9). His naturalistic and naturalising stance must, therefore, rest on a definition of the concept of freedom capable of keeping it within the natural order. He defines freedom as a natural phenomenon grounded in intentional action: “the enveloping, enabling, life-shaping, conceptual atmosphere of intentional action, planning and hoping and promising—and blaming, resenting, punishing, and honoring” (Dennett, 2003, p. 10), always in accordance with his materialist theory of consciousness (Dennett, 1991) and the Darwinian algorithms that would govern evolution (Dennett, 1995). Once we refer to “freedom” as the phenomenon of intentional action, it becomes a phenotypic trait, where approximate degrees of intentionality are possible and where freedom follows an evolutionary course in which it makes sense, emerging and being made possible by the dynamics of that course.

This being so, the close relationship between freedom and the possession of real options to choose from changes completely. Dennett argues, with respect to probability and based on the idea of possible worlds, that it is a neutral issue with respect to the debate between determinism and indeterminism. This is because, when we say that something could have been otherwise, we do not normally mean, in a limited sense, that, given *the same* initial conditions, the consequence could have been otherwise, but that, under similar but slightly different initial conditions, other consequences could have followed. This counterfactual aptitude is useful to us, and is evolutionarily relevant because we do not

operate knowing the exact description of the initial conditions in which we live. As a result, playing with counterfactual variations is essential to us and imagining different possibilities is, therefore, part of the causal fabric of reality, whether or not it is deterministic. Possibility can only be ruled out in a deterministic world if it is interpreted in its narrow or limited sense. Therefore, "the truth or falsity of determinism should not affect our belief that certain unrealized events were nevertheless 'possible', in an important everyday sense of the word" (Dennett, 2003, p. 77). Following this line, what increases freedom is the increase in the number of possibilities that we are capable of seriously considering, a matter entirely foreign to the metaphysical debate. One begins with an entirely necessary course of action in the deterministic sense, that is to say, characterised by the absence of possibilities, and from then on develops degrees of freedom: "When life began, there was just one way of being alive. It was do A or die. Now there are options: do A or B or C or D or... die" (Dennett, 2003, p. 144). To move from one state to the other requires intentional action and this can only be based, Dennett says, on information regarding the future course of your life, without the need for it to be psychological (Dennett, 2003, pp. 154-155). Such information would allow us to move from a "mere opportunity" (bare opportunity) to a real opportunity (Dennett, 1984, pp. 116-117), so that, refining the above, our degree of freedom will be directly proportional to the amount of real possibilities available, which links the notion of freedom with the notion of control, its lower range being the model of a switch, "a simple switch, turned on and off by some environmental change" (Dennett, 2003, p. 162). As the real possibilities increase, the need for control will increase, generating an increasingly need for refinement of the information that can be captured, which in turn paves the evolutionary path for our brains (Dennett, 2003, p. 162).

A set of neural networks emerging from this need depends heavily on both the speed and reliability of their representations. However, such neural networks will inevitably make mistakes, gathering information that is not useful or missing information that is. In response to these errors, it will be adaptive for the organism to adjust its behaviour within an acceptable period of time. Thus, the concept of freedom is linked not only to control and the range of real possibilities, but also to a capacity for intentional learning aimed at perfecting the organism-environment adaptation. Therefore, from the

perspective of an evolutionary epistemology, the mind does not generate freedom, but it does refine and amplify it.

The next amplifier, obviously, lies in language. The difference between, say, birds and humans lies in our ability to, based on the uncertainties inherent in the possibility of choice, articulate doubts through language and thought, so that when our innate equipment malfunctions or its proper functioning proves adaptively insufficient, we can, in a sense, rewrite it and thus expand our available set of strategies, and with it, our freedom: "That sort of bridging of the gap between appearance and reality is a wrinkle that we human beings alone have mastered" (Dennett, 2003, p. 165).

Thus, freedom becomes a matter of knowledge, for through these abilities, we can, through culture, accumulate and share information—representations of reality and of our own interoceptive and behavioural processes to confront it—and this not only results in better strategies for tackling immediate or distant problems, but also in comparing different methods to attain truth or adaptive success, overcoming mere appearance, depending on their performance. In this way, the future projected by our genetics or our environment can be taken into account, represented, and changed once we understand how it works. In other words, we can say that freedom consists in the ability to intentionally exploit the widest achievable spectrum of real possibilities in terms of control based on a dual and interconnected representation of the present and the different possible futures arising from that representation of the present. It is in this sense that we "dwell" in the possibility of making mistakes, of falling into error, for we can misrepresent any of these realities, making learning an essential part of freedom.

Consequently, an indeterministic world does not provide us with more possibilities than a deterministic one, as the notion of possibility understood in this sense is independent of both. That is, once we place freedom within the Darwinian framework, libertarianism becomes irrelevant insofar as determinism loses its fatalistic sense of unconditional determination of phenomena, and is also detached from the principle of causality. If the world were indeterministic, as libertarians claim, we would retain possibilities in the same sense. Therefore, the fact that determinism implies a preestablished future does not exclude the human being from the realm of possibility, self-improvement, and choice. On the contrary, these elements would be part of the causal fabric of reality for a being burdened by ignorance, living by gathering the

necessary information to remedy it, since it has evolved to do precisely that. The belief in an incompatibility between determinism and freedom would arise from simultaneously adopting a cosmic view—derived from taking the perspective of God’s eye on the entire universal history as already given, whether it is a single deterministic course or “a static branching tree of trajectories” (Dennett, 2003, p. 93)—and the particular experiences that we constantly undergo.

If we were merely beings embedded in a deterministic world, we would be beings living in a world determined to present sets of possibilities for us to choose from when such situations arise. It would be irrelevant whether the choice was determined, for in any case, we would be choosing between possibilities, between options, and we would have evolved “to” choose in the best possible manner. A libertarian world—that is, an indeterministic one—does not, in this naturalistic sense, provide us with a greater number of possibilities or a wider range of choice. In both worlds, we would have evolved to consider a set of real possibilities, that is, those we are aware of and that are available, and whose use and control depend on the quality of our intellectual representations, which will vary depending on the needs required to make such representations. Freedom, from Dennett’s Darwinian perspective, is the ability to act by choosing between real possibilities based on representations of the present connected to representations of possible future outcomes, whether short-term or long-term. And this is independent of both the libertarian and the deterministic positions.

A more interesting philosophical question is whether Dennett’s arguments succeed in cancelling or dissolving the need for the metaphysical debate. We might agree with him that the problem can be naturalised without sharing the anti-metaphysical flavour of his proposal, assuming that each problem has its own place. Dennett, we might say, has dissociated the issues without dissolving any of them, such that a Darwinist need not reject libertarianism, but rather simply know how to distinguish between freedom as a natural phenomenon—which, against authors such as Immanuel Kant (1724–1804), would indeed be possible, not confined to the mere realm of the noumenal, nor should it be assumed as the domain of a kind of faith born from the needs of practical reason, since, as we saw, freedom would be feasible in a deterministic world, just as responsibility would be—and freedom as a cosmic or transcendent reality. More specifically, the freedom that is possible in the natural world would be an entirely biological

phenomenon, that is, “a natural freedom,” as the title of the first chapter of *Freedom Evolves* (Dennett, 2003) states, and would be based on action that seeks control, which relies on foresight, in one form or another and to varying degrees.

However, if this is the case, then it seems that with the problem of freedom on a cosmic scale we are truly only discussing the classic debate between determinism and indeterminism. That is, there would indeed be a dissolutive component in Dennett’s position: freedom would not be a cosmic problem but rather one relative to organisms, while the debate between determinism and indeterminism would apply to both scales without involving the issue of freedom. In other words, there would be, as Dennett (2005) points out, ways of defining freedom that are logically or conceptually possible but simultaneously incompatible with a deterministic world. The point, according to Dennett, is that they do not matter, insofar as they fall outside of a naturalistic view, since they remove themselves from nature itself, introducing an unnecessary ontological duality into our world. Darwinism, as Dennett puts it, “in a single stroke [...] unites the world of meaning and purpose and design with the world of meaningless matter and mechanism, cause and effect” (2005, p. 449). In other words, it is incorrect to think that the mere logical possibility of a concept is equivalent to its relevance, as among the set of logical possibilities regarding freedom, what is significant is the subset of logical possibilities that are physically and biologically realizable. It is in this sense that Dennett dismisses as irrelevant those concepts of freedom that fall outside this select subset. What matters to us is, to paraphrase Dennett himself, “the realm of genuine possibility” (2005, p. 451).

The debate, therefore, hinges on what is included in genuine possibility and what is not, leading us to a very different discussion: that of the merit of the naturalistic approach itself, as well as the best way to understand and modulate it. It is, we might say, an ideological debate about the best way to do philosophy, and it is evident, both from the tone and wording of some of its critics—Tim O’Connor (2005), for example, accuses Dennett of setting up a “church”—as well as from Dennett’s sharp irony—continuing with the example, he describes O’Connor’s article as “a generally funny article” (Dennett, 2005, p. 455)—that multiple psychological and emotional elements are at play, which in turn reveals the traditional importance of this debate for the moral life of human beings.

It is here that Darwinism can adopt a paradoxical relationship with libertarianism: could it be that the tendency to adopt a libertarian view has, in general, been adaptive? Or, more finely, could it be that our propensity for reward-punishment, for blaming and punishing or praising and rewarding actions, is itself adaptive, and more easily explained in libertarian terms, even though this may be erroneous or unnecessary, such that it indirectly emerges from this adaptivity as the predominant belief? What, in sum, is the Darwinian explanation for libertarianism?

For example, Sarkissian et al. (2010) demonstrate how the belief that the world is indeterministic and that a deterministic world is incompatible with the attribution of responsibility spans at least four different cultural contexts based on a sample of 231 undergraduate students living and studying in the United States, Hong Kong, India, and Colombia. Certainly, their sample is small and their study insufficient to determine why this predominance occurs in psychological terms, but it is not dismissible that, more or less paradoxically, it may have been the Darwinian structure of the biological dynamics of our world that has led to the predominance of the tendency towards libertarianism.

This points to the need for an experimental philosophy that studies these notions, not only by naturalising their content, as concepts that must be shaped from their genuine possibility, but also by naturalising the ideas themselves as elements with their own framework within the world woven by the Darwinian algorithm. If Dennett's approach is to naturalise, then accepting it means that philosophers and their debates must also be naturalised, and the failure to achieve this is clearly reflected in the bitterness of the expressions exchanged between them. Dennett may find the irony that it is the Darwinism that requires no libertarianism whatsoever that insists on the birth and flourishing of libertarianism, and then, what purpose do our debates serve, other than perpetuating a similar distribution of positions over the centuries?

Also, and against the homogenising aspirations of traditional philosophy, which is determined to provide a singular answer to every question posed, we may find that naturalising our investigations perhaps requires assuming heterogeneous answers depending on the situation (Nichols, 2011), as they address a function whose fulfilment does not respond to philosophy's drive to find the truth.

5. Conclusions

Against actualism—the view that only what is real is possible—we concur with Dennett in defending compatibilism, maintaining that the simultaneous belief in the notions of free will and determinism within the philosophy of biology entails no logical inconsistency, as the two are perfectly compatible. Nonetheless, the coherence of compatibilism crucially depends upon the definition of possibility, and we take it that the position would gain in rigour by clarifying how possibility may be simultaneously related to the notions of causal necessity and causal sufficiency.

When analysing the issue of temporal priority in the cause-effect relation, it is necessary to distinguish among the different theories established within physics. In classical mechanics, as illustrated by chaos theory, some typical systems are unpredictable due to their stochastic or chaotic causal behaviour. Consequently, it is problematic to describe them as deterministic in the strict sense, even though they are governed by deterministic mechanistic laws: temporal precedence is generally the rule, yet admits exceptions. In relativistic mechanics, a cause precedes its effect temporally only under inertial observation—that is, in a situation free from any external force—a condition that is in fact physically impossible, since gravity would be exactly zero only under the equally impossible hypothesis of infinite distance. It may therefore be said that, in relativistic causality, temporal precedence between cause and effect ceases to be a necessary requirement for explaining the ordinary behaviour of physical systems, as it suffices for the effect to lie within the future light cone of its corresponding cause. Hence the development of probabilistic theories of causality.

Finally, in quantum mechanics, temporal precedence between cause and effect appears even less indispensable, since, given Heisenberg's uncertainty principle, there is no logical or conceptual obstacle to admitting that a cause may be simultaneous with, or even posterior to, its effect. Let us not forget that quantum—or non-commutative—probability (commutativity referring to the interchangeability of operands without alteration of result) accommodates both quantum and classical states, and that the epistemological nature of quantum mechanics is intrinsically probabilistic, relying on notions such as probability amplitude to describe the behaviour of systems. Accordingly, since the same cause

does not always produce the same effect, but a range of possible effects, it becomes impossible—and, in a sense, irrelevant—to link the cause-effect relation to necessary temporal precedence.

Within Dennett's distinction among deterministic, indeterministic, and nihilistic universes—where the latter are defined by the absence of any regularity among past, present, and future states of matter—the difference between logical and empirical causality is not made sufficiently explicit. At the logical level of analysis, a nihilistic universe could indeed be conceivable alongside deterministic and indeterministic ones, but such an assertion seems far less plausible at the physical or factual level. This circumstance allows us to question the assumption that the sets of deterministic and indeterministic universes are mutually exclusive, and to examine the properties of their intersection. Thus, to account for chaotic systems within an operational framework of deterministic rules would require accepting that the laws of nature governing their behaviour are genuinely random—a *contradictio in terminis*—or, more simply and consistently, admitting that such systems possess a genuinely chaotic character despite being governed by deterministic laws.

Finally, if we remove the debate concerning freedom from a primarily metaphysical context and confine it within the framework of an evolutionary epistemology, referring it to the phenomenon of intentional action, then freedom may be regarded as a phenotypic trait associated with the control and range of real possibilities, as well as with an intentional learning capacity aimed at enhancing organism–environment adaptation. That is, freedom becomes an epistemic matter conferring adaptive advantage. It consists in the capacity to intentionally exploit the broadest attainable range of real possibilities in terms of control, by means of a dual and interconnected representation of the present and of the various possible futures, thereby acting through choices among real possibilities based on representations of the present linked with representations of future states—whether short or long term. From this perspective, human beings would not be deprived of the domain of possibility, self-improvement, and choice, despite the implications of a future determined by natural law.

Bibliography

- Álvarez Toledo, S. (1999). Notas sobre la causalidad probabilística. In E. Abalde Paz & J. M. Muñoz Cantero (eds.), *Metodología educativa I* (pp. 21-28). Universidad de La Coruña.

- Aristóteles. (1988). *Tratados de lógica. (Órganon). II. Sobre la interpretación, Analíticos primeros, Analíticos segundos*. M. Candel Sanmartín (trans.). Gredos.
- Batterman, R. (1993). Defining Chaos. *Philosophy of Science*, 60, 43-66. <https://doi.org/10.1086/289717>
- Curie, P. (1894). Sur la symétrie des phénomènes physiques : symétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique*, 3, 393-415. <https://doi.org/10.1051/jphysap:018940030039300>
- Dennett, D. C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. The MIT Press-Oxford University Press.
- Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Company.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea*. Simon & Schuster. <https://doi.org/10.1002/j.2326-1951.1995.tb03633.x>
- Dennett, D. C. (1999). *La peligrosa idea de Darwin. Evolución y significados de la vida*. C. Pera Blanco-Morales (trans.). Galaxia Gutenberg.
- Dennett, D. C. (2003). *Freedom Evolves*. Viking Penguin.
- Dennett, D. C. (2004). *La evolución de la libertad*. R. Vila Vernis (trans.). Paidós.
- Dennett, D. C. (2005). Natural Freedom. *Metaphilosophy*, 36(4), 449-459. <https://doi.org/10.1111/j.1467-9973.2005.00383.x>
- Eddington, A. (1928). *The Nature of the Physical World*. Cambridge University Press. <https://doi.org/10.5962/bhl.title.5859>
- Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3), 172-198. <https://doi.org/10.1007/BF01397280>
- Hempel, C. (1942). The Function of General Laws in History. *The Journal of Philosophy*, 39, 35-48. <https://doi.org/10.2307/2017635>
- Hoefer, C. (2024). Causal Determinism. *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition). E. N. Zalta & U. Nodelman (eds.). <https://plato.stanford.edu/archives/sum2024/entries/determinism-causal/>
- Hume, D. (1994). *Investigación sobre el conocimiento humano*. J. de Salas Ortueta (trans.). Alianza.
- Laplace, P. (1985). *Ensayo filosófico sobre las probabilidades*. P. Castrillo (trans.). Alianza.
- López Cuétara, J. (2005). El aristotelismo en el pensamiento de Robert Grosseteste. *Verdad y vida*, 63(242), 49-92.
- Monod, J. (1993). *El azar y la necesidad. Ensayo sobre la filosofía natural de la biología moderna*. F. Ferrer Lerín (trans.). Tusquets Editores.

- Nichols, S. (2011). Experimental Philosophy and the Problem of Free Will. *Science*, 331(6023), 1401-1403. <https://doi.org/10.1126/science.1192931>
- O'Connor, T. (2005). Pastoral Counsel for the Anxious Naturalist: Daniel Dennett's Freedom Evolves. *Metaphilosophy*, 34(4), 436-448. <https://doi.org/10.1111/j.1467-9973.2005.00382.x>
- Ostalé, J. y Pradier, A. (2010). Los futuros contingentes en Roberto Grosseteste, con una traducción inédita de su *De veritate propositionis*. *Daimon. Revista Internacional de Filosofía*, suplemento 3, 29-38.
- Poincaré, H. (1892). *Les méthodes nouvelles de la mécanique céleste*. Éditions Gauthiers-Villars.
- Quine, W. V. O. (1974). *Relatividad ontológica y otros ensayos*. M. Garrido y J. L. Blasco (trans.). Tecnos.
- Sarkissian, H., Chatterjee, A., DeBrigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is Belief in Free Will a Cultural Universal? *Mind & Language*, 25(3), 346-358. <https://doi.org/10.1111/j.1468-0017.2010.01393.x>
- Suppes, P. (1970). *A Probabilistic Theory of Causality*. North-Holland.
- Suppes, P. (1993). The Transcendental Character of Determinism. *Midwest Studies in Philosophy*, 18, 242-257. <https://doi.org/10.1111/j.1475-4975.1993.tb00266.x>
- Swartz, N. (2004). *Foreknowledge and Free Will*. Internet Encyclopedia of Philosophy. <https://iep.utm.edu/foreknow/>
- Van Fraassen, B. (1989). *Laws and Symmetry*. Clarendon Press. <https://doi.org/10.1093/0198248601.001.0001>
- Van Inwagen, P. (1983). *An Essay on Free Will*. Clarendon Press.

<https://doi.org/10.21555/top.v760.3248>

Dennett, Gould y la polémica sobre el adaptacionismo cuatro décadas después

Dennett, Gould and the Controversy over Adaptationism Four Decades Later

Rodrigo Sebastián Braicovich

Consejo Nacional de Investigaciones Científicas y Técnicas
Argentina

rbraicovich@gmail.com

<https://orcid.org/0000-0003-2785-7293>

Recibido: 18 - 09 - 2024.

Aceptado: 15 - 03 - 2025.

Publicado en línea: 15 - 06 - 2026.

Cómo citar este artículo: Braicovich, R. S. (2026). Dennett, Gould y la polémica sobre el adaptacionismo cuatro décadas después. *Tópicos, Revista de Filosofía*, 76, 187-218. <https://doi.org/10.21555/top.v760.3248>



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

Resumen

El objetivo de este artículo es reconstruir de forma general los acuerdos y divergencias entre Daniel Dennett y Stephen Jay Gould subyacentes a la polémica que mantuvieron durante el último cuarto del siglo pasado. Muestro que en los cuestionamientos teóricos y metodológicos iniciales de Gould al adaptacionismo radical se encuentran ya esbozadas las razones que condujeron a la crisis tanto de la sociobiología como de la Psicología Evolucionista, y que la incapacidad de Dennett de comprender las motivaciones de los ataques de Gould al adaptacionismo radical se desprende de su decisión de abordar la polémica en torno a las credenciales epistemológicas de la Psicología Evolucionista desde una perspectiva exclusivamente teórica.

Palabras clave: evolución; adaptacionismo; Psicología Evolucionista; naturaleza humana; Daniel Dennett; Stephen Jay Gould; sociobiología; polémica epistemológica; cuestionamientos metodológicos; filosofía de la biología.

Abstract

The aim of this paper is to provide a general reconstruction of the agreements and divergences between Daniel Dennett and Stephen Jay Gould that underlay the discussion between them during the last quarter of the 20th century. I show that Gould's initial theoretical and methodological critiques of radical adaptationism already outline the reasons that led to the crisis of both sociobiology and Evolutionary Psychology, and that Dennett's inability to understand the motivations behind Gould's attacks on radical adaptationism stems from his decision to approach the controversy surrounding the epistemological credentials of Evolutionary Psychology from an exclusively theoretical perspective.

Keywords: evolution; adaptationism; evolutionary psychology; human nature; Dennett; Stephen Jay Gould; sociobiology; epistemological controversy; methodological questions; philosophy of biology.

1. Introducción

Hacia finales del siglo pasado, Daniel Dennett y Stephen Jay Gould se embarcaron, en una serie de libros, artículos, reseñas de libros y notas breves, en una de las polémicas contemporáneas más célebres al interior del mundillo académico interesado en la capacidad explicativa de la teoría de la evolución por selección natural. La publicación por parte de Gould en 1977 de “The Spandrels of San Marco and the Panglossian Paradigm” (escrito en coautoría con Richard Lewontin) y en 1980 de “Sociobiology and the Theory of Natural Selection” recibió una primera respuesta por parte de Dennett en un extenso artículo titulado “Intentional Systems in Cognitive Ethology: The ‘Panglossian Paradigm’ Defended” (1983), en el que Dennett proponía una defensa limitada del adaptacionismo criticado por Gould y Lewontin, que luego habría de volverse la columna vertebral de uno de los libros más célebres de Dennett, *Darwin’s Dangerous Idea* (1993). El momento más álgido de la controversia se produjo, sin embargo, en 1997 con la publicación en *The New York Review of Books*¹ de un intercambio entre ambos autores en torno, entre otras cosas, al surgimiento de la Psicología Evolucionista² como variante de reemplazo de la sociobiología en el paradigma adaptacionista, y a las credenciales metodológicas del nuevo modelo teórico.³

¹ Un intercambio previo había sido publicado por la misma revista en 1992 en torno a la reseña por parte de Gould del libro de Helena Cronin, *The Ant and the Peacock*, en la cual aquel atacaba la concepción genecéntrica de la evolución defendida por Cronin —junto con Richard Dawkins y (de acuerdo con Gould) el propio Dennett—.

² En lo sucesivo, seguiré la propuesta de Buller (2005) de distinguir, a través del uso de mayúsculas iniciales, entre, por un lado, la “psicología evolucionista” (en minúsculas) como la adopción de una perspectiva evolutiva en el estudio de la conducta y la psicología humana, y la Psicología Evolucionista como el abordaje teórico específicamente defendido por autores como David Buss, Leda Cosmides, Martin Daly, Steven Pinker, Margo Wilson, John Tooby y Daniel Dennett.

³ Si bien Gould (2002) es importante para terminar de reconstruir el intercambio general entre ambos autores, el foco exclusivo de la discusión en ese texto concierne al concepto de “equilibrio puntuado”, por lo que habrá de quedar en un segundo plano en las próximas páginas.

En el libro de memorias que publicó antes de su fallecimiento, *I've Been Thinking*, Dennett (2023, pp. 395-404), sin embargo, ofrecía una reconstrucción de la polémica que procedía por un camino alternativo, poniendo el foco no tanto en los argumentos y posiciones que poblaron una polémica que duró cerca de veinticinco años, sino en las razones de índole subjetiva que habrían motorizado los desencuentros entre él y Gould. Si nos atenemos a la versión que ofrece Dennett de la historia del gradual alejamiento teórico y personal entre ambos (enmarcada por encuentros en campus universitarios y casas de campo en las que discutían los detalles de la controversia entre ambos, mientras compartían champaña con *sailing buddies* y académicos de renombre), el motor de la polémica debería ser buscado en ciertas inseguridades afectivas de Gould que le habrían llevado a adoptar una posición defensiva (y académicamente censurable) frente a los cuestionamientos de Dennett. Como sugeriré en las próximas páginas, no obstante, esa misma historia puede ser mirada desde un ángulo muy distinto, que pone sobre la escena la importancia de retornar sobre algunas advertencias claves de Gould que atañen al debate actual acerca de las credenciales metodológicas de la psicología evolucionista.

2. Algunos acuerdos de base

Cuando se emprende una lectura secuencial y detallada de los textos involucrados de manera directa o indirecta en el intercambio entre Dennett y Gould, se pone en evidencia inmediatamente que un número importante de los puntos de la polémica tiene su raíz, o bien en una mala lectura (a veces mutua, a veces unidireccional; a veces producto de una lectura descuidada de algún pasaje, a veces quizás atribuible a cierta mala fe), o bien en un intento por parte de ambos autores de responder no tanto a las críticas dirigidas de forma mutua, sino a las críticas que los lectores les atribuyeron.⁴ Una vez que desmalezamos los textos y traemos la luz las verdaderas posiciones asumidas en el debate, encontramos un terreno común entre ambos autores en torno a los siguientes principios:

⁴ Esto se hace particularmente claro, por ejemplo, en la evaluación que Dennett realiza en 1996 de *Spandrels* —un texto “oft-cited, oft-reprinted, but massively misread” (Dennett, 1996a, p. 238)—, en donde desliga explícitamente a Gould y Lewontin de la mala interpretación de la que aquel texto era víctima con frecuencia, la cual atribuía a ambos autores la intención última de impugnar por completo las pretensiones explicativas del darwinismo.

1. *La insuficiencia explicativa del mecanismo de la selección natural*: la selección natural no puede explicar la estabilización en una especie de la totalidad de los rasgos fenotípicos de todos los organismos que la componen.
2. *La legitimidad de ciertas explicaciones adaptacionistas*: en ciertos casos, una explicación adaptacionista de un rasgo determinado de un organismo puede ser tomada como la interpretación más plausible (al menos hasta que se vea refutada o surja una hipótesis con mejores credenciales).
3. *La falibilidad de las explicaciones adaptacionistas*: las explicaciones adaptacionistas que han sido propuestas a lo largo de la historia (o que pueden ser propuestas en el futuro) pueden estar equivocadas, lo cual puede conducir o bien a la necesidad de revisarlas o de abandonarlas por completo.

Si bien cuarenta años después del inicio de la controversia estos tres puntos de acuerdo pueden aparecer como verdades de Perogrullo que no ameritan atención, intentaré mostrar que desatender a las razones que condujeron a Gould a enfatizar dichos puntos es lo que ha conducido (entre otras cosas) a la psicología evolucionista a su paradójico estado actual. Para mostrar por qué esto es así, me detendré brevemente en los dos primeros principios mencionados.

3. La insuficiencia del mecanismo de la selección natural y el Adaptacionismo Radical

Una parte importante de la obra de Gould (o al menos de su obra más polémica) puede ser leída como una crítica a una variante reduccionista del darwinismo, a la que podemos denominar (por razones meramente prácticas) Adaptacionismo Radical (o Panadaptacionismo, o fundamentalismo ultra-darwiniano, en términos de Gould), que puede ser caracterizada como la *tendencia a atribuir al mecanismo de la selección natural un poder omniexplicativo*. Según esta postura, cada uno de los rasgos fenotípicos de una especie puede, y *debe* (eventualmente al menos), ser explicado mostrando cómo la posesión de dicho rasgo tiene

que haber conferido al portador de este una ventaja adaptativa frente a los organismos que carecían de dicho rasgo. A partir de esta premisa, Gould y Lewontin deducían un corolario adicional (no necesariamente asumido explícitamente por los adaptacionistas radicales), a saber, la idea de que existe un ajuste (*fit*) perfecto entre la totalidad de los organismos existentes en un momento dado y el entorno, haciendo del mundo el mejor (o más “ajustado”) de los mundos posibles.⁵

El cuestionamiento al Adaptacionismo Radical, sin embargo, no condujo a Gould a abandonar (como lo hará el ictiólogo Donn Rosen en esa misma época) toda esperanza de ofrecer una reconstrucción evolutiva de la genealogía de los rasgos de un organismo (Rosen, 1982), sino solamente a denunciar los excesos y abusos de los adaptacionistas radicales y a defender la necesidad de un pluralismo explicativo —más acorde, por otra parte, con el espíritu inicial de Darwin, quien había finalizado la introducción a *El origen de las especies* con palabras taxativas al respecto: “Estoy convencido de que la Selección Natural ha sido el medio de modificación principal *pero no el único*” (Darwin, 2008, p. 8)—.⁶ El desarrollo más sustancial de *Spandrels*, en efecto, está representado por las múltiples alternativas explicativas que pueden ser exploradas al intentar comprender la genealogía de un determinado rasgo: deriva

⁵ Gustavo Caponi sintetiza con precisión las alternativas en juego: “Una cosa es mostrar que la selección natural puede construir un ojo, y otra es analizar los más nimios detalles morfológicos y etológicos, muchos de ellos sin ninguna funcionalidad evidente, con base en la sospecha de que esa selección natural está siempre, directa o indirectamente, involucrada en la evolución de los mismos. Es esto último, y no lo primero, lo que funda y orienta al programa adaptacionista” (Caponi, 2011, p. 120).

⁶ “Aunque Darwin consideraba la selección como el más importante de los mecanismos evolutivos (al igual que nosotros), ningún argumento de sus oponentes le enfureció más que el intento común de caricaturizar y trivializar su teoría afirmando que se basaba exclusivamente en la selección natural” (Gould y Lewontin, 1979, p. 155). “Como Darwin insistió a gritos, y en contra de la mitología que se ha creado sobre él, la evolución es mucho más que selección natural. Darwin era un pluralista coherente que consideraba la selección natural como el agente más importante del cambio evolutivo, pero que aceptaba una serie de otros agentes y especificaba las condiciones de su presunta eficacia” (Gould, 2019, p. 310). En lo sucesivo, tanto las traducciones como las eventuales itálicas de énfasis son mías.

genética, pleiotropía, limitantes de orden filético, adaptación cultural, subproductos de otros rasgos adaptativos (*spandrels*), etc.

Ante la energía del embate de Gould, cabe preguntarse en este punto cuán real es (o era) el riesgo del Adaptacionismo Radical, dado que parecería improbable que un investigador formado atribuya a la selección natural un carácter omniexplicativo, negando con ello cualquier lugar a la deriva genética, las catástrofes geológicas, etc., en la reconstrucción de la historia de la evolución de las especies. En ocasiones, Dennett desestima por completo el riesgo en cuestión, calificando al Adaptacionismo Radical como un mero hombre de paja jamás encarnado por un evolucionista real: “La tesis de que cada propiedad de cada rasgo de cada cosa en el mundo viviente es una adaptación no es una tesis que nadie haya tomado en serio seriamente, o implicada por lo que nadie ha tomado seriamente, que yo sepa” (Dennett, 1996a, p. 276). Gould, por su parte, admite que aun los adaptacionistas más radicales conceden a factores como la deriva genética, la alometría, etc., un lugar en la genealogía de los rasgos fenotípicos, pero denuncia el carácter puramente teórico o formal de ese reconocimiento:

Por supuesto, se reconocen las limitaciones al poder omnipresente de la selección natural (la inercia filética, principalmente, entre ellas, aunque rara vez se reconocen las limitaciones arquitectónicas inmediatas [...]). Pero normalmente se les resta importancia o, lo que es más frustrante, simplemente se reconocen y luego no se toman en serio ni se invocan. [...] ¿No hemos oído todos el catecismo sobre la deriva genética?: solo puede ser importante en poblaciones tan pequeñas que es probable que se extingan antes de desempeñar un papel evolutivo sostenido [...]. La admisión de alternativas *en principio* no implica su seria consideración *en la práctica diaria* (Gould y Lewontin, 1979, pp. 150-152).

Como se hace evidente en este contrapunto, el problema fundamental no se vincula entonces con la existencia de evolucionistas que defiendan o hayan defendido *explícitamente* el Adaptacionismo Radical (aunque Gould ofrece ejemplos convincentes al respecto y el propio Dennett

parece defender esa posición en algunos momentos específicos),⁷ sino con las prácticas concretas de los investigadores. Y aquí es Dennett mismo quien lleva incesantemente agua para el molino de Gould, admitiendo la existencia de abusos de interpretaciones adaptacionistas (Dennett, 1996a, p. 242) —que son defendidas con una “despreocupada confianza” (Dennett, 1996a, p. 249)—, y que, en consecuencia, el Adaptacionismo Radical representa una amenaza real (Dennett, 1983, p. 353). La insistente alusión por parte de Dennett a la denominada “ley de Sturgeon” al momento de evaluar la solidez metodológica de las interpretaciones adaptacionistas puede ser vista, en este sentido, como una concesión importante a la posición de Gould: “Como en cualquier otro campo, la ley de Sturgeon se cumple: el noventa por ciento de todo es basura” (Dennett, 2023, p. 397).

4. La legitimidad de las explicaciones adaptacionistas

A partir de lo anterior, podemos afirmar que la polémica entre Dennett y Gould se halla estructurada sobre un acuerdo de base respecto del diagnóstico de que el Adaptacionismo Radical es tanto *teóricamente insostenible* como *realmente existente*. Como el propio Dennett se encarga de enfatizar, por otra parte, lejos está de las intenciones de Gould impugnar

⁷ Cfr. a modo de ejemplo: “[Gould y Lewontin] afirman que, aunque el adaptacionismo desempeña un papel importante en la biología, en realidad no es tan central ni omnipresente y, de hecho, deberíamos intentar equilibrarlo con otras formas de pensar. Sin embargo, he estado demostrando que desempeña un papel crucial en el análisis de todos los acontecimientos biológicos a todas las escalas, desde la creación de la primera macromolécula autorreplicante en adelante. [...] *El razonamiento adaptacionista no es opcional; es el corazón y el alma de la biología evolutiva*. Aunque pueda ser complementado y sus defectos reparados, pensar en desplazarlo de su posición central en biología es imaginar no solo la caída del darwinismo, sino el colapso de la bioquímica moderna y de todas las ciencias de la vida y la medicina” (Dennett, 1996a, p. 238). Si bien Dennett (1997) afirmará más adelante que en este pasaje no estaba defendiendo el Adaptacionismo Radical cuestionado por Gould y Lewontin, sino simplemente el uso de las interpretaciones adaptativas en general, parece evidente que se trata de una relectura forzada del texto. David Buss y colegas, por otra parte, no han dudado en mantener la apuesta del adaptacionismo radical en respuesta explícita a Gould: “La selección natural es el único proceso causal conocido capaz de producir mecanismos orgánicos funcionales complejos” (Buss *et al.*, 1998, p. 533).

la validez de las explicaciones adaptacionistas en su totalidad, dado que su objetivo se limita a enmarcar dichas explicaciones en un contexto pluralista como el defendido por el propio Darwin. La impugnación de Gould del Adaptacionismo Radical, en suma, no equivale a la impugnación del adaptacionismo como herramienta hermenéutica, sino solo de sus excesos. “¿Entonces cuál es el problema? El problema es cómo distinguir el buen —e insustituible— adaptacionismo del mal adaptacionismo” (Dennett, 1996a, p. 241). Siguiendo a George Williams, Dennett establece tres reglas básicas para distinguir —en principio— entre buenas y malas prácticas hermenéuticas: 1) no apelar al criterio de la adaptación en los casos en los que un rasgo puede ser explicado por otros factores de orden más primario (como la física); 2) no recurrir a explicaciones adaptativas cuando un rasgo es el producto de algún requisito de desarrollo general; 3) no recurrir al criterio de la adaptación cuando un rasgo es un subproducto de otra adaptación.⁸ Si atendemos a este listado, se hace evidente que Gould no podría estar más de acuerdo con las limitaciones impuestas por Dennett al pensamiento adaptacionista, dado que es precisamente la desatención a esas advertencias metodológicas (entre otras) las que él y Lewontin habían denunciado en *Spandrels*.

Partiendo de esta coincidencia (al menos parcial) en cuanto a los criterios que permiten distinguir el buen adaptacionismo del malo, o, mejor, la aplicación legítima del adaptacionismo de la ilegítima, podemos atender a la pregunta central —en cuanto a lo que me interesa destacar en estas páginas—: ¿es legítima la aplicación del adaptacionismo al estudio de la mente y las conductas humanas? Tomando en cuenta la defensa sistemática que a lo largo de toda su obra Dennett realizó de la sociobiología y de la Psicología Evolucionista, es evidente que en su caso la respuesta es decisivamente afirmativa: aun a pesar de las críticas que eventualmente dirigiera a ambos abordajes, la persistencia en su defensa evidencia que Dennett estuvo convencido hasta el final de que las buenas versiones de la sociobiología y de la Psicología Evolucionista estaban bien encaminadas en términos teóricos y metodológicos y que representaban el camino adecuado para comprender las raíces

⁸ La formulación de Williams era particularmente clara: “La regla fundamental —o quizás sería mejor llamarla una doctrina— es que la adaptación es un concepto especial y oneroso que debería ser usado solo cuando es realmente necesario” (Williams, 1996, p. 4).

evolutivas de nuestra mente, en la medida en que podían liberarse de los cargos elevados por los críticos contra el Adaptacionismo Radical.

En el caso de Gould, por el contrario, sus críticas a los abordajes sociobiológicos en particular fueron sistemáticas y persistentes, y no se centraron solamente en el hecho de que dichos abordajes representaran versiones de Adaptacionismo Radical y, en cuanto tal, postularan el carácter omniexplicativo del mecanismo de la selección natural (desatendiendo otros mecanismos explicativos como la deriva genética, la pleiotropía, etc.). La crítica central consistió en el hecho de que el adaptacionismo radical de los sociobiólogos los condujo a especulaciones carentes de otro fundamento que la plausibilidad narrativa y a conformarse con la “mera consistencia con la selección natural como criterio de aceptación” (Gould, 2019, pp. 312-313). La negativa a considerar la posibilidad de cualquier otra vía de explicación que no fuera la de la adaptación se conjugaba con una total despreocupación por indicar la evidencia (ya sea de tipo arqueológica o neurológica) que podría apoyar las interpretaciones propuestas, con lo cual la soberbia terminaba aunándose a cierta ceguera hermenéutica:

El rechazo de una historia adaptativa suele conducir a su sustitución por otra, más que a la sospecha de que podría ser necesario otro tipo de explicación. Dado que el abanico de historias adaptativas es tan amplio como fértiles son nuestras mentes, siempre pueden postularse nuevas piedras. Y si una historia no está disponible de inmediato, siempre se puede alegar ignorancia temporal y confiar en que llegará (Gould y Lewontin, 1979, p. 153).

Ahora bien: ¿es verdaderamente un problema que las interpretaciones sociobiológicas carezcan de apoyo empírico y que no sean más que *lazy guesswork*? Aparentemente no lo era para Dennett, dado que los peligros implícitos en la especulación adaptacionista se veían compensados, para él, por su (eventual) potencial explicativo, y esto a tal grado que el peligro de *no* aventurarse en intentos de explicación adaptacionista podía ser considerado mayor aun que el de hacerlo:

Los peligros señalados por Skinner, Gould y Lewontin son reales. Los adaptacionistas, al igual que los mentalistas, corren el riesgo de construir edificios

teóricos a partir de casi nada, y de hacer el ridículo cuando estos castillos de naipes se derrumban, como ocurre de vez en cuando. Ese es el riesgo que siempre se corre cuando se adopta la postura intencional o la postura adaptacionista, pero puede ser sensato correr el riesgo, ya que la recompensa suele ser muy alta y la tarea a la que se enfrenta el teórico más cauto y abstemio es extraordinariamente difícil. [...] Los peligros de la ceguera en el pensamiento adaptacionista, señalados tan vívidamente por Gould y Lewontin, tienen su imagen especular en cualquier enfoque que rehúya la curiosidad adaptacionista (Dennett, 1983, p. 353).

La defensa de Dennett parece atendible, al menos en principio, en el caso de la sociobiología, dado que una especulación adaptacionista imaginativa y creativa puede transformarse en un excelente punto de partida para una investigación que después proceda por la vía de la búsqueda de correlaciones entre distintas líneas taxonómicas o de la formulación de predicciones cuantitativas testeables (Gould, 2019, p. 314). Pero si esto puede valer para las investigaciones realizadas en el ámbito de la etología, el abordaje sociobiológico de la psicología y las conductas humanas requiere una concepción atomizada de los rasgos que serían objeto de la selección natural (la agresividad, la tendencia a la cooperación, etc.) que implica desestimar el carácter sistémico del desarrollo de un organismo y de sus bases genéticas:

Si la selección puede romper cualquier correlación y optimizar las partes por separado, entonces la integración de un organismo cuenta poco. Con demasiada frecuencia, el programa adaptacionista nos dio *una biología evolutiva de partes y genes, pero no de organismos*. Daba por sentado que todas las transiciones podían producirse paso a paso y subestimaba la importancia de los bloques de desarrollo integrados y las omnipresentes limitaciones de la historia y la arquitectura. Una visión pluralista podría devolver a los organismos, con toda su complejidad recalcitrante pero inteligible, a la teoría evolutiva (Gould y Lewontin, 1979, p. 163).

Este problema se vuelve más complejo aun con el surgimiento de la Psicología Evolucionista (PE, de aquí en más), cuyos méritos explicativos Daniel Dennett defendió a lo largo de toda su obra posterior a *Darwin's Dangerous Idea*. Las razones de esta mayor complejidad se deben a que la atomización de los rasgos sobre la que se construyen las interpretaciones sociobiológicas adopta en la PE la vestimenta de la tesis de la modularidad masiva, un abordaje que se vio atravesado por controversias desde sus primeras formulaciones. La PE se basa, en efecto, en una concepción de la mente que plantea que esta está compuesta por una cantidad masiva de módulos psicológicos innatos, inconscientes e informacionalmente encapsulados, que poseerían una base genética. El carácter innato (y genético) de dichos módulos se debería a que la posesión de cada uno de ellos otorgó (de manera independiente) una ventaja adaptativa a sus poseedores, al permitirles resolver un problema específico de supervivencia enfrentado de forma regular en el período en el cual nuestra especie se conformó como tal (el denominado EEA: *Environment of Evolutionary Adaptation*). De allí se deriva un postulado adicional, para la PE, que consiste en la negación de la doctrina culturalista de la *tabula rasa* y en la postulación de la existencia de una Naturaleza Humana, constituida por el conjunto de módulos que se halla presente ya en forma homogénea en todos los miembros de nuestra especie y que no está sujeto a variaciones a “corto plazo” (dado que la selección natural opera con temporalidades mucho más extensas que las que suponía la sociobiología).⁹ Esta atomización de los rasgos fenotípicos conductuales se vuelve más problemática que en el modelo sociobiológico por el hecho de que los problemas adaptativos que los módulos psicológicos ayudarían a resolver no son los actuales, sino los que enfrentó nuestra especie durante el Pleistoceno (entre 2.59 millones de años y 12 000 años atrás), lo cual permite a la PE eludir el cargo de panglossianismo (dado que ya no es necesario mostrar cómo las conductas humanas actuales se hallan ajustadas adaptativamente al entorno), pero conduce a una pregunta crucial: ¿cómo podemos estar seguros de cuáles fueron las presiones adaptativas para nuestra especie en aquel entorno, dado el estado absolutamente fragmentario del registro fósil? Y aquí el problema es cualitativamente distinto: si la

⁹ Entre los textos fundacionales o devenidos canónicos de la PE cabe mencionar, entre otros, Barkow *et al.* (1992), Buss (1999), Cosmides (1989) y Pinker (1999).

sociobiología podía (al menos en principio) volverse eventualmente una disciplina científica (en la medida en que nada impedía que produjera predicciones testeables y fuera confrontada con evidencia empírica), la PE clausura de antemano esa alternativa, puesto que las predicciones respecto del carácter adaptativo de los módulos innatos no pueden ser testeadas por el hecho de que han desaparecido las condiciones en las que operaron como respuesta a problemas adaptativos,¹⁰ por lo que la PE se vuelve una teoría estructuralmente infalsable y con ello, en el mejor de los casos, mera literatura entretenida y provocadora.

5. Cuatro décadas más tarde

Dos décadas después de la publicación de *Spandrels*, Massimo Pigliucci y Jonathan Kaplan (2000) realizaron un relevamiento del impacto que había tenido aquella publicación —junto con las críticas de Richard Lewontin (1979), Steven Rose (2000), Philip Kitcher (1985) y Marshall Sahlin (1973), entre otros— en los abordajes adaptacionistas, y llegaron a la conclusión de que el único avance verificable consistía en el reconocimiento consensuado entre los adaptacionistas de la necesidad de tomar en cuenta tanto la importancia de las distintas restricciones genéticas y epigenéticas (o de desarrollo) como el carácter dinámico del entorno (lo cual complejiza las hipótesis adaptacionistas). Esto no implicaba, no obstante, que el resto de las objeciones de Gould y de la primera ola de críticos del adaptacionismo hubieran sido tenidas en cuenta, fundamentalmente en lo que concierne a la necesidad de aportar evidencia empírica que pudiera alejar al adaptacionismo de la mera especulación imaginativa.

¹⁰ “Pero, ¿cómo podemos saber con detalle lo que hacían pequeñas bandas de cazadores-recolectores en África hace dos millones de años? Estos antepasados dejaron algunas herramientas y huesos, y los paleoantropólogos pueden hacer algunas inferencias ingeniosas a partir de esas pruebas. Pero, ¿cómo podemos obtener la información clave que se necesitaría para demostrar la validez de los relatos adaptativos sobre un EEA: las relaciones de parentesco, las estructuras sociales y el tamaño de los grupos, las diferentes actividades de machos y hembras, el papel de la religión, la simbología, la narración de historias y otros cien aspectos centrales de la vida humana que no pueden rastrearse en los fósiles? Ni siquiera conocemos el entorno original de nuestros antepasados: ¿permanecieron los humanos ancestrales en una misma región o se desplazaron? ¿Cómo variaron los entornos a lo largo de los años y los siglos?” (Gould, 1997, § 23).

Cuatro décadas después, y habiéndose ya probado tanto el vestido de la sociobiología como el de la Psicología Evolucionista, el adaptacionismo no parece poder ofrecer mejores credenciales epistemológicas que las que tenía para mostrar en sus inicios. Por el lado de la sociobiología, en efecto, los viejos hábitos metodológicos no parecen haber muerto, como podemos ver, a modo de ejemplo, en la controversial interpretación sociobiológica de la WHR (*waist-hip ratio*): desde los polémicos estudios iniciales de Devendra Singh (1993; Singh y Young, 1995), que postulaban la existencia de una preferencia masculina *universal* por una relación específica entre las caderas y la cintura (0.7) del sexo opuesto que habría tenido una función adaptativa (a saber: seleccionar una pareja con las mejores posibilidades de éxito reproductivo), dicha preferencia ha seguido siendo considerada hasta la actualidad en ciertos estudios como parámetro de investigación de las preferencias innatas en la selección de pareja (v. g. Cloud *et al.*, 2023; Pazhoohi *et al.*, 2024), a pesar del cúmulo de críticas que dicho abordaje ha recibido desde una perspectiva metodológica.

Por el lado de la Psicología Evolucionista, si tomamos en cuenta las objeciones no solo metodológicas sino también teóricas a las que ha sido sometida, la afirmación de Dennett respecto de que dicha disciplina comenzaba en 1996 a mostrarse como “una escuela polifacética y robusta” (Dennett, 1996b, p. 170) parece haber perdido todo sustento, y no solo por el hecho de no haber logrado ni en la teoría ni en la práctica liberarse del lastre del Adaptacionismo Radical, sino por una serie de falencias teóricas y metodológicas decisivas. Para fundamentar dicha conclusión esquematizaré, en primer lugar, los principios fundamentales de la PE, para luego sintetizar algunas de las objeciones más importantes que han sido presentadas a esta.

5.1. Postulados centrales de la PE y ejemplos

Haciendo a un lado por cuestiones prácticas ciertos matices y variantes ocasionales, podemos sintetizar el núcleo de la PE como sigue: la rica diversidad que observamos en las conductas de los hombres es el resultado del encuentro entre las condiciones sociales en las que el sujeto actúa y un substrato mental universal, homogéneo y genéticamente codificado, lo cual explica tanto la diversidad (cultura) como la universalidad (biología) —esto último se debe a que todos los seres humanos comparten una arquitectura mental específica—. La razón de la universalidad de dicha arquitectura mental obedece al hecho

de que esa arquitectura en particular otorgó una ventaja adaptativa durante el período en el cual nuestra especie se conformó como tal (es decir, durante el Pleistoceno). La especificidad de dicha arquitectura radica en el hecho de que está compuesta por un número de rasgos de diseño (“módulos” o, alternativamente, “mecanismos psicológicos” en general), cada uno de los cuales permitió a nuestra especie responder a los desafíos adaptativos más críticos durante aquel período (obtención de recursos, supervivencia, reproducción, etc.).

Sobre la base de este modelo ha surgido una cantidad enorme (y creciente) de explicaciones adaptacionistas de ciertos rasgos fenotípicos o conductuales. David Buss, uno de los Psicólogos Evolucionistas más prolíficos y reconocidos, por ejemplo, ha postulado que los seres humanos nacemos con “mecanismos mentales evolucionados [...] que nos motivan a matar” (2006, p. 240). Randy Thornhill y Craig Palmer han concluido que la tendencia masculina a violar a las hembras es un producto de la “sexualidad evolucionada masculina [*men’s evolved sexuality*]” (2000, pp. 190-191), ya sea porque constituye una adaptación en sí misma o porque representa un subproducto de otros rasgos adaptativos. Tomando como premisa esta conclusión, Gavin Vance y su equipo han propuesto como corolario que en la medida en que los hombres sigan enfrentando situaciones que hagan posible llevar adelante una violación sin perspectivas (reales o percibidas) de castigo —tal como sucede en las situaciones de conflicto bélico— “las violaciones en épocas de guerra van a persistir” (Vance *et al.*, 2024, p. 152). En un orden menos dramático de cosas (aunque conectado por la misma lógica patriarcal), Max Krasnow *et al.* (2011, p. 9) han postulado que cuando una mujer recorre los mercados en busca de comida, muestra un desempeño superior al de los hombres a causa de que se activa en ella una adaptación para la navegación específica de recolección (*gathering-specific navigational adaptation*) que habría surgido a lo largo de la historia evolutiva de nuestra especie. Para concluir un muestreo breve que termine de evidenciar cómo la perspectiva marcadamente sexista del (“mal”) adaptacionismo no cesó con la sociobiología, Anya Hurlbert y Yazhu Ling (2007, p. 625) han defendido la existencia de preferencias universales (transculturales) de los hombres por el color azul y de las mujeres por el color rosa, argumentando que esta última puede derivar del desarrollo adaptativo de las mujeres de una preferencia por los colores que sería el resultado de la división natural del trabajo propia de las comunidades cazadoras-recolectoras, en las que la mujer debía

afinar su capacidad de detectar frutas maduras (en el rango del rojo) en un entorno de follaje verde.

5.2. Principales objeciones teóricas y metodológicas a PE

Entre las objeciones más frecuentes que han sido presentadas tanto al núcleo teórico de la PE como a las investigaciones concretas llevadas adelante desde dicho modelo, cabe destacar al menos las siguientes cuatro. La primera de ellas consiste en que la PE postula explícitamente (contra la sociobiología) que los tiempos que la evolución impone para que un rasgo adaptativo sea incorporado en forma mayoritaria a una población obligan a suponer que los mecanismos adaptativos humanos tienen que haber surgido durante un período extremadamente largo y estable (como el del Pleistoceno). Como señala Buller, sin embargo, “la idea de que estamos estancados en una psicología adaptada al Pleistoceno subestima enormemente el ritmo al que la selección natural y sexual puede impulsar el cambio evolutivo” (Buller, 2009, p. 79).¹¹ La premisa acerca de las altas demandas cronológicas del proceso de la selección natural ha sido cuestionada teóricamente (Hawks *et al.*, 2007; Reznick *et al.*, 1997) y se ha presentado evidencia que sugiere la existencia de adaptaciones recientes de diversa índole (Barreiro *et al.*, 2008; Lao *et al.*, 2007; Nielsen *et al.*, 2005; Pickrell *et al.*, 2009; Simonson *et al.*, 2010; Tishkoff *et al.*, 2007). Esta objeción, desde ya, no afecta al adaptacionismo como estrategia de investigación, sino solamente a la premisa de la PE que afirma el linaje exclusivamente pleistocénico de nuestra arquitectura mental (para una crítica detallada de la *ancient provenance thesis*, cfr. Buller, 2005, pp. 107-111).

De modo complementario, se ha objetado que la datación de dicho linaje desatiende el hecho de que, si la mente humana está efectivamente compuesta por mecanismos psicológicos adaptativos, es más que probable que muchos de ellos posean un origen anterior al surgimiento de nuestra especie como tal (Lloyd, 1999, p. 229). Esto implicaría que las adaptaciones pleistocénicas conviven en nuestra mente con mecanismos filogenéticamente anteriores (quizás compartidos con otras especies de

¹¹ Cabe agregar una observación importante de Buller (2005, p. 108): lo que importa no es qué porcentaje de la historia de nuestra especie haya transcurrido en el Pleistoceno y qué porcentaje en un entorno más semejante a nuestro entorno actual, sino qué tipo de cambios ocurrieron durante cada uno de los dos períodos.

animales no humanos), con lo que, como señalan Gray *et al.* (2003, p. 265), “la evolución conductual y cognitiva ni comenzó ni terminó de forma abrupta en el Pleistoceno”.

Vinculada a esta objeción se halla la observación (ya constatada por Gould) de que la PE oscila sin fundamentos explícitos entre un adaptacionismo “desfasado” (como la maladaptativa tendencia al suicidio después de una ruptura de pareja o de una infidelidad, postulada por Saad, 2007) y un adaptacionismo “ajustado” —como en la tendencia innata a matar postulada por Buss (2006, p. 239), que todavía tendría un efecto adaptativo—, lo cual le permite alternar estratégicamente entre las dos posiciones según sus necesidades: cuando no se puede conjurar una narrativa que “demuestre” el carácter adaptativo de un rasgo en el presente (ya sea porque ya no confiere una ventaja adaptativa o porque, de hecho, posee un carácter maladaptativo), se postula que este fue adaptativo durante el Pleistoceno. La PE oscila, de esta manera, entre afirmaciones *fácticamente* carentes de evidencia respecto del carácter actualmente adaptativo, y afirmaciones *estructuralmente* carentes de evidencia respecto del carácter adaptativo en el pasado de cada rasgo.

A ello cabe agregar dos elementos que, en rigor, pueden no ser considerados como objeciones teóricas sino más bien como dificultades —sumamente complejas— inherentes a los programas adaptacionistas. El primero de ellos concierne a la dificultad de especificar exactamente los límites de un rasgo: atribuirle carácter adaptativo a un determinado rasgo presupone no solo atribuirle una base genética (sobre lo que volveré más adelante) sino también un carácter discreto, límites precisos y efectos claramente demarcados (que son los que, en rigor, le conferirían al rasgo un carácter adaptativo) —y ello, como lo muestra el clásico ejemplo analizado por Gould y Lewontin de la morfología de la mandíbula humana, puede ser extremadamente difícil (para una reciente revisión sistemática de este caso en particular y del problema teórico de la definición de un rasgo, cfr. Meneganzin *et al.*, 2024)—. A esto cabe sumar el problema de la potencial arbitrariedad implícita en la definición del problema adaptativo que un determinado rasgo se supone que estaría llamado a resolver: como señalan Sterelny y Griffiths (1999), el problema de la elección de pareja, por ejemplo, puede reducirse a la mera obtención casual de pareja para asegurar la reproducción, o puede, alternativamente, ser enfocado como un mosaico de problemas que incluye la infidelidad, el acceso a recursos, etc. El problema que un rasgo se supone que debe resolver, en términos de los autores, “no

está ahí afuera”, sino que es en buena medida resultado de un proceso hermenéutico complejo (Sterelny y Griffiths, 1999, p. 239).

Si asumiéramos que estas últimas dos dificultades podrían ser eventualmente sorteadas mediante definiciones sólidas respecto de cada uno de los rasgos investigados y mediante una reconstrucción precisa del problema (o del mosaico de problemas) a cuya solución dichos rasgos colaboraban en el pasado, permanecen en pie tres objeciones adicionales que atañen a una triada de elementos que estructura todo el paradigma: rasgo adaptativo, módulo psicológico y base genética. Respecto del primer elemento, la objeción concierne a la premisa acerca de la pretendida universalidad de los rasgos, y es aquí donde la PE ha presentado su costado más débil desde una perspectiva metodológica, dado que ha tendido a tomar como evidencia de universalidad de un rasgo muestras extremadamente reducidas, en absoluto representativas y culturalmente sesgadas (provenientes en general de entornos WEIRD: Apicella *et al.*, 2020; Pollet y Saxton, 2019; Veillard, 2017). Respecto del tercer término de la triada, se ha objetado con frecuencia que la actitud de los defensores de la PE al momento de precisar las bases genéticas de dichos rasgos (que son las encargadas de hacer de estos un elemento hereditario) ha tendido a optar por una de dos alternativas: o bien permanecer en silencio al momento de ofrecer evidencia en favor de la existencia de dicha base genética (pero dejando entrever que efectivamente existe o que eventualmente podrá ser descubierta), o bien negar, como hace Steven Pinker, que exista una relación directa entre *un* rasgo y *un* gen, postulando a cambio una relación más compleja entre genes y rasgos:

Los órganos mentales complejos, como los órganos físicos complejos, seguramente se construyen mediante recetas genéticas complejas, con muchos genes que cooperan de formas aún insondables. Un defecto en cualquiera de ellos podría corromper todo el dispositivo, del mismo modo que un defecto en cualquier pieza de una máquina complicada (como un cable de distribución suelto en un coche) puede hacer que la máquina se pare (Pinker, 1999, p. 35).

Si bien esta última alternativa parece ser la más sensible en virtud de la comprensión actual de la relación existente entre genes y rasgos fenotípicos complejos, el problema que enfrenta esta estrategia consiste

en el hecho de que postular que un determinado rasgo adaptativo se halla codificado por una pluralidad de genes actuando en forma articulada parece poner serias restricciones a las posibilidades de transmisión por vía de la replicación genética, dado que, parafraseando a Pinker, una falla en la replicación de uno solo de los genes involucrados podría hacer que la máquina (el rasgo codificado por la pluralidad articulada) ya no arranque nunca más en la historia de esa especie.¹²

El eslabón intermedio de la triada señalada, sin embargo, es el que ha recibido la munición más gruesa por parte de los críticos: la tesis de la modularidad masiva de la mente (TM, de aquí en más), en efecto, ha sido objeto de cuestionamientos tanto conceptuales como metodológicos, que parecen haber sumido en una crisis casi definitiva a la versión adaptacionista de la modularidad. Esto es particularmente importante porque la TM representó el mejor intento de especificación conceptual sistemática de una mera hipótesis general, a saber, la de la existencia de “mecanismos psicológicos” hereditarios que conferirían una ventaja adaptativa. La apropiación por parte de los psicólogos evolucionistas del concepto fodoriano de “módulo” permitió precisar con (relativa) exactitud qué era aquello que mediaba entre los genes y las conductas: si las conductas en sí mismas no pueden hallarse codificadas genéticamente, los módulos que son responsables de esas conductas *sí* podrían estarlo, dado que la TM especifica un modelo de arquitectura cerebral. Esto abría a la posibilidad de diseñar experimentos que permitieran testear y, eventualmente, ofrecer evidencia en apoyo de la existencia efectiva de una organización modular de la mente. Si la mente humana se hallaba compuesta de módulos con las características específicas que la PE les atribuye, cabía esperar 1) que los desarrollos en el ámbito de las

¹² A esto cabe agregar el problema (ya no teórico sino práctico) de que la literatura proveniente de la PE se halla plagada de alusiones efectivas a la existencia de una base genética *unitaria* para cada rasgo propuesto como adaptativo: “La mayoría de los psicólogos evolucionistas afirman claramente que no existe una correlación de uno a uno entre ningún gen específico y ningún órgano, módulo o mecanismo mental determinado. Sin embargo, su narrativa está repleta de conjeturas que involucran genes y módulos con nombres que se correlacionan con comportamientos notablemente específicos, estrategia narrativa que, en el mejor de los casos, induce a error. A medida que uno va leyendo se le olvida la advertencia y solo oye la repetición de la especificidad de uno a uno con el vínculo genético-conductual” (McKinnon, 2012, p. 25).

neurociencias revelarían gradualmente las redes neuronales implicadas específicamente en la activación de cada uno de los módulos, y 2) que los desarrollos en el ámbito de la genética revelarían la base genética que codifica la conformación, durante el desarrollo primario del cerebro, de dichas redes. Esto implicaba (voluntariamente o no) proporcionar, al menos, un criterio de falsabilidad, dado que si resultaba ser el caso de que los desarrollos en el ámbito de las neurociencias tendieran a favorecer hipótesis no localizacionistas, la tesis de la modularidad de la mente debería ser abandonada o, cuando menos, revisada.¹³

En este frente de la PE, los problemas que enfrentó la TM fueron insuperables: a las escasas precisiones que los psicólogos evolucionistas ofrecieron respecto de la base neuronal y fisiológica de los módulos que postulaban, se sumó, en primer lugar, el hecho de que la concesión de la posible existencia de módulos psicológicos no implicaba atribuirles una base genética (y, por lo tanto, hereditaria), dado que estos podían ser explicados como productos de otros factores —como la *downward engineering* (Sterelny, 2003) o la plasticidad neuronal durante el período de desarrollo del cerebro (Buller y Hardcastle, 2000; Peters, 2013; Woodward y Cowie, 2003)—; en segundo lugar, el hecho de que

¹³ Cabe enfatizar que el origen de la TM fue tan especulativo como todo el resto de los postulados de la PE, dado que los psicólogos evolucionistas no fueron conducidos inicialmente a la TM por vía de la evidencia, sino que la dedujeron como corolario de ciertas premisas (indemostradas), a saber, la tesis de que la selección natural no puede haber favorecido el desarrollo de mecanismos psicológicos de propósito general, dado que estos serían demasiado lentos o establecerían exigencias informacionales y de procesamiento demasiado costosas. Como señalan James Woodward y Fiona Cowie, sin embargo, “es simplemente erróneo suponer que los módulos son invariablemente (o incluso normalmente) superiores a los dispositivos de propósito general. Qué tipo de organización mental se verá favorecida por la selección depende totalmente de los detalles de las presiones de selección a las que está sometido un organismo y de su estructura genética. [...] Factores como la variabilidad del entorno, los costes de cometer distintos tipos de errores, los costes de incorporar al organismo distintos tipos de capacidades discriminativas, etc., pueden tener grandes efectos sobre la aptitud relativa de las estrategias de propósito general frente a las más especializadas. Además, la capacidad de adaptarse rápidamente (es decir, durante la vida del individuo) a circunstancias cambiantes es vital para los organismos que viven en entornos inestables” (Woodward y Cowie, 2003, p. 313).

ciertos fenómenos que habían sido presentados como evidencia del procesamiento modular de la mente podían ser explicados recurriendo a procesos de dominio general (Eliasmith, 2013, pp. 231-235); y, en tercer lugar, que la integración de la información proveniente de los distintos módulos obligaba a postular la existencia de un metamódulo o de un mecanismo de dominio general (Carruthers, 2006, pp. 211-276; Buller, 2005, pp. 158-160). El golpe definitivo para la TM provino, no obstante, del hecho de que las investigaciones en neurología comenzaron, ya desde principios de siglo (Uttal, 2001 se volvería un punto de referencia ineludible), a abandonar los abordajes estrictamente localizacionistas en los que la TM se había inspirado y comenzaron a favorecer abordajes basados en redes funcionales distribuidas (Gonsalves y Cohen, 2010; Kaskan *et al.*, 2005; Oosterwijk *et al.*, 2012; Palecek, 2017; Pessoa, 2008; Prinz, 2006; Thagard y Schröder, 2015), en los cuales cada estado mental es el resultado de “procesos mentales más básicos de dominio general” (Lindquist y Barrett, 2012, p. 533).

Dado que la suerte de la PE se hallaba atada a la TM —asumiendo que realmente aquella aspiraba a convertirse en un abordaje científico mediante la formulación de hipótesis testeables y falsables, y no meramente en un dispositivo heurístico que “guía el descubrimiento biológico, [pero] no la justificación biológica” (Resnik, 1996, p. 49)—, entonces las alternativas abiertas a los psicólogos evolucionistas no eran muchas. La más sensata consistía en abandonar la TM¹⁴ y partir en busca de un modelo teórico de arquitectura cerebral que fuera a la vez compatible con los presupuestos adaptacionistas y con los desarrollos más actuales en neurociencias. Otra alternativa, que es la que de hecho primó, consistió en simplemente abandonar el repertorio conceptual de la modularidad y replegarse en un vocabulario neutro que no

¹⁴ Cabe admitir que el abandono de la TM no representaba, en rigor, la única alternativa, dado que esta podía ser revisada sustancialmente, como propuso Martin Palecek (2017). El desafío, sin embargo, es complejo, dado que lo que exige ser revisado parece ser aquello que es intrínseco a la versión adaptacionista de la TM. Es por ello que la alternativa de Palecek parece tirar al bebé junto con el agua de la bañera: “La forma de hacer que la modularidad vuelva a ser productiva es modificar la modularidad postfodoriana: abandonar la ingeniería inversa y la especulación sobre la adaptación, y entender la modularidad dentro de una perspectiva científica de redes” (Palecek, 2017, p. 140).

comprometía a la PE con ninguna hipótesis falsable. De allí que ante las primeras críticas teóricas y metodológicas y ante los avances sostenidos de las neurociencias que favorecían un modelo de redes funcionales integradas, la mayoría de los psicólogos evolucionistas retornó, después de la primera década de este siglo, al concepto indefinido de “mecanismos psicológicos”, abandonando todo compromiso explícito con la TM.

6. Algunas consideraciones finales sobre Dennett, los intelectuales y la “Psicología Evolucionista pop”

Si bien es cierto, como dijo el genetista de poblaciones Theodosius Dobzhansky, que nada en biología tiene sentido si no es a la luz de la evolución, su afirmación requiere una ampliación. Nada en biología tiene sentido si no es a la luz de la historia: la historia evolutiva de la especie, la historia del desarrollo del organismo vivo individual y, en el caso de los humanos, por supuesto, la historia social, cultural y tecnológica. A esto hay que añadir *la historia de nuestras propias ciencias*, que proporciona los supuestos en los que intentamos ver e interpretar el mundo.

Rose (2000, p. 279).

El medio siglo que abarca (aproximadamente) el devenir de la socio-biología y la Psicología Evolucionista puede ser visto como la historia del surgimiento y caída de una de las posibles estrategias para poner la teoría de la evolución por selección natural al servicio de la exploración de la psicología humana. A pesar del consenso relativamente extendido entre la comunidad científica de relegar ambas teorías a un rincón olvidado de la historia de la ciencia —o de las pseudociencias, a los ojos de Tattersall (2001, p. 657)—,¹⁵ ninguna de ambas propuestas ha desapa-

¹⁵ En un ejercicio de fino sarcasmo, Johan Bolhuis *et al.* resumían la situación que la PE afrontaba hace ya más de una década: “Los conceptos clave de la PE han dado lugar a una serie de suposiciones muy extendidas (por

recido del horizonte académico, y menos aún del horizonte mediático: si hace casi veinte años David Buller señalaba que la PE había sido tan efectiva en el terreno del *marketing* que había logrado que todo intento de abordaje de la psicología humana desde una perspectiva evolucionista pasara a ser identificado directamente con el paradigma de la PE (Buller, 2005, p. 12), la situación en el presente se ha radicalizado: las revistas de divulgación científica y los medios masivos de comunicación se hallan inundados de noticias respecto del “descubrimiento” de mecanismos psicológicos que explicarían las diferencias entre hombres y mujeres al momento de abordar ciertas situaciones o procesar determinada información, o que explicarían ciertas tendencias de la “naturaleza humana” que habían permanecido hasta ahora ocultas o inexploradas—todo ello enmarcado en narrativas evolucionistas reduccionistas, esquemáticas, completamente especulativas e infalsables—.

En la medida en que las fuentes de financiamiento de la investigación sigan estando regidas por la lógica de mercado que mide la calidad de las investigaciones (y decide, en función de ello, a qué investigadores o a qué líneas de investigación financiar) en términos de impacto académico y mediático (como lo hace, fundamentalmente, el sistema norteamericano), no cabe esperar que desaparezca el estímulo a seguir publicando *papers* y libros producidos desde un paradigma que hace tiempo atraviesa una crisis terminal. En el caso específico de la PE, esto adquiere una faceta adicional, que se vincula con la coincidencia que existe entre la tendencia general de los corolarios que la PE ha extraído respecto de la supuesta naturaleza humana (y particularmente a la diferencia entre la “psiquis femenina” y la masculina) y la cosmovisión neoliberal y patriarcal del individuo y de su relación con la comunidad:

Pese a la fragilidad de su argumentación, la psicología evolucionista es atractiva porque reúne en una sola gran narración un número de creencias que resultan centrales para la cultura euroamericana respecto al carácter innato de las diferencias de género y de la doble moral

ejemplo, que es improbable que el comportamiento humano sea adaptativo en los entornos modernos, que la cognición es específica de un dominio, que existe una naturaleza humana universal), que con la ventaja de la retrospectiva ahora sabemos que son cuestionables. Una PE moderna adoptaría un marco teórico más amplio, abierto y multidisciplinario. [...] La evidencia de las disciplinas adyacentes sugiere que, si la PE puede reconsiderar *sus principios básicos*, florecerá como disciplina científica” (Bolhuis *et al.*, 2011, p. 6).

sexual; acerca de la naturalidad de los valores económicos neoliberales del interés personal, la competencia, la elección racional y el poder del mercado de crear relaciones sociales; sobre la supervivencia del más apto y la fuerza determinante de los genes; en relación con los orígenes evolutivos y el hombre, el cazador, y respecto a que la complejidad de la vida tiene una única clave. Al combinar tales creencias populares en una narración atractiva y encubrirla con el manto de la ciencia, los psicólogos evolucionistas dan a las que de otra manera serían “verdades” culturalmente específicas el aura de una verdad única, fundamental y universal (McKinnon, 2012, p. 78).

Curiosamente, la relación entre las investigaciones llevadas adelante en el marco de la PE y lo que David Buller ha denominado (parafraseando a Philip Kitcher) la “Psicología Evolucionista pop” —la cual “ofrece afirmaciones grandiosas y abarcadoras sobre la naturaleza humana para consumo popular” (Buller, 2009, p. 79)— ha comenzado a devenir hace tiempo en una relación asintótica: mientras que la PE ha sido relegada al plano de las pseudociencias, las publicaciones enmarcadas en un modelo laxamente vinculado con esta han proliferado exponencialmente desde 2010 en adelante (aun en editoriales universitarias reconocidas), y las repercusiones mediáticas de dichas publicaciones han contribuido a cimentar un imaginario sumamente difundido basado en la idea de que la agresividad, la violencia, la competitividad, etc., son adaptaciones que forman parte de la naturaleza (masculina) y que cualquier intento de poner límites a dichas tendencias está destinado al fracaso. Una *Realpolitik* científicamente informada, desde esta perspectiva, debería tomar como punto de partida del diseño de políticas públicas las lecciones acerca de los límites que la evolución ha establecido a la maleabilidad humana. Las supuestas demostraciones de la PE pop no son, en este sentido, ideológicamente neutras: en la medida en que el carácter competitivo de la naturaleza humana se considere demostrado, todo proyecto de construcción de una sociedad anclada en valores como la cooperación y el altruismo, por ejemplo, será considerado estructuralmente inviable, tal como lo será toda pretensión de alcanzar una igualdad absoluta entre géneros si partimos de la pretendida constatación de diferencias esenciales y socialmente relevantes entre el cerebro masculino y el femenino. Aun cuando admitamos que las hipótesis más extremas de la PE pop —como, por ejemplo, que poseemos “mecanismos mental evolucionados [...] que nos motivan a matar” (Buss, 2006, p. 240)— no pueden en rigor ser refutadas (dado su carácter esencialmente

infalsable), adoptar una postura crítica respecto del carácter ideológico que se juega en los usos de dichas hipótesis puede ser considerado como una obligación moral y política, tanto como epistemológica (o, quizás mejor: *a la vez* que epistemológica, dado que la separación de las dos dimensiones parece ser precisamente el problema en este caso en particular).

¿Podemos suponer que Dennett fue consciente de esto último y que compartió tal preocupación? Es imposible saberlo, desde ya, pero cabe aventurar que, aun cuando no es necesario suponer que compartía la panglossiana convicción de su amigo Steven Pinker (2018, p. 19) de que vivimos en el mejor de los mundos posibles hasta el momento, quizás fue su sistemática renuencia a aventurarse en el terreno de la filosofía política (más allá de su defensa de ciertos valores sociopolíticos generales como la laicidad del Estado y políticas educativas de cuño racionalista) lo que lo habría llevado a mantenerse al margen del problema de los malos usos de la (mala) ciencia. Es quizás por ello que a pesar de que en su obra podamos encontrar críticas ocasionales a ciertas expresiones del adaptacionismo más demoledoras que las del propio Gould (v. g. Dennett, 1996a, pp. 452-493), y a pesar de que eventualmente llegó a admitir —contra el espíritu central de PE— que los entornos rápidamente cambiantes [como el entorno humano a partir del desarrollo cultural] son enemigos de respuestas conductuales programadas [*hard-wired behavioral responses*],¹⁶ la posición de Dennett frente a las distintas variantes del adaptacionismo fue intransigente a lo largo de su obra: su defensa de las virtudes explicativas de la “buena” sociobiología permaneció inalterada hasta 2023 (aun cuando ya en la década de los noventa gran parte de los sociobiólogos habían comenzado

¹⁶ Cabe agregar a esta enumeración la curiosa argumentación que Dennett ofrece en “The mythical threat of genetical determinism” (en contra, nuevamente, de Gould, pero reiterando en su mayor parte la igualmente falaz argumentación que encontramos en Dawkins. 1982, Capítulo 2), en donde plantea que la postulación de la existencia de una naturaleza humana (compuesta por una serie de tendencias innatas) no compromete al adaptacionismo a un determinismo genético, dado que dichas tendencias pueden ser suprimidas por la cultura, el desarrollo, la educación, etc. (Dennett, 2003, 2004, p. 480). Se trataría, en otras palabras, de un extraño caso de un rasgo en cuya posesión se juega la supervivencia (comparativa) de un organismo pero que puede ser desactivado a voluntad por su portador.

a abandonar un barco que se hundía), y su entusiasmo por el potencial explicativo de la PE jamás decayó,¹⁷ si bien jamás ofreció una defensa específica y sistemática de ninguna de las explicaciones provenientes de dicho campo de las tendencias conductuales que conforman la supuesta naturaleza humana. Si releemos la obra de Dennett en búsqueda de las razones de su apoyo ineludible a la sociobiología y la PE, podríamos aventurar que veía en ambos abordajes un intento de mantener viva la esperanza de Darwin (2008, p. 359) de que la psicología algún día habría de estar basada en el estudio de los poderes y capacidades mentales favorecidos por la selección natural. Una actitud más crítica frente a la PE y más abierta a otras vías de exploración quizás lo habrían llevado a apostar por el desarrollo de un abordaje evolucionista de la psicología humana que no estuviese atravesado por sesgos antropocéntricos y etnocéntricos: un abordaje que, lamentablemente, todavía aparece como una mera promesa.

Referencias

- Apicella, C., Norenzayan, A. y Henrich, J. (2020). Beyond WEIRD: A Review of the Last Decade and a Look Ahead to the Global Laboratory of the Future. *Evolution and Human Behavior*, 41(5), 319-329. <https://doi.org/10.1016/j.evolhumbehav.2020.07.015>
- Barkow, J. H., Cosmides, L. y Tooby, J. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press.
- Barreiro, L. B., Laval, G., Quach, H., Patin, E. y Quintana-Murci, L. (2008). Natural Selection Has Driven Population Differentiation in Modern Humans. *Nature Genetics*, 40(3), 340-345. <https://doi.org/10.1038/ng.78>
- Bolhuis, J. J., Brown, G. R., Richardson, R. C. y Laland, K. N. (2011). Darwin in Mind: New Opportunities for Evolutionary Psychology. *PLOS Biology*, 9(7), e1001109. <https://doi.org/10.1371/journal.pbio.1001109>
- Buller, D. J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. MIT Press.

¹⁷ Todavía en 2017 Dennett sostenía: “Contrariamente a la opinión ampliamente engendrada por [Spandrels], el adaptacionismo está vivo y goza de buena salud; la ingeniería inversa, cuando se lleva a cabo con la debida atención a los riesgos y obligaciones, sigue siendo el camino real hacia el descubrimiento en biología y el único camino hacia el descubrimiento en el exigente mundo de la química prebiótica del origen de la vida” (Dennett, 2017, p. 43).

- Buller, D. J. (2009). Four Fallacies of Pop Evolutionary Psychology. *Scientific American*, 300(1), 74-81. <https://doi.org/10.1038/scientificamerican0109-74>
- Buller, D. J. y Hardcastle, V. G. (2000). Evolutionary Psychology, Meet Developmental Neurobiology: Against Promiscuous Modularity. *Brain and Mind*, 1, 307-325. <https://doi.org/10.1023/A:1011573226794>
- Buss, D. (1999). *Evolutionary Psychology: The New Science of the Mind*. Allyn and Bacon.
- Buss, D. (2006). *The Murderer Next Door: Why the Mind Is Designed to Kill*. Penguin Books.
- Buss, D., Haselton, M. G., Shackelford, T. K., Bleske, A. L. y Wakefield, J. C. (1998). Adaptations, Exaptations, and Spandrels. *American Psychologist*, 53(5), 533-548. <https://doi.org/10.1037/0003-066X.53.5.533>
- Caponi, G. (2011). *La segunda agenda darwiniana: Contribucion preliminar a una historia del programa adaptacionista*. Centro de Estudios Filosóficos, Politicos y Sociales Vicente Lombardo Toledano.
- Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Clarendon Press.
- Cloud, J. M., Stone, A. M. y McCarthy, J. D. (2023). No Time to “Waist:” Low Waist-to-Hip Ratios Pop out in Visual Search. *Evolutionary Behavioral Sciences*, 17(3), 345-356. <https://doi.org/10.1037/ebbs0000308>
- Cosmides, L. (1989). The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task. *Cognition*, 31(3), 187-276.
- Darwin, C. (2008). *The Origin of Species*. Oxford University Press.
- Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press.
- Dennett, D. C. (1983). Intentional Systems in Cognitive Ethology: The “Panglossian Paradigm” Defended. *Behavioral and Brain Sciences*, 6(3), 343-355. <https://doi.org/10.1017/S0140525X00016393>
- Dennett, D. C. (1996a). *Darwin’s Dangerous Idea: Evolution and the Meaning of Life*. Touchstone.
- Dennett, D. C. (1996b). *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfeld & Nicolson.
- Dennett, D. C. (1997, 14 de agosto). ‘Darwinian Fundamentalism’: An Exchange. *The New York Review of Books*, 44(13). <https://www.nybooks.com/articles/1997/08/14/darwinian-fundamentalism-an-exchange/>
- Dennett, D. C. (2003). The Mythical Threat of Genetic Determinism. *Chronicle of Higher Education*, 49(12), B7-B9.

- Dennett, D. C. (2004). Commentary on John Dupré's "Human Nature and the Limits of Science". *Philosophy and Phenomenological Research*, 69(2), 473-483.
- Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. Norton.
- Dennett, D. C. (2023). *I've Been Thinking*. Norton & Co.
- Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press.
- Gonsalves, B. D. y Cohen, N. J. (2010). Brain Imaging, Cognitive Processes, and Brain Networks. *Perspectives on Psychological Science*, 5(6), 744-752. <https://doi.org/10.1177/1745691610388776>
- Gould, S. J. (1997, 26 de junio). Evolution: The Pleasures of Pluralism. *The New York Review of Books*, 44(11). <https://www.nybooks.com/articles/1997/06/26/evolution-the-pleasures-of-pluralism/>
- Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Harvard University Press.
- Gould, S. J. (2019). Sociobiology and the Theory of Natural Selection. En G. W. Barlow y J. Silverberg (eds.), *Sociobiology: Beyond Nature/Nurture?* (pp. 257-269). Routledge. <https://doi.org/10.4324/9780429306587-10>
- Gould, S. J. y Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161), 581-598. <https://doi.org/10.1098/rspb.1979.0086>
- Gray, R. D., Heaney, M. y Fairhall, S. (2003). Evolutionary Psychology and the Challenge of Adaptive Expansion. En K. Sterelny y J. Fitness (eds.), *From Mating to Mentality: Evaluating Evolutionary Psychology* (pp. 247-268). Psychology Press.
- Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C. y Moyzis, R. K. (2007). Recent Acceleration of Human Adaptive Evolution. *Proceedings of the National Academy of Sciences*, 104(52), 20753-20758. <https://doi.org/10.1073/pnas.0707650104>
- Hurlbert, A. C. y Ling, Y. (2007). Biological Components of Sex Differences in Color Preference. *Current Biology*, 17(16), R623-R625. <https://doi.org/10.1016/j.cub.2007.06.022>
- Kaskan, P. M., Franco, E. C. S., Yamada, E. S., de Lima Silveira, L. C., Darlington, R. B. y Finlay, B. L. (2005). Peripheral Variability and Central Constancy in Mammalian Visual System Evolution. *Proceedings of the Royal Society B: Biological Sciences*, 272(1558), 91-100. <https://doi.org/10.1098/rspb.2004.2925>

- Kitcher, P. (1985). *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. MIT Press.
- Krasnow, M. M., Truxaw, D., Gaulin, S. J. C., New, J., Ozono, H., Uono, S., Ueno, T. y Minemoto, K. (2011). Cognitive Adaptations for Gathering-Related Navigation in Humans. *Evolution and Human Behavior*, 32(1), 1-12. <https://doi.org/10.1016/j.evolhumbehav.2010.07.003>
- Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A. y Kayser, M. (2007). Signatures of Positive Selection in Genes Associated with Human Skin Pigmentation as Revealed from Analyses of Single Nucleotide Polymorphisms. *Annals of Human Genetics*, 71(3), 354-369. <https://doi.org/10.1111/j.1469-1809.2006.00341.x>
- Lewontin, R. C. (1979). Sociobiology as an Adaptationist Program. *Behavioral Science*, 24(1), 5-14. <https://doi.org/10.1002/bs.3830240103>
- Lindquist, K. A. y Barrett, L. F. (2012). A Functional Architecture of the Human Brain: Emerging Insights from the Science of Emotion. *Trends in Cognitive Sciences*, 16(11), 533-540. <https://doi.org/10.1016/j.tics.2012.09.005>
- Lloyd, E.A. (1999). Evolutionary Psychology: The Burdens of Proof. *Biology and Philosophy*, 14(2), 211-233. <https://doi.org/10.1023/a:1006638501739>
- McKinnon, S. (2012). *Genética neoliberal. Mitos y moralejas de la psicología evolucionista*. M. Lamas (trad.). Fondo de Cultura Económica.
- Meneganzin, A., Ramsey, G. y DiFrisco, J. (2024). What Is a Trait? Lessons from the Human Chin. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 342(2), 65-75. <https://doi.org/10.1002/jez.b.23249>
- Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J., Sninsky, J., Adams, M. D. y Cargill, M. (2005). A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *PLoS Biology*, 3(6), e170. <https://doi.org/10.1371/journal.pbio.0030170>
- Oosterwijk, S., Lindquist, K., Anderson, E., Dautoff, R., Moriguchi, Y. y Barrett, L. F. (2012). States of Mind: Emotions, Body Feelings, and Thoughts Share Distributed Neural Networks. *NeuroImage*, 62, 2110-2128. <https://doi.org/10.1016/j.neuroimage.2012.05.079>
- Palecek, M. (2017). Modularity of Mind: Is It Time to Abandon This Ship? *Philosophy of the Social Sciences*, 47(2), 132-144. <https://doi.org/10/gm5rzb>
- Pazhoohi, F., Afhami, R., Chegeni, R., Dubrov, D., Gałasińska, K., Garza, R., Moharrampour, N. G., Grigoryev, D., Kowal, M., Pallesen, S. y

- Pfuhl, G. (2024). Cross-Cultural Preferences for Women's Waist to Hip Ratio and Men's Shoulder to Hip Ratio: Data From Iran, Norway, Poland, and Russia. *Adaptive Human Behavior and Physiology*, 10(1), 1-17. <https://doi.org/10.1007/s40750-024-00232-7>
- Pessoa, L. (2008). On the Relationship Between Emotion and Cognition. *Nature Reviews. Neuroscience*, 9, 148-158. <https://doi.org/10.1038/nrn2317>
- Peters, B. M. (2013). Evolutionary Psychology: Neglecting Neurobiology in Defining the Mind. *Theory & Psychology*, 23(3), 305-322. <https://doi.org/10.1177/0959354313480269>
- Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W. y Pritchard, J. K. (2009). Signals of Recent Positive Selection in a Worldwide Sample of Human Populations. *Genome Research*, 19(5), 826-837. <https://doi.org/10.1101/gr.087577.108>
- Pigliucci, M. y Kaplan, J. M. (2000). The Fall and Rise of Dr Pangloss: Adaptationism and the Spandrels Paper 20 Years Later. *Trends in Ecology & Evolution*, 15(2), 66-70. [https://doi.org/10.1016/S0169-5347\(99\)01762-0](https://doi.org/10.1016/S0169-5347(99)01762-0)
- Pinker, S. (1999). *How the Mind Works*. Norton.
- Pinker, S. (2018). *Los ángeles que llevamos dentro. El declive de la violencia y sus implicaciones*. J. Soler Chic (trad.). Paidós.
- Pollet, T. V. y Saxton, T. K. (2019). How Diverse Are the Samples Used in the Journals 'Evolution & Human Behavior' and 'Evolutionary Psychology'? *Evolutionary Psychological Science*, 5(3), 357-368. <https://doi.org/10.1007/s40806-019-00192-2>
- Prinz, J. J. (2006). Is the Mind Really Modular? En R. Stainton (ed.), *Contemporary Debates in Cognitive Science* (pp. 22-36). Blackwell.
- Resnik, D. B. (1996). Adaptationism: Hypothesis or Heuristic? *Biology and Philosophy*, 12(1), 39-50. <https://doi.org/10.1023/A:1017936706537>
- Reznick, D. N., Shaw, F. H., Rodd, F. H. y Shaw, R. G. (1997). Evaluation of the Rate of Evolution in Natural Populations of Guppies (*Poecilia reticulata*). *Science*, 275(5308), 1934-1937. <https://doi.org/10.1126/science.275.5308.1934>
- Rose, S. P. R. (2000). Escaping Evolutionary Psychology. En H. Rose (ed.), *Alas, Poor Darwin: Arguments against Evolutionary Psychology* (pp. 272-291). Vintage.
- Rosen, D. E. (1982). Teleostean Interrelationships, Morphological Function and Evolutionary Inference. *American Zoologist*, 22(2), 261-273. <https://doi.org/10.1093/icb/22.2.261>

- Saad, G. (2007). Suicide Triggers as Sex-Specific Threats in Domains of Evolutionary Import: Negative Correlation Between Global Male-To-Female Suicide Ratios and Average per Capita Gross National Income. *Medical Hypotheses*, 68(3), 692-696. <https://doi.org/10.1016/j.mehy.2006.08.012>
- Sahlins, M. (1973). *The Use and Abuse of Biology: An Anthropological Critique of Sociobiology*. University of Michigan Press.
- Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., Bai, Z., Lorenzo, F. R., Xing, J., Jorde, L. B., Prchal, J. T. y Ge, R. (2010). Genetic Evidence for High-Altitude Adaptation in Tibet. *Science*, 329(5987), 72-75. <https://doi.org/10.1126/science.1189406>
- Singh, D. (1993). Body Shape and Women's Attractiveness. *Human Nature*, 4(3), 297-321. <https://doi.org/10.1007/BF02692203>
- Singh, D. y Young, R. K. (1995). Body Weight, Waist-To-Hip Ratio, Breasts, and Hips: Role in Judgments of Female Attractiveness and Desirability for Relationships. *Ethology and Sociobiology*, 16(6), 483-507. [https://doi.org/10.1016/0162-3095\(95\)00074-7](https://doi.org/10.1016/0162-3095(95)00074-7)
- Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell.
- Sterelny, K. y Griffiths, P. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press.
- Tattersall, I. (2001). Evolution, Genes, and Behavior. *Zygon*, 36(4), 657-666. <https://doi.org/10.1111/0591-2385.00389>
- Thagard, P. y Schröder, T. (2015). Emotions as Semantic Pointers: Constructive Neural Mechanisms. En L. F. Barrett y J. A. Russell (eds.), *The Psychological Construction of Emotion* (pp. 144-167). The Guilford Press.
- Thornhill, R. y Palmer, C. T. (2000). *A Natural History of Rape: Biological Bases of Sexual Coercion*. MIT Press.
- Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghoris, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. y Deloukas, P. (2007). Convergent Adaptation of Human Lactase Persistence in Africa and Europe. *Nature Genetics*, 39(1), 31-40. <https://doi.org/10.1038/ng1946>
- Uttal, W. R. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. MIT Press.
- Vance, G. S., Zeigler-Hill, V., Vonk, J. y Shackelford, T. K. (2024). Inevitable or Preventable? The Biosocial Theory of Wartime

- Rape. *Evolutionary Psychological Science*, 10(2), 135-154.
<https://doi.org/10.1007/s40806-024-00387-2>
- Veillard, N. (2017). *WEIRD Sampling in Crosscultural Psychology, Should It Not Be Less WEIRD and More Representative? The Overrepresentation of Individuals from Western Educated Industrialized and Democratic Countries as Sample Populations in Cross-Cultural Psychology Research* [Tesis de maestría, Universidad de Leiden]. Leiden University Student Repository. <https://hdl.handle.net/1887/56361>
- Williams, G. C. (1996). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.
- Woodward, J. y Cowie, F. (2003). The Mind Is Not (Just) a System of Modules Shaped (Just) by Natural Selection. En C. R. Hitchcock (ed.), *Contemporary Debates in Philosophy of Science* (pp. 312-334). Blackwell.

<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 junturas 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 podamos 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 desenvolviendo. 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 covariaban 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.

<https://doi.org/10.21555/top.v760.3904>

Dennett y el materialismo cultural

Dennett and Cultural Materialism

Fernando Broncano

Universidad Carlos III de Madrid

España

fernando.broncano@uc3m.es

<https://orcid.org/0000-0002-7705-2947>

Recibido: 20 - 10 - 2025.

Aceptado: 10 - 12 - 2025.

Publicado en línea: 15 - 06 - 2026.

Cómo citar este artículo: Broncano, F. (2026). Dennett y el materialismo cultural. *Tópicos, Revista de Filosofía*, 76, 255-279.
<https://doi.org/10.21555/top.v760.3904>



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

Resumen

En este trabajo confronto el materialismo de Dennett con el enactivismo postcognitivo. Aunque ambos son formas naturalistas y eliminacionistas en lo que se refiere al teatro cartesiano, el funcionalismo de Dennett aún queda limitado por su adscripción al representacionalismo en el funcionamiento interno de la mente. La principal diferencia que se observa entre estas dos formas de materialismo afecta al lugar de la experiencia y la fenomenología, un lugar central aceptado por el enactivismo como marca de la vida al precio de abandonar la idea de representación y contenido.

Palabras clave: cognitivismo; materialismo; enactivismo; representacionalismo; cultura material; heterofenomenología; actitud intencional; autómatas celulares; eliminacionismo; Daniel Dennett.

Abstract

In this paper, I contrast Dennett's materialism with post-cognitive enactivism. Although both are naturalist and eliminativist approaches concerning the Cartesian Theater, Dennett's functionalism remains limited by its adherence to representationalism regarding the mind's internal workings. The main difference observed between these two forms of materialism concerns the place of experience and phenomenology, a central position accepted by enactivism as a hallmark of life, at the cost of abandoning the idea of representation and content.

Keywords: cognitivism; materialism; enactivism; representationalism; material culture; heterophenomenology; intentional stance; cellular automata; eliminativism; Daniel Dennett.

Dennett y el funcionalismo¹

Dennett significó en el último tercio del siglo pasado un renacimiento del materialismo en filosofía de la mente contra el dualismo que había significado el cognitivismo. Pues el cognitivismo, basado en la metáfora del procesamiento simbólico, resucitaba el dualismo cartesiano enmascarado tras la distinción entre *hardware* (húmedo, cerebral) y software mental, organizado en competencias y módulos. Fodor era el representante más combativo de la idea de la mente representacional, desbordando a los intentos mediadores de Davidson con su propuesta del monismo anómalo. Parecía hacerse realidad el arrogante aviso de Fodor de que el cognitivismo computacional era “el único juego en el pueblo”. El materialismo de los esposos Churchland, neurocéntrico, no tenía por el momento la potencia conceptual suficiente para desbancar al computacionalismo. Aún no se habían popularizado las redes neuronales y una mayoría entendía que el único modelo posible de lo mental era el de una semántica composicional basada en algún todavía desconocido modo de procesamiento simbólico neuronal.

Dennett irrumpió en este paisaje con una luz nueva al difundir una estrategia metodológica que minaba el dominio del cognitivismo tanto por arriba como por abajo. Por arriba, impulsaba la *heterofenomenología*, basada en una escalada explicativa que se iniciaba en modelos mecánicos para los sistemas muy simples, que ascendía a la *actitud de diseño*, para el caso de sistemas de complejidad intermedia como ordenadores o seres vivos con una conducta previsible, para terminar en la *actitud intencional* o hermenéutica, para aquellos seres dotados de espontaneidad de conducta. Aquí no había un compromiso ontológico con el computacionalismo, ni tampoco su negación: se trataba de reconocer que las modalidades explicativas saltaban de uno a otro nivel dependiendo de la complejidad del sistema que se tratase de entender. Hilary Putnam (1975) lo había expresado con una idea muy gráfica: no se necesita descender al nivel cuántico para explicar por qué un tornillo cuadrado no entra en una tuerca redonda.

Mediante este movimiento Dennett se liberaba de responder a la tediosa controversia sobre la naturaleza de la mente. La actitud de

¹ Este artículo recupera ideas y pasajes de Broncano (2025).

diseño del computacionalismo cognitivista estaba condenada a la pura especulación. Era lo que Lakatos habría llamado un programa de investigación regresivo, condenado a rehacerse cada vez que los estudios empíricos proporcionasen evidencia en contra (por ejemplo, contra la hipótesis modular). La heterofenomenología prometía salvar la superficie de los conceptos mentales y, en general, de la cultura humanística basándose en una hipótesis que de hecho vaciaba su potencia explicativa: se trataba de un reconocimiento de imposibilidad de explicación de otra forma. Así, ¿cómo interpretar textos y personas sino reconociendo que no había otro camino que usar un vocabulario de intenciones, razones y deseos? Pero este vocabulario, por otra parte, se reconocía como una simple actitud de derrota explicativa.

Por abajo, la estrategia de Dennett era mucho más agresiva. Frente a las “ciencias especiales” de Fodor, promovía la unificación explicativa de todo fenómeno vivo o artificial basándose en el inmenso poder explicativo del adaptacionismo por selección. Richard Dawkins había publicado en 1976 *El gen egoísta* y con ello había popularizado una suerte de programa de hegemonía del darwinismo monista que había iniciado John Maynard Smith con su libro *The Theory of Evolution* (1958). Se trataba de un funcionalismo de nuevo cuño, basado en la teoría sintética de la evolución que combinaba el adaptacionismo con la genética de poblaciones. Un funcionalismo en realidad muy abstracto que remitía a los inicios del computacionalismo basado en Turing. El *Juego de la vida*, un autómata celular diseñado por el matemático británico John Horton Conway en 1970, cuyo poder computacional es equivalente al de una máquina de Turing, producía formas por selección que tenían la apariencia de objetos vivos moviéndose y reproduciéndose en el espacio abstracto del juego. El juego se había convertido en el pasatiempo de mucha gente y había producido una enorme cantidad de especímenes que vagamente podrían pasar por xenomorfos de algún planeta extraño. Había dado origen, por otra parte, a una extraña nueva disciplina autodenominada “vida artificial”. Dennett unió inteligentemente el adaptacionismo darwinista con la abstracción de los autómatas celulares de Conway. La analogía estaba clara: los genes podían pasar por autómatas celulares que se autorreproducían y generaban una cobertura protectora en la forma de células y organismos. Aquellas formas que lograban reproducirse generaban poblaciones estables y con el tiempo especies. Todo lo que necesitaba la Naturaleza para producir el árbol de la vida eran las reglas “elementales” de la reproducción de

los autómatas egoístas que, en la teoría de Dawkins y Dennett, eran los genes individuales.

Toda célula (agente diminuto que puede llevar a cabo un limitado número de tareas) carece de mente, más o menos igual que un virus. ¿Es posible que si se colocan juntos suficientes de estos estúpidos homúnculos, hombrecillos, el resultado sea una persona auténtica y consciente, con una mente de verdad? Según la ciencia moderna, no hay otro modo de formar una auténtica persona. Ahora bien: desde luego del hecho de que descendamos de robots no se sigue el hecho de que nosotros lo seamos. Después de todo, también somos descendientes directos de los peces y no somos peces; somos descendientes directos de las bacterias pero no somos bacterias. Pero a menos que haya algún ingrediente secreto en nosotros (que es lo que solían creer los dualistas y los vitalistas), estamos hechos de robots o, lo que viene a ser lo mismo, cada uno de nosotros es una colección de billones de máquinas macromoleculares. Y todas ellas descienden, en último extremo, de las macromoléculas con capacidad de duplicarse que hubo en un principio. De manera que hay cosas hechas con robots que pueden mostrar genuina conciencia, habida cuenta de que nosotros somos el mejor ejemplo (Dennett, 1997, pp. 33-34).²

La diferencia de este funcionalismo adaptacionista con el funcionalismo del computacionalismo cognitivo está en esta pinza de lo abstracto del concepto de “autómata celular”, que carga toda la potencia computacional Turing, junto con la hipótesis de la adaptación por variación más selección natural como fuerza productora de diseño. Nos encontramos aquí con una especie de ascenso epistemológico desde la seguridad científica de los mecanismos moleculares, pasando por la fisiología, ascendiendo, y al mismo tiempo perdiendo seguridad epistemológica, hasta el escepticismo de la heterofenomenología de la actitud intencional. Dennett responderá que la heterofenomenología captura

² A lo largo de este artículo, las traducciones son mías.

“patrones reales” en el nivel de la psicología folk, pero este concepto de realidad ya no es el del diseño ni mucho menos el de los mecanismos causales subyacentes. Es un modo de eliminacionismo que sustituye la vida mental por una interpretación que no es sino una ilusión de usuario ignorante de los mecanismos que ocurren en el cerebro. En su haber está el materialismo que se asienta en la potencia creativa de la evolución; en su deber está la justificación de estos ascensos (descensos, más bien) epistémicos desde la física a la psicología y la antropología.

El alcance del adaptacionismo

La noción de “diseño” está en el corazón del materialismo desde el famoso argumento del reloj que se encuentra en *Teología natural o Evidencias de la existencia y atributos de la Deidad* de 1802 por William Paley (1743-1805). La adecuación forma/función señalaría como causa el diseño inteligente: la explosión de posibilidades combinatorias de la materia, en ausencia de una inteligencia, no podría dar cuenta de la producción de sistemas en los que forma, función y acción estuviesen tan bien adaptados. Responder a este argumento le resultó difícil incluso a Darwin.

Ni Dennett ni Dawkins ocultan que su adaptacionismo tiene una pretensión filosófica mucho más general que la biología como ciencia: se trata de responder a todas las variantes del argumento del diseño y contradecir toda pretensión de buscar explicaciones milagrosas (“ganchos celestes” de los que colgarse para ascender en la escala del ser, afirma Dennett con una de sus más conocidas metáforas). La selección natural proporciona “grúas” suficientes para explicar tanto la variedad como la complejidad de los fenotipos y diseños de los seres vivos. Esta es la marca del materialismo de Dennett y nos lleva a la reiterada pregunta de si, aun aceptando que la variación más selección es una buena explicación naturalista, acaso es la única forma de materialismo suficiente para toda la escala de la complejidad de la vida, de su homeostasis, autopoiesis, cultura, pero también de sus múltiples lagunas, errores e inconsistencias de diseño, como por ejemplo, en el caso humano, las complejas (y peligrosas) relaciones del sistema afectivo y el cognitivo. No son pocos los autores que han pedido más prudencia a las tesis aparentemente más ortodoxamente darwinianas de la omniexplicación seleccionista y adaptacionista.

Godfrey-Smith (2001) aclara la confusión generalizada entre tres formas de adaptacionismo: el adaptacionismo empírico, el explicativo y el metodológico. Define el empírico de esta forma:

La selección natural es una fuerza poderosa y omnipresente y existen pocas restricciones, excepto las generales y obvias, sobre la variación biológica que la alimenta. En gran medida, es posible predecir y explicar el resultado de los procesos evolutivos prestando atención únicamente al papel que desempeña la selección. Ningún otro factor evolutivo tiene este grado de importancia causal (Godfrey-Smith, 2001, p. 336).

El explicativo, a su vez, lo define en estos términos:

El aparente diseño de los organismos y las relaciones de adaptación entre los organismos y sus entornos son las grandes preguntas, los hechos sorprendentes de la biología. Explicar estos fenómenos es la misión intelectual central de la teoría evolutiva. La selección natural es la clave para resolver estos problemas; la selección es la gran respuesta. Debido a que responde a las preguntas más importantes, la selección tiene una importancia explicativa única entre los factores evolutivos (Godfrey-Smith, 2001, p. 336).

Por último, el metodológico es definido de esta forma: “La mejor manera de que los científicos aborden los sistemas biológicos es buscar características de adaptación y buen diseño. La adaptación es un buen ‘concepto organizativo’ para la investigación evolutiva” (Godfrey-Smith, 2001, p. 337).

Godfrey-Smith concede que tal vez la intención de Dennett y Dawkins haya sido defender el adaptacionismo explicativo, aunque de muchas de sus expresiones parece concluirse que lo que tienen en la cabeza es una modalidad mucho más fuerte que combina las tres. El punto está en si se sostienen (a) que el diseño es ubicuo en la naturaleza incluso cuando no parece obvio a la mirada del biólogo y (b) que el diseño es siempre producto de la selección natural. Corresponde a la forma más fuerte de darwinismo, en el que se escala desde (1) la selección natural es una fuerza importante en la evolución a (2) la selección natural es

la fuerza más importante en la evolución o, más allá, a (3) la selección natural es la única fuerza relevante en la evolución.

La filosofía de la biología estuvo llena de ácidas controversias en el último tercio del siglo pasado en las que se debatieron estas afirmaciones junto con otras, especialmente la cuestión de cuáles son los niveles sobre los que opera la selección: los genes, el organismo o los grupos. La cuestión de si la evolución biológica y la cultural siguen o no las mismas pautas sería la tercera gran controversia.

El centro del debate tiene que ver con la idea de que la evolución tiene que resolver problemas de diseño a través de la selección natural. No es este el lugar para recordar todas estas controversias (Ayala y Arp, 2010 es una magnífica reconstrucción de estos enfrentamientos). Diseño es lo que parece diferenciar el mundo de lo vivo (incluyendo las producciones artefactuales) del mundo fisicoquímico. El diseño es un modo de orden de lo natural que entraña el acoplamiento de materia, energía, información y, por último, una acción que pueda ser descrita en términos funcionales. Dennett sostiene que el diseño es el problema más importante de la biología, el que centra la mayoría de las necesidades explicativas. Como sabemos, sin embargo, la idea de “función”, que es la que articula el diseño, puede ser entendida sistémica o etiológicamente, es decir, por su contribución causal al funcionamiento del sistema del que forma parte o por la historia causal que explica que ese diseño sea tal como es y forme parte del organismo en que está o que constituye.

La primera forma de diseño nos lleva a una explicación fisiológica, que en la biología contemporánea desciende a niveles más profundos que el del gen, pues de alguna manera el propio éxito de la genética ha conducido a que la biología se convierta crecientemente a un paradigma molecular, en donde las expresiones de los genes se entrecruzan con las enzimas, proteínas y todo lo que compone la fábrica del organismo. Entendido así, tiene razón Dennett en que el problema del diseño es el problema más importante de la biología. Lo es casi trivialmente porque podría decirse que todas las ramas de la biología son componentes de las teorías correspondientes del diseño que explica la reproducción, ciclo de vida y conducta de los seres vivos en todos sus niveles de arquitectura jerárquica.

La biología es una ciencia de sistemas y subsistemas anidados y de ensamblamientos de organismos, parásitos y simbioses sin los que no es posible la vida. La fábrica de lo vivo está organizada por múltiples dinámicas cuasi independientes y por complejos sistemas

de control que se encargan de las varias modalidades de homeostasis que garantizan la estabilidad del ser vivo. Un ser que está en continuo intercambio y coconstitución con su entorno, algo que extiende la complejidad del diseño a los múltiples *feedbacks* con el medio, desde los puramente metabólicos a los informacionales. ¿Nos puede servir el algoritmo darwiniano como paradigma metodológico para todas las disciplinas del diseño biológico? Bueno, sí en un sentido trivial: muchas dinámicas pueden ser representadas por un algoritmo de generación y selección, incluida tal vez la evolución química. Al fin y al cabo, los nuevos modelos de inteligencia artificial pueden crear proteínas e incluso virus. La cuestión es que el concepto abstracto no nos sirve para estudiar concretamente las innumerables funciones que desarrollan los subsistemas orgánicos. Y lo mismo puede decirse respecto la concepción etiológica de las funciones. Explicar la historia causal de las funciones, aunque la investigación siga la hipótesis de la selección natural, exige comprobar que la aparición de un rasgo no es un caso de exaptación, de atavismo de diseño, de producto del desarrollo o de la deriva genética, que son algunas de las otras posibilidades de historia causal de una función.

Las controversias sobre el adaptacionismo (y sobre el nivel en el que opera la selección) tienen una dimensión científica, que no es el caso analizar aquí, y una dimensión filosófica. Dennett participa, quizás es el mayor representante, del materialismo filosófico, con una agenda que no está solo orientada a la metodología científica sino que, en unión con Dawkins, pretende una dimensión práctica en la militancia contra las creencias religiosas y a favor de un eliminacionismo ontológico de todo lo que no sea explicable por el mecanismo poderoso de replicación, variación y selección.

El problema que detectan varios autores como Sterelny, Godfrey-Smith o Sober, que no participan de la polarización de las posiciones, es que hay cuestiones sobre la potencia explicativa del adaptacionismo que están ligadas a la cuestión de los niveles de selección, en particular, sobre qué sistemas se tienen en cuenta para aplicar las presiones selectivas, si hay o no que tener en cuenta el desarrollo del organismo en la expresión e incluso modificación del acervo genético. En el materialismo de Dennett podemos diferenciar el problema general del diseño, como respuesta a la presión del argumento del diseño de Paley, de la cuestión más específica de cómo se extiende el adaptacionismo a la emergencia de las particularidades de la especie humana y de su cultura, comprendiendo

en esta cuestión la posible coevolución biológica y cultural de la especie. Aunque la teoría del fenotipo extendido que abarca lo que Dennet llama metafóricamente “grúas” puede dar cuenta de muchos aspectos de la evolución cultural, es razonable pensar que la teoría sintética de la evolución deba tener en cuenta algunos puntos que, aunque son característicos de la especie humana, podemos quizás extenderlos a otras especies culturales. Me refiero a la posibilidad de que especies que transforman culturalmente el entorno creando nichos generan también escenarios de presión selectiva si esos nichos evolucionan o coevolucionan con la especie.

De la actitud de diseño a la actitud intencional: el caso de la cultura y la memética

La actitud de diseño en lo que respecta a la evolución de las mentes fue tratada por Dennett en *Kinds of Minds* (1997), que puede ser considerado como una extensión de *La idea peligrosa de Darwin* (Dennett, 1995). El objetivo era proporcionar una hipótesis evolucionista y naturalista sobre la cuestión de la emergencia de la mente en la naturaleza. Las mentes serían como otro cualquiera de los rasgos fenotípicos, producciones de “robots” interactuando entre ellos y con el entorno. El materialismo de Dennett es una modalidad de eliminativismo que sostiene que la conciencia, la comprensión y los *qualia* son ilusiones de usuario producidas por la acción de múltiples sistemas no accesibles al punto de vista de la primera persona o del “teatro cartesiano”, una expresión que forma parte ya del diccionario dennettiano. En lo que puede considerarse como la recopilación más reciente y acaso completa de sus puntos de vista, *From Bacteria to Bach and Back* (2017), esta posición se reafirma. El debate sobre el eliminativismo ha sido casi el centro de gravedad de la filosofía de la mente desde el conocido artículo de Nagel sobre qué sería sentir como un murciélago. Desde mi punto de vista, que comparte muchas intuiciones con Dennett, la cuestión interesante del materialismo no es si hay o no algo más que el mundo físico o si las facultades mentales de los humanos son producto de la acción de las neuronas, o son la acción de las neuronas, ni tampoco si la evolución da cuenta conceptualmente de la posibilidad de esta emergencia. Se trata más bien de si estas características intencionales son propiedades reales de los seres vivos. Del mismo modo que la información tiene siempre una base material y energética, y puede transmitirse en

diversas realizaciones materiales, la cuestión es si todo lo que llamamos cualidades intencionales son o no propiedades emergentes que tienen una forma de realidad similar a la que tiene la información respecto al mundo físico.

La conciencia y los *qualia* pueden ser entendidos, en términos cartesianos, como algo privado, no expresable ni accesible más que en primera persona, pero también pueden serlo como propiedades emergentes de la forma de vida sobre la Tierra, en donde los organismos están en interacción continua con el entorno. En la filosofía de la mente y las ciencias cognitivas contemporáneas, se han desarrollado dos líneas de respuesta al funcionalismo computacional de Dennett. Una es la de la intencionalidad intrínseca de Searle, la otra es la del cognitivismo y la teoría de la mente encarnada y extendida, que conlleva un mantenimiento de la fenomenología pero un abandono de lo que para Dennett es esencial: el contenido mental, entendido como un producto de procesamiento de información en el entorno interno del cerebro.

El argumento de Searle sigue siendo vigente en la discusión sobre el materialismo y no es necesario considerar estas propiedades nada “misterioso”. Que haya procesos algorítmicos en la base de estas cualidades no determina sus propiedades semánticas, como el hecho de que una máquina de Turing procese símbolos no agota la semántica. Searle y Dennett sostuvieron una agria polémica en varios foros, especialmente en las páginas del *New York Review of Books* a propósito de las recensiones que Searle hizo de varias de sus obras. Fue un choque de dos personalidades aficionadas a la controversia. Sin entrar ahora en los detalles de esta polémica, creo sin embargo que Searle tiene razón en considerar que el problema de la conciencia sigue siendo un problema central, al menos en los términos que expresa esta cita que no resisto a traer aquí:

Hay un único problema general que me ha preocupado desde que empecé a trabajar en filosofía hace casi medio siglo: ¿cómo podemos tener una explicación unificada y teóricamente satisfactoria de nosotros mismos y de nuestras relaciones con otras personas y con el mundo natural? ¿Cómo podemos conciliar nuestra concepción común de nosotros mismos como agentes conscientes, libres, conscientes, capaces de realizar actos de habla y racionales en un mundo que creemos que consiste

enteramente en partículas físicas brutas, inconscientes, sin mente, sin sentido y mudas en campos de fuerza? En resumen, ¿cómo podemos hacer que nuestra concepción de nosotros mismos sea plenamente coherente con la descripción del mundo que hemos adquirido de las ciencias naturales, especialmente la física, la química y la biología? Las preguntas que más me han preocupado —¿qué es un acto de habla? ¿Qué es la conciencia? ¿Qué es la intencionalidad? ¿Qué es la sociedad? ¿Qué es la racionalidad?— se han abordado, de una forma u otra, en esta problemática más amplia. Creo que este problema —o conjunto de problemas— es el más importante de la filosofía y, de hecho, en cierto sentido, en nuestra época concreta, es el único problema importante de la filosofía (Searle, 2010, p. 1).

No es mi intención adoptar un bando en la controversia, como tampoco lo es respecto a la polémica con Steven Jay Gould sobre el adaptacionismo, pero me parece necesario considerar ambas polémicas como hitos de la respuesta a las preguntas sobre cómo articular nuestras concepciones de la vida y de la vida consciente con una visión del mundo basada en complejas dinámicas fisicoquímicas, es decir, cómo es posible una concepción materialista del mundo.

Los dos aspectos de la conciencia y la intencionalidad son, por un lado, la dimensión fenoménica o experiencial; por otro, el contenido y en particular el contenido considerado como una representación en el cerebro. La controversia es entre dos formas de naturalismo: Dennett desea salvar el contenido representacional, en tanto que producto algorítmico generado como otros aspectos del fenotipo por selección natural. La conciencia y la intencionalidad serían aquí modos de hablar, ilusiones de usuario o estrategias interpretativas, pero no genuinas propiedades. Searle, por su parte, considera la conciencia como una propiedad intrínseca de la vida, generada causalmente por la química y física de la composición fisiológica, y el contenido un producto de la interacción social. Se trata en ambos casos no de una posición materialista contra otra dualista sino de formas que puede adaptar el materialismo en relación en lo que parecen ser las características más prominentes de lo humano.

Tiene razón Dennett en que no puede concebirse una posición materialista que no tenga en cuenta la evolución biológica. Pero en la evolución humana está ya la cuestión de qué fuerza productiva, evolutiva, tiene la cultura y su particular modo de evolución. Kim Sterelny, uno de los más reputados filósofos actuales de la biología, afirma a este respecto:

Sostendré que la singularidad humana se explica por la confluencia de tres factores en la evolución humana. Uno de ellos, la cooperación, fue el tema central del último capítulo. Allí sostuve que la selección de grupo ha desempeñado un papel especialmente importante en la evolución humana. Esto explica la amplia cooperación humana. Pero también interactúa con otras dos características de la evolución de los homínidos: la construcción de nichos y la plasticidad fenotípica (o variabilidad). La intersección de la selección de grupo, la construcción de nichos y la plasticidad nos convierte de animales sociales en animales culturales. El tema central de este capítulo es la construcción de nichos humanos, que es nuestra propensión a rehacer nuestro entorno físico, social y epistémico. Los grupos de humanos diseñan sus hábitats. Además, transmiten ese entorno alterado a la siguiente generación. Así, la siguiente generación hereda un hábitat modificado y, a menudo, lo modifica aún más. Este proceso de modificación sucesiva es tan importante y omnipresente que John Odling-Smee, Kevin Laland y sus colegas lo consideran un sistema de herencia ecológica que funciona en paralelo a la herencia genética (Sterelny, 2003, p. 8).

En el capítulo 13 de *La peligrosa idea de Darwin* y luego en el capítulo 14 de *Tipos de mente*, Dennett establecía una jerarquía ascendente de mentes: las darwinianas, que tienen patrones de acción fijos que solo transforma la evolución cultural; las skinnerianas, dotadas de aprendizaje por premios y castigos; las popperianas, que consideran antes de actuar varias alternativas y eliminan las más peligrosas, y, por último las gregorianas, las mentes que toman instrumentos de la creación colectiva y los asimilan en un comportamiento más complejo

que las anteriores. Usando la misma metáfora de Dennett, la evolución ha construido una “torre” de generación y prueba. Las mentes aparecen en esta escala como sistemas plásticos de producción de conductas que son eliminadas (o permanecen si aciertan) por el medio. La plasticidad darwiniana lleva generaciones; la skinneriana lleva solamente un primer fracaso, que, si no mata, conlleva la memoria negativa/positiva del resultado de la acción; la popperiana (que presuntamente estaría presente ya en algunos mamíferos) permite que mueran las hipótesis antes que el organismo, para lo cual el animal usa su cuerpo como una especie de predictor (vómitos, mareos... sensaciones malas). En lo alto de la torre está la asimilación de artefactos que aumentan la potencia predictiva: palabras, imágenes y artefactos.

Godfrey-Smith (2016 y 2018) ha discutido esta “torre” de Dennett recordándole que la evolución no edifica torres, sino árboles. Es una crítica suave donde Godfrey-Smith sugiere que puede haber en la historia de la mente muchos más “experimentos” que los que establece la escalera dennettiana: puede haber mentes conductistas con memoria,³ puede haber inicios de aprendizaje más débiles que las criaturas skinnerianas. El punto de Godfrey-Smith es teóricamente mucho más relevante de lo que parece. La emergencia de la inteligencia ha sido una senda sinuosa y aún mal explorada (su conocido trabajo con pulpos muestra cuán lejos estamos aún de catalogar los modos de inteligencia animal y vegetal). Uno de los puntos centrales de esta emergencia ha sido la construcción de nichos que presionan sobre formas de comportamiento inteligente en direcciones que adaptativamente no parecerían muy promisorias. Así, por ejemplo, Sterelny escribe a propósito de la vida social: “La vida social es a veces una forma de ingeniería epistémica, ya que una de las formas de ingeniería ecológica es la modificación por parte de los agentes de su entorno epistémico. Así pues, una de las funciones de la

³ La extensión de lo que se denomina “cognición mínima” es uno de estos casos que desbordan la jerarquía de Dennett: “Se ha demostrado que el *Physarum polycephalum* presenta características similares a las observadas en organismos unicelulares e insectos eusociales. Por ejemplo, un equipo de investigadores japoneses y húngaros ha demostrado que el *P. polycephalum* puede resolver el problema del camino más corto. Cuando se cultiva en un laberinto con avena en dos puntos, el *P. polycephalum* se retira de todas las partes del laberinto, excepto de la ruta más corta que conecta las dos fuentes de alimento” (*Physarum polycephalum*, 2026; mi traducción).

vida social, como demuestran las suricatas, es compartir los beneficios de la vigilancia de los demás” (Sterelny, 2003, p. 8). La construcción (y no simplemente el habitar pasivo) de nichos, continúa Sterelny, no es un elemento externo al proceso evolutivo (variación, reproducción, selección), sino que puede modificar activamente las presiones selectivas:

La construcción de nichos por parte de los homínidos, a diferencia de la construcción de madrigueras por parte de los conejos, sí tiene efectos en la población en su conjunto. La domesticación de plantas y animales, un proceso de incalculable importancia en la historia de la humanidad, fue un proceso acumulativo de este tipo. Antes de la domesticación, los recolectores modificaron de forma acumulativa el perfil biológico de su hábitat. Por ejemplo, a medida que mejoraban las armas y la cooperación de los homínidos, los depredadores peligrosos se volvieron escasos y cautelosos. De hecho, algunos se volvieron extremadamente raros (Sterelny, 2003, p. 154).

Nada de esto afecta de modo catastrófico al programa funcionalista de Dennett, quien acepta claramente que algunos artefactos son “andamios”, “grúas” en su tecnología, que modifican el fenotipo del organismo. Sin embargo tanto las sugerencias de Godfrey-Smith como las de Sterelny abren la puerta a una recomposición del programa darwinista en el que las contingencias y caminos ramificados de la evolución tengan que ver con otros componentes, como el desarrollo fisiológico y, sobre todo, las constricciones concretas que generan entornos técnicos creados activamente por los organismos y las poblaciones. En el caso particular de la antropogénesis, todos los trabajos de la “arqueología de la mente”, es decir, de paleontólogos que han estudiado la cultura material de los ancestros de *Homo sapiens*, sugieren tomarse mucho más en serio de lo que lo hace Dennett las mediaciones entre los medios materiales y sociales en la creación de trayectorias híbridas de evolución biológica y cultural. En lo que sigue expondré, en primer lugar, las formas de materialismo que significan los programas de cultura material y de conocimiento distribuido para, en segundo lugar, plantear la alternativa enactivista radical al modelo computacionalista de Dennett.

Antropogénesis y cultura material

Paleontólogos como Steven Mithen (1996) popularizaron la idea de la ingeniería inversa, mediante la que podían conjeturarse las capacidades cognitivas de las especies culturales a través de los artefactos construidos por ellas de modo sistemático, no casual, en los respectivos nichos ecológicos y técnicos en los que evolucionaron. Dennett expresa en *Tipos de mentes* una idea luminosa que, desgraciadamente, no llega a explotar suficientemente:

[...] el uso de herramientas es una señal de inteligencia de ida y vuelta: no solo requiere inteligencia reconocer y mantener una herramienta (y más aún fabricarla), sino que una herramienta confiere inteligencia a aquellos suficientemente afortunados a los que se les ha dado una. Cuanto mejor diseñada esté la herramienta (cuanta más información suponga su fabricación), mayor inteligencia potencial confiere a su usuario. Y, nos recuerda Gregory, entre las herramientas más destacadas están aquellas que él denomina herramientas mentales: las palabras (Dennet, 1997, p. 101).

Mientras que Merlin Donald (1991) supone que en el estadio superior de la antropogénesis y la construcción de la mente los artefactos son un medio de descarga de tareas informacionales, para Dennett la dirección es solamente de los artefactos a la mente: los artefactos son asimilados, pero no —tal como piensan quienes sostienen la tesis de la mente extendida— sirven de complemento y descarga de información. Si en el extremo menos complejo de la cognición tenemos casos de inteligencia cooperativa como la del moho *Physarum polycephalum* en la antropogénesis, en la que la complejidad de interacciones internas y externas se mueve en una escala exponencialmente superior, la inteligencia emerge, mucho más claramente que en la cognición mínima, de la doble dirección entre un entorno dotado de posibilidades de acción e incluso de diseños orientados a completar la inteligencia del organismo.

La formación de la mente tanto en animales como en humanos es, desde el punto de vista de los defensores de la construcción de nichos de cultura material, un proceso basado en la agencia transformadora que,

a su vez, es mediada por la influencia que tienen las transformaciones artefactuales en la configuración de las mentes. Así, el paleontólogo Malafouris (2013 y 2019) cree que esta concepción dialéctica de mente-artefactos entraña preguntarse por los límites de la mente y el mundo. Aunque se ha hablado de “mente extendida”, la opinión generalizada es que la mente ocupa el cerebro y solo en contadas ocasiones deposita parte de su procesamiento en objetos externos; sin embargo, Malafouris tiene una visión más ecuménica y ecológica de la mente:

Basándome en trabajos recientes sobre la cognición enactiva, distribuida y extendida, sugeriré que, contrariamente a lo que cree la ciencia cognitiva clásica y a lo que la arqueología cognitiva a menudo reitera implícitamente, lo que está fuera de la cabeza no tiene por qué estar necesariamente fuera de la mente. Por muy natural que pueda parecer (especialmente desde la perspectiva de otras disciplinas) señalar al cerebro humano como la sede de todo lo que es verdaderamente mental, desde un punto de vista arqueológico no veo ninguna razón de peso por la que el estudio de la mente deba detenerse en la piel o el cráneo. Sugiero que sería más productivo explorar la hipótesis de que la inteligencia humana “se extiende” más allá de la piel, hacia la cultura y el mundo material (Malafouris, 2013, p. 3).

El bastón de los ciegos funciona como metáfora de la cultura material: “Yo llamo zona gris al compromiso material, es decir, la zona en la que los cerebros, los cuerpos y las cosas se fusionan, catalizándose y constituyéndose mutuamente” (Malafouris, 2013, p. 5). El bastón convierte el tacto en vista y se convierte en parte del cuerpo. Constituye un componente de la zona gris en donde el mundo el cuerpo y la mente intersecan generando la cognición. Como ejemplos de esta dinámica mente-mundo, Malafouris ofrece el dibujo de líneas o la talla de bifaces: no es correcta una descripción del proceso en el que la intención mental precede a la acción. Son las “intenciones en la acción” en un continuo vaivén de direcciones de ajuste entre la mente y el mundo donde se constituye la agencia humana material.

Los seres humanos se convierten “a través de una implicación saturada y situada del pensamiento y el sentimiento con las cosas y los materiales generadores de formas” [...]. A ese proceso lo denomino “coseimiento [*thinging*] creativo” y volveré a ejemplificar su significado más adelante. Basta por ahora con decir que desde las primeras ecologías líticas hasta las últimas ontologías digitales, este proceso se encuentra en el centro de la evolución humana. Los seres humanos piensan construyendo signos, trazando líneas y dejando huellas de memoria. Lo hacen principalmente por medio de sus cuerpos en movimiento, especialmente sus manos. Esto no quiere decir que los signos que hacemos o las líneas que trazamos simplemente “representen” o “reflejen” la inteligencia. La inteligencia “reflejada” no está oculta en algún reino “mental” separado dentro del cráneo. La mano en movimiento y sus huellas materiales no solo exteriorizan el funcionamiento interno de la mente. En cambio, la inteligencia se manifiesta a través de ellos; procede a lo largo de líneas y signos materiales de un tipo u otro. Por ejemplo, la fabricación de una herramienta de piedra no es el producto del pensamiento, sino una forma de pensar. Cuando miramos una herramienta de piedra, no vemos simplemente la exteriorización de la forma, la habilidad o la memoria; más bien observamos cómo las posibilidades que ofrece la piedra hacen posible que los cuerpos humanos aprendan y recuerden habilidades, perciban la causalidad o pongan en práctica intenciones. En resumen, dentro de una ecología lítica, las herramientas de piedra generan y limitan las posibilidades de acción e imaginación del organismo. En ese sentido, el proceso de pensar se invierte efectivamente. Nuestras formas de extensión corporal y compromiso material no son simplemente marcadores externos de una arquitectura mental humana distintiva. Más bien participan de forma activa y significativa en el proceso que llamamos mente (Malafouris, 2013, p. 2).

Este entrelazamiento de objetos y mentes evoluciona a lo largo del tiempo también en un entrelazamiento entre cultura material y evolución biológica. Se crean en él trayectorias que presentan contingencias vueltas irreversibles por la cultura, que transforma el entorno con ellas, así como la colonización de los cereales y las formas de vida urbana fueron una de las trayectorias que siguió una parte de la población humana y en la que se crearon entornos nuevos donde operaron, por ejemplo, virus y bacterias que modificaron los cuerpos de las gentes agrupadas en espacios pequeños.

Hasta aquí todavía puede mantenerse una cierta compatibilidad entre el materialismo de Dennett y la teoría del compromiso material de Malafouris. Tanto Dennett como Andy Clark, próximo a él en tantos aspectos, mantendrían que la zona gris es un entorno del fenotipo extendido que evoluciona todavía por los algoritmos de la variación, reproducción y selección, incorporando a la mente estos agentes materiales como los bastones de ciego o las técnicas de escritura, pero ambos consideran que la cognición pertenece básicamente a la parte del fenotipo que es el cerebro, en donde residen las representaciones, que ocasionalmente pueden complementarse con medios externos de procesamiento. La propuesta de la mediación de la cultura material y de la coconstrucción de la mente y el entorno que ha postulado el enactivismo radical abre las puertas a una forma distinta de materialismo postcognitivist, que no abandona la fenomenología y sí prescinde de lo que es el núcleo del funcionalismo computacional: las representaciones como sedes del significado.

La concepción dinámica de la mente

En los años cincuenta, cuando estaban dando los primeros pasos la tecnología de los sistemas de control, la automática y los primeros ordenadores, se formaron dos corrientes que tomaban la biología como horizonte ontológico para entender el nuevo campo de la información. Por un lado, la teoría computacional, que sostiene que la inteligencia es procesamiento de símbolos y aboga por una teoría representacional de la mente. Su origen, al menos en la leyenda de la historia del cognitivismo, fue la formación de una comunidad de investigación nacida de la reunión de Dartmouth de 1956, convocada por John McCarthy. Quizás el trabajo más avanzado presentado en esta reunión fue el de Allen Newell y Herb Simon (1972). La unión de Turing y Darwin que propone Dennett

pertenece a esta tradición, en donde el concepto de “representación” sostiene la carga metafísica del proyecto.

La segunda gran tradición nace de la convergencia de varias fuentes. En primer lugar, de lo que popularmente en los años cincuenta se conoció como “cibernética”. El enfoque inicial de la cibernética se centró en los paralelismos entre los procesos de retroalimentación reguladora en los sistemas biológicos y tecnológicos. En 1943 se publicó “Behavior, Purpose and Teleology”, de Arturo Rosenblueth, Norbert Wiener y Julian Bigelow, basado en la investigación sobre organismos vivos que Rosenblueth llevó a cabo en México. Wiener introdujo el neologismo *cibernética* para denotar el estudio de los “mecanismos teleológicos” y lo popularizó a través del libro *Cybernetics: Or Control and Communication in the Animal and the Machine* (1948). La primera era de la cibernética tuvo mucho impacto tanto en la ingeniería de control como en la teoría de sistemas aplicada a las ciencias sociales, pero su influencia en la idea de lo mental se debe en gran medida al libro *Autopoiesis and Cognition: The Realization of the Living*, de los biólogos chilenos Humberto Maturana y Francisco Varela (1972), que estudia la química autosuficiente de las células vivas. La segunda línea convergente es la concepción encarnada de lo mental, que remite a la fenomenología de Merleau-Ponty y que actualmente está representada en parte por la así llamada “postfenomenología” defendida por Don Ihde, una forma de pensar que analiza cómo experimentamos el mundo, especialmente cómo nuestras experiencias se ven influidas por las herramientas y tecnologías que utilizamos. Toma de la fenomenología la comprensión de las experiencias desde las propias perspectivas del agente, pero sostiene que la tecnología moldea la experiencia y las herramientas que utilizamos cambian la forma en que vemos y comprendemos el mundo que nos rodea, que las interacciones con el mundo están mediadas por estas tecnologías y, sobre todo, hace hincapié en que los cuerpos no están separados de las mentes, sino que funcionan juntos. Otra aportación fue el influyente libro de Edward Hutchins, *Cognition in the Wild*, que, a través de un estudio etnográfico de la circulación de conocimiento en un sistema complejo (como una nave), propugnó una concepción distribuida del conocimiento entre mentes, artefactos y especializaciones sociales. En la concepción postcognitivista y dinámica de la mente fueron muy influyentes los experimentos con robots no programados, dotados de sensores, capacidad de movimiento y una red neuronal interna, tal que lograban explorar el medio y desarrollar

habilidades formadas por sus movimientos. El texto de Brooks (1991) sobre inteligencia sin representación basado en estos dispositivos supuso algo así como un apoyo empírico a lo que estaba ya comenzando a intuirse como una nueva concepción dinámica de la mente en la que el conocimiento emerge del sistema dinámico formado por el tejido nervioso, los sensores y actuadores, la interacción con un medio con cierta estructura que varía desde un espacio experimental a un nicho ecológico o, en el caso humano, un entorno de cultura material y de zona de desarrollo próximo, en donde la respuesta del medio es social. El conocimiento en este modelo fue llamado “4E” por las dimensiones que lo caracterizan (*embodied, extended, embedded, enactive*).

El complejo de todas estas aproximaciones conduce a la concepción enactivista radical en la que se establece una radical separación con el programa materialista de Dennett. Di Paolo, Buhrmann y Barandiaran (2017), conocidos representantes de esta línea, junto con Daniel Hutto o Anthony Chemero (2009), expresan así el abandono del representacionalismo:

El enfoque enactivo, por el contrario, no es representacional. Esto no solo se debe a que las críticas al representacionalismo son demasiado numerosas y graves como para ignorarlas, sino también porque el representacionalismo no es la única opción disponible (como se intentará demostrar en el resto de los capítulos de este libro) y, por lo tanto, los argumentos por defecto (“no hay alternativa”) no funcionan [...] para conocer las formas plausibles que pueden adoptar las explicaciones no representacionales. El enfoque enactivo busca proporcionar una base conceptual sólida para el aspecto de la materialidad que limita la validez de las descripciones funcionales positivas de los términos clave de las ciencias de la mente (significado, agencia, identidad, interacción, entorno, normas, acción, percepción, etc.). Por muchas de las razones ya mencionadas (disonancia fenomenológica, falta de operatividad, confusiones entre los niveles personal y subpersonal, y círculos viciosos homúnculos) una base conceptual naturalista no es posible para la noción de “representación interna”. En cambio, los enactivistas conciben a los agentes dando

sentido a su entorno al acoplar procesos precarios de autoindividuación (en diferentes niveles) con la dinámica ambiental. Estos procesos están imbuidos de valor y sensibilidad hacia los efectos potenciales sobre el mantenimiento de la identidad del agente. El concepto de autonomía en el centro de esta actividad no puede abordarse en términos funcionalistas porque se basa en la metaestabilidad precaria de los procesos constitutivos, un aspecto negativo de la materialidad que limita la validez de las descripciones funcionales positivas (Di Paolo *et al.*, 2017, pp. 26-27).

Este párrafo resume muy bien la forma de naturalismo enactivista radical en la que el abandono de las extrañas entidades que son las representaciones no entraña el abandono del significado, la experiencia, la autoidentidad del sujeto y el mundo de sentido en el que discurre el tiempo y el espacio de los seres vivos, y en particular los humanos. En la dinámica inacabable del cuerpo, la mente y el entorno, emerge el mundo de sentido sin que sea necesario postular la dicotomía representación/contenido. La mente no puede ser reducida a procesos computacionales: es fundamentalmente relacional y la cognición se entiende en términos de prácticas y habilidades en lugar de representaciones internas. Permítaseme citar otro pasaje de Di Paolo *et al.* que resume muy bien el materialismo alternativo del enactivismo:

La metáfora de la computadora impone una gramática implícita; promueve ciertas preguntas, pero no otras. La atención se dirige hacia el funcionamiento de las cosas, más precisamente, hacia cómo se resuelven o deberían resolverse los problemas cognitivos. Sin embargo, existen muchas preguntas importantes que no pueden ser respondidas únicamente, o en absoluto, de esta manera. Una teoría de la cognición, la acción, la emoción y la percepción, es decir, una teoría de la mente, debe ser capaz de indicarnos con cierta precisión qué es un sistema cognitivo y cuáles sistemas no son cognitivos, sin importar su complejidad. Debe proporcionar una buena explicación de cómo un animal o una persona es individuado como tal, cómo el cuerpo se transforma a través del desarrollo y se moldea por

influencias socioculturales. Fundamentalmente, una teoría de la mente debería ser capaz de aclararnos en qué sentido somos seres autónomos que generan significado, qué significa cuando nos proclamamos agentes y por qué nuestras interacciones con el mundo son experimentadas como imbuidas de significancia e importancia. ¿Por qué un sistema cognitivo y vivo se preocupa por algo? ¿Por qué nos importa? No existen respuestas algorítmicas a estas preguntas, ya que no se trata de preguntas sobre cómo se resuelve un problema o sobre cómo debe diseñarse o implementarse algún cálculo, ni sobre cuál fuente de información es la más provechosa. Sin embargo, esto no implica que no existan respuestas científicas válidas que podamos considerar (Di Paolo *et al.*, 2017, p. 2).

A modo de conclusión: el materialismo limitado de Dennett

El eliminativismo computacional de Dennett tiene un componente teórico que ha sido audaz, iluminador, que ha situado la mente y la cultura en el espacio darwiniano de la emergencia de la vida, y tiene un componente un tanto propagandístico, que socava el lugar de la fenomenología en la dinámica de la vida (vegetal, animal, humana), la cuestión de la libertad de decisión en su modalidad de compatibilismo y la propia autoconcepción humana, pero sobre todo el lugar material de la experiencia y la agencia. Es un materialismo en parte retórico que abandona el cartesianismo (o el estereotipo del cartesianismo) para crear una nueva división entre materia e información. La relectura de sus libros, especialmente de *La peligrosa idea de Darwin*, provoca, al menos en quien suscribe estas líneas, un sentimiento de nostalgia por lo iluminadora que fue su obra y la necesidad de abandonar el dualismo entre mente y cuerpo, implícito aún en ciertas formas de funcionalismo. Quizás el impulso retórico anticartesiano llevó a Dennett a no entender lo que estaban significando las nuevas consideraciones de la fenomenología o postfenomenología y a no valorar las nuevas concepciones dinámicas y extendidas de la mente. Puede que, en un cierto sentido, haya que considerar que Dennett aún se mueve en lo que podríamos llamar, con una expresión del marxismo, un “materialismo vulgar”.

Referencias

- Ayala, F. J. y Arp, R. (2010). *Contemporary Debates in Philosophy of Biology*. Wiley-Blackwell. <https://doi.org/10.1002/9781444314922>
- Broncano, F. (2025, 20 de octubre). Materialismos. *El laberinto de la identidad. Reflexiones en las fronteras de la cultura y la ciencia, la filosofía y la literatura, la melancolía y la esperanza*. <https://laberintodelaidentidad.blogspot.com/2025/10/?m=0>
- Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence*, 47(1-3), 139-59. [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M)
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. The MIT Press. <https://doi.org/10.7551/mitpress/8367.001.0001>
- Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster.
- Dennett, D. (1997). *Kinds of Minds: Towards an Understanding of Consciousness*. Basic Books.
- Dennett, D. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. W. W. Norton & Company.
- Di Paolo, E., Buhrmann, T. y Barandiaran, X. E. (2017). *Sensory Motor Life: An Enactive Proposal*. Oxford University Press.
- Donald, M. (1991). *Origins of the Modern Mind*. Harvard University Press.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- Godfrey-Smith, P. (2001). Three Kinds of Adaptationism. En S. H. Orzack y E. Sober (eds.), *Adaptationism and Optimality* (pp. 335-357). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609084.012>
- Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus and Giroux.
- Godfrey-Smith, P. (2018). Towers and Trees in Cognitive Evolution. En B. Huebner (ed.), *The Philosophy of Daniel Dennett* (pp. 225-249). Oxford University Press. <https://doi.org/10.1093/oso/9780199367511.003.0015>
- Hutchins, E. (1995). *Cognition in the Wild*. The MIT Press. <https://doi.org/10.7551/mitpress/1881.001.0001>
- Ihde, D. (1993). *Postphenomenology: Essays in the Postmodern Context*. Northwestern University Press.
- Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. The MIT Press. <https://doi.org/10.7551/mitpress/9476.001.0001>
- Malafouris, L. (2019). Mind and Material Engagement. *Phenomenology and the Cognitive Sciences*, 18(1), 1-17. <https://doi.org/10.1007/s11097-018-9606-7>

- Maturana, H. R. y Varela, F. J. (1972). *Autopoiesis and Cognition: The Realization of the Living*. Reidel.
- Maynard Smith, J. (1958). *The Theory of Evolution*. Penguin Books.
- Newell, A. y H. A. Simon. (1972). *Human Problem Solving*. Prentice-Hall.
- Physarum polycephalum*. (2026, 30 de abril). En Wikipedia. https://en.wikipedia.org/w/index.php?title=Physarum_polycephalum&oldid=1351818378
- Putnam, H. (1975). Philosophy and Our Mental Life. En H. Putnam, *Mind, Language and Reality: Philosophical Papers. Volume 2*. (pp. 291-303). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625251.016>
- Rosenblueth, A., Wiener, N. y Bigelow, J. (1943). Behavior, Purpose and Teleology. *Philosophy of Science*, 10(1), 18-24. <https://doi.org/10.1086/286788>
- Searle, J. (2010). *Consciousness and Language*. Cambridge University Press
- Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Cambridge University Press.
- Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press.

<https://doi.org/10.21555/top.v760.3580>

Simulación y justificación en la era del *deep learning*: AlphaFold como ejemplar paradigmático de modelos opacos

Simulation and Justification in the Age of Deep Learning: AlphaFold as a Paradigmatic Case of Opaque Models

Erick Manuel Rubio
Fundación Universitaria San Martín
Colombia
Universidad del Norte
Colombia
erick.rubio@sanmartin.edu.co
<https://orcid.org/0009-0005-1900-2369>

Leandro Giri
Universidad Nacional de Tres de Febrero
Argentina
Consejo Nacional de Investigaciones
Científicas y Técnicas
Argentina
leandrogiri@gmail.com
<https://orcid.org/0000-0002-7068-9750>

Miguel Fuentes
Consejo Nacional de Investigaciones
Científicas y Técnicas
Argentina
fuentesmig@gmail.com
<https://orcid.org/0000-0003-0833-1245>

Elkin Navarro
Fundación Universitaria San Martín
Colombia
elkin.navarro@sanmartin.edu.co
<https://orcid.org/0000-0001-7567-6409>

Recibido: 25 - 09 - 2025.

Aceptado: 27 - 02 - 2026.

Publicado en línea: 10 - 07 - 2026.

Cómo citar este artículo: Rubio, E. M., Giri, L., Fuentes, M. y Navarro, E. (2026). Simulación y justificación en la era del *deep learning*: AlphaFold como ejemplar paradigmático de modelos opacos. *Tópicos, Revista de Filosofía*, 76, 283-313. <https://doi.org/10.21555/top.v760.3580>



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

Resumen

Este artículo examina el estatuto epistémico de los modelos computacionales contemporáneos basados en *deep learning* a partir del caso de AlphaFold, con el objetivo de evaluar en qué medida estos sistemas pueden comprenderse dentro del marco tradicional de las simulaciones computacionales o si requieren categorías epistémicas diferenciadas. A partir del análisis del marco propuesto por Paul Humphreys —en particular, las nociones de “plantilla” y “modelo computacional”— se explora la posibilidad de una continuidad formal y epistémica entre las simulaciones clásicas y los sistemas de aprendizaje profundo, atendiendo críticamente a los problemas de opacidad, verificación y validación pragmática, y al papel constitutivo de la representación visual. Se muestra que, si bien las concepciones representacionistas clásicas resultan insuficientes para dar cuenta del funcionamiento interno de modelos como AlphaFold, estos sistemas tampoco pueden reducirse sin más a meros dispositivos algorítmicos. En este sentido, el artículo argumenta que AlphaFold sitúa a la filosofía de la ciencia ante la necesidad de repensar los criterios de justificación epistémica y el alcance mismo de la noción de “simulación” en el contexto del *deep learning*.

Palabras clave: *deep learning*; simulaciones de *deep learning* epistémicamente opacas; aprendizaje profundo; opacidad epistémica; representación científica; validación; verificación; justificación; simulación computacional; modelo computacional; plantilla computacional.

Abstract

This paper examines the epistemic status of contemporary computational models based on deep learning through the case of AlphaFold, with the aim of assessing to what extent such systems can be understood within the traditional framework of computational simulations or whether they require distinct epistemic categories. Drawing on Paul Humphreys' analytical framework—particularly the notions of computational templates and computational models—the paper explores the possibility of a formal and epistemic continuity between classical simulations and deep learning systems, while critically addressing issues of opacity, pragmatic validation, and the constitutive role of visual representation. The analysis shows that, although classical representationalist accounts are insufficient to capture the internal functioning of models such as AlphaFold, these systems cannot be straightforwardly reduced to mere algorithmic devices. In this sense, it is argued that AlphaFold confronts philosophy of science with the need to rethink both the criteria of epistemic justification and the scope of the notion of simulation in the era of deep learning.

Keywords: deep learning; epistemically opaque deep-learning simulations; epistemic opacity; scientific representation; validation; verification; justification; computer simulation; computational model; computational template.

Introducción

Las simulaciones computacionales se han consolidado como una herramienta científica de fuste, surgiendo y afianzándose notablemente a partir de la Segunda Guerra Mundial (Galison, 1996; Winsberg y Alvarado, 2022). En ese período, su impacto fue decisivo en campos como la meteorología, con figuras pioneras como Edward Lorenz y su Royal McBee LGP-30,¹ y en la física nuclear, donde matemáticos como John von Neumann realizaron aportes determinantes en el marco del Proyecto Manhattan, especialmente en el desarrollo de armas nucleares. De hecho, fue en este contexto donde el método de simulación de Monte Carlo, propuesto en gran medida por Stanislaw Ulam, tuvo su génesis en la investigación científica, marcando una nueva forma de trabajar con la aleatoriedad como pilar fundamental. Para Peter Galison (1996, p. 119), las simulaciones computacionales, particularmente el método de Monte Carlo, transformaron las tradiciones de investigación existentes —teórica, experimental e instrumental (Lenhard, 2016, pp. 729 y 736)— y sentaron las bases para un nuevo modo de producción de conocimiento científico.

Desde entonces, se ha vuelto indispensable en un número creciente de disciplinas. La lista de ciencias que hacen un uso extensivo de la simulación por computadora ha crecido hasta incluir la astrofísica, la física de partículas, la ciencia de materiales, la ingeniería, la mecánica de fluidos, la ciencia del clima, la biología evolutiva, la ecología, la economía, la teoría de la decisión, la medicina, la sociología, la epidemiología y muchas otras. Incluso hay algunas disciplinas, como la teoría del caos y la teoría de la complejidad, cuya propia existencia ha surgido junto con el desarrollo de

¹ El Royal McBee LGP-30, adquirido por Lorenz en 1960, constituyó una de las primeras computadoras de propósito general aplicadas al modelado atmosférico; permitió formalizar el estudio de la sensibilidad a condiciones iniciales en sistemas dinámicos no lineales.

los modelos computacionales que estudian (Winsberg y Alvarado, 2022).²

Se consideran no solo instrumentos, sino métodos que han abierto nuevas vías para comprender el mundo y sus fenómenos (Durán, 2017, p. 125). En muchos contextos, el modelado computacional se ha consolidado como una herramienta insustituible para explorar escenarios complejos que desafían el análisis experimental o lógico-formal (Vicsek, 2002). Este tipo de modelado a menudo se enmarca en lo que la literatura reciente denomina “modelado modal”, que busca proveer información sobre lo posible, lo contrafactual o las potencialidades de sistemas; sirve para la exploración de hipótesis y la comprensión de fenómenos que aún no son actuales (Sjölin Wirling y Grüne-Yanoff, 2021). Así, la simulación computacional, desde sus inicios, no solo amplía el alcance del conocimiento científico hacia dominios tradicionalmente inaccesibles, sino que también desafía a la filosofía de la ciencia a repensar las formas de justificación y explicación. Este trasfondo permite comprender por qué la aparición de sistemas como AlphaFold representa no solo una continuidad en la sofisticación y centralidad del modelado computacional, sino también una transformación cualitativa en los modos de producción y validación del conocimiento científico.

Un ejemplo reciente y de gran relevancia de la capacidad predictiva y transformadora de los sistemas de simulación es AlphaFold, un sistema de inteligencia artificial desarrollado por DeepMind que, mediante arquitecturas de tipo *deep learning* basadas en *transformers*,³ predice

² A menos que la bibliografía final indique lo contrario, las traducciones son nuestras.

³ El *transformer* es una arquitectura de red neuronal para modelos de transducción de secuencias, como la traducción automática, que se basa únicamente en mecanismos de atención para establecer dependencias globales entre la entrada (*input*) y la salida (*output*). A diferencia de modelos anteriores, el *transformer* prescinde por completo de la recurrencia y las convoluciones. Su estructura principal consiste en un codificador (*encoder*) y un decodificador (*decoder*) apilados, ambos contruidos con capas de “atención de múltiples cabezales” (*multi-head self-attention*) y redes neuronales de avance (*feed-forward*). Esta arquitectura permite un mayor grado de paralelización en el entrenamiento y reduce significativamente el tiempo necesario para aprender dependencias a larga distancia dentro de las secuencias (Vaswani *et al.*, 2017).

la estructura tridimensional de proteínas a partir de su secuencia de aminoácidos.

En 2024, el Premio Nobel de Química fue otorgado a Demis Hassabis y John Jumper por el desarrollo precisamente de AlphaFold, y a David Baker por sus contribuciones al diseño computacional de proteínas. La importancia científica de AlphaFold radica en que ha alcanzado, en numerosos casos, niveles de precisión comparables a los obtenidos mediante técnicas experimentales consolidadas como la cristalografía de rayos X o la criomicroscopía electrónica, resolviendo así, en gran medida, un problema abierto durante más de cinco décadas en biología estructural (Jumper *et al.*, 2021).

La posibilidad de abordar sistemas complejos mediante simulaciones computacionales permite identificar patrones y comportamientos replicables, incluso cuando se trata de conjuntos masivos de datos (Chirimuuta, 2021). Sin embargo, el caso de AlphaFold marca un punto de inflexión. Ya no se trata únicamente de representar dinámicamente procesos físicos conocidos, sino de generar conocimiento a partir de la detección de correlaciones empíricas que no necesariamente están mediadas por teorías explícitas o modelos representacionales. En este sentido, AlphaFold pone en tensión los marcos clásicos con los que la filosofía de la ciencia ha comprendido los modelos, la representación y la justificación epistémica.

A primera vista, podría parecer más apropiado considerarlo una nueva clase de artefactos científicos: opacos, no representacionales, pero epistémicamente fértiles. En este punto, se puede establecer un paralelo con el modelado modal, donde los modelos a menudo no satisfacen los criterios estándar de precisión representacional, pero aun así ofrecen un conocimiento epistémico valioso sobre las posibilidades (Sjölin Wirling y Grüne-Yanoff, 2021). AlphaFold, al predecir estructuras, puede entenderse como un medio para explorar lo que es objetivamente posible en el plegamiento de proteínas, incluso en ausencia de una teoría causal explícita.

No obstante, también podría sostenerse —como alternativa que será examinada críticamente más adelante— que casos como AlphaFold invitan a repensar la noción de “simulación” en un sentido formal-computacional. Esta lectura, sin embargo, no se apoya en la mera presencia de estructuras matemáticas o en la algoritmicidad del sistema, ya que estos rasgos, por sí solos, resultarían insuficientes para distinguir una simulación de otros dispositivos computacionales. Más bien, lo que

estaría en juego es si la resolución algorítmica de ciertas estructuras, integrada en prácticas de validación y uso experimental, puede cumplir funciones epistémicas análogas a las tradicionalmente atribuidas a las simulaciones. En este artículo nos ocuparemos de AlphaFold no como un objeto aislado, sino como un caso paradigmático de una nueva clase de artefactos computacionales opacos sustentados en redes neuronales profundas, que denominaremos EODS (*Epistemically Opaque Deep-learning Simulations*). La cuestión de si estos sistemas pueden entenderse mediante la noción clásica de “simulación” o no será objeto de análisis a lo largo del trabajo.

Casos análogos —como los grandes modelos de lenguaje (por ejemplo, ChatGPT) o los generadores de imágenes (p. ej., DALL-E)— sugieren que lo que está en juego no se restringe a la biología estructural; porque comparten con AlphaFold un mismo perfil epistémico, en tanto que producen resultados empíricamente fiables mediante arquitecturas algorítmicas opacas, carecen de una teoría mediadora explícita del fenómeno que tratan y se integran a la práctica científica (o técnico-científica) a través de criterios de validación pragmática y de interfaces representacionales diseñadas para el uso humano. En este sentido, estos sistemas señalan una transformación más general en la arquitectura epistémica de la ciencia contemporánea, en la que la producción de conocimiento se apoya crecientemente en sistemas algorítmicos opacos cuya fiabilidad se establece por vías distintas de la comprensión causal tradicional.

Otro ejemplo potente se encuentra en la química cuántica y el diseño de fármacos. Tradicionalmente, predecir las propiedades de una molécula (como su energía de enlace o su reactividad) requiere costosas simulaciones basadas en primeros principios, como la teoría del funcional de la densidad (DFT) (Jensen, 2017). Hoy en día, redes neuronales de grafos (GNNs) son entrenadas con enormes bases de datos de estructuras moleculares y sus propiedades cuánticas calculadas. Estos modelos aprenden a predecir dichas propiedades directamente a partir de la topología 2D de la molécula (Gilmer *et al.*, 2017; Yang *et al.*, 2019), obviando la necesidad de resolver las ecuaciones de Schrödinger. Al igual que AlphaFold, su funcionamiento interno es opaco, pero su precisión predictiva es tan alta que se han convertido en herramientas indispensables para el cribado virtual (*virtual screening*) de millones de moléculas candidatas a fármacos, acelerando drásticamente el proceso de descubrimiento (Goh *et al.*, 2017; Stokes *et al.*, 2020).

En suma, este trabajo defiende una doble tesis: (1) existe una continuidad estructural y epistémica significativa entre las simulaciones computacionales clásicas y los sistemas actuales de *deep learning* —ambos constituyen tecnologías de producción de conocimiento sustentadas cada vez más en la validación empírico-pragmática, la opacidad inferencial y, en muchos casos, la centralidad de la visualización de resultados—; y (2) el reconocimiento filosófico de esta continuidad exige matizar y expandir los marcos conceptuales clásicos de modelo, representación y justificación epistémica, dada la irrupción de retos inéditos (como la radicalización de la opacidad, la falta de teoría mediadora y la centralidad de la interfaz humano-algoritmo) en sistemas como AlphaFold.

Destacar la relación entre ambas dimensiones —más que yuxtaponerlas— permite identificar con precisión qué aspectos se preservan, transforman o desafían en este nuevo paisaje epistémico. El análisis histórico y funcional de la continuidad solo cobra pleno sentido filosófico si revela los límites de los marcos clásicos; la propuesta de un cambio conceptual, a su vez, se fundamenta en el reconocimiento situado de las verdaderas novedades que trae el *deep learning* en el contexto de producción científica. Esta mirada doble evita tanto el rupturismo —que vería a AlphaFold como incomunicable con la tradición filosófica— como el simple continuismo —que reduciría la novedad a mera sofisticación técnica—, y justifica la urgente tarea de expandir las categorías filosóficas para comprender el estatuto epistémico del conocimiento científico, justamente en este nuevo paisaje epistémico.

Entre las principales objeciones cabe mencionar: (a) que la continuidad entre simulación y *deep learning* sería superficial dada la opacidad radical e inexplicabilidad de este último; (b) que los llamados a nuevos marcos conceptuales solo reeditan debates previos sobre justificación pragmática en la modelización. Nuestra estrategia es mostrar, a través del enfoque doble, tanto los elementos que justifican la comparación (validación empírica, creatividad modal, función exploratoria) como aquellos que exigen ampliaciones filosóficas genuinas (ausencia de teoría mediadora, nuevo *locus* para la representación y la justificación en la interfaz humano-algoritmo).

El trabajo se organiza en cuatro partes. En la primera se contextualiza el lugar de las simulaciones dentro de la filosofía de la ciencia atendiendo a las posiciones que las consideran una metodología epistémicamente novedosa y a aquellas que adoptan una postura continuista. En la segunda, se analiza la propuesta de Paul Humphreys

y su noción de “plantillas” y “modelos computacionales” como unidad de análisis evaluando en qué medida AlphaFold puede comprenderse en este marco. La tercera parte aborda la tensión entre las concepciones representacionistas clásicas y los modelos opacos basados en *deep learning*. Finalmente, se examinan los procedimientos de verificación y validación que consolidan el estatuto epistémico de AlphaFold, para concluir con una reflexión sobre las implicaciones que este tipo de artefactos tiene para la filosofía de la ciencia contemporánea.

1. Debate sobre el estatuto epistémico de las simulaciones computacionales

A pesar de su creciente ubicuidad e impacto transformador en la práctica científica, las simulaciones computacionales han suscitado una respuesta ambivalente dentro de la filosofía de la ciencia (Durán, 2017, p. 125; Lenhard, 2016; Morrison, 2012). Mientras en la práctica son cada vez más centrales en el descubrimiento, validación y justificación científica, su estatuto epistémico ha sido objeto de debate. ¿Son simplemente herramientas auxiliares que extienden nuestras capacidades de cálculo y visualización, o constituyen una forma cualitativamente nueva de conocimiento científico que exige repensar nuestras categorías filosóficas tradicionales? Este debate constituye el trasfondo necesario para evaluar hasta qué punto artefactos recientes como AlphaFold pueden ser pensados como parte de esa tradición o, más bien, como un caso que desborda sus categorías.

Autores como Eric Winsberg (2001 y 2010), Peter Galison (1996), Fritz Rohrlich (1990) y Paul Humphreys (2004 y 2009) han defendido la idea de que las simulaciones computacionales no pueden ser reducidas ni comprendidas adecuadamente desde los marcos clásicos de la filosofía de la ciencia. Para ellos, las simulaciones introducen una metodología con características propias, distintas de la teoría, el experimento y la instrumentación. Rohrlich (1990), por ejemplo, sostenía que el poder de cálculo asociado a las simulaciones supera las capacidades cognitivas humanas, y por ello constituyen una “metodología cualitativamente nueva” para las ciencias naturales (citado en Durán, 2017, p. 125). Winsberg (2001, p. 447) argumenta que las simulaciones poseen una “epistemología distintiva”, no reductible a las categorías tradicionales, como los modelos o las leyes (citado en Durán, 2017, p. 126).

Juan M. Durán (2017), en esta misma línea, sostiene que las simulaciones computacionales abren un campo filosófico específico que desafía el modo clásico de concebir la explicación y la justificación en ciencia. A su juicio, la epistemología que requieren las simulaciones no es solo una variación técnica o metodológica dentro del paradigma clásico, sino una ruptura conceptual que obliga a pensar en nuevas unidades de análisis.

Estas posiciones comparten una convicción fundamental, a saber, que la novedad de las simulaciones no es meramente empírica ni tecnológica, sino también filosófica. Tal novedad se expresa en distintos planos: ontológico, al introducir artefactos que no se limitan a representar entidades físicas, sino que operan mediante correlaciones estadísticas; epistémico, al producir conocimiento sin comprensión teórica explícita; metodológico (Winsberg, 2001 y 2010; Durán, 2017), al desplazar los procedimientos clásicos de validación y verificación. Por último, representacional, al cuestionar la centralidad de la transparencia conceptual como criterio de valor científico.

En contraste con estas posturas afirmativas, otros filósofos de la ciencia han mostrado escepticismo frente a la necesidad de revisar las categorías tradicionales. Roman Frigg y Julian Reiss (2009) reconocen que las simulaciones computacionales han introducido elementos novedosos en la práctica científica, pero se resisten a considerar que esta novedad implique una disrupción filosófica significativa. Según ellos, las cuestiones que plantean las simulaciones pueden ser tratadas, sin grandes ajustes conceptuales, desde los marcos ya existentes en la filosofía de los modelos científicos y de la experimentación.

A su juicio, muchas de las características distintivas de las simulaciones, como su componente visual, su carácter iterativo o su dependencia de algoritmos complejos, no constituyen problemas filosóficos genuinos, sino más bien cuestiones de tipo técnico, matemático o incluso psicológico. Así, por ejemplo, afirman que la preferencia por la visualización de resultados no es un tema epistemológico en sí mismo, sino una cuestión de percepción. “Parece ser más bien una cuestión de psicología antes que de epistemología el hecho de que preferimos lo visual a otras formas de representación” (Frigg y Reiss, 2009, p. 607).

Este tipo de argumentación refleja una postura que podríamos denominar continuista, en la medida en que no considera que las simulaciones computacionales exijan una reelaboración sustantiva del aparato conceptual de la filosofía de la ciencia. Aunque Frigg y Reiss

(2009) admiten que las simulaciones “introducen algo nuevo y excitante para la *ciencia*” (p. 595), concluyen que no hay razones suficientes para “reescribir la *filosofía de la ciencia*” (p. 595). Es importante señalar, no obstante, que esta postura fue formulada en un contexto previo a la irrupción masiva de tecnologías basadas en aprendizaje profundo —*deep learning*—. En ese momento, la mayoría de las simulaciones computacionales seguían estando estructuradas por modelos representacionales relativamente transparentes, apoyados en principios físicos o ecuaciones diferenciales explícitas. Así, el escepticismo de Frigg y Reiss puede entenderse como una defensa de la capacidad explicativa de los marcos tradicionales dentro de un paisaje científico todavía dominado por modelos interpretablemente anclados en teorías.

La novedad epistémica de las simulaciones que proponen Frigg y Reiss —al considerarlas meramente una forma de modelos matemáticos computacionalmente ejecutables— resulta problemática incluso en el contexto en el que escribieron. Ya entonces, prácticas como la climatología computacional, con sus mapas de temperatura, precipitación o circulación atmosférica, o la cosmología con simulaciones como la *Millennium Simulation* (Particle Physics & Astronomy Research Council, 2005), mostraban que las visualizaciones no eran simples ayudas perceptivas, sino instrumentos epistémicos centrales para la validación y la exploración teórica (Lenhard, 2016). Estos casos históricos bastaban para cuestionar la tesis de que lo visual es un asunto meramente psicológico. Ahora bien, la fuerte irrupción de EODS, como AlphaFold, intensifica aún más esta dificultad, pues en ellos la visualización constituye prácticamente la única vía de acceso epistémico a modelos internos opacos, lo cual refuerza el carácter indispensable de la representación visual para la práctica científica.

Diversos autores han destacado que la visualización cumple funciones que no pueden ser reemplazadas por descripciones formales o numéricas. Morrison (2015b), por ejemplo, ha argumentado que la forma en que los científicos interactúan con modelos complejos es inseparable de las maneras en que estos modelos se presentan visualmente. En este sentido, la visualización se convierte en una condición epistémica de posibilidad para el trabajo científico en entornos computacionalmente intensivos donde la complejidad se hace manifiesta (Lenhard, 2016). Aunque, podría objetarse, claro está, que la visualización, por más central que sea en la práctica científica, no altera el estatuto epistemológico de

fondo, sino que simplemente facilita el acceso humano a resultados ya generados algorítmicamente.

En efecto, los argumentos de Frigg y Reiss pueden ser refutados en sus propios términos, pues incluso antes del *deep learning* la visualización desempeñaba funciones epistémicas centrales que excedían lo meramente perceptivo. Sin embargo, el punto decisivo no está en esa disputa histórica, sino en el impacto de los desarrollos contemporáneos. AlphaFold representa un quiebre porque introduce una transformación significativa, en tanto que su validez epistémica no puede explicarse apelando a modelos representacionales clásicos ni a simulaciones dinámicas tradicionales ni a herramientas visuales subordinadas a teorías explícitas. Se trata de un tipo de artefacto científico —EODS— que obliga a reconsiderar la manera en que entendemos la justificación y el valor de las simulaciones en la ciencia actual.

Su potencia reside en la capacidad de generar predicciones altamente precisas sin apoyarse en una teoría explicativa explícita ni en una representación interpretable del proceso físico estudiado. En lugar de describir cómo ocurre el plegamiento proteico, infiere directamente estructuras tridimensionales a partir de correlaciones estadísticas inscritas en un espacio de alta dimensionalidad sin semántica biológica explícita. Desde esta perspectiva, pone en tensión las concepciones tradicionales de “simulación”, “modelo” y “representación” para abrir una zona epistémica distinta —cercana a la noción de “régimenes de objetividad” de Daston y Galison (2007)— en la que el conocimiento científico no descansa en la inteligibilidad causal, sino en la eficiencia inferencial y en la validación empírica de sistemas opacos.

No obstante, también puede sostenerse que AlphaFold prolonga, en un sentido formal-computacional, la tradición de la simulación. En efecto, su núcleo operativo se apoya en una estructura matemática precisa: la arquitectura *transformer* (Vaswani *et al.*, 2017; Jumper *et al.*, 2021), cuya lógica de atención y actualización de parámetros constituye un aparato formal-matemático que puede interpretarse como el “modelo” subyacente. En este sentido, AlphaFold no abandona del todo la lógica de las simulaciones; más bien, la radicaliza y la desplaza hacia un plano en el que lo decisivo ya no es la representación transparente de un proceso físico, sino la ejecución algorítmica de una estructura matemática que produce resultados empíricamente validados.

En lo que sigue, analizaremos cómo este doble registro —AlphaFold como ruptura epistémica y AlphaFold como simulación en sentido

formal— fue anticipado, al menos parcialmente, por propuestas como la de Paul Humphreys, quien buscó dotar a las simulaciones computacionales de una unidad de análisis propia, distinta de categorías como “teoría”, “modelo” o “ley”. Su aproximación servirá como punto de partida para mostrar por qué AlphaFold no debería entenderse simplemente como un ejemplo sofisticado de simulación, pero tampoco puede desligarse por completo de esta tradición; se sitúa, más bien, en la frontera misma donde la noción de simulación se redefine.

2. AlphaFold y las *Epistemically Opaque Deep-learning Simulations* (EODS)

Una de las propuestas filosóficas más influyentes para comprender el estatuto epistémico de estas prácticas es la de Paul Humphreys. A diferencia de los enfoques que buscan subsumirlas dentro del esquema tradicional de teorías científicas, ya sea en su versión sintáctica o semántica, Humphreys sostiene que requieren una nueva unidad de análisis filosófico capaz de integrar sus elementos distintivos: opacidad, manipulación algorítmica, temporalidad computacional y representación de resultados. Esta propuesta resulta especialmente relevante en la medida en que ofrece un marco intermedio: permite reconocer la especificidad epistémica de las simulaciones sin tratarlas como una ruptura completamente ajena a la tradición filosófica previa.

Humphreys argumenta que los conceptos clásicos como “teoría”, “modelo”, “ley” o “paradigma” no son adecuados para capturar la especificidad de las simulaciones computacionales modernas. Desde esta perspectiva, las unidades básicas de análisis que la filosofía de la ciencia ha utilizado históricamente resultan insuficientes para caracterizar estas prácticas. Por ello, Humphreys (2004, p. 60) introduce la noción de “plantilla” (*template*) como una herramienta conceptual destinada a reformular, más que simplemente reemplazar, esas unidades tradicionales. Una plantilla se define como un conjunto de restricciones muy generales y abstractas que luego pueden ser especificadas para generar una formulación concreta aplicable a un fenómeno. Como explica Durán siguiendo a Humphreys (2004, p. 60):

La segunda ley de Newton, tal y como la conocemos, se describe para cualquier fuerza, masa y aceleración. Sin embargo para hacer uso de ella aplicándola en un caso particular es necesario primero especificar una función

de fuerza particular, ya sea la fuerza gravitacional, electroestática, magnética o cualquier otra; algo similar ocurre con los restantes elementos de los que habla la ley: masa, aceleración, etc. [...]. A partir de esto es posible interpretar la segunda ley de Newton como una “plantilla teórica”, esto es, como restricciones muy generales que no son aplicables a ningún fenómeno particular sin la previa especificación de algunos de sus componentes. Variando de especificación, entonces, será posible dar cuenta de los distintos fenómenos que puede “contener” la plantilla (Durán, 2017, p. 131).

Del mismo modo, la teoría de la probabilidad funciona como una plantilla cuyos elementos específicos, como funciones de densidad o de distribución, deben completarse en contextos particulares. Estas plantillas, en tanto construcciones sintácticas abstractas, poseen dos funciones principales, según Humphreys, “una de ellas consiste en representar el mundo; la otra, en facilitar el cálculo” (Durán, 2017, p. 132).

El paso decisivo ocurre cuando una plantilla teórica se especifica de tal forma que produce un resultado computacionalmente tratable; en ese momento, se convierte en una plantilla computacional, esto es, en una estructura formal que puede ser ejecutada algorítmicamente. Esta transición —de lo abstracto a lo computacionalmente ejecutable— resulta central para comprender el estatuto de sistemas como AlphaFold. En efecto, su arquitectura *transformer* puede entenderse como una estructura matemática general que, al ser entrenada y especificada mediante grandes volúmenes de datos proteicos, se convierte en un aparato ejecutable capaz de producir resultados empíricamente contrastables. En este sentido, AlphaFold puede interpretarse como una simulación en un sentido formal-computacional, aun cuando sus vínculos con teorías biológicas explícitas sean débiles o incluso inexistentes. Humphreys (2004, p. 78) caracteriza estas plantillas computacionales mediante una quintupla de supuestos de construcción, citada por Durán (2017, p. 134):

©Ontología, idealizaciones, abstracciones, restricciones, aproximaciones©

En esta quintupla, el primer elemento, la ontología, especifica el tipo de objetos, propiedades y relaciones que son representados. Esta especificación sirve de base para los demás componentes. Los

otros cuatro conforman lo que Humphreys denomina el *conjunto de corrección*. Como señala Durán (2017, p. 134), “este conjunto [...] tiene como propósito principal indicar cómo las plantillas deben ser ajustadas cuando los resultados no coinciden con los datos empíricos”.

Dicho conjunto puede implicar desde la relajación de restricciones y el refinamiento de aproximaciones hasta modificaciones más profundas en la ontología de base. Un aspecto especialmente relevante para nuestra discusión es que estas plantillas computacionales no requieren derivarse de teorías científicas tradicionales, lo cual amplía significativamente el espectro de artefactos que pueden ser comprendidos dentro de este marco. Un ejemplo paradigmático de este punto es el proceso de Poisson, que puede funcionar como una plantilla de gran generalidad aplicable a fenómenos muy diversos —desde llamadas telefónicas hasta desintegraciones radioactivas— sin presuponer una teoría causal común que los unifique. Lo que legitima su uso no es la pertenencia a un mismo dominio teórico, sino la posibilidad de ser representados y computados mediante una misma estructura formal (Durán, 2017, p. 135).

Esta idea prepara el terreno para pensar modelos como AlphaFold, que no se derivan de ninguna teoría causal explícita, sino que operan sobre estructuras inferenciales entrenadas estadísticamente. Sin embargo, para comprender plenamente el estatuto epistémico de este tipo de artefactos, es necesario avanzar hacia una unidad de análisis más amplia. Si bien las plantillas computacionales capturan adecuadamente el paso de lo abstracto a lo computacionalmente ejecutable, Humphreys sostiene que no bastan para caracterizar por completo lo que ocurre en una simulación. Por ello, propone una unidad de análisis más robusta, a saber, el modelo computacional, definido como una séxtupla que integra no solo la plantilla y sus supuestos, sino también elementos vinculados con la interpretación, la justificación epistémica y la representación.

Así, el modelo computacional se presenta como una estructura formal compuesta por los siguientes elementos (Humphreys, 2004, pp. 102-103, citado en Durán, 2017, p. 137):

©Plantilla, supuestos de construcción, conjunto de corrección, interpretación, justificación inicial, representación©

La plantilla computacional provee el andamiaje formal básico —ya sea en forma de ecuaciones u otros aparatos matemáticos—, mientras que los supuestos de construcción y el conjunto de corrección se refieren,

respectivamente, a las decisiones necesarias para operacionalizar el modelo y a los ajustes que permiten mantenerlo en sintonía con los datos empíricos. A estos elementos se añaden tres componentes que amplían la dimensión propiamente epistémica del modelo computacional: la interpretación, que especifica cómo deben entenderse sus elementos en relación con el sistema estudiado; la justificación inicial, que explica por qué resulta adecuado emplear ese modelo en un contexto determinado; y la representación, que comprende tanto los resultados del proceso de simulación como las formas de su presentación.

La novedad de esta propuesta radica en que la unidad de análisis no es ninguna de estas partes por sí sola, sino su combinación funcional. En este sentido, los modelos computacionales no pueden ser comprendidos adecuadamente desde una perspectiva reduccionista o parcial. En particular, la representación de los resultados no constituye un aspecto meramente estético o auxiliar, sino una dimensión metodológica central para la interpretación y la validación del modelo. En este punto, la propuesta de Humphreys se aleja tanto de la concepción sintáctica de las teorías científicas —que las concibe como conjuntos de sentencias organizadas deductivamente— como de la concepción semántica —que las identifica con familias de modelos matemáticos abstractos—.

Sin embargo, podría objetarse que este desplazamiento hacia la combinación de elementos corre el riesgo de convertirse en una taxonomía acumulativa más que en una solución conceptual: ¿es suficiente enumerar componentes para dar cuenta de la novedad epistémica, o se requiere un marco más radicalmente distinto? Este interrogante resulta especialmente pertinente en el caso de AlphaFold, donde algunos de los elementos de la séxtupla —en particular la interpretación y la representación— adquieren un estatuto problemático, pues no se apoyan en teorías subyacentes claras, sino en correlaciones opacas aprendidas algorítmicamente. En este contexto, la inclusión de la temporalidad computacional adquiere un papel decisivo: para Humphreys, el modelo computacional no solo representa algo del mundo, sino que ejecuta un proceso computacional que produce resultados en el tiempo.

Humphreys denomina “núcleo de la simulación” a la dimensión temporal del proceso computacional, entendido como un proceso que produce soluciones a un modelo computacional en el tiempo. De este modo, la simulación no se reduce a una estructura formal o representacional, sino que implica una transformación activa de *inputs* en *outputs* mediante algoritmos y procedimientos numéricos. Aunque

toda computación transcurre en el tiempo, en las simulaciones científicas el tiempo de ejecución no constituye un mero subproducto técnico, sino el medio a través del cual se generan dinámicas que sustituyen, representan o anticipan procesos del mundo físico. Es precisamente en este punto donde AlphaFold tensiona el marco de Humphreys; su núcleo temporal no simula el plegamiento proteico paso a paso, sino que realiza inferencias estadísticas masivas que desembocan directamente en configuraciones estructurales finales.

A diferencia de otros enfoques, Humphreys sostiene que las peculiaridades metodológicas de las simulaciones emergen con especial claridad en los procesos de representación de sus resultados. El modo en que los *outputs* de una simulación son presentados —ya sea en forma de tablas numéricas, gráficos estáticos o visualizaciones dinámicas— no es metodológicamente neutro, pues incide directamente en la interpretación del modelo computacional y en las inferencias que pueden extraerse de él. En este sentido, la representación cumple una función epistemológica central en la medida en que media entre los resultados computacionales y la comprensión humana, aun cuando no siempre facilite un entendimiento causal del fenómeno estudiado (Durán, 2017, pp. 137-138).

Esto resulta especialmente relevante en el caso de las EODS como AlphaFold, donde la comprensión humana del proceso inferencial interno es limitada, pero la presentación visual de los resultados permite la producción de inferencias científicas robustas. En estos casos, la representación visual no se limita a comunicar un resultado previamente constituido, sino que participa activamente en su validación epistémica. De este modo, AlphaFold produce una forma de conocimiento que no se apoya en una comprensión teórica o causal del fenómeno ni en una inteligibilidad transparente de su funcionamiento interno. Como señalan Gao y Ganguli (2015), incluso cuando se conocen la arquitectura de la red, sus reglas de actualización y sus parámetros, “aún nos elude una comprensión significativa de cómo funcionan estas redes” (p. 151).

Ahora bien, podría objetarse que la dependencia de AlphaFold respecto de visualizaciones tridimensionales para la validación de sus resultados no lo diferencia de manera categórica de otras prácticas científicas en las que las representaciones gráficas también funcionan como mediaciones indispensables —por ejemplo, en la astrofísica computacional o en la biología de sistemas—. La diferencia, sin embargo, no es meramente instrumental: en AlphaFold la visualización no cumple

un rol auxiliar respecto del modelo, sino que se convierte en la principal vía de acceso epistémico a un sistema cuyo funcionamiento inferencial interno permanece en gran medida opaco.

Esta opacidad epistémica ha sido señalada por diversos autores como una característica distintiva del aprendizaje profundo. Omri Barak (2017, p. 5), por ejemplo, advierte que “el aprendizaje automático nos proporciona niveles de rendimiento cada vez mayores, acompañados de un aumento paralelo de la opacidad”. En estos casos, lo que Humphreys denomina el *predicamento antropocéntrico* se vuelve especialmente relevante. “Las capacidades computacionales superan las humanas, forzando una reevaluación de cómo entendemos y evaluamos estos métodos científicos” (Humphreys, 2009, p. 617). Por supuesto, la opacidad epistémica de las EODS no es absoluta. En principio, es posible auditar los algoritmos, examinar su arquitectura y analizar sus parámetros e hiperparámetros (Rubio *et al.*, 2024). Sin embargo, este acceso técnico no elimina la opacidad relevante desde el punto de vista epistémico. La complejidad combinatoria de estos sistemas y la naturaleza distribuida de sus representaciones internas hacen que cualquier reconstrucción causal del proceso inferencial sea, en el mejor de los casos, parcial e incompleta. En consecuencia, aunque pueda conocerse *cómo* está construido el sistema, sigue siendo difícil —cuando no imposible— explicar de manera inteligible *por qué* llega a un resultado específico en un caso particular.

Esta situación conduce a una redefinición de los criterios de validez epistémica en la ciencia actual. En el contexto de las EODS, ya no ocupan un lugar central la transparencia, la comprensión causal ni la fidelidad representacional, sino la precisión empírica, la eficiencia inferencial y la utilidad predictiva. El valor epistémico de estos sistemas no se fundamenta en su capacidad para hacer inteligibles los mecanismos subyacentes, sino en la estabilidad, robustez y reproducibilidad de sus resultados frente a evidencia independiente. En este sentido, la justificación epistémica se desplaza desde criterios explicativos tradicionales hacia formas de validación orientadas al desempeño.

Esta redefinición de los criterios de validez epistémica encuentra un marco conceptual adecuado en las propuestas recientes sobre modelado modal. En este enfoque, los modelos no se evalúan principalmente por su capacidad de representar fielmente procesos causales reales, sino por su aptitud para generar información confiable acerca de lo que es posible, probable o esperable bajo determinadas condiciones contrafácticas.

Así, la justificación epistémica se apoya en la consistencia interna del modelo, en su desempeño predictivo y en la plausibilidad física de los escenarios que produce, aun cuando estos no se correspondan con una representación literal del sistema estudiado (Sjölin Wirling y Grüne-Yanoff, 2021).

Desde esta perspectiva, AlphaFold puede entenderse como un sistema que genera información modal en forma de predicciones estructurales altamente confiables más que como un modelo que representa dinámicamente el proceso físico del plegamiento proteico. Su aporte epistémico no consiste en describir cómo ocurre dicho proceso, sino en delimitar un espacio de posibilidades estructurales empíricamente plausibles, cuya validez se evalúa mediante su concordancia con datos experimentales independientes. En este sentido, la justificación epistémica de AlphaFold no depende de una explicación causal subyacente, sino de la estabilidad, precisión y utilidad predictiva de los escenarios que produce, en línea con algunas de las estrategias justificatorias propuestas para el modelado modal (Sjölin Wirling y Grüne-Yanoff, 2021).

En este marco, la representación adquiere un estatuto epistémico particular. En AlphaFold, la representación no opera principalmente como una correspondencia interna entre un modelo y un sistema físico, sino como un punto de articulación entre un proceso inferencial opaco y las capacidades cognitivas humanas. Las estructuras tridimensionales generadas por el sistema no hacen visible el mecanismo interno de inferencia, sino que traducen resultados algorítmicos en formatos que permiten su contraste empírico, su evaluación comparativa y su integración en prácticas científicas preexistentes. De este modo, la representación no se sitúa dentro del modelo, sino en la interfaz donde los *outputs* computacionales se vuelven epistémicamente utilizables.

Considerado desde esta perspectiva, AlphaFold no puede ser entendido ni como una mera extensión técnica de las simulaciones computacionales tradicionales ni como una ruptura completamente desvinculada de ellas. Por un lado, comparte con las simulaciones clásicas la ejecución algorítmica de una estructura matemática formal, la dependencia de criterios de validación empírica y el papel metodológicamente central de la representación de resultados. Por otro lado, radicaliza rasgos ya presentes —como la opacidad inferencial y la distancia respecto de teorías explicativas explícitas— hasta el punto de tensionar los marcos conceptuales heredados. En este sentido,

AlphaFold ocupa una posición fronteriza: prolonga la tradición de la simulación computacional al tiempo que revela sus límites conceptuales, preparando el terreno para una reconsideración más amplia del estatuto epistémico de los sistemas basados en aprendizaje profundo.

3. Humphreys y el marco de plantillas y modelos computacionales

En la tradición filosófica sobre modelos científicos, el valor epistémico de un modelo se ha medido, en gran parte, por su capacidad de representar fielmente algún aspecto estructural o causal del sistema estudiado. Bajo este ideal representacionista, un modelo científicamente valioso actúa como mediador que preserva —al menos parcialmente— la estructura relevante del fenómeno, permitiendo así comprenderlo a través de su forma. Como afirma Morrison (2015b, p. 390), esta perspectiva entiende que el mérito de un modelo radica en “salvar” la estructura del fenómeno objetivo, lo cual hace de la correspondencia estructural un criterio central de legitimidad epistémica.

Este ideal de correspondencia está presente también en la caracterización que Barberousse *et al.* (2009) hacen de la simulación como “una correspondencia obvia entre el proceso que opera en el dispositivo simulador y los que operan en el sistema modelado” (p. 559). En la misma línea, Hughes (1999) muestra cómo en ciertos casos —como un modelo hidráulico de una rueda real, un modelo eléctrico de un péndulo amortiguado, una simulación basada en agentes, o un autómata celular concebido como sistema físico— la relación entre procesos del modelo y procesos del sistema representado es clara y trazable.

Hughes contrasta estos ejemplos con una simulación computacional de las interacciones atómicas entre un “pine” y una placa, descrita en Rohrlich (1991). En esta última, aunque la correspondencia directa entre procesos físicos y computacionales no es evidente, se mantiene el ideal representacional. Como advierten Barberousse *et al.* (2009, p. 559), Hughes señala que: “al igual que los sistemas examinados [los tres primeros sistemas simuladores], la computadora es un sistema material, y al utilizarla para responder nuestras preguntas estamos confiando en los procesos físicos que ocurren dentro de ella”. La legitimidad epistémica de la simulación descansa así en la idea de que los procesos computacionales, por su fisicalidad, pueden fungir como sustitutos representacionales de los procesos del sistema estudiado.

En este marco, el uso de la simulación computacional se justifica en la medida en que se postula la existencia de una relación estructural

o formal que vincula tres niveles: el modelo base (teórico), el modelo implementado (o “modelo agrupado”) y la computadora programada. Norton y Suppe (2001) denominan a esta condición “realización” y se exige que los tres niveles “compartan una estructura común” (Barberousse *et al.*, 2009, p. 563). Este requisito opera como un ideal normativo de transparencia estructural, según el cual el modelo no solo debe producir resultados correctos, sino hacerlo mediante procesos que puedan ser entendidos como reflejos, transformaciones o idealizaciones explícitas de los procesos del sistema físico.

Este enfoque está íntimamente ligado a un ideal de inteligibilidad. Un modelo es científicamente valioso en la medida en que permite a los científicos entender el sistema estudiado a través de su estructura representacional. Incluso cuando el cálculo se vuelve demasiado complejo para resolverse de manera analítica, subsiste como ideal regulativo la aspiración a una solución comprensible, capaz de mostrar de manera al menos parcial cómo las propiedades del sistema se reflejan en las propiedades del modelo.

En la práctica, las simulaciones computacionales a menudo requieren aproximaciones no analíticas: el modelo matemático de partida no es directamente implementable en un computador digital y, por tanto, debe sufrir transformaciones, discretizaciones y ajustes que modifican su estructura formal. Esto crea lo que podríamos llamar una brecha de implementación estructural entre el modelo teórico ideal y el artefacto computacional efectivo, una brecha que no constituye una anomalía ocasional, sino una característica constitutiva del modelado computacional.

Esta brecha pone en tensión el ideal representacionalista clásico, pues implica que la validez epistémica de una simulación no depende únicamente de su fidelidad estructural a un modelo base, sino también de la robustez de los procedimientos numéricos, de la calidad de los datos de entrada y de las estrategias de verificación y validación (V&V) empleadas. En consecuencia, la noción misma de “solución científica” se reconfigura (Lenhard, 2016): deja de identificarse exclusivamente con una derivación analítica exacta y pasa a entenderse como una solución operativamente adecuada, cuya legitimidad se mide por su capacidad de producir resultados eficaces para la investigación, incluso cuando los procesos internos permanecen opacos o irreductiblemente complejos.

Esta reconfiguración es una consecuencia directa de la naturaleza de los problemas que abordan las EODS. La opacidad no es un velo

accidental, sino una característica constitutiva del método. Estos modelos operan en espacios de parámetros de altísima dimensionalidad (millones o miles de millones de dimensiones), donde la tarea consiste en aprender una función no lineal que mapea un *input* complejo (e. g., una secuencia de aminoácidos) a un *output* igualmente complejo (e. g., coordenadas atómicas en 3D). La solución que encuentra el algoritmo de optimización —como el descenso de gradiente estocástico (SGD) o el backpropagation (BP)— puede caracterizarse como una región funcionalmente estable en este espacio de alta dimensionalidad. Dicha solución es en gran medida irreductible a una secuencia de pasos lógicos o a una ecuación directamente comprensible para la cognición humana. En este sentido, el modelo no “deduce” la estructura en un sentido clásico, sino que la reconstruye inferencialmente mediante una representación distribuida y correlacional aprendida de los datos. Así, la explicación causal del plegamiento no es eliminada, pero queda desplazada por una inferencia paralela que se implementa en la arquitectura computacional del sistema.

Ahora bien, aquí emerge el dilema central de nuestro trabajo. Para algunos autores, este desplazamiento no exige abandonar la noción de simulación, sino reformularla o ampliarla; lo decisivo ya no sería la correspondencia representacional estricta, sino la resolución algorítmica de estructuras matemáticas (una lectura continuista en sentido amplio). Sin embargo, puede sostenerse también —y AlphaFold lo ilustra con particular fuerza— que estamos ante una transformación más profunda: un nuevo tipo de artefacto científico —las EODS— cuya opacidad y fertilidad epistémica tensionan seriamente las categorías heredadas (una lectura disruptiva, aunque no necesariamente rupturista). Frente a este dilema, la epistemología de los modelos modales cobra un papel relevante, en la medida en que ofrece estrategias conceptuales para justificar los *outputs* de modelos que no son tradicionalmente representacionales sin por ello desligarlos por completo de la práctica científica establecida (Sjölin Wirling y Grüne-Yanoff, 2021).

Este desplazamiento se acentúa en las simulaciones contemporáneas basadas en redes neuronales profundas. Aquí, la carga heurística y el trabajo de ajuste se redistribuyen; ya no recaen exclusivamente en el investigador humano, sino que son parcialmente delegados a la infraestructura computacional, entendida como arquitecturas con millones de parámetros optimizadas mediante BP o SGD. En estos casos, incluso la noción de “correspondencia estructural” entre

modelo y sistema físico se vuelve problemática, pues el proceso interno del modelo ya no refleja, ni siquiera de manera idealizada, el comportamiento causal del fenómeno que estudia, sino regularidades estadísticas funcionalmente eficaces.

Ahora bien, el hecho de que AlphaFold opere sobre arquitecturas formales y ejecute algoritmos matemáticos no es, por sí solo, suficiente para considerarlo una simulación computacional en sentido filosófico. Un criterio basado únicamente en la ejecución algorítmica o en la existencia de una estructura matemática subyacente sería excesivamente laxo, pues llevaría a clasificar como “simulaciones” a artefactos puramente computacionales, como una calculadora o un algoritmo de ordenamiento, trivializando así la noción de “simulación científica”. Lo que distingue a AlphaFold no es simplemente que compute, sino el rol epistémico que desempeña dentro de una práctica científica específica: explora un espacio de configuraciones físicamente posibles, produce información modal relevante sobre sistemas del mundo, y sus *outputs* son sometidos a procedimientos sistemáticos de verificación y validación empírica independientes. En este sentido, AlphaFold puede entenderse como una simulación no por su ontología computacional, sino por su función epistémica, en tanto sustituye, complementa y orienta prácticas experimentales en biología estructural.

Desde esta perspectiva, AlphaFold prolonga la noción de “simulación”, pero solo bajo una concepción ampliada y no trivial de esta; esto es, no como mera ejecución de una estructura formal, sino como un artefacto computacional integrado en un régimen de producción, validación y uso del conocimiento científico, donde la estructura matemática subyacente es condición necesaria, pero no suficiente, para su estatuto epistémico.

4. Opacidad, representación y visualización en las simulaciones contemporáneas

En la práctica científica, las V&V constituyen procedimientos clave para evaluar la fiabilidad de las simulaciones. La verificación busca asegurar que el modelo teórico ha sido implementado correctamente en el código, mientras que la validación consiste en contrastar los resultados de la simulación con datos experimentales u observacionales (Morrison, 2015b). Sin embargo, en sistemas complejos, esta validación se ve limitada por la escasez o inaccesibilidad de datos

empíricos, precisamente la razón que motiva el recurso a simulaciones. Como señalan Morrison (2015a y 2015b) y Winsberg y Alvarado (2022), esta tensión ha impulsado el desarrollo de métricas y protocolos de validación que no se reducen a la comparación directa con datos, sino que se configuran en estrecha dependencia de la propia simulación, poniendo en cuestión la idea de un criterio de validación completamente externo e independiente del artefacto computacional.

Este escenario difumina los límites entre simulación y experimentación convencional. La simulación ya no se presenta solo como una herramienta auxiliar de la experimentación, sino como una parte constitutiva del proceso experimental en sí mismo. Así, el valor epistémico de una simulación no depende exclusivamente de su ajuste a un modelo teórico previo, sino también de su capacidad para integrarse en un ciclo iterativo de predicción, prueba y ajuste que involucra tanto al modelo computacional como a los dispositivos experimentales, desplazando el foco desde la representación fiel de un fenómeno hacia la eficacia operativa del conjunto modelo-simulación-experimento.

Un ejemplo representativo de esta integración lo ofrece el Gran Colisionador de Hadrones (LHC). En este caso, la simulación no es un simple complemento, sino la fundación misma de los experimentos (Morrison, 2015a, p. 288). Las simulaciones son indispensables para interpretar datos, diseñar detectores y seleccionar eventos relevantes, aunque estas simulaciones permanecen ancladas en modelos teóricos explícitos de la física de partículas. De hecho, en palabras de Morrison, “nada se sabe sin simulación” (2015a, p. 288).

Esto implica que los datos denominados “reales” y los generados por simulaciones se procesan de forma conjunta, lo que hace “prácticamente sin sentido cualquier distinción tajante entre simulación y ‘experimento’” en ciertos contextos altamente mediados computacionalmente (Morrison, 2015a, p. 289). La validación, en este marco, se convierte en un proceso continuo e integrado, en el que las simulaciones no solo contrastan hipótesis, sino que participan activamente en la definición de los experimentos y en la interpretación de sus resultados. Esta simbiosis transforma la simulación en una nueva forma de experimentación sin eliminar por completo las diferencias conceptuales entre ambos dominios.

Se consolida así la práctica de la experimentación *in silico*, entendida no como un sustituto de la experimentación empírica tradicional, sino como un dispositivo heurístico que estructura y restringe el espacio

de hipótesis científicas. En el caso de AlphaFold, un biólogo puede obtener una estructura tridimensional plausible para una proteína de función desconocida. Esta estructura no constituye una explicación del proceso de plegamiento, pero sí una hipótesis geométrica altamente informada, capaz de revelar sitios de unión potenciales o similitudes estructurales con otras proteínas. A partir de esta conjetura generada computacionalmente, es posible diseñar experimentos de laboratorio (*wet lab*) más enfocados y eficientes —por ejemplo, mediante mutagénesis dirigida— orientados a evaluar la función inferida.

De este modo, las EODS no solo validan o predicen resultados, sino que reconfiguran ciertos aspectos de la lógica del descubrimiento científico. En lugar de una exploración empírica relativamente abierta, se generan conjeturas altamente restringidas mediante procedimientos computacionales, que luego se someten a corroboración o falsación experimental. El modelo funciona, así, como un dispositivo generativo de conjeturas más que como una representación del fenómeno subyacente. Un caso paradigmático de esta dinámica —aunque no basado en aprendizaje profundo— puede encontrarse en la búsqueda del bosón de Higgs, donde las simulaciones desempeñaron un papel decisivo en la interpretación de los datos y en el diseño experimental. En estos contextos, los llamados “datos experimentales” y los “datos de simulación” se entrelazan de tal modo que resulta difícil sostener una jerarquía epistémica tajante entre ambos.

El caso de AlphaFold encarna con particular claridad este nuevo régimen de validación en las simulaciones actuales. Dada la opacidad interna de su funcionamiento, la posibilidad de una verificación en sentido clásico —esto es, el seguimiento y control inferencial de los procesos computacionales que conducen a los resultados— se ve drásticamente limitada. En consecuencia, su integración en la práctica científica se apoya fundamentalmente en la corroboración empírica sistemática de sus *outputs* más que en la comprensión de los procesos que los generan. Su legitimidad epistémica no proviene de la inteligibilidad de su arquitectura interna ni de una correspondencia representacional con el proceso de plegamiento proteico, sino de la concordancia estable y reproducible entre sus predicciones y las estructuras proteicas determinadas experimentalmente mediante cristalografía de rayos X o criomicroscopía electrónica.

En contextos donde la validación directa resulta difícil o imposible, AlphaFold es evaluado mediante métricas especializadas y a través de

competencias comunitarias como CASP (*Critical Assessment of Protein Structure Prediction*). En estos escenarios, las predicciones se comparan con resultados experimentales aún no publicados, lo que permite evaluar su desempeño predictivo de manera controlada e independiente. Los éxitos reiterados de AlphaFold en estas evaluaciones han mostrado niveles de exactitud comparables a los métodos experimentales (Jumper *et al.*, 2021), lo que consolida su fiabilidad como herramienta predictiva, sin que ello implique una validación del proceso inferencial interno ni una caracterización representacional de su funcionamiento.

La discusión en torno a la validación nos devuelve al punto de partida: la necesidad de contar con categorías filosóficas capaces de dar cuenta de prácticas científicas que ya no se ajustan plenamente a los marcos tradicionales. En este sentido, la propuesta de Humphreys resulta especialmente fecunda como punto de transición. Al concebir las simulaciones como unidades epistémicas específicas, permite desplazar la atención desde la reducción a teorías o modelos clásicos hacia la articulación funcional de plantillas, supuestos, procesos computacionales y formas de representación.

Sin embargo, como el caso de AlphaFold pone de manifiesto, incluso este marco requiere ser revisado en un punto decisivo: la representacionalidad ya no puede seguir operando como una propiedad definitoria de las simulaciones computacionales. En EODS, la validez epistémica se apoya menos en la correspondencia representacional con el fenómeno *target* que en regímenes de validación empírica y en la capacidad de integrar outputs confiables en la práctica científica, lo que obliga a repensar el estatuto mismo de la representación dentro de la simulación.

De este modo, AlphaFold puede entenderse, en un primer registro, como la manifestación paradigmática de una nueva clase de artefactos científicos (EODS): opacos, no representacionales y, sin embargo, epistémicamente fértiles. Alternativamente —y de manera más controvertida— podría sostenerse que estos sistemas radicalizan la noción misma de “simulación”, siempre que esta se redefina en un sentido estrictamente formal-computacional, desvinculado del ideal clásico de correspondencia estructural con un fenómeno *target*. En este segundo registro, lo decisivo ya no sería la representación del proceso físico, sino la ejecución algorítmica de estructuras matemáticas capaces de producir *outputs* empíricamente validados. No obstante, esta ampliación conceptual no es inocua y exige precisiones filosóficas sustantivas, precisamente para

evitar que toda computación exitosa sea trivialmente clasificada como simulación. Ambas lecturas convergen, así, en un mismo diagnóstico: la filosofía de la ciencia contemporánea debe ampliar —pero también afinar— su marco conceptual si pretende comprender cómo se produce y se justifica el conocimiento en la era de los sistemas opacos basados en inteligencia artificial.

Comentarios de cierre

El recorrido realizado en este trabajo ha mostrado que la discusión filosófica sobre las simulaciones computacionales no puede permanecer inalterada ante la irrupción de artefactos como AlphaFold. Las posiciones afirmativas (Winsberg, Galison, Rohrlich, Humphreys, Durán) han defendido, con distintos matices, que las simulaciones constituyen una metodología epistémicamente significativa y, en ciertos sentidos, novedosa (Rohrlich, 1990; Winsberg, 2001). Por su parte, las posturas más escépticas (Frigg y Reiss) han sostenido que estas prácticas pueden comprenderse adecuadamente dentro de los marcos tradicionales de la filosofía de los modelos y la experimentación científica.

El análisis desarrollado sugiere, sin embargo, que la emergencia de sistemas estadísticos de gran escala como las *Epistemically Opaque Deep-learning Simulations* (EODS) introduce tensiones que no pueden resolverse sin ajustes conceptuales. AlphaFold no reconstruye el proceso físico del plegamiento proteico en sentido dinámico o causal; en su lugar, aprende regularidades complejas en espacios de alta dimensionalidad y produce resultados cuya fiabilidad depende de su validación empírica y de su integración efectiva en la práctica científica. Esto plantea un desafío directo a las concepciones representacionistas clásicas, sin que ello implique, automáticamente, la necesidad de abandonar la noción de “simulación”.

En efecto, si bien podría sostenerse que AlphaFold prolonga ciertos rasgos de la simulación en un sentido formal-computacional —en la medida en que ejecuta estructuras matemáticas precisas mediante procedimientos algorítmicos—, este rasgo, por sí solo, no es suficiente para justificar su clasificación como simulación computacional. De lo contrario, cualquier artefacto computacional exitoso podría ser trivialmente subsumido bajo dicha categoría. El problema filosófico relevante no consiste, entonces, en identificar la presencia de cálculo o formalismo, sino en determinar qué condiciones epistémicas y metodológicas permiten distinguir entre herramientas de cómputo y

artefactos capaces de generar conocimiento científicamente significativo sobre sistemas del mundo.

En este contexto, la noción de “representación” que recupera Humphreys adquiere un papel central. La representación ya no se entiende como una correspondencia interna entre modelo y fenómeno, sino como una interfaz funcional que conecta procesos computacionales opacos con las capacidades cognitivas y prácticas de los científicos. Las visualizaciones tridimensionales producidas por AlphaFold no son meros recursos ilustrativos, sino componentes constitutivos de su uso epistémico: es a través de ellas que los resultados se interpretan, se validan y se integran en programas de investigación más amplios.

Los procedimientos de verificación y validación refuerzan este desplazamiento. Al igual que en otros contextos de experimentación altamente mediada por simulaciones —como el Large Hadron Collider, donde “nada se sabe sin simulación” (Morrison, 2015a, p. 288)—, la fiabilidad de AlphaFold se sostiene en ciclos iterativos de contrastación empírica, evaluación comunitaria y métricas especializadas. La justificación epistémica no descansa aquí en la transparencia de los mecanismos internos, sino en la robustez de los resultados y en su capacidad para integrarse productivamente en prácticas experimentales posteriores.

Este escenario invita a repensar la categoría de modelo más allá de los moldes representacionalistas tradicionales. AlphaFold ilustra un régimen epistémico en el que es posible generar conocimiento científicamente valioso sin una comprensión causal inmediata del fenómeno estudiado, pero de manera sistemática, evaluable y socialmente controlada. Comprender este régimen exige reconocer que la inteligibilidad no siempre es condición necesaria de la justificación y que la representación puede concebirse —siguiendo a Humphreys— como una mediación funcional entre cálculo automatizado y uso humano. En este sentido, el análisis converge con los objetivos del modelado modal, orientado a justificar cómo los modelos pueden proveer información fiable sobre lo posible o lo contrafactual incluso en ausencia de teorías causales completas (Sjölin Wirling y Grüne-Yanoff, 2021).

En definitiva, AlphaFold no debe entenderse simplemente como un logro computacional excepcional ni como una ruptura absoluta con las prácticas científicas previas. Más bien constituye un caso límite que obliga a refinar las categorías con las que la filosofía de la ciencia analiza los modelos, las simulaciones y la justificación epistémica.

Enfrentar este desafío conceptual es una tarea central para comprender cómo se produce y se valida el conocimiento científico en un contexto crecientemente mediado por sistemas opacos de inteligencia artificial

Referencias

- Barak, O. (2017). Recurrent Neural Networks as Versatile Tools of Neuroscience Research. *Curr Opin Neurobiol*, 46, 1-6. <https://doi.org/10.1016/j.conb.2017.06.003>
- Barberousse, A., Franceschelli, S. y Imbert, C. (2009). Computer Simulations as Experiments. *Synthese*, 169(3), 557-574. <https://doi.org/10.1007/s11229-008-9430-7>
- Chirimuuta, M. (2021). Prediction Versus Understanding in Computationally Enhanced Neuroscience. *Synthesis*, 199, 767-790. <https://doi.org/10.1007/s11229-020-02713-0>
- Daston, L. y Galison, P. (2007). *Objectivity*. Zone Books.
- Durán, J. M. (2017). Simulaciones computacionales. *Principia*, 21(1), 125-140. <https://doi.org/10.5007/1808-1711.2017v21n1p125>
- Frigg, R. y Reiss, J. (2009). The Philosophy of Simulation: Hot New Issues or Same Old Stew? *Synthese*, 169, 593-613. <https://doi.org/10.1007/s11229-008-9438-z>
- Galison, P. (1996). Computer Simulations and the Trading Zone. En P. Galison y D. J. Stump (eds.), *The Disunity of Science: Boundaries, Context and Power* (pp. 118-157). Stanford University Press.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. y Dahl, G. E. (2017). Neural Message Passing for Quantum Chemistry. *Proceedings of the 34th International Conference on Machine Learning*, 70, 1263-1272. <https://proceedings.mlr.press/v70/gilmer17a.html>
- Gao, P. y Ganguli, S. (2015). On Simplicity and Complexity in the Brave New World of Large-Scale Neuroscience. *Curr Opin Neurobiol*, 32, 148-155. <https://doi.org/10.1016/j.conb.2015.04.003>
- Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Deep Learning for Computational Chemistry. *Journal of Computational Chemistry*, 38(16), 1291-1307. <https://doi.org/10.1002/jcc.24764>
- Hughes, R.I.G. (1999). The Ising Model, Computer Simulation, and Universal Physics. En M. Morgan y M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 97-145). Cambridge University Press. <https://doi.org/10.1017/CBO9780511660108.006>

- Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford University Press. <https://doi.org/10.1093/0195158709.001.0001>
- Humphreys, P. (2009). The Philosophical Novelty of Computer Simulation Methods. *Synthese*, 169, 615-626. <https://doi.org/10.1007/s11229-008-9435-2>
- Jensen, F. (2017). *Introduction to Computational Chemistry*. Wiley. <https://doi.org/10.14293/S2199-1006.1.SOR-CHEM.CLYWQK5.v1>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... y Hassabis, D. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596, 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Lenhard, J. (2016). Computer Simulation. En P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science* (pp. 710-736). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199368815.013.42>
- Morrison, M., (2012). Models, Measurement and Computer Simulation: The Changing Face of Experimentation. *Philosophical Studies*, 143, 33-57. <https://doi.org/10.1007/s11098-008-9317-y>
- Morrison, M. (2015a). Legitimizing Simulation: Methodological Issues of Verification and Validation. En M. Morrison, *Reconstructing Reality: Models, Mathematics, and Simulations* (pp. 248- 286). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199380275.003.0008>
- Morrison, M. (2015b). Without It There's Nothing: The Necessity of Simulation in the Higgs Search. En M. Morrison, *Reconstructing Reality: Models, Mathematics, and Simulations* (pp. 287-316). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199380275.003.0009>
- Norton, S. y Suppe, F. (2001). Why Atmospheric Modeling Is Good Science. En P. Edwards y C. Miller (eds.), *Changing the Atmosphere* (pp. 67-106). The MIT Press. <https://doi.org/10.7551/mitpress/1789.003.0006>
- Particle Physics & Astronomy Research Council (2005, 4 de junio). Millennium Simulation—The Largest Ever Model of the Universe. *ScienceDaily*. www.sciencedaily.com/releases/2005/06/050604061156.htm
- Rohrlich, F. (1990). Computer Simulation in the Physical Sciences. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 507-518. <https://doi.org/10.1086/psaprocbienmeetp.1990.2.193094>
- Rubio, E., Giri, L. y Ilcic, A. (2023). Desafíos epistemológicos en la era de las redes neuronales artificiales: abordando sistemas complejos

- desde una perspectiva computacional. *Argumentos de Razón Técnica*, 26, 145-178. <https://doi.org/10.12795/Argumentos/2023.i26.06>
- Sjölin Wirling, Y. y Grüne-Yanoff, T. (2021). The Epistemology of Modal Modeling *Philosophy Compass*, 16(10), e12775. <https://doi.org/10.1111/phc3.12775>
- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., ... y Collins, J. J. (2020). A Deep Learning Approach to Antibiotic Discovery. *Cell*, 180(4), 688-702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. y Polosukhin, I. (2017). Attention Is All You Need. En U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus (eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 6000-6010). Curran Associates Inc. <https://doi.org/10.48550/arXiv.1706.03762>
- Vicsek, T. (2002). The Bigger Picture. *Nature*, 418(6894), 131. <https://doi.org/10.1038/418131a>
- Winsberg, E. (2001). Simulations, Models, and Theories: Complex Physical Systems and their Representations. *Philosophy of Science*, 68, S442-S454. <https://doi.org/10.1086/392927>
- Winsberg, E. (2010). *Science in the Age of Computer Simulation*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226902050.001.0001>
- Winsberg, E. y Alvarado, R. (2022). Computer Simulations in Science. *The Stanford Encyclopedia of Philosophy*. E. N. Zalta y U. Nodelman (eds.). <https://plato.stanford.edu/archives/spr2026/entries/simulations-science/>
- Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H. Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K. y Barzilay, R. (2019). Analyzing Learned Molecular Representations for Property Prediction. *Journal of Chemical Information and Modeling*, 59(8), 3370-3388. <https://doi.org/10.1021/acs.jcim.9b00237>

<https://doi.org/10.21555/top.v760.3390>

Análisis de algunas críticas contemporáneas en torno a la eventual propuesta política de Alasdair MacIntyre

Analysis of Some Contemporary Criticisms of Alasdair MacIntyre's Eventual Political Proposal

Alberto Díaz Rivero

Universidad Gabriela Mistral

Chile

alberto.diaz@ugm.cl

<https://orcid.org/0000-0002-0934-697>

Sebastián Buzeta Undurraga

Universidad Gabriela Mistral

Chile

Universidad de los Andes

Chile

sebastian.buzeta@ugm.cl

<https://orcid.org/0000-0001-9166-0490>

Recibido: 26 - 02 - 2025.

Aceptado: 03 - 04 - 2025.

Publicado en línea: 10 - 07 - 2026.

Cómo citar este artículo: Díaz Rivero, A. y Buzeta Undurraga, S. (2026). Análisis de algunas críticas contemporáneas en torno a la eventual propuesta política de Alasdair MacIntyre. *Tópicos, Revista de Filosofía*, 76, 315-343. <https://doi.org/10.21555/top.v760.3390>



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

Resumen

La filosofía política de Alasdair MacIntyre ha mostrado ser sumamente influyente en los debates actuales sobre las instituciones políticas modernas y su relación con la ética, pero también ha sido foco de numerosas críticas, provenientes de diversas corrientes de pensamiento. De ellas, las más documentadas apuntan a la falta de explicitación, claridad y viabilidad en su propuesta política. En este artículo analizamos tres de estas críticas, expresadas por Perreau-Saussine, Pierre Manent, Keith Breen, Ronald Beiner y Daniel Edward Young. El primer objetivo, por tanto, es exponer y destacar brevemente los aspectos principales de la teoría política macintyreana, mientras que el segundo es analizar las críticas de dichos autores. Finalmente, exploramos la existencia de una eventual propuesta política viable desde la teoría política de MacIntyre incorporando en el análisis, a su vez, aquellas críticas que parezca que tengan asidero.

Palabras clave: Alasdair MacIntyre; propuesta política; liberalismo; Estado moderno; neutralidad estatal; teoría política; comunidades locales; modernidad; comunitarismo; aristotelismo.

Abstract

Alasdair MacIntyre's political philosophy has proven to be highly influential in current debates concerning modern political institutions and their relationship to ethics. However, it has also been the focus of numerous criticisms from various schools of thought. Of these, the most documented ones point to a lack of explicitness, clarity, and viability throughout his political proposal. This paper analyzes three of these criticisms, expressed by Perreau-Saussine, Pierre Manent, Keith Breen, Ronald Beiner, and Daniel Edward Young. Firstly, it presents and briefly highlights the main aspects of MacIntyrean political theory. It then analyzes the criticisms of these authors. Finally, it explores the possible existence of a viable political proposal based on MacIntyre's political theory while also incorporating into the analysis those criticisms that seem to be well-founded.

Keywords: Alasdair MacIntyre; political proposal; liberalism; modern state; state neutrality; political theory; local communities; modernity; communitarianism; Aristotelianism.

Introducción

En el presente artículo analizaremos una crítica contemporánea a la teoría política de Alasdair MacIntyre en la que se afirma la inexistencia de una propuesta institucional clara por su parte, principalmente a partir de las afirmaciones de Perreau-Saussine, Pierre Manent, Ronald Beiner, Daniel Edward Young y Keith Breen. Dada la variedad de argumentos para sostener dicha crítica, nos ocuparemos específicamente de las que, a nuestro juicio, pueden comprometer gran parte del pensamiento político de MacIntyre: en primer lugar, que la eventual inexistencia de una propuesta política ocurre porque MacIntyre se distancia de la política aristotélica al focalizarse más en la ética; la segunda, que esta aparente imposibilidad de plantear un sistema político viable nace de su adscripción no reconocida al comunitarismo; y, finalmente, que una política macintyreana sería inviable por una contradicción central, a saber, que, según estos críticos, MacIntyre pareciera fomentar los mismos males que critica del liberalismo, lo que impediría establecer una propuesta política desde bases antiliberales. Debido a la relevancia de MacIntyre, urge mostrar la actualidad de esta discusión y la importancia que tiene en sí misma, al menos como objetivo general de este trabajo.

Así, para lograr el objetivo propuesto, a saber, revelar eventualmente la existencia de una propuesta política institucional macintyreana y su correspondiente aplicación concreta, dividiremos el texto en dos partes: primero, examinaremos las ideas principales de su teoría política para, posteriormente, exponer las críticas de los comentaristas a dicha teoría, que sostienen la inexistencia de una propuesta política en su pensamiento, realizando al mismo tiempo un análisis de estas.

1. Elementos fundamentales de la teoría política macintyreana

La teoría política de MacIntyre se inserta en una crítica general a los aspectos más importantes de la modernidad, por lo que no puede ser expuesta en toda su extensión en el presente trabajo. No obstante, es importante que demos cuenta de las características que posteriormente serán criticadas por los autores que mencionamos más arriba, de modo que podamos analizar sus postulados. Así, será necesario dividir la teoría política de MacIntyre en tres partes. En primer lugar, abordaremos de manera general su crítica a la modernidad; luego, cuál es su visión

específica acerca de la institución política por excelencia, el Estado; y, finalmente, expondremos la eventual propuesta política que se pueda hallar en estos postulados.

1.1. Crítica a la modernidad

Alasdair MacIntyre ha elaborado una teoría centrada en los aspectos éticos y políticos de la modernidad influenciado por el famoso artículo “Modern Moral Philosophy” de Elizabeth Anscombe (1958) en el que se da cuenta de la fragmentación del uso del lenguaje moral en nuestra época, en la que ha perdido su sentido original.¹ Con este fin, ha logrado mantener discusiones con los más variados filósofos contemporáneos, desde Hans-Georg Gadamer hasta John Rawls, lo que demuestra, además, una gran erudición en el estudio y problematización de los textos de la filosofía clásica y medieval. Estos últimos han servido a MacIntyre (1981, pp. 170 y 245) de materia prima para suscribir un tipo de neoaristotelismo que fomenta una política de comunidades locales, y una ética centrada en la búsqueda del florecimiento humano por medio de la virtud y la excelencia. A grandes rasgos, toda la teoría macintyreana se propone como un neoaristotelismo tomasiano que informa su trabajo no solo de una crítica política y ética, sino también económica de nuestra época.²

Al mismo tiempo, es importante desmentir, desde un principio, que la visión macintyreana sobre la modernidad sea simplemente pesimista. En *After Virtue* expone su unidad intelectual y la intención de dar respuesta a los problemas que se planteaban en ese entonces;³ ahí

¹ Al final de su artículo “Hume on ‘Is’ and ‘Ought’”, MacIntyre (1959) reconoce la importancia del artículo de Anscombe para comprender gran parte de las dificultades del lenguaje moral en la modernidad.

² Para una detallada distinción entre el neoaristotelismo macintyreano y otros, cfr. Knight (2007, pp. 91-106) y Beadle y Moore (2011, p. 87).

³ MacIntyre (1991, p. 67) afirmó, en su entrevista a la revista *Cogito*, que previo a 1971 sus trabajos habían estado dispersos en cuanto a su contenido intelectual y que les faltaba unidad. Sin embargo, quienes han leído de cerca su biografía intelectual coinciden en que, desde los años cuarenta, cuando era *lecturer* de filosofía de la religión, hasta su definitivo quiebre con el marxismo hacia 1968, la teoría macintyreana gozó de una contundente fuerza. La única objeción que puede haber acerca del poder o de la unidad en esta etapa inicial de su pensamiento es que se trató de una carrera más crítica que propositiva y que “[i]f his early work raised important questions, his later work provides important

une gran parte de esa crítica a la ética y la política “liberal” propia de la modernidad con un proyecto en los términos que mencionamos, con el que pretende responder como una alternativa a este esquema.

Ya sea desde el agustinismo o del aristotelismo tomasiano, MacIntyre ha desarrollado un trabajo filosófico que aúna sus críticas a la modernidad con una propuesta de florecimiento humano basada en el fortalecimiento de la excelencia, por medio de la virtud, *en* la época actual. A lo largo de *After Virtue*, se destaca una continua comparación de los desastres de la modernidad con la Antigüedad, como puede verse en la famosa cita hacia el final del texto:

Si la tradición de las virtudes ha podido sobrevivir a los horrores de los últimos tiempos oscuros, no estamos completamente sin bases para la esperanza. En este tiempo, sin embargo, los bárbaros no están esperando tras las fronteras; ellos han estado ya gobernándonos por un tiempo ya. [...] Estamos esperando no por un Godot, sino por otro —sin duda muy diferente— San Benito (1981, p. 245).⁴

En consideración de este “lado pesimista” de MacIntyre frente a la modernidad y la esperanza optimista por una alternativa a sus males, se hace evidente la tentación de interpretarlo como un filósofo nostálgico de las sociedades antiguas, tal como lo hace notar Terry Pinkard.⁵ Sin embargo, del mismo modo en que ocurrió con el advenimiento del debate sobre la opción benedictina,⁶ los constantes paralelos con la

answers” (Knight, 1998, p. 1) Sin embargo, es completamente aceptable concebir que esos problemas planteados en su primera etapa sí estaban unificados en tanto aparato crítico y, por lo tanto, afirmar que sí estaban ordenados. Para esta última afirmación, cfr. Gregson (2019, p. 2), Blackledge y Davidson (2008, p. xvi) y Beadle y Moore (2020, p. 2). Esto se puede reafirmar con lo expuesto previamente acerca de su interés por dotar al cristianismo de una crítica a la modernidad desde los postulados del “joven Marx”.

⁴ Las traducciones de los textos en inglés son propias.

⁵ “To understand MacIntyre’s dissatisfaction with modernity, one therefore has to look in places other than the Romantic nostalgia for a nonexistent, unified and harmonious past in order to find the other sources of his distaste for modernity, in particular the rather politically conservative Max Weber and the non-Marxist socialist Karl Polanyi” (Pinkard, 2003, p. 177).

⁶ Para una guía completa sobre este debate, cfr. Thompson (2021).

Antigüedad forman parte de un modo de análisis necesario para dar cuenta del carácter efímero que tienen las formas institucionales en las que se vive.

Así, para MacIntyre, la normalidad moderna no es *natural*,⁷ tanto en cuanto a la asociatividad humana ni mucho menos en cuanto al individuo. Las concepciones del individuo y de la comunidad modernas constituyen un “descalabro” (*melange*)⁸ del lenguaje moral moderno, que no solo indica un estado de desorden, sino que permite dar cuenta de una realidad ética pretérita (cfr. MacIntyre, 1981, p. 10), previa a la Ilustración. Asimismo, el carácter fragmentario es evidente por el abandono de los esquemas de pensamiento aristotélico, debido a lo cual todo el aparato lingüístico moral se simplificó y perdió sus distinciones necesarias a la hora de responder a las problemáticas éticas de la comunidad (cfr. MacIntyre, 1981, p. 150).⁹ Esto no significa que las nociones actuales de “individuo” y “comunidad” no tengan un sentido *en sí*, sino que han perdido la relación que tenían con un contexto moral que las dotaba de inteligibilidad. Por eso es que MacIntyre considera necesario dar cuenta de la situación actual *desde* el prisma antiguo a partir de su preocupación en torno a la falta de claridad de las primeras premisas de los razonamientos éticos de la modernidad.

⁷ Cuando nos referimos a “lo natural” en MacIntyre, es importante tener en cuenta su adscripción al aristotelismo tomasiano, que comprende lo natural como lo que se identifica con la naturaleza humana. El estado actual del vocabulario moral y las concepciones de individuo y comunidad modernas atentan contra la naturaleza humana en la medida en que desestructuran al agente en diversas esferas sociales, dejando en el olvido cuestiones fundamentales como el sentido de la vida y otras preguntas “aristotélicas” necesarias para la plenitud humana.

⁸ Cfr. MacIntyre (1981, p. 210). El autor lo llama de este modo para dar cuenta del nivel de desorden del lenguaje que se emplea en términos morales.

⁹ MacIntyre da cuenta de cómo esto generó la confusión en torno al término “virtud” que expresan las obras de Lutero y Hobbes: “When eighteen or nineteen hundred years after Aristotle the modern world came systematically to repudiate the classical view of human nature—and with it in the end a great deal that had been central to morality—it did repudiate it very precisely as Aristotelianism. ‘That buffoon who has misled the church’ said Luther of Aristotle, setting the tone; and when Hobbes explained the Reformation he saw it as partly due to ‘the fayling of Venue in the Pastors’, but partly ‘from bringing of the Philosophy, and doctrine of Aristotle into Religion’ (*Leviathan*, I, 12)” (MacIntyre, 1981, p. 154).

En definitiva, para MacIntyre, la modernidad tiene una serie de características que la hacen una época peculiar: un descalabro moral y una falta de asidero objetivo para dirimir en cuestiones éticas han marcado profundamente el *ethos* occidental con sus consecuentes correlatos de orden político, resumidos en la falta de poder autoritativo en relación con el Estado moderno.

1.2. Crítica y propuesta política en MacIntyre

MacIntyre sostiene que toda narración que se realice de la modernidad debe considerar el marco económico-político que ha llevado a las grandes desigualdades de las que hemos sido testigos en nuestra historia, a pesar de aceptar los logros y avances de la época (cfr. MacIntyre, 2016, p. 124). No podemos entender lo moderno sin ese marco y, sin embargo, ha de ser “trascendido”,¹⁰ cuestión que MacIntyre ha trabajado desde sus primeros escritos sobre las tesis de Feuerbach en K. Marx.¹¹ Para ello, es imprescindible buscar modos de investigación filosófica antagónicos a ese mismo marco moderno (cfr. MacIntyre,

¹⁰ Cfr. la terminología utilizada en MacIntyre (1988, p. 7): “It will be important in trying to answer this question not to trap ourselves by, perhaps inadvertently, continuing to accept the standards of the Enlightenment. We already have the best of reasons for supposing that those standards cannot be met, and we know in advance, therefore, that from the standpoint of the Enlightenment and its successors any account of an alternative mode of understanding will inescapably be treated as one more contending view, unable to vindicate itself conclusively against its Enlightenment rivals”.

¹¹ Frente a la distancia que comenzó a tomar Marx con Hegel y Feuerbach en torno a la relación entre teoría y práctica, el primero (el “joven Marx”) consideró que no es “desde dentro” de la teoría que es posible revertir los problemas sociales en general, sino desde la práctica. MacIntyre se hace parte de esta posición marxiana y condena a los marxistas, que han sido más hegelianos que apegados a Marx: “For Marx had identified and found reason to reject the hierarchical structure of that type of reformist theory and practice. Those who without abandoning the standpoint of civil society take themselves to know in advance what needs to be done to effect needed change are those who take themselves to be therefore entitled to manage that change. Others are to be the passive recipients of what they as managers effect” (MacIntyre, 1994, p. 286). Se ve cómo MacIntyre propone trascender la filosofía política y moral de la modernidad, en tanto esta solo representa una serie de intentos de cambio “desde dentro”.

1988, p. 7). Evidentemente, con este fin se debe dar expresión a estos tipos de frustraciones y desencuentros, en términos modernos, desde la modernidad (cfr. MacIntyre, 1998, p. 252), tal como puede apreciarse en sus palabras:

El liberalismo, mientras impone a través del poder estatal regímenes que declaran a todos libres de perseguir lo que sea que consideren su propio bien, priva a la mayoría de la posibilidad de entender sus vidas como una búsqueda para el descubrimiento y el alcance del bien, especialmente del modo en que intenta desacreditar aquellas formas tradicionales de comunidad humana dentro de las cuales este proyecto tiene que ser encarnado (MacIntyre, en Borradori, 1994, p. 143).

En efecto, es desde un determinado tipo de asociaciones que es posible encontrar salida al descalabro moral moderno. Solo en la comunidad la comprensión compartida del bien común otorga inteligibilidad a las nociones (hoy inconexas) de la moral y la virtud.¹² Asimismo, son las estructuras sociales las que permiten o debilitan este desarrollo y, por lo tanto, la ética moderna *es* un asunto político. En efecto, el “yo” de la modernidad está dividido entre distintos tipos de esquemas éticos, por lo que defiende ciertos bienes últimos en una esfera social y posteriormente otros en una distinta. El efecto ético-político que tiene esto en el desarrollo individual es la falta de una real racionalidad práctica, una vida que actúe siguiendo un ideal de perfección humana. De este modo, es la misma noción de “agente moral” la que está en juego. MacIntyre considera que esto se debe a la estructura de compartimentalización que existe entre las esferas sociales, que no permite ser constante e integral en cuanto a lo que se busca como fin último. Así, la modernidad fomenta un tipo de individuo que no es plenamente responsable de sus actos, en la medida en que solo lo es para determinado tipo de esfera. Para que esto no sea así, es necesario que existan determinados tipos de estructuras institucionales que permitan y fomenten un tipo de discurso integral y crítico que posibilite a un agente ser tal, es decir, un portador

¹² MacIntyre (1990b, p. 48) dice que “intelligibility is nothing other than translability into ‘our’ language and conceptual idiom”.

de su propia moralidad (cfr. MacIntyre, 2006b, p. 196) cuestión que, por su misma naturaleza, es política.

Como ya se advierte, las características éticas del pensamiento macintyreano están íntimamente ligadas con las dimensiones políticas. Por eso es por lo que, gradualmente, comenzando desde los aspectos éticos, se puede ir revelando que el problema central de la modernidad, para el pensamiento político macintyreano, es el poder corruptor del Estado moderno para el florecimiento humano como tal. Como dijimos, la propuesta de MacIntyre fomenta como fin último del ejercicio de la política, la idea de que debe llevarse a cabo dentro de cierto tipo de comunidades (cfr. MacIntyre, 1998, p. 248). Esto último es central, siempre que se tenga en cuenta su noción de “acción humana” en términos tomasianos de herencia aristotélica, es decir, con una estructura finalista, teleológica en el orden de la naturaleza, comprendida en términos aristotélicos.¹³ En este sentido, si entendemos al ser humano como político por naturaleza, entonces un obstáculo en la práctica misma de la política puede llevar al ser humano a fracasar, imposibilitándole la obtención de la plenitud que por naturaleza está destinado a obtener (cfr. Widow, 1984). Para que el ser humano pueda efectivamente perfeccionarse según su propia naturaleza, incluso dentro de las esferas institucionales de la modernidad, es imprescindible entrar en conflicto con los valores enarbolados por estas mismas, ya que poseen un poder corruptor en cuanto a la dirección de los deseos humanos.

El problema que presenta la modernidad es que sus mismas estructuras sociales impiden el desarrollo de preguntas genuinamente aristotélicas, es decir, promueven una vida que *desnaturaliza* al ser humano. De modo que se vuelve imperioso que la crítica a la moralidad moderna sea necesariamente política, lo que daría sentido, desde estos postulados, al proyecto macintyreano que propone el perfeccionamiento humano por medio de las virtudes humanas, con un natural efecto de orden político. Lograr que esto suceda es una tarea orgánica en la que están implicadas las instituciones, la virtud y lo que MacIntyre denomina *prácticas*, comprendidas como actividades compartidas enfocadas en alcanzar bienes que les son propios y para cuya obtención es necesaria determinada concepción de la excelencia humana (cfr. MacIntyre, 1981,

¹³ Para comprender mejor la noción de “naturaleza” heredada del pensamiento de Aristóteles y Tomás de Aquino, cfr. Spaemann (1989), González (2006) y Buzeta (2013).

p. 175), a diferencia de las instituciones que están destinadas a alcanzar bienes que son comunes a otras prácticas (dinero, honor, etc.), para lo que es necesaria alguna pericia, habilidad o perfeccionamiento técnico.

Estas dos dimensiones, de las prácticas y las instituciones, se relacionan de modo causal y cada una tiene un modo particular de abordar la relación entre medios y fines. Así, en este esquema, las instituciones, destinadas para obtener ciertos bienes limitados o —en términos aristotélicos— externos, son necesarias para el desarrollo de las prácticas, que se planifican para alcanzar los bienes internos. Como ambas dimensiones conviven constantemente, las virtudes necesarias para las prácticas siempre están bajo el riesgo de ser corrompidas por el éxito que pueda tener la institución (cfr. MacIntyre, 1981, p. 181).

De este tipo de prácticas se puede alimentar la imaginación política, fundamental para los momentos en que los esquemas entran en crisis (cfr. MacIntyre, 2006a, p. 160). Sin embargo, no solo es necesario esto, sino también, como señalamos antes, estructuras sociales que sean portadoras de cierta flexibilidad para adaptarse a nuevas imágenes políticas. El problema radica, según MacIntyre, en que en la modernidad la concepción de la moral se encuentra tan distorsionada que esta necesaria relación entre prácticas, virtudes e instituciones se malentiende.

1.3. Crítica a la institución del Estado moderno

Desde este marco institucional que comporta la modernidad, es necesario focalizar la atención en la pretensión central de la Ilustración, es decir, el saber universal, para luego observar su relevancia ética: el intento de llevar a cabo una moral universal, frente a la fractura moderna, se concentra en la articulación política, pero también soberana, del Estado-nación moderno. El ápice de la crisis moral de nuestro tiempo tiene su representante en la política, dado que fomenta y se presenta como un remedio a las dificultades éticas contemporáneas por medio de un esquema abstracto pero universal. El resultado de esta pretensión corresponde a lo que MacIntyre llamó el *fracaso ilustrado*, que —entre otras cosas— multiplicó *ad infinitum* las corrientes éticas supuestamente portadoras de la definitiva objetividad moral.

Para MacIntyre (2016, p. 134), una característica resalta cuando nos referimos al Estado moderno: la pretensión de mantenerse *neutral* frente a las concepciones del bien humano. Esta particularidad, reflejo del intento ilustrado por universalizar la normatividad de la ética, está

en el centro de su crítica a la dimensión “liberal” de la política, tanto en su variante teórica como en la institucional. Por ello, y en un intento de encontrar una alternativa al individualismo liberal, MacIntyre (2016, p. 135) sostiene que el Estado moderno, al proponerse neutral frente a las nociones de bien último humano, permite el desarrollo y fomento del tipo de vida de esta modalidad de liberalismo individualista.

No obstante, al analizar detenidamente la crítica macintyreana a la neutralidad estatal (cfr. MacIntyre, 2006d), se evidencia la conocida tesis en la que propone aceptar este carácter liberal del Estado moderno a pesar de ser en sí misma una mera ficción. Es ficticio, afirma, dado que resulta irrealizable un desarrollo *de facto* de una imparcialidad institucional estatal frente al hecho plural (*the fact of pluralism*)¹⁴ constituido por diversas —y muchas veces rivales— concepciones sobre el bien humano (cfr. MacIntyre, 2006d, p. 210). En ese sentido, MacIntyre destaca el papel histórico que ha mantenido esta institución política en la resolución de conflictos religiosos y económicos durante los siglos XVII y XVIII, en los que, lejos de ser un ente externo e imparcial en cada antagonismo, se presentó como una de las partes contendientes, y victoriosa —para bien o para mal—. Con una metodología antagónica respecto de la de los liberales citados por MacIntyre (2006d, p. 214) —Mill y Locke—, llega a una conclusión similar: el Estado no debe imponerse desde una visión particular del bien humano al conjunto de la sociedad.

No podemos negar que la visión macintyreana de la modernidad en general, de las instituciones políticas de esta época, y del Estado moderno en específico, es bastante crítica y culmina, como vemos, en una dificultad si se propone como antiliberal. Sin embargo, como hemos visto, esto no reviste una *ausencia* de propuesta clara en estos términos. Para justificar estas afirmaciones, es necesario contrastarlas con algunas críticas contemporáneas que señalan una ausencia en este punto.

2. Crítica a la existencia de una propuesta política en MacIntyre

Conforme a los objetivos planteados en el presente trabajo, centraremos el análisis de las críticas a la propuesta política macintyreana desde tres objeciones: una supuesta “amputación” política en el trabajo de MacIntyre, una eventual adscripción no reconocida al comunitarismo, y una aparente imposibilidad de llevar a cabo las posibles prácticas

¹⁴ Para distinguir esta noción del “hecho plural” de otras versiones, cfr. Jones (2006, p. 191).

macintyreanas desde un antiliberalismo, dado que parece adscribir una postura eminentemente liberal.

2.1. “Amputación” aristotélica en MacIntyre. La crítica de Perreau-Saussine y de Manent

La primera crítica puede expresarse del siguiente modo: Perreau-Saussine (1999) comienza su artículo “Alasdair MacIntyre between Aristotle and Marx” haciendo un paralelo entre la esclavofilia de Lenin y los bolcheviques, que fue transformada en una doctrina revolucionaria, y el supuesto marxismo-leninismo macintyreano, convertido en una esclavofilia aristotélico-escocesa. Perreau-Saussine considera que MacIntyre trata de reivindicar algo del marxismo de su juventud gracias a su tipo peculiar de aristotelismo tomasiano señalando que la ley natural es subversiva y que la virtud de la templanza es “sediciosa”. Toda la importancia que MacIntyre les otorga al concepto de “comunidad”, a las sociedades preindustriales y a su tendencia antiestatista son entendidas, para Perreau-Saussine (1999, pp. 1-2), en el contexto de la *New Left* británica a la que el filósofo escocés perteneció. Las corrientes marxistas orientadas hacia el estatismo solo se entenderían desde la influencia del *fabianismo*, pero para el resto de los movimientos socialistas británicos la tendencia ha sido marcadamente anarcosindicalista (cfr. Perreau-Saussine, 1999, p. 1).

Para Perreau-Saussine, esta importancia dada a las comunidades junto con el rechazo del Estado como eje político hacen que MacIntyre comience a divorciarse de la política. El análisis de esta tesis se detalla mucho más en su destacado libro sobre MacIntyre, *Alasdair MacIntyre: Une biographie intellectuelle*, en el que afirma dos cosas: 1) MacIntyre “sacrifica la filosofía política a la filosofía moral” (Perreau-Saussine, 2022, p. 38),¹⁵ y en términos aristotélicos, opta por la *virtud cívica* más que por la *virtud moral*,¹⁶ y 2) MacIntyre tiene una (supuesta) tendencia

¹⁵ Utilizaremos la traducción al inglés del libro, realizada por Nathan J. Pinkoski (Perreau-Saussine, 2022).

¹⁶ En cuanto a su lado ético, Perreau-Saussine considera que MacIntyre se inclina hacia las teorías de Charles Taylor y Michael Walzer, es decir, el comunitarismo. Sin embargo, reconoce explícitamente que MacIntyre no es comunitarista (cfr. Perreau-Saussine, 2022, p. 40) porque se distancia de sus implicancias políticas. En este sentido, Perreau-Saussine realiza su interpretación

“apolítica” (Perreau-Saussine, 2022, pp. 43-44). Analizaremos esta crítica con un poco más de profundidad.

Pierre Manent (2022, p. xiii), en la presentación del citado libro, afirma coincidir con Perreau-Saussine acerca de que el Aristóteles de MacIntyre es uno “mutilado” de la *Política* y que las afirmaciones políticas que realiza son guiadas por la *Ética a Nicómaco*. Considera que pensar desde ese Aristóteles mutilado es lo que ha permitido que MacIntyre nunca se haya distanciado de la “postura política opositora” en la que se mantiene siempre *a favor de* “la comunidad subpolítica amenazada por la comunidad política que se alza en el poder, y ‘en contra’ de esta última” (Manent, 2022, p. xiii). En ese sentido, Manent (2022, p. xiii-xiv) cree que esto permite entender la ausencia de cuestiones referidas al mejor régimen de gobierno y de justicia política en el trabajo de MacIntyre, pero también que MacIntyre sí se haya centrado en la labor política de las agrupaciones de pescadores, monasterios benedictinos y otras.

Si lo analizamos de cerca, podremos dar cuenta de un malentendido en la lectura de Manent. La crítica sobre la mutilación de la política aristotélica por parte de la teoría macintyreana fue reconocida también por Vincent Descombes (2013, pp. 136 y 140), quien se suma a Perreau-Saussine y Sandra Laugier, asegurando un “conservadurismo” en MacIntyre dada la interpretación wittgensteiniana del “lenguaje privado”. Y, aunque MacIntyre (2013) se hace cargo explícitamente en “Replies” de estas críticas,¹⁷ es importante que analicemos a qué se refieren ambos (Perreau-Saussine y Manent) con la imputación central de estas: la ausencia de política en MacIntyre.

desde un prisma profundamente moderno en el que se encuentran divididas ética y política.

¹⁷ MacIntyre (2013, p. 203) reconoce que se trata de una crítica similar que se da en el grupo de sus “críticos franceses”, dentro de los que se encuentra Perreau-Saussine, que había fallecido tres años antes. Nathan J. Pinkoski, traductor del libro de Perreau-Saussine, destaca que la respuesta de MacIntyre ocurre después de la muerte (accidental y muy lamentable) de Perreau-Saussine, de modo que hay que leerlo en su justa medida. Tampoco debe olvidarse que todavía no estaban publicadas obras fundamentales de MacIntyre, como *Ethics in the Conflicts of Modernity*, en la que aborda gran parte de los cuestionamientos sobre su viabilidad política. En el presente artículo buscamos hacernos cargo de la crítica específica en torno a un aspecto del debate sobre la inviabilidad institucional política de la teoría de MacIntyre, por lo que hemos tenido en cuenta la vigencia de esta discusión.

Perreau-Saussine (2022, p. 9) reconoce que el antiliberalismo de MacIntyre hace que no esté de acuerdo con las democracias liberales, que es el régimen que el liberalismo encarna. Su problema es creer que esto es una contradicción: MacIntyre (2006d, p. 220) continuamente ha destacado su amistad no solo con los principios liberales de autores como Mill, sino también con las formas democráticas existentes hoy. El alejamiento de la democracia liberal por parte de MacIntyre es cierto, pero se debe entender, contrariamente, como un intento de acercarse a la política “real”, o sea, como una búsqueda del bien común en términos aristotélicos. En esa búsqueda, MacIntyre mantiene una relación con la democracia liberal y, al contrario, intenta unir los ámbitos que estos autores presuponen separados, a saber, lo social y lo político. Separación que, por otro lado, ha sido diagnosticada (como dijimos previamente) como una fragmentación moderna. Quizás el malentendido de Perreau-Saussine (1999, p.6) se deba a que encuentra esta misma línea argumentativa en Hegel, Marx, Sartre y Marcuse, a saber, la búsqueda del Uno frente a la fragmentación moderna, una suerte de hegemonía moral universal. Sin embargo, concebir a MacIntyre como un autor que busca unificar la ética en un solo esquema es tan erróneo como creer que es un autor comunitarista. Ambos extremos son foco de la crítica macintyreana, y parece ser que Perreau-Saussine prescinde de su análisis. De hecho, el autor francés cree que MacIntyre habría abrazado el comunitarismo, al menos en su variante ética (no en la política, tal como vimos más arriba, en la cita 36), como una reacción al empobrecimiento del lazo social humano y en contra de los límites de las democracias liberales y el fracaso del comunismo.

Perreau-Saussine (2022, p. 30) entiende la crítica al individualismo y el énfasis en lo local como una vuelta al corporativismo de G. D. H. Cole cincuenta años después. De acuerdo con este autor, aparte de la crítica a la carencia de los lazos humanos que adopta el individualismo liberal, los comunitaristas están desilusionados con el estalinismo y el socialismo. Por esta razón, MacIntyre habría —como Charles Taylor— adoptado formas de marxismo cristiano. En este aspecto, sus palabras son las siguientes:

Por su sensibilidad, MacIntyre es fuertemente atraído al comunitarismo. Él da pie, a veces, a una irritante retórica, esperando por una renovación de comunidades de pescadores y artesanos, en un tono de romanticismo

envejecido. Fácilmente critica el Estado-nación. “Que se pida que mueran por él”, escribe, “sería como si te pidieran que mueras por una compañía telefónica”. En contra de este malestar, MacIntyre responde menos con virtud cívica y más con virtud moral (Perreau-Saussine, 2022, p. 38).

En este caso es evidente cómo el recorrido que Perreau-Saussine realiza en ambos textos necesita de ambas dimensiones del desenvolvimiento de MacIntyre. Si el foco solo se fija en los textos, es imposible llegar a la conclusión de que MacIntyre “deja la política”, puesto que el tema en cuestión es constante. Solo si se une su propuesta localista con el abandono del partidismo político como tema *biográfico* es posible entender qué está haciendo Perreau-Saussine: identifica *política* —y la participación de MacIntyre en la política partidista— con *Estado-nación*, y no, entonces, con *comunidad local de virtudes*. Cree, en efecto, que MacIntyre opta por la *virtud moral* más que por la *virtud civil*, utilizando conceptos de la *Política* de Aristóteles. De esta manera, para Perreau-Saussine, es posible construir el razonamiento necesario para afirmar dos cosas: primero, que MacIntyre ha optado por abandonar “la política” y, en segundo lugar, que, en rigor, se ha alejado de Aristóteles.

Perreau-Saussine no se equivoca al afirmar que MacIntyre prefiere, como modo de respuesta al malestar moderno, más una política local que una basada en la *participación política*. No obstante, es necesario insistir en que este giro macintyreano es realizado, como dijimos, en búsqueda de las preguntas legítimamente políticas. Concebir el Estado como epítome de lo político, al mismo tiempo que entenderlo como el que provisiona de los bienes públicos *mínimos* para una sana convivencia, es una contradicción. Lo político ha de entenderse bajo un esquema distinto del meramente instrumental. Pero esto *no quiere decir* que el Estado, entendido de modo instrumental, sea completamente descartado, porque estos bienes mínimos son necesarios. En concreto, quiere decir, principalmente, que esta relación que se tiene con el Estado tiene *muy poco de política*. En relación con este punto, MacIntyre declara lo siguiente:

En la investigación política, importa mucho cómo uno comienza. Si uno comienza dando por descontadas las categorías del orden cultural y social dominante en el presente, que son también las categorías del currículo

académico del presente, uno estará en peligro de esquematizar sus preguntas acerca de los pensadores políticos del pasado en términos que presuponen las compartimentalizaciones del orden presente: lo político, lo social, lo económico, lo moral, lo religioso, lo estético. Pero algunas de estas, al menos, son categorías y distinciones que necesitan ser puestas en cuestión. Un camino para ponerlas en cuestión es, considerar órdenes sociales y culturales donde esas categorías y distinciones fueron significativamente diferentes (MacIntyre, 2013, p. 206).

Al respecto, lo que hace Perreau-Saussine es analizar a MacIntyre con el prisma de la modernidad, del Estado moderno, lo que no resultaría equivocado si en eso no se incluyera que, además, lo *evalúa* según la terminología política de la modernidad: el vocabulario de las separaciones liberales.¹⁸ De esta manera, para Perreau-Saussine, es fácil confundir “comunidad” con aquello presentado por los socialistas o comunistas y pensar que MacIntyre debería ser analizado bajo esas categorías. Según Perreau-Saussine (2022, p. 42), Marx dio cuenta de que es imposible conciliar igualdad con diferencia: las comunidades no pueden convivir con el Estado y solo en la sociedad comunista se podrá eliminar la diferencia entre lo público y lo privado. ¿Por qué habría de haber, siguiendo a Perreau-Saussine, una relación de *analogía* entre la dicotomía público-privado y la de Estado-comunidades? ¿No es el mismo MacIntyre (1990a) quien, en “The Privatization of Good”, criticó fuertemente la privatización de la noción de “bien humano” *al interior* de la esfera institucional estatal? *Lo privado* dentro de *lo social* no es lo que MacIntyre busca, en modo alguno. Si se quisiera llamar comunitario al carácter local al que MacIntyre adhiere, debemos recordar que señala precisamente un ámbito *político* de comunidades, en el que se comparte una visión explícita del bien humano. Lo que se da en lo local, entendido en clave aristotélica, es, en este caso, *lo público*.

Esto genera que, para Perreau-Saussine, los comunitaristas sean apolíticos, puesto que proponen algo inalcanzable. Señala que MacIntyre

¹⁸ Esta respuesta no es solo nuestra. Para una caracterización detallada de las diferencias entre las visiones de Perreau-Saussine y Manent, por un lado, y la de MacIntyre, por otro, sobre su supuesta apoliticidad, cfr. Pinkoski (2019).

no les hace esta crítica a los comunitaristas (la de su efectividad política), y la razón de esto es su *sensibilidad comunitaria*, que lo hace parcialmente apolítico (cfr. Perreau-Saussine, 2022, p. 42). Por este motivo, MacIntyre (2013, p. 205) afirma lo siguiente: “Cuando Perreau-Saussine y Manent toman mi aristotelismo como no-político, es en parte porque ellos y yo estamos en desacuerdo sobre el Estado moderno. Pero también estamos en desacuerdo acerca del aristotelismo”. En ese sentido, MacIntyre entiende que la política aristotélica no se encuentra separada de las esferas éticas, que nacen del bien común de la comunidad. Dado que el Estado no es una comunidad, entonces la propuesta macintyreana cuida de no identificar política con Estado, ni comunidad con Estado-nación. Se puede ver, entonces, la importancia que puede llegar a tener lograr distinguir el trabajo de MacIntyre de las corrientes llamadas comunitaristas. A continuación, expondremos un breve ejemplo al interior del trabajo de Ronald Beiner.

2.2. Crítica al comunitarismo y eventual inconsistencia en MacIntyre

En la crítica de Ronald Beiner (2000) se señala directamente a los comunitaristas; se afirma que son continuamente mal evaluados por carecer de propuestas políticas propias, o por no representar una alternativa real al liberalismo, al que suelen denostar abiertamente. Beiner (2000, p. 460) afirma que, cuando han logrado plantear propuestas interesantes, han sido justamente liberales. Transformarse en una teoría liberal resulta un contrasentido al menos desde esa óptica.¹⁹ Su afirmación es que “todos los filósofos políticos de Occidente saben” (Beiner, 2000, p. 460) que la más común de las críticas que deben afrontar los comunitaristas es que carecen de propuesta política. Esta afirmación nos importa, porque Beiner considera que MacIntyre forma parte del grupo comunitarista, en gran medida por dicha ausencia.

Aunque la afirmación de Perreau-Saussine acerca de la similitud (solo) ética entre el comunitarismo y la propuesta de MacIntyre tiene otras causas, es importante dar cuenta de los límites de las afirmaciones

¹⁹ El comunitarismo contemporáneo nace como una crítica a los males del liberalismo, lo que se ve reflejado en el famoso debate entre ambos. Para una visión más completa sobre este debate, cfr. Mulhall y Swift (1996).

de Beiner.²⁰ MacIntyre reconoce que hasta 1968 carecía de una propuesta alternativa al capitalismo que no fuera la marxista. No sabía cómo formularla porque no lograba dar expresión a su marxismo y a su cristianismo *a la vez* (cfr. MacIntyre, 2006c, p. 154). Sin embargo, posteriormente reconoció que al negar el liberalismo político debe haber una propuesta, es decir, un “set” de afirmaciones propositivas (cfr. MacIntyre, 2006c, pp. 153-154). Como vimos anteriormente, esto pudo lograrlo hacia 1971 (ver nota 4)..

Esta crítica al marxismo será muy similar a la que realizó contra los comunitaristas, puesto que no presentan, según lo que afirma MacIntyre, un real cambio en cuanto a las formas de pensar la política moderna. Pretenden que el individuo es constituido por las comunidades, pero en la medida en que esa comunidad es el Estado (cfr. MacIntyre, 1998). En ese sentido, la imputación que hacía Perreau-Saussine a los comunitaristas es cierta: consideran que la comunidad sí puede llevarse a cabo desde las categorías institucionales del Estado moderno. Esto es lo que nos permite entender por qué Beiner considera que se trata solo de variantes del liberalismo. En esto, también MacIntyre está de acuerdo:

[...] los valores del comunitarismo también pueden ser encontrados en el bolso de valores del Estado, y estaban ahí mucho tiempo antes que el nombre “comunitarismo” les fuera dado a ellos. Así, junto con los compromisos de los gobiernos modernos a principios universales que salvaguardan derechos y confieren libertades, están los compromisos de los mismos gobiernos para defender los lazos familiares y las solidaridades de una variedad de grupos (MacIntyre, 1998, p. 245).

²⁰ Esta crítica a Beiner hay que entenderla en contexto. Desde la introducción a *Marxism and Christianity*, MacIntyre (2006c, p. 153) ha dado cuenta de las fallas de los teólogos en exponer las injusticias del capitalismo, pero también de los marxistas en proporcionar una solución a los problemas modernos, justamente por la adscripción a un tipo de moral moderna, que fue lo mismo que ocurrió en el caso de los teólogos provenientes del pensamiento de Karl Barth que pretendieron separar las dimensiones éticas y políticas, promoviendo una constante especialización compartimentalizada. Tal como ocurre con Perreau-Saussine, Beiner realiza estas consideraciones suponiendo que se puede pensar en alternativas desde un lenguaje moral moderno.

Por eso es que, para MacIntyre (1998, p. 244), el comunitarismo señala, no un antagonismo respecto del liberalismo, sino una corriente que destaca *falencias* en los autores liberales. Algunas de las implicancias que puede tener la interpretación de MacIntyre desde la lectura de Beiner pueden verse expresadas con más detalle en la siguiente problematización de la teoría macintyreana en dos destacados autores contemporáneos.

2.3. Crítica al supuesto liberalismo de MacIntyre. La crítica de Breen y Young

Daniel Edward Young (2022, p. 1) plantea la tesis de que, si atendemos a las sugerencias de MacIntyre, traducidas en un Estado no interventor respecto del bien común local y su foco en la descentralización del poder, podría ser caracterizado como pluralista conservador, liberal conservador o libertario.²¹ Para Young, la noción de “comunidad” macintyreana, referida al bien común, no puede señalar tipos de sociedades o comunidades que promulguen leyes, debido, en gran parte, a la separación moderna entre Iglesia y Estado. Young no explica muy bien la identificación entre comunidad macintyreana e Iglesia, pero, al hacerlo, la toma como premisa del siguiente silogismo:

P1) Las comunidades locales macintyreanas no promulgan leyes, sino que se tratan de localidades de orden social y-o cultural.

P2) En la cultura occidental, las comunidades locales no pueden coincidir con quien promulga las leyes, porque, entre otras cosas, el Estado y la Iglesia están separados.

²¹ A pesar de lo que pueda parecer, esta tesis no es novedosa. Es conocida la supuesta paradoja del pensamiento político macintyreano en la que, por un lado, MacIntyre es enemigo del liberalismo y por el otro es amigo de muchas de las ideas de J. S. Mill. Esta aparente discrepancia fue trabajada años antes por Keith Breen (cuestión que analizamos un poco más adelante), quien también la concibe como tal aunque mucho más matizada. Sin embargo, no es de ningún modo ajeno al pensamiento macintyreano encontrar una serie de semejanzas y trabajos muy parecidos al de MacIntyre en filósofos como Joseph Raz, por ejemplo.

Conclusión: para MacIntyre, el Estado es ilegítimo, de modo que su teoría implica anarquismo (cfr. Young, 2022, p. 2).

Si a la ecuación agregamos la aceptación, por parte de MacIntyre, de *cierta* injerencia estatal, entonces más bien deberíamos estar hablando de la teoría minimalista o minarquista de Alasdair MacIntyre, muy similar a las teorías de autores como Robert Nozick (cfr. Young, 2022, p. 2).

Sin embargo, Young no explica acabadamente el origen de su identificación entre una comunidad que busca el bien común y una parroquia católica o una congregación protestante. Independientemente de que ambos ejemplos sí pueden ser concebidos como comunidades virtuosas que comparten una noción de “bien común humano”, parece no tener en cuenta que MacIntyre habla de bien común en un sentido mucho más amplio. Evidentemente, una noción aristotélica del ser humano implica la tendencia natural hacia una realización trascendental, *más allá*, hacia bienes que son más altos. Sin embargo, tal como vimos previamente, el tipo de comunidad abordada por MacIntyre en su propuesta política comprende cualquier tipo de actividad (práctica) gracias a la cual un ser humano pueda florecer en cuanto a sus facultades, al interior de la que no haya una diferencia entre lo que busquen los individuos y lo que busque la comunidad (cfr. Young, 2022, p. 5). Esto no es excluyente de una agrupación religiosa, pero no se puede reducir solo a ello.

Para eso, es cierto, es necesario poner límites a las injerencias estatales sobre las comunidades locales macintyreanas, de modo que efectivamente cree que el Estado no debería ser presentado ni concebido como el representante o el baluarte de alguna visión sustantiva del bien común humano (cfr. Young, 2022, p. 6). Si solo entendemos la política de MacIntyre bajo este esquema, efectivamente podríamos sostener que el autor de *After Virtue* promueve algún tipo de neutralidad estatal. Sin embargo, esto no es así, por el carácter de la propuesta política macintyreana. La respuesta a la supuesta neutralidad macintyreana puede ser realizada acabadamente desde la crítica de Keith Breen (que discutimos más adelante). Para eso, es necesario abordar previamente las siguientes afirmaciones realizadas por Young.

Para Young (2022, p. 6), otra equivocación de MacIntyre sería la de no hacerse cargo del alto grado de incompatibilidad (*diversidad*, según Young) de visiones del bien humano, incluso al interior de las

comunidades contemporáneas, con excepción de los *amish*. Luego, la crítica de Young apunta un —supuesto— paralelo entre MacIntyre y Nisbet, dado que ambos coincidirían en reconocer las amenazas del Estado moderno a las comunidades locales. Young (2022, p. 11) considera que, dadas las consecuencias similares de los planteamientos de Nisbet y MacIntyre, puede que ambos sean conservadores. Sin embargo, hacia el final de su artículo, afirma que, en el contexto norteamericano actual, los aliados más cercanos a MacIntyre podrían ser los libertarios, que sostienen que el Estado debe existir en una mínima expresión (cfr. Young, 2022, p. 12). Si no fuera así, sigue Young, MacIntyre debería ser concebido como anarquista por las mismas razones expuestas más arriba, a saber, su antiestatismo.

Para responder a este conjunto de afirmaciones, debemos tener en cuenta lo que analizamos previamente. Young (2022, p. 12) da a entender que su noción política de la realidad moderna es muy similar a la de Perreau-Saussine y Manent, en el sentido de que asocian política con Estado, y lo no estatal con el anarquismo. Así, MacIntyre estaría equivocado al suponer que todo tipo de esquema local debe ser antiliberal (cfr. Young, 2022, p. 13). Sin embargo, lo que Young no logra ver, a nuestro juicio, es que MacIntyre no está en contra de todo tipo de liberalismo, como vimos previamente (ver nota 47). Al leer el artículo de Young, podemos notar que falla al no captar que MacIntyre ya ha aceptado las posibles acusaciones previamente y se ha hecho cargo de ellas. Además, tampoco comprende que el filósofo escocés está preocupado por la institucionalización *de facto* de las teorías liberales en las arenas reales de la política más que por los textos de los filósofos liberales.

Young nombra, entre los críticos más importantes a la teoría de MacIntyre, a Keith Breen, quien ha realizado variados trabajos en torno al escocés, elaborando una interesante síntesis de su pensamiento político. Al respecto, en un espléndido análisis, Breen (2005) acierta al identificar una dificultad en la teoría de MacIntyre, a saber, la de una similitud entre algunos liberales y él. Ambos (liberales y MacIntyre) acusan que una política limitada meramente a entregar seguridad carece de poder autoritativo y es vulnerable a *freeriders*, además de no tomar en cuenta la importancia intrínseca de una vida política. Por otro lado, tanto los liberales como MacIntyre tienen razón al temer que los Estados se sostengan como baluartes de valores sagrados, puesto que —para MacIntyre— atentaría contra la conformación inicial de lo local

entendido desde sus propias nociones del bien humano (cfr. Breen, 2005, p. 492).

En este momento podemos hacernos cargo de la supuesta neutralidad macintyreana. Las críticas macintyreanas a la neutralidad moral del Estado son interpretadas por Breen (2005, p. 492) como *neutralidad ostensible*, dado que es una ficción, pero muy importante para la protección de ciertas libertades civiles (liberales) a las que MacIntyre adscribe. En esta aceptación “a medias” de la neutralidad estatal habría un fomento a la compartimentalización, cuestión que es vehementemente criticada por MacIntyre. Las comunidades necesitan negociar con el Estado sobre lo necesario para el bien común local. De este modo, Breen se plantea si desde la teoría de MacIntyre es consistente y sostenible en el tiempo una política de comunidades locales que se relacione con el Estado de esta forma antagónica pero *ad hoc* según las necesidades de la comunidad, así como también si es posible concebir la comunidad local como el *locus* de la actividad y la teorización política.

A pesar de ser atendibles estas preguntas de Breen, sus respuestas no son acabadas. Para abordar el problema de las comunidades locales, cita a MacIntyre en su elogio a la labor que realizó la marina estadounidense en su lucha contra el racismo como un ejemplo paradigmático de comunidad local basada en virtudes (cfr. Breen, 2005, p. 494). La perplejidad de Breen frente a esto estriba en que precisamente esta comunidad es estatista. Sin embargo, este argumento no es suficiente, puesto que MacIntyre ejemplifica, con esto, que hay dimensiones del Estado que es preferible mantener a pesar de que sea corruptor de prácticas locales. A pesar de que en ocasiones es necesario negociar con el Estado para obtener bienes necesarios para la comunidad (estatal o no), el poder con que se presenta consistentemente intenta corromper las prácticas locales *en cuanto* políticas.

Como adelantábamos más arriba, Breen (2005, p. 495) considera que la política local macintyreana, al entregar tan poco al Estado, fomenta la reproducción de la figura del *freerider*. Es cierto que la condescendencia que MacIntyre ha mostrado respecto del Estado implica la existencia del *freerider*, que desprecia por ser un síntoma de la estructura institucional liberal. Sin embargo, el planteamiento de MacIntyre no pretende que, por tener que negociar de vez en cuando con el Estado, *necesariamente* nos libremos de ello, sino que, de ocurrir, al menos no sería reflejo de un síntoma del esquema de comunidades locales. La propuesta macintyreana sostiene que debemos promover tipos de vida locales, cambiar el *locus*

de lo político a lo local, tal como Breen lo acepta, de modo que el poder vaya pasando gradualmente a modos de vida locales. Es necesario, como Breen nota, asegurar una base institucional, un contexto de instituciones políticas que permitan el desarrollo de virtudes humanas, para que ese tipo de ética liberal —MacIntyre (1991, p. 70) mismo reconoce que el término es difuso— bajo la cual todos los occidentales modernos vivimos pueda comenzar a cambiar.

Un tercer punto tocado por Breen en su crítica asegura que lo local, por ser tal, no goza de una intercomunicación amplia que pueda dar cabida a la arena de la discusión real en términos científicos y éticos. Sin este componente transversal, la crítica a la estructura mercantil del Estado no tendría ningún sentido y, por lo mismo, tampoco toda la teoría macintyreana. Sin embargo, esta crítica debe ser entendida desde un componente normativo previo consistente en la aceptación de las prescripciones derivadas de la ley natural en un sentido tomasiano de origen aristotélico.²²

Por último, creo que podemos encontrar una crítica acertada en el artículo de Breen (2005, p. 495), en la que afirma que al parecer MacIntyre no tiene teorizado el mecanismo económico para poder financiar aquello que sí necesitamos del Estado, que a la vez sea justamente distribuido y no acaparado por quienes tienen acceso al poder político y económico en la modernidad. Eso es algo que todavía debe ser elaborado por la teoría macintyreana. Precisamente en esto nos encontramos trabajando algunos lectores del filósofo escocés, que consideramos necesario un cambio macintyreano en la política moderna.

En este sentido, MacIntyre no entrega ningún “esquema” (*blueprint*) de cómo ha de darse esta sociedad de comunidades locales desde las cuales se generen las relaciones de política en términos de virtud (cfr. Pinkoski, 2019, p. 132). Sin embargo, la virtud de la esperanza no juega solo un papel teológico en MacIntyre, sino también político, puesto que está focalizado en comunidades nacientes, que existen (algunas antiguas y otras nuevas) y que, pensadas desde la propuesta:

[...] deben ser de pequeña escala y, en lo posible, lo más autosuficientes que necesiten ser para protegerse a sí

²² Cfr. Petersen C. (2020). El problema fue planteado por Elizabeth Frazer y Nicola Lacey (1994) en torno a las “prácticas perversas”. La dificultad surge, como siempre, al confundir lo políticamente transversal con la figura del Estado.

mismas de las incursiones destructivas del Estado y de la economía de mercado en general. Necesitan ser de escala pequeña para que, cuando sea necesario, aquellos que ocupan cargos políticos puedan ser cuestionados por los ciudadanos y los ciudadanos cuestionados por quienes ocupan cargos políticos en el curso de un debate deliberativo prolongado en el cual haya una participación amplia y desde la cual nadie de quien algo se pueda aprender sea excluido. Esto es, desde la cual nadie es excluido. El propósito [...] es llegar a una concepción común. [...] [N]o estoy describiendo algo ajeno a la experiencia del día a día. Este es un tipo de participación familiar de muchas empresas locales a través de la cual la comunidad local se realiza. Lo que es menos familiar es la proposición de que estas arenas locales son ahora los lugares donde la comunidad política puede ser construida, una comunidad política mucho más en desacuerdo con la política del Estado-nación (MacIntyre, 1998, p. 248).

Conclusión

Si tomamos un poco de distancia respecto de las críticas que hemos expuesto hasta este punto, nos podemos dar cuenta de que tienen una unidad: la Moralidad (con mayúsculas según MacIntyre), la cual se presenta *a través* de un esquema institucional protagonizado por el Estado. Esta no concibe pensar una política desde otro tipo de organización que no sea la estatal. Aunque no quiere decir que —para MacIntyre— sea necesario pensar “desde fuera” del Estado, sino simplemente cambiar el *locus* de la actividad política hacia la comunidad local.

En una primera sección, hicimos una exposición de esta parte de la teoría filosófica de MacIntyre, con énfasis en su crítica a la ética y política modernas. Con ello, pudimos dar cuenta de la existencia de una propuesta política, cuya eventual viabilidad ha de ser desarrollada al interior de lo que MacIntyre denomina “prácticas” humanas. Para esto, hemos podido ver que un entramado institucional resulta fundamental, y no al revés, como si las instituciones fueran el *locus* de la política. Es este carácter pragmático el que ha sido criticado por los autores que

analizamos en la segunda parte del presente artículo, a saber, el de la medida de su *aplicabilidad*.

Bajo ese respecto, analizamos tres críticas muy fundadas a la viabilidad de la propuesta política macintyreana. En primer lugar, estudiamos la imputación, por parte de Perreau-Saussine y Pierre Manent, de que la teoría política de MacIntyre se encuentra “amputada” de su elemento político. Pudimos dar cuenta de que el supuesto “abandono” que MacIntyre hiciera de la política aristotélica solo puede ser entendido bajo un esquema que presupone un divorcio entre ética y política, y a esta última como *necesariamente* representada bajo la figura del Estado. En segundo lugar, analizamos la crítica de Ronald Beiner, que considera que MacIntyre se encuentra bajo el paradigma comunitarista, razón que permitiría entender la ausencia de una propuesta política y la adscripción, por parte de MacIntyre, a ciertas líneas liberales. Sin embargo, pudimos constatar que estas limitaciones pueden ser aclaradas recordando la crítica macintyreana a los comunitaristas, cuestión que resulta fundamental para una correcta interpretación, ya que busca destacar la continua confusión, por parte de dicha corriente, entre los conceptos de “comunidad” y “Estado”. El antagonismo entre comunitaristas y liberales no es tal.

Por último, abordamos la crítica que, a nuestro juicio, resulta más sólida contra la postura macintyreana, dado que es cierta bajo algunos aspectos. A la luz de Daniel E. Young y Keith Breen, la afirmación es que, a pesar de mostrarse siempre detractor del liberalismo político, MacIntyre sostiene una teoría en la que defiende formas institucionales comprendidas solo desde esa tradición. Nuestra respuesta a Young estuvo basada en evidenciar la —recurrente— confusión que el propio Young tiene entre los conceptos de “comunidad” macintyreana y religiosa, entre Estado y comunidad, y lo “sin Estado” respecto del anarquismo. Para responder a las imputaciones de Breen, en cambio, debimos ser selectivos, puesto que, bajo ciertos aspectos, su crítica resulta certera. Breen considera que, a pesar de sus intenciones, la propuesta macintyreana fomentaría el desarrollo de los *freeriders* en el plano humano, pero también, a nivel institucional, propondría una “neutralidad ostensible” dado el interés, por parte de MacIntyre, en que el Estado no interfiera en lo local. A esto se sumaría el apego que MacIntyre muestra por los logros del liberalismo, como cierto grupo de libertades y derechos. La defensa macintyreana, decíamos, fue selectiva, puesto que efectivamente Breen también confunde el *locus*

de la política con el Estado. El hecho de que MacIntyre reniegue de la interferencia estatal tiene como fundamento ese *locus*, lo que permite configurar una política que no nace del Estado, pero que no es “anti”-estatal. Breen apunta, no obstante, a la ausencia de planificación real de algún mecanismo económico bajo un esquema macintyreano. En ese sentido tiene razón, al igual que Perreau-Saussine y Manent al culpar al escocés por no entregar una hoja de ruta sobre cómo ha de ser el sistema político. Esta es una tarea política, fundamentada por la teoría macintyreana de raigambre aristotélica, pero no realizada del todo. Por último, consideramos coherentes las imputaciones de Breen a la ausencia de interconexión ético-política entre comunidades locales, exponiendo las respuestas de MacIntyre, que recurre a las prescripciones emanadas de la ley natural en sentido tomasiano.

Con esto, hemos podido defender la existencia de una propuesta política *viable* desde la teoría de MacIntyre frente a las críticas que expusimos. Dicho de otro modo, aunque exista la posibilidad cierta de que se desacredite la existencia de una propuesta política *viable* en el trabajo del filósofo escocés, esto no ocurriría por las razones esgrimidas por los autores que analizamos.

Bibliografía

- Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124), 1-19. <https://doi.org/10.1017/S0031819100037943>
- Beadle, R. y Moore, G. (2011). MacIntyre, Neo-Aristotelianism and Organization Theory. *Research in the Sociology of Organizations*, 32, 85-121. <https://doi.org/10.1108/s0733-558x%282011%290000032006>
- Beadle, R. y Moore, G. (2020). *Learning from MacIntyre*. Pickwick Publications.
- Beiner, R. (2000). Community versus Citizenship: MacIntyre's Revolt against the Modern State. *Critical Review*, 14(4), 459-479. <https://doi.org/10.1080/08913810008443569>
- Blackledge, P. y Davidson, N. (2008). *Alasdair MacIntyre's Engagement with Marxism: Selected Writings 1953-1974*. Brill. <https://doi.org/10.1163/ej.9789004166219.i-443>
- Borradori, G. (1994). Nietzsche or Aristotle? Alasdair MacIntyre. En G. Borradori, *The American Philosopher: Conversations with Quine, Davidson, Putnam, Nozick, Danto, Rorty, Cavell, MacIntyre, and Kuhn* (pp. 137-152). R. Crocitto (trad.). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226066493.001.0001>

- Breen, K. (2005). The State, Compartmentalization and the Turn to Local Community: A Critique of the Political Thought of Alasdair MacIntyre. *The European Legacy*, 10(5), 485-501. <https://doi.org/10.1080/10848770500173722>
- Buzeta, S. (2013). *Sobre el conocimiento por connaturalidad*. Servicio de Publicaciones de la Universidad de Navarra. <https://hdl.handle.net/10171/37680>
- Descombes, V. (2013). Alasdair MacIntyre en France. *Revue Internationale de Philosophie*, 264(2), 135-156. <https://doi.org/10.3917/rip.264.0135>
- Frazer, E. y Lacey, N. (1994). MacIntyre, Feminism and the Concept of Practice. En J. Horton y S. Mendus (eds.), *After MacIntyre: Critical Perspectives on the Work of Alasdair MacIntyre* (pp. 265-282). University of Notre Dame Press.
- González, A. M. (2006). *Moral, razón y naturaleza: una investigación sobre Tomás de Aquino*. EUNSA.
- Gregson, J. (2019). *Marxism, Ethics and Politics—The Work of Alasdair MacIntyre*. PalgraveMacmillan. <https://doi.org/10.1007/978-3-030-03371-2>
- Jones, P. (2006). Toleration, Value-pluralism, and the Fact of Pluralism. *Critical Review of International Social and Political Philosophy*, 9(2), 189-210. <https://doi.org/10.1080/13698230600655016>
- Knight, K. (1998). Editor's Introduction. En K. Knight (ed.), *The MacIntyre Reader* (pp. 1-30). University of Notre Dame Press. <https://doi.org/10.2307/j.ctv19m62gb.4>
- Knight, K. (2007). *Aristotelian Philosophy: Ethics and Politics from Aristotle to MacIntyre*. Polity.
- MacIntyre, A. (1959). Hume on "Is" and "Ought". *The Philosophical Review*, 68(4), 451-468. <https://doi.org/10.2307/2182491>
- MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. Duckworth.
- MacIntyre, A. (1988). *Whose Justice? Which Rationality?* University of Notre Dame Press.
- MacIntyre, A. (1990a). The Privatization of Good: An Inaugural Lecture. *The Review of Politics*, 52(3), 344-377. <https://doi.org/10.1017/S0034670500016922>
- MacIntyre, A. (1990b). *Three Rival Versions of Moral Enquiry: Encyclopaedia, Genealogy, and Tradition*. Duckworth.
- MacIntyre, A. (1991). An Interview with Alasdair MacIntyre. *Cogito*, 5(2), 67-73. <https://doi.org/10.5840/cogito19915230>

- MacIntyre, A. (1994). The Theses on Feuerbach: A Road Not Taken. En C. C. Gould y R. S. Cohen (eds.), *Artifacts, Representations and Social Practice* (pp. 277-290). Springer. https://doi.org/10.1007/978-94-011-0902-4_16
- MacIntyre, A. (1998). Politics, Philosophy and the Common Good. En K. Knight (ed.), *The MacIntyre Reader* (pp. 235-254). University of Notre Dame Press.
- MacIntyre, A. (2006a). Poetry as Political Philosophy: Notes on Burke and Yeats. En A. MacIntyre, *Ethics and Politics: Selected Essays, Volume 2* (pp. 159-171). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606670.010>
- MacIntyre, A. (2006b). Social Structures and Their Threats to Moral Agency. En A. MacIntyre, *Ethics and Politics: Selected Essays, Volume 2* (pp. 186-204). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606670.012>
- MacIntyre, A. (2006c). Three Perspectives on Marxism: 1953, 1968, 1995. En A. MacIntyre, *Ethics and Politics: Selected Essays, Volume 2* (pp. 145-158). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606670.009>
- MacIntyre, A. (2006d). Toleration and the Goods of Conflict. En A. MacIntyre, *Ethics and Politics: Selected Essays, Volume 2* (pp. 205-223). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606670.013>
- MacIntyre, A. (2013). Replies. *Revue Internationale de Philosophie*, 67(264) (2), 201-220. <https://doi.org/10.3917/e.rip.264.0201>
- MacIntyre, A. (2016). *Ethics in the Conflicts of Modernity: An Essay on Desire, Practical Reasoning, and Narrative*. Cambridge University Press. <https://doi.org/10.1017/9781316816967>
- Manent, P. (2022). Foreword. En É. Perreau-Saussine, *Alasdair MacIntyre: An Intellectual Biography* (pp. xi-xvii). N. J. Pinkoski (trad.). University of Notre Dame Press. <https://doi.org/10.2307/jj.21995673.4>
- Mulhall, S. y Swift, A. (1996). *Liberals and Communitarians*. Blackwell.
- Perreau-Saussine, É. (2022). *Alasdair MacIntyre: An Intellectual Biography*. N. J. Pinkoski (trad.). University of Notre Dame Press. <https://doi.org/10.2307/jj.21995673>
- Perreau-Saussine, É. (1999). *Alasdair MacIntyre between Aristotle and Marx*. [Manuscrito sin publicar]. Committee on Social Thought, University of Chicago.
- Petersen C., M. (2020). El aristotelismo revolucionario de Alasdair MacIntyre. En F. J. de la Torre y J. M. Loria (eds.), *Alasdair MacIntyre:*

- relecturas iberoamericanas: recepción y proyecciones* (pp. 267-286). Dykinson. <https://doi.org/10.2307/j.ctv17hm9hm.19>
- Pinkard, T. (2003). MacIntyre's Critique of Modernity. En M. C. Murphy (ed.), *Alasdair MacIntyre* (pp. 176-200). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609961.008>
- Pinkoski, N. J. (2019). Manent and Perreau-Saussine on MacIntyre's Aristotelianism. *Perspectives on Political Science*, 48(2), 125-135. <https://doi.org/10.1080/10457097.2018.1563402>.
- Spaemann, R. (1989). *Lo natural y lo racional: ensayos de antropología*. J. Olmo (trad.). Rialp.
- Thompson, C. M. (2021). *The Original Benedict Option Guidebook: Benedict of Nursia's Own Rules for Living Christian Community in a Post-Christian Society*. Marchese di Carabas.
- Widow, J. A. (1984). *El hombre, animal político: el orden social: principios e ideologías*. Academia Superior de Ciencias Pedagógicas de Santiago.
- Young, D. E. (2022). A Dead End in Alasdair MacIntyre's Political Thought? Reconsidering Social Institutions, Pluralism, and the Liberal Tradition. *Journal of Religion, Culture & Democracy*, 9, 1-16.

<https://doi.org/10.21555/top.v760.3370>

El reconocimiento identitario en la filosofía de Hannah Arendt: la primacía del espectador y los límites del agente

The Recognition of Identity in Hannah Arendt's Philosophy: The Primacy of the Spectator and the Limits of the Agent

María Camila Sanabria

Pontificia Universidad Javeriana
Cali, Colombia

maria.sanabriac@upb.edu.co

<https://orcid.org/0000-0001-5506-4106>

Jesús María Carrasquilla Ospina

Pontificia Universidad Javeriana
Cali, Colombia

carrasquilla@javerianacali.edu.co

<https://orcid.org/0000-0003-0778-1754>

Tatiana Saavedra Flórez

Pontificia Universidad Javeriana
Cali, Colombia

tsaavedra@javerianacali.edu.co

<https://orcid.org/0000-0001-7875-1785>

Recibido: 27 - 01 - 2025.

Aceptado: 27 - 04 - 2025.

Publicado en línea: 10 - 07 - 2026.

Cómo citar este artículo: Sanabria, M. C., Carrasquilla Ospina, J. M. y Saavedra Flórez, T. (2026). El reconocimiento identitario en la filosofía de Hannah Arendt: la primacía del espectador y los límites del agente. *Tópicos, Revista de Filosofía*, 76, 345-371. <https://doi.org/10.21555/top.v760.3370>



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

Resumen

En este artículo abordamos el problema del reconocimiento identitario a partir de la filosofía de Hannah Arendt. Analizamos la posible incompatibilidad entre la autopercepción de un actor (individual o colectivo) y la percepción que los demás tienen de él. Sostenemos que este problema puede explorarse tanto desde la *vita activa* como desde la *vita contemplativa*. Argumentamos que el reconocimiento, en tanto fenómeno político, depende más de la mirada del espectador que de la voluntad del actor. Aunque no siempre responde a una intención deliberada, su búsqueda requiere coherencia en la conducta y apelación a la persuasión, evitando la violencia. Esta, al estar ligada a la lógica de la fabricación en Arendt, tiende a anular el carácter político del reconocimiento y las condiciones necesarias para que tenga lugar.

Palabras clave: Hannah Arendt reconocimiento; acción; juicio; espectador; *vita contemplativa*; *vita activa*; actor; identidad; trabajo.

Abstract

In this paper, we address the problem of recognition in relation to identity through the lens of Hannah Arendt's philosophy. We analyze the possible incompatibility between an actor's self-perception (whether individual or collective) and the perception others have of her. We argue that this problem can be explored through the lens of both *vita activa* and *vita contemplativa*. We maintain that recognition, as a political phenomenon, depends more on the gaze of the spectator than on the will of the agent. Although it does not always stem from deliberate intention, its pursuit requires coherence in conduct and an appeal to persuasion, avoiding violence. Violence, insofar as it is tied to the logic of work in Arendt's thought, tends to nullify the political character of recognition and the necessary conditions for it to take place.

Keywords: Hannah Arendt; recognition; action; judgment; spectator; *vita contemplativa*; *vita activa*; agent; identity; work.

1. Introducción¹

Estamos en un auge de teorías del reconocimiento y de éticas de la comunicación (Quintana, 2010, p. 430). De acuerdo con Fraser (2006), independientemente de los distintos asuntos que se abordan desde el término “reconocimiento” (como el matrimonio homosexual o las reivindicaciones territoriales indígenas), las concepciones sobre este buscan develar las bases normativas de las reivindicaciones políticas. Las distintas concepciones sobre dicho término, indica Fraser (2006), remiten a su relación con el concepto de “redistribución”. Al respecto, Honneth (1997) propone que la distribución se deriva de la lucha por el reconocimiento, y entiende a este último como categoría moral; Fraser (2006), por el contrario, no considera que la distribución se pueda subsumir en el reconocimiento y propone la irreductibilidad de ambas categorías.

En este artículo no abordamos la relación entre reconocimiento y distribución puesto que nuestro interés no apunta a proponer un modelo de justicia identificando sus bases normativas.² No obstante, entendemos la veta hegeliana que está en el trasfondo del problema del reconocimiento e indicamos la influencia del pensador alemán en dos obras cruciales de nuestra época: la de Charles Taylor y la de Judith Butler.³

¹ Artículo derivado como producto del proyecto de investigación “La narración de sí como camino de reconciliación en el marco de las violencias en Colombia”, código CUT03, avalado y financiado por la Pontificia Universidad Javeriana, Cali, Colombia. Este artículo es un producto del grupo de investigación *De Humanitate* de la Facultad de Humanidades y Ciencias Sociales de la misma Universidad.

² No obstante, la propuesta de Axel Honneth sobre la lucha por el reconocimiento sigue siendo sugerente en términos de comprender la veta hegeliana para entender los conflictos sociales, pues el reconocimiento se puede leer en clave de justicia: cómo reivindicar nuestro ejercicio de libertad, nuestro *Derecho de la libertad*, según Honneth (2014), en términos de identidad moral y psicológica y de intersubjetividad política y democrática.

³ El problema del reconocimiento tiene hoy una veta clara en Hegel en su *Fenomenología del espíritu*, que enfatiza que la conciencia como autoconciencia busca ser reconocida en otro o en otra autoconciencia. Esto les da un sello intersubjetivo a las relaciones humanas, que no solo busca reivindicar una

Nuestro propósito es ofrecer una aproximación al fenómeno del reconocimiento que cuestionaría la plausibilidad de la lucha por este mediante la violencia, y, en su lugar, propondría la *constancia* en la autopresentación y la *persuasión* como alternativas para su búsqueda. Para ello, encontramos en los escritos de Hannah Arendt unos desarrollos conceptuales descriptivo-ontológicos alrededor de la pluralidad, la acción, el juicio del espectador y el reconocimiento mutuo en un espacio de aparición que se hacen importantes a la hora de comprender dicha noción. Aunque en la filosofía arendtiana no hay propiamente un desarrollo explícito de algo que pudiera llamarse una teoría del reconocimiento, se concuerda con que “[...] sus planteamientos pueden ofrecer elementos que podrían leerse a la luz de esta noción” (Quintana, 2010, p. 430). Una de estas lecturas es realizada por Muñoz (2020), quien ofrece aportes que vinculan las nociones de “identidad” y “reconocimiento”. La autora retoma el planteo arendtiano según el cual el espacio de aparición es condición para el reconocimiento mutuo como iguales y plurales, y afirma que “aparecer es afirmar la identidad propia gracias al reconocimiento” (Muñoz, 2020, p. 67).

identidad, sino también entrar en interacción con otras y poder construir así un proyecto común que impacte en la sociedad. Esta veta hegeliana ha tenido repercusiones en las obras de Charles Taylor y de Judith Butler. En el caso de Taylor (1993), se explora en *Política del reconocimiento* la dimensión dialógica de la construcción de la identidad humana pero también se destaca el reconocimiento como una necesidad humana vital, no como una mera cortesía y esto tiene implicaciones hermenéuticas (fusión de horizontes) y también ético-políticas para la praxis social. Para el caso de Butler (2012), en *Sujetos del deseo* se indaga el deseo como punto de encuentro entre la filosofía y el sujeto, pero ese mismo deseo trastoca la armonía interna del yo y esto tiene implicaciones para el mismo deseo de reconocimiento. Según Butler, el sujeto de Hegel despliega una movilidad crítica y una ambigüedad, que se expresa en un caso como el de Antígona (que daría para múltiples interpretaciones). No obstante, la falta de reconocimiento de los sujetos en la sociedad y una lucha que no garantiza este reconocimiento, pues los sujetos no están alineados en su condición intersubjetiva, puesto que no establecen un vínculo con el otro, así como este desgaste del sujeto en procurar el reconocimiento del otro en los ámbitos moral y político, pueden provocar que el mismo sujeto se pierda a sí mismo en este acontecer vital. Creemos que Hannah Arendt aporta una lectura diferente que puede enriquecer el problema del reconocimiento.

En este texto ofrecemos una lectura del fenómeno del reconocimiento visto tanto a la luz de la *vita activa* como de la *vita contemplativa* y lo delimitamos dentro del marco conceptual de la identidad, bien sea personal o colectiva.⁴ Entendemos el fenómeno del reconocimiento como la expectativa de correspondencia entre la autopercepción del individuo-grupo y la percepción que tienen los espectadores sobre aquellos. En su obra *Caminos del reconocimiento*, Ricœur advierte sobre la identidad: “Diré, en primer lugar, que se trata, ahora y siempre, de identificación. Ser reconocido, *si alguna vez acontece*, sería para cada uno recibir la plena garantía de su identidad gracias al *reconocimiento por parte de otro* de su dominio de capacidades” (Ricœur, 2006, p. 312; nuestro énfasis).

Cabría señalar dos ideas a partir de este fragmento. Primero, nos interesa llamar la atención sobre la aclaración “*si alguna vez acontece*”, la cual encontraríamos en este planteamiento a partir de la relación que se establece entre el reconocimiento y la categoría arendtiana de “acción”. Como veremos en el desarrollo del artículo, el reconocimiento no podría caracterizarse desde la predecibilidad enmarcada en la categoría arendtiana de la fabricación. Por el contrario, en tanto impredecible e incontrolable por parte del agente, correspondería con la categoría de la acción y podría comprenderse desde la noción

⁴ Honneth (1997) distingue tres formas de reconocimiento: el amor, el derecho y la solidaridad. Aunque los tres modelos están interrelacionados con la configuración de la identidad personal, la categoría de la solidaridad podría aproximarse a la concepción de “reconocimiento” de la cual parte nuestra propuesta. Respecto a la solidaridad, indica el autor: “Tanto Hegel como Mead han diferenciado, respecto del amor y de la relación jurídica, una más amplia forma de reconocimiento de la que se eligen descripciones ocasionalmente distintas, pero en cuya determinación de función siempre han coincidido: para poder conseguir una ininterrumpida autorrelación, los sujetos humanos necesitan, más allá de la experiencia de la dedicación afectiva y del reconocimiento jurídico, *una valoración social que les permite referirse positivamente a sus cualidades y facultades concretas*” (Honneth, 1997, p. 148; nuestro énfasis). La valoración social se aproximaría en tanto apunta a la correspondencia entre lo que el individuo percibe de sí y lo que la sociedad le reconoce. No obstante, el énfasis de la propuesta de Honneth parecería enfocarse en el objetivo práctico del reconocimiento, a saber, el compromiso y la participación de los individuos en objetivos compartidos por un grupo que reconoce y valora las capacidades de sus miembros.

de “acontecimiento”. Segundo, su referencia al reconocimiento de la identidad correspondería con el planteamiento que desarrollamos en el presente texto, a saber, la primacía del espectador sobre el actor en el fenómeno del reconocimiento. Este, como menciona Ricœur, consistiría en una identificación *recibida* por parte de *otro*. En este texto hacemos referencia a este “otro” desde la noción arendtiana del espectador, quien contaría con la posición privilegiada tanto para juzgar como para identificar el *quién* de la acción. Ambas ideas serán desarrolladas en el segundo y tercer apartado de este texto.

Para esta aproximación, es preciso abordar el fenómeno desde la *vita activa* y desde la *vita contemplativa* teniendo en cuenta el siguiente problema: en la filosofía arendtiana, como veremos, el reconocimiento estaría relacionado con la apariencia. Por una parte, reconocer implica que el espectador comparta la percepción que el reconocido tiene de sí; esta condición nos remitiría a la *vita activa*. En otras palabras, tendríamos que comprender el reconocimiento desde la categoría de “acción”, dada la similitud con la revelación del *quién* en la esfera pública.

Por otra parte, reconocer es decir algo sobre nuestra percepción de un fenómeno. En este sentido, reconocer también es juzgar, lo que nos remitiría a la *vita contemplativa*. Esta concepción se sostiene, al igual que la anterior, en la apariencia como precondition originaria del reconocimiento. La apariencia sería, en la filosofía arendtiana, el único elemento que el espectador tendría a la mano para reconocer algo en el fenómeno; asimismo, la propia apariencia sería la única *arma* para *luchar* por el reconocimiento. Puesto que todo juicio es un *dokei moi*, al estar basado en la apariencia, esto nos permitiría delimitar los alcances del agente como una posible lucha por el reconocimiento.⁵

⁵ Si bien no nos proponemos mantener un diálogo con los argumentos de Honneth que resaltan la importancia de la lucha por el reconocimiento o la ubican como consecuencia de la negación de este, cabría considerar que en su argumentación podríamos identificar el reconocimiento como el acto del espectador: “Aquí está el consenso dominante en el objetivo práctico, que súbitamente engendra un horizonte de valor intersubjetivo en el que cada uno aprende a reconocer la significación de las capacidades y cualidades del otro” (Honneth, 1997, p. 157). El reconocimiento, desde esta perspectiva, implica la reciprocidad. No obstante, la posición de aquel que reconoce podría coincidir con la figura del espectador; nos situamos como espectadores de los otros, los percibimos para poder ver en ellos sus capacidades y cualidades.

Cuando pensamos el reconocimiento desde la categoría arendtiana de la acción, parecería que la lucha por el reconocimiento, aunque legítima, podría no lograr sus objetivos. El *quién* sería reconocido en un ejercicio de juicio retrospectivo. El reconocimiento se caracterizaría por ser impredecible y frágil. Estaría relacionado con la forma de la tragedia (Arendt, 2006a). Pero cuando lo pensamos como un juicio, el agente (como agente y como espectador) parecería tener la posibilidad de lograr ser percibido como se percibe a sí mismo gracias a la persuasión y a la constancia.

Para desarrollar lo anterior, dividimos el texto en tres momentos. En el primero, exponemos lo que podría ser una noción arendtiana de “reconocimiento” a la luz de la categoría de “acción”. En el segundo momento, reconstruimos la noción de “reconocimiento” a partir de las características de la facultad de juicio. En tercer y último lugar, presentamos las conclusiones retomando la apariencia como condición originaria de ambas perspectivas.

2. Caracterización de la noción de “reconocimiento” a partir de la categoría de “acción”

En este apartado proponemos que el reconocimiento podría ser entendido a partir de algunos elementos de la categoría arendtiana de “acción” y no desde la categoría de la fabricación, y concluimos que la violencia no es un medio adecuado en la lucha por el reconocimiento. En primer lugar, mostramos por qué este puede ser entendido a partir de la acción y no de la fabricación. En segundo lugar, advertimos la inevitabilidad del conflicto en una concepción del reconocimiento que surge en la trama de las relaciones humanas. En tercer y último lugar, retomando la caracterización de la fabricación, concluimos que, a pesar del conflicto inherente a la vida política, la violencia no es un medio legítimo en la lucha por el reconocimiento, precisamente porque anula el carácter político que lo posibilita.

Comenzaremos retomando las características de la categoría arendtiana de la fabricación para sostener por qué, si entendemos el reconocimiento desde esta, estaríamos legitimando la violencia como forma de lucha y esto, como última consecuencia, sería el final de la vida política. La primera característica de la fabricación, que no correspondería con el reconocimiento, es la posibilidad de realizarse en aislamiento. De acuerdo con Arendt (2011), aunque todas las actividades

están condicionadas por el hecho de que los seres humanos viven juntos, tanto la labor como la fabricación de objetos pueden darse en el aislamiento. Sin embargo, no podría haber acción —y según nuestra propuesta, tampoco reconocimiento— fuera del cuerpo político. La acción, indica Arendt (2011), depende de la constante presencia de los demás.

La segunda característica de la fabricación es la violencia. Esta se aplica a la materia, se la transforma con el uso de instrumentos para producir objetos: “Este elemento de violación y de violencia está presente en toda la fabricación, y el *homo faber* creador del artificio humano siempre ha sido un destructor de la naturaleza” (Arendt, 2011, p. 168). Desde esta perspectiva, el reconocimiento sería algo que se impone usando los medios que hagan falta para ello.

Esto nos lleva a la tercera característica de la fabricación, la categoría “medios-fines”. Según Arendt (2011, p. 170), la cosa trabajada es un producto final porque en esta culmina el proceso de producción y porque este proceso solo es un medio para la producción de un fin. La mirada del *homo faber* nunca va hacia atrás, sino que siempre mira hacia el futuro. Opera con un modelo aplicado violentamente a la materia y prevé el resultado a partir de la guía dada por el modelo en el proceso de fabricación (Arendt, 2011). Esto significa que comprender el reconocimiento como un fin al que se llega por ciertos medios, luchar y forzar hasta conseguirlo es, por definición, un acto de violencia.

A diferencia de la fabricación, la categoría arendtiana de “acción” hace referencia a la posibilidad que tienen todos los seres humanos de comenzar algo nuevo, de tomar una iniciativa, esto por el hecho de haber nacido, ser únicos y diferentes a todos los demás que hayan vivido, siempre y cuando se revelen en el espacio de aparición mediante palabra o acto (Arendt, 2011, p. 206)

Arendt (2011) caracteriza la acción desde la espontaneidad y la no controlabilidad. Ambas están relacionadas con la posibilidad de reacción de los otros. Estos pueden re-accionar espontáneamente a un acto o palabra, tejiendo una trama de relaciones humanas cuyo final no sería predecible ni controlable, como lo es la fabricación. En su texto *Sobre la violencia*, Arendt (2006b) aborda la distinción entre acción y fabricación indicando que la finalidad de la acción humana, a diferencia de los bienes fabricados, nunca puede ser fiablemente prevista. Las acciones y los discursos que tienen lugar en la esfera política tienen reacciones en los otros que, a su vez, también actúan, y estas consecuencias no se

pueden revertir o deshacer como con los productos de la fabricación. Así se teje una trama de relaciones humanas en la que algún agente, *espontáneamente*, puede hacer o decir algo sin tener control sobre las consecuencias y sin saber que ha comenzado algo nuevo.

El reconocimiento, aunque no esté necesariamente relacionado con el carácter de novedad de la acción, podría ser entendido en términos de espontaneidad y no controlabilidad. Se trataría de un fenómeno en el que los agentes se revelan, es decir, manifiestan quiénes son mediante acto y palabra, pero no pueden predecir cómo van a ser percibidos y reconocidos. Con respecto a la acción, un agente puede hacer o decir algo esperando conducir a cierta reacción en los otros, pero no podrá estar seguro de que así será. Del mismo modo, los agentes pueden procurar ser reconocidos como se perciben a sí mismos, sin embargo, no podrán controlar la percepción ni los discursos interpretativos de los otros.

Esta no controlabilidad se comprende aún más desde el concepto de “revelación”, perteneciente a la categoría de “acción”. Lo que se revela en la esfera pública es la identidad de los agentes. Los agentes no son puramente algo, sino que son alguien. A partir de allí se comprende la vida humana como un proceso permanente de revelación mutua entre todos aquellos que habitan y comparten el mundo humano. En esta línea, Muñoz (2020) interpreta que, según Arendt, la identidad es un proceso llevado a cabo en el espacio público, donde los individuos muestran su singularidad mediante palabra y acto.

En otras palabras, el espacio de aparición resulta ser condición tanto para la acción como para la revelación de los agentes y de la realidad misma. Según Birulés (2007, p. 73), la esfera pública es, para Arendt, el espacio donde la realidad se revela. Se trata de un espacio fenoménico porque allí los seres humanos se exponen, aparecen, se hacen visibles ante los demás y los demás aparecen, se exponen y se hacen visibles ante cada uno como en una comunidad de mutua revelación: “[...] el espacio de aparición cobra existencia siempre que los hombres se agrupan por el discurso y la acción” (Arendt, 2011, p. 225).

Con este elemento de la revelación insistimos en la plausibilidad de entender el reconocimiento a partir de la acción. Ambos, reconocimiento y acción, emergen espontáneamente en el mundo de las apariencias⁶ sin

⁶ En la filosofía arendtiana encontramos una doble terminología para referir a la noción de “mundo”: *world of appereances* y *common world*. Intérpretes como Campillo (2005), Forti (1999), Garrido (2020), Narváez Álvarez (2020),

tratarse de un proceso controlado, como ya vimos, y ambos comparten una doble condición: la manifestación en el mundo de las apariencias, así como la revelación del quién mediante el discurso y la acción. En efecto: “El descubrimiento de ‘quién’ en contradicción al ‘qué’ es alguien [...] está implícito en todo lo que ese alguien dice y hace” (Arendt, 2011, pp. 208-209). Sin embargo, como señala Arendt, el quién parecería ocultarse parcialmente al sí mismo, pero revelarse con más claridad a los otros. Es como si la respuesta a la pregunta “¿quiénes somos?” tuviera en cuenta la perspectiva de quienes nos ven y no solamente la nuestra:

[...] es más que probable que el “quién”, que se presenta tan claro e inconfundible a los demás, permanezca oculto para la propia persona, como el *daimon* de la religión griega que acompañaba a todo hombre a lo largo de su vida, siempre mirando desde atrás por encima del hombro del ser humano y por lo tanto sólo visible a los que éste encontraba de frente (Arendt, 2011, p. 209).

Ahora bien, partir de las relaciones entre revelación, acción y discurso en el espacio de aparición que se ha descrito, podría plantearse que el fenómeno del reconocimiento comparte las mismas características de la acción. La revelación del quién, oculta para el agente y visible para los otros, apunta a la relación entre el reconocimiento y la posibilidad del conflicto.⁷ En efecto, cuando la imagen que un agente tiene de sí no

Rodríguez Suárez (2009) y Sorrentino (2009) abordan esta distinción. A partir de estas lecturas, la característica fundamental del mundo entendido como *world of appearances* parecería ser el carácter de apariencia, mientras que la idea central en los estudios del mundo entendido como *common world* parecería ser el concepto de “pluralidad”. Ambas acepciones, sin embargo, comparten el carácter fenoménico del mundo, el cual es retomado en esta investigación.

⁷ Si bien Arendt no ofrece una definición sistemática de “conflicto”, proponemos esbozar esta noción tomando como fundamento las nociones de “pluralidad” y “realidad” como espacio de aparición. En el acápite “La esfera pública y la privada”, Arendt (2021) afirma que la realidad de la esfera pública radica en la presencia simultánea de una pluralidad de perspectivas en las que se presenta el mundo común. La autora enfatiza que quienes están presentes ocupan diferentes posiciones y expresa que “el puesto de uno puede no coincidir más con el de otro que la posición de dos objetos. De lo visto y oído por otros deriva su significado del hecho de que todos ven y oyen desde una posición

coincide con la que es percibida por los demás, puede generarse una tensión que afecta directamente su identidad. En este sentido, Honneth define como “menosprecio” al acto del no reconocimiento, y señala que esta negación por parte del espectador repercute en la identidad personal de quien es percibido.⁸

Un agente en particular puede advertir que la narrativa de sí no coincide con el discurso que se da en el espacio de aparición. En este punto puede surgir el conflicto de las interpretaciones, ya que la revelación de los agentes ocurre en un espacio de aparición que se configura a partir de la pluralidad. Al respecto, Sánchez (2005, p. 224) entiende el conflicto como una posibilidad que siempre está presente debido a la pluralidad, unicidad y diferencia de los individuos.

Parece que la categoría del conflicto es inherente a la vida política, pues es consecuencia lógica de la pluralidad de seres humanos que se revelan mutuamente en el espacio de aparición, pero también de la pluralidad de espectadores. En este sentido, habría una proximidad entre ambos fenómenos políticos, entre el conflicto y el reconocimiento, pues este último está condicionado por el espacio de aparición, lo que implica la pluralidad de perspectivas desde las que se percibe un mismo fenómeno:

Nuestra sensación de la realidad depende por entero de la apariencia y, por lo tanto, de la existencia de una esfera pública. Solo donde las cosas pueden ser vistas por muchos en una variedad de aspectos y sin cambiar su identidad, de manera que quienes se agrupan a su alrededor saben que ven lo mismo en total diversidad, solo allí aparece auténtica y verdaderamente la realidad mundana (Birulés, 2007, p. 73).

diferente” (Arendt, 2021, p. 218). Partiendo de esta concepción de la realidad mundana, comprendemos el conflicto, desde la perspectiva arendtiana, como la posibilidad inherente al disenso que surge inevitablemente de la pluralidad de puntos de vista.

⁸ Honneth (1997) identifica tres formas de menosprecio: el maltrato físico, la exclusión de derechos y la negación del valor social. Nuestra interpretación es que las dos últimas podrían situarse dentro de la pluralidad de narrativas que surgen como consecuencia de la pluralidad de espectadores. En este sentido, si bien no negamos la legitimidad de la lucha por el reconocimiento, insistimos en que el conflicto entre narrativas es consecuencia del aparecer ante los otros.

Además de esta multiplicidad de perspectivas, que, contrario a un obstáculo, es la condición de la misma realidad, la identidad de los agentes en la esfera pública es una identidad narrada. La narrativa teje una historia que se cuenta acerca de los agentes o grupos que han dicho o hecho algo y se cuentan historias sobre ello; en palabras de Simona Forti: “La narrativa es en sustancia un artificio lingüístico que reconstruye aquello que ha sucedido en la historia a través de una trama que privilegia a los agentes humanos más que los procesos impersonales [...]” (Forti, citada en Birulés, 2007, p. 181). En las historias se da la revelación de los agentes. Las historias aparecen en la esfera pública y las reacciones a ellas no son previsibles. El quién de la acción, pues, la identidad de los seres humanos como seres únicos, se revela en esa identidad narrada. Pero una identidad implícita en estos discursos está siempre sujeta a interpretación.

Ahora bien, dada la pluralidad y la no controlabilidad del reconocimiento como narración del proceso de revelación, es inevitable el desbalance entre la narrativa que se da en el espacio de aparición y la propia narrativa de la persona que también se narra a sí misma porque también quiere saber quién es. Ello no indica que la narrativa personal sea privada, sino que hay pluralidad de narrativas y ellas pueden entrar en conflicto.

En suma, el reconocimiento comparte con la acción el ser frágil, imprevisible y no controlable, además de su carácter político que implica el revelarse entre los otros y ser percibido por estos. Por último, acción y reconocimiento coinciden en su carácter retrospectivo, pues tanto el significado de la historia como el del quién de los agentes se revela desde la mirada que vuelve hacia el pasado: “Solo de modo retrospectivo, solo desde el punto de vista del historiador que mira hacia atrás, desde la perspectiva privilegiada del presente, la masa caótica de sucesos pasados se convierte en un relato coherente” (Birulés, 2007, p. 180).

Concluimos esta sección insistiendo en que, a pesar de la inevitabilidad del conflicto, la violencia no es un medio legítimo en la lucha por el reconocimiento y que aceptarlo significaría partir de una concepción del reconocimiento que concuerde con la categoría de la fabricación. En una concepción del reconocimiento vista desde la categoría de la acción, como la que aquí proponemos, las narrativas sobre los diferentes grupos componentes del cuerpo político no se pueden fabricar ni controlar, como los objetos del mundo. Entonces ¿en

qué consistiría una lucha por el reconocimiento? De esto nos ocupamos en el siguiente apartado.

3. El reconocimiento como forma de juicio

Hasta aquí hemos abordado el tema del reconocimiento desde la *vita activa*, a la vez que hemos advertido el paso hacia la *vita contemplativa*. Desde esta perspectiva, el reconocimiento es una actividad en la que el espectador expresa algo sobre lo percibido, esto es, un juicio. De ahí que, aunque comparta características con la acción, también lo hace con la facultad de juzgar.⁹ Para limitar nuestra investigación, en esta segunda parte responderemos a la pregunta por los límites del agente: ¿podemos conseguir que los demás nos reconozcan tal y como nos reconocemos?

Antes de continuar con el desarrollo argumentativo, presentaremos una clarificación sobre dos presupuestos conceptuales: la facultad de juicio y la noción arendtiana de “dos-en-uno”. El siguiente fragmento resume algunas características de la concepción arendtiana de “juicio”: “La manifestación del viento del pensar no es el conocimiento; es la capacidad de distinguir lo bueno de lo malo, lo bello de lo feo” (Arendt, 2002b, p. 215). El juicio es la manifestación del pensamiento, es decir, su expresión en el mundo de las apariencias. Al pensar nos retiramos en una aparente distracción, pero con el juicio, que para Arendt (2002b) es la más política de las facultades mentales, regresamos a este mundo y damos cuenta de nuestras opiniones hablando con los otros o actuando conforme a ellas. Para Arendt, el juicio moral y estético se da del mismo modo, de manera reflexionante y no determinante, por lo que la actividad de pensar es su precondition. Se juzga después de haber pensado, cuando se han puesto en cuestión los prejuicios: “La consecuencia de esta peculiaridad es que el pensamiento tiene inevitablemente un efecto destructivo, socava todos los criterios establecidos, todos los valores y las pautas del bien y del mal, en suma, todos los hábitos y reglas de conducta que son objeto de la moral y de la ética” (Arendt, 2002b, p. 198).

⁹ Intérpretes como Beiner (2003) y Sidorova (2019) vinculan los últimos estudios arendtianos sobre el juicio con la figura del espectador. Sidorova (2019) hace énfasis en el carácter narrativo del juicio, destacando el vínculo entre el discurso y la narración como medios de interpretación del mundo del cual se es parte. Esta presencia del espectador en el espacio de aparición reivindica la articulación entre *vita activa* y *vita contemplativa* que ofrecemos en la presente investigación.

La condición del pensamiento, a su vez, es la autoconciencia (Arendt, 2002b). Arendt toma la concepción socrática del pensar y lo entiende como un diálogo del yo consigo mismo. Para pensar, el yo debe ser consciente de sí mismo. Al pensar, el yo actualiza la diferencia inserta en su unicidad. Mientras que al estar con los otros es solamente uno, durante el pensamiento esa unicidad se escinde y permite un diálogo, siempre y cuando el yo (*self*) y el yo (pensante) estén en armonía; dicho de un modo simple, siempre y cuando la persona no haya transgredido sus propios límites. Como argumenta Arendt (2002b), esta coherencia o incoherencia solo es posible, precisamente, porque la unicidad está constituida por un dos-en-uno. Hasta aquí la clarificación.

Ahora bien, refiriéndose a los premios, Arendt expresa: “El hecho de que, en general, una persona aparezca en público y que este público la reciba y confirme no es algo evidente” (Arendt, 2006a, p. 13). Si nos detenemos, toda la afirmación soporta la caracterización del reconocimiento a la que habíamos llegado. Primero, el agente debe *aparecer en público*: el espacio de aparición es una condición para el reconocimiento. Segundo, el público *confirma*, es decir, el reconocimiento parecería ser algo que está en manos de los espectadores. Tercero, esta confirmación no es algo evidente, no es previsible como lo es la producción en la categoría de la fabricación. En contraposición a esta última, que, por tener un “fin” en el sentido de una finalidad, mira hacia el futuro, el reconocimiento mira hacia el pasado en un ejercicio retrospectivo. Por ello, como vimos, no cumple con la categoría “medio-fin” de la fabricación ni usa la violencia aplicada a la materia para producir un objeto. El reconocimiento no se consigue con las manos ni con los instrumentos, no es un fin previsto o proyectado, sino que se revela tras volver sobre lo ocurrido:

Mencionamos esta tragedia¹⁰ deliberadamente porque representa, más que las otras formas literarias, un proceso de reconocimiento. El héroe trágico, al volver a vivir lo que se ha cometido en forma de sufrimiento, llega a conocerlo, y en este *pathos*, sufriendo nuevamente el pasado, la red de los actos individuales se transforma en acontecimiento, en un todo significativo (Arendt, 2006a, p. 31).

¹⁰ Arendt hace referencia a la novela *Una fábula*, de William Faulkner.

El hecho de que Arendt (2006a) equipare el “efecto trágico” con el proceso de reconocimiento es un indicador de los límites del agente. Si el reconocimiento es algo que llega después, cuando se ve la historia como un todo y su significado se revela, entonces los esfuerzos del agente en su “lucha”, aunque legítimos, podrían no ser tan efectivos. De nuevo, porque el reconocimiento —al menos visto desde la filosofía arendtiana— dependería más de quien reconoce que de quien pide ser reconocido tal y como se percibe a sí mismo o, en su defecto, de la relación de percepción que se construye entre ambos, en la que entran en juego la proyección de las imágenes de sí mismo y del otro, como en una especie de prisma.

No obstante, cuando nos aproximamos a una concepción de “reconocimiento” entendida como una forma de juicio, aparece una opción para el actor y su lucha por el reconocimiento. Para situarnos en esta perspectiva, es preciso iniciar abordando la noción arendtiana de “apariencia”, ya que todos nuestros elementos (actores, espacio de aparición, percepción del espectador) se relacionan con ella. En el apartado de *La vida del espíritu* que Arendt titula “La naturaleza fenoménica del mundo”, expresa:

El mundo en el que nacen los hombres abarca muchas cosas, naturales y artificiales, vivas y muertas, efímeras y eternas; todas tienen en común que *aparecen*, lo que significa ser vistas, oídas, tocadas, catadas, y olidas, ser percibidas por criaturas sensitivas dotadas de órganos sensoriales adecuados. Nada puede aparecer; el término *apariencia* carecería de sentido si no existiesen receptores para las apariencias, criaturas vivas capaces de percibir, reconocer y reaccionar [...]. No existe nada ni nadie en este mundo cuya misma existencia no presuponga un *espectador* (Arendt, 2002b, p. 43).

Este fragmento nos señala, al menos, dos ideas. Primero, la relación entre apariencia y reconocimiento. Esto mismo lo habíamos visto al acercarnos desde la perspectiva de la *vita activa*. El espacio de aparición y la apariencia misma son condición para el reconocimiento. Lo que nos lleva a la segunda idea: que el reconocimiento es un acto de los receptores de apariencia, es decir, es un proceso realizado por quien percibe aquello que aparece y *no* por lo que aparece. Esto nos remite a

la distinción entre la posición del espectador y la del actor, para lo cual cabe tener en cuenta la siguiente caracterización:

En primer lugar, el actor debe poder representar el papel en la obra, mientras que el espectador se vincula al significado de lo mostrado. En segundo lugar, el actor depende de la opinión del espectador otorgándole a este último su capacidad autónoma; en tercer lugar, el ejercicio de la autonomía hace posible la participación activa del espectador en tanto que trasmite lo que ve. Los espectadores no están solos, pero tampoco dependen de alguien, pues elaboran críticamente, por ellos mismos, el juicio respecto a lo obrado. Hasta aquí aparecen las funciones de las actividades mentales que hacen posible pensar en un mundo de apariencias (Barrior, 2010, p. 109).

Estas diferencias apuntan a la primacía del espectador, frente al actor, en el fenómeno del reconocimiento. El actor se muestra, aparece, pero depende del espectador. Este es quien percibe el significado. En tanto hay pluralidad de espectadores, también hay pluralidad de percepciones que son posteriormente comunicadas. La dependencia del actor nos regresa al aspecto de ocultamiento que habíamos visto en el apartado anterior. Como indica Arendt (2002b), parte del quién de la persona es como el *daimon* de la religión griega: visible a los otros, pero no a sí mismo. Tal vez por ello el reconocimiento sea algo que dependa más de los espectadores que del actor mismo, aunque sea este quien lo reclame. Es decir, la última palabra sobre quién es alguien no la tiene ese alguien, sino quienes lo ven. Pero ¿qué es lo que ven los espectadores?

Los seres vivos *hacen su aparición* como actores en un escenario preparado para ellos. El escenario es el mismo para todos los que están vivos, pero *parece* distinto para cada especie, diferente incluso para cada individuo. El “parecer”—el “me parece”, *dokei moi*— es el modo, quizá el único posible, de reconocer y percibir un mundo que se manifiesta (Arendt, 2002b, p. 45).

Ven el fenómeno de una forma específica. Es decir, no el todo, no la Verdad, ningún absoluto. Lo que cada espectador percibe y reconoce es la forma en la que ese objeto —el mismo que se aparece ante los

demás— se le aparece a él. Los espectadores reconocen lo que *les parece*. De ahí que, en este apartado, veamos el reconocimiento desde la *vita contemplativa* como una forma de juicio. El reconocimiento, como el juicio, es un “me parece” de carácter político.

Aparecer siempre implica parecerle algo a los otros, y este parecer cambia según el punto de vista y la perspectiva de los espectadores. En otras palabras, todo objeto que aparece adquiere, en virtud de su propia condición de aparecer, una suerte de disfraz que puede, pero no tiene por qué, ocultarlo o desfigurarlo. El “parecer” se corresponde con el hecho de que cada apariencia, a pesar de su identidad, es percibida por una pluralidad de espectadores (Arendt, 2002b, pp. 45-46).

Es de carácter político porque tiene lugar en la esfera pública donde un mismo fenómeno se manifiesta ante una pluralidad de espectadores que lo perciben desde distintas perspectivas. Con la metáfora del disfraz, Arendt hace referencia a la variabilidad de las apariencias que puede tener un mismo objeto sin que estas manifestaciones refieran a una suerte de falsificación u ocultamiento.

Ahora bien, ¿qué nos dice esta metáfora en relación con el reconocimiento visto desde la perspectiva del juicio? Que la apariencia se muestre como un disfraz no alude a la posibilidad de que los agentes puedan controlar la percepción de los espectadores. Al no tratarse de un ocultamiento o desfiguración, el disfraz no cumple la función de que un objeto A logre verse como un objeto B. La metáfora del disfraz muestra que lo que es percibido nunca es percibido de un solo modo.

La lucha por el reconocimiento, desde la filosofía arendtiana, no se encontraría en un intento de engaño a los espectadores por parte de los actores, pues cada disfraz corresponde a lo que “le parece” a cada espectador. Coincidimos con el planteo de Muñoz (2020), según el cual la identidad no es el resultado de una fabricación, una conducta regida por medios-fines, sino que es el resultado de la aparición en público y, por tanto, estaría condicionada “por el reconocimiento de nuestra singularidad que hacen los otros” (Muñoz, 2020, p. 112). Ahora bien, aunque el agente no cuente con la posibilidad de “disfrazarse”, siguiendo a Arendt (2002b), “hasta cierto punto” puede decidir cómo desea aparecer:

Además de la necesidad de autoexhibirse, mediante la cual los seres vivos se adecuan en un mundo de apariencias, los humanos también se *presentan*, de obra y palabra, y así indican cómo *desean* aparecer, qué es según ellos apropiado para ser visto y qué no lo es. Este elemento de elección deliberada de lo que se puede mostrar y lo que hay que ocultar parece específicamente humano. *Hasta cierto punto* se puede escoger cómo aparecer ante los otros, y esta apariencia no es, en modo alguno, la manifestación externa de una disposición interior; si así fuera, probablemente todos actuaríamos y hablaríamos de manera idéntica (Arendt, 2002b, p. 58).

Al elegir lo que dice y lo que hace, el agente en cierto sentido elige cómo aparece ante los otros. De ahí que Arendt (2002b) señale la diferencia entre autoexhibición y autopresentación, relacionando a la primera con un tipo de exposición de cualidades que tiene el ser vivo. Esta podría ser la exposición del animal *laborans*, mientras que la segunda, requeriría cierto grado de autoconciencia y autorreflexión.

La relación entre el reconocimiento y la *vita contemplativa* se presenta, ahora, con más claridad. El hecho de que el reconocimiento, la posibilidad de elegir cómo ser visto, dependa de la autopresentación y que esta esté condicionada por la autoconciencia, justifica el enfoque que tomamos en este apartado. Hasta el momento podríamos afirmar lo siguiente: la lucha por el reconocimiento tendría dos condiciones, el hábito de la reflexibilidad para poder conocerse a sí mismo conociendo lo que se dice y lo que se hace y, segundo, juzgar y tener un criterio propio sobre la comunicabilidad y la publicidad, y así elegir cómo mostrarse a los otros. En suma, las facultades de pensamiento y de juicio aparecen como una precondition para la lucha por el reconocimiento.

Pero ¿basta con aparecer de determinada manera para que los demás nos reconozcan así? Desde la perspectiva de la *vita activa*, viendo el fenómeno del reconocimiento mediante las características de la acción, diríamos que no. No solo por la pluralidad de discursos que pueden darse sobre una misma apariencia, sino por la fragilidad de la interpretación y la imprevisibilidad del significado revelado. Desde la perspectiva de la *vita contemplativa* diríamos, también, que no basta pero que es la opción que tenemos.

Aquí podríamos preguntarnos si el reconocimiento consiste en una correspondencia entre la forma en la que se percibe un agente o un grupo, y la forma en la que lo perciben los demás. ¿Hablar de elegir cómo ser visto podría traducirse en engaño? Algo así como lo que llamamos “falsas apariencias” o, coloquialmente, “aparentar lo que no es”. Sobre esto Arendt dice: “[...] la única forma de distinguir el engaño y la presunción de lo real y verdadero es la incapacidad de los primeros para mantenerse consecuentes” (Arendt, 2002b, p. 60), y más adelante nos recuerda el dicho de Sócrates con estas palabras: “[...] aparece *siempre* tal y como deseas aparecer ante los otros, incluso si estas solo y no te muestras ante nadie más que ante ti mismo” (Arendt, 2002b, p. 61).

El punto está, entonces, en mantener la apariencia. En, como invita Sócrates, ser *siempre* como queremos que nos vean. En esto consistiría la lucha por el reconocimiento. ¿Cómo hacer para que nos reconozcan de determinada manera? Apareciendo *siempre* así. Hasta aquí llegarían los límites del agente. Sin embargo, como vimos, el reconocimiento es una actividad del espectador, de ahí que finalicemos este apartado aproximándonos a esta actividad entendida como una forma de juicio.

Si entendemos el reconocimiento como una forma de juicio, no nos referimos al acto de discriminar sino, por una parte, a las condiciones previas que darían validez a esa discriminación. Por la otra, al hecho de que ambos se expresan como un “me parece”, es decir, que tanto el juicio (bien sea moral o estético) como la representación dan cuenta de la percepción que tiene un espectador de un fenómeno sobre el cual ha reflexionado.

Estas dos características en común están conectadas y la primera da lugar a la segunda. La pluralidad de perspectivas sobre un mismo fenómeno, la pluralidad de “me parece”, nos lleva a la pregunta por la validez de estos. ¿Todos los juicios son válidos? O, en este caso, ¿todos los reclamos de reconocimiento serían válidos, así como los discursos de sus defensores y objetores? En lo concerniente al juicio, Arendt (2007, p. 144) toma la concepción kantiana del juicio estético y la extiende hacia el juicio moral, proponiendo que este se realiza sin la aplicación de reglas generales que sean apodícticamente verdaderas. Este tipo de juicio es de carácter reflexionante. A diferencia del determinante, en donde se subsume el particular bajo una regla general, el juicio reflexionante consiste en percibir lo universal en el particular (Arendt, 2003).

A pesar de la ausencia de reglas, la publicidad o comunicación serían los criterios de validez del juicio. Estas mismas pautas podrían ser un

criterio para evaluar los discursos sobre el reconocimiento. Quien piensa como miembro de una comunidad tiene en cuenta a los otros al evaluar lo apropiado o inapropiado que podría ser el hacer pública su opinión:

No se siente demasiada ansia por expresar alegría ante la muerte de un padre o los sentimientos de odio y envidia; en cambio, no se tendrá escrúpulo alguno al comunicar que se ama el trabajo científico ni se ocultara la pena ante la muerte de un excelente esposo. El criterio es, pues, la comunicabilidad, y la pauta para decidir sobre ello es el sentido común (Arendt, 2003, pp. 128-129).

Seguindo a Arendt (2003), las tres máximas del sentido común según Kant son: 1) pensar por uno mismo, 2) situarse en el lugar del otro (mentalidad amplia) y 3) estar de acuerdo con uno mismo. Según Camps (2006), Arendt retoma la tesis kantiana que propone la mentalidad amplia como característica del juicio. Esta condición del juicio refiere al ejercicio de sopesar la propia opinión con la de los demás, es decir, tener en cuenta a los otros que también juzgan. Así, ambas condiciones, la publicidad y la mentalidad ampliada, pretenden la imparcialidad del juicio. En pocas palabras, parecería que la validez del juicio está condicionada por lo parcial o imparcial que pueda ser quien da cuenta de su “parecer”.

Esto nos lleva de regreso al inicio de este apartado. Habíamos indicado que es del espectador, más que del actor, de quien depende el reconocimiento. La posición del espectador resulta ser otra de las condiciones para el juicio, no solo porque por definición su tarea es percibir el fenómeno, sino porque su posición le garantiza el desinterés y la distancia necesaria para la reflexión (Arendt, 2002b). El espectador, explica Arendt (2003) siguiendo a Kant, es quien tiene la capacidad de distanciarse para poder reflexionar sobre el fenómeno de forma desinteresada. Este distanciamiento refiere a la relación entre el juicio y la facultad de la imaginación¹¹ que es retomada por Camps. De acuerdo

¹¹ Si bien en este artículo no abordamos la facultad de la imaginación, mencionamos que, desde la perspectiva de Arendt (2002a), esta facultad refiere tanto a la capacidad de representar lo ausente como a la de relocalizarse. Este segundo sentido muestra la relación entre la imaginación y la mentalidad

con Arendt, la distancia permite ver la realidad desde otras perspectivas y tener en cuenta los posibles juicios de los demás (Camps, 2006).

Lo que otorga validez al juicio es la imparcialidad, garantizada por la posición del espectador, y también la mentalidad amplia cuyo límite será el de la comunidad de la cual somos miembros en virtud del sentido común. Arendt (2018) entiende el sentido común como la capacidad de ajustar los sentidos privados y subjetivos a un mundo objetivo que tenemos en común y compartimos con otros. Por ello, para juzgar se debe apelar al sentido común, el cual nos integra en una comunidad con los otros (Arendt, 2007).

Proponemos que estas mismas condiciones pueden servir de criterio para la validez de los discursos que buscan, otorgan o niegan el reconocimiento. La interpretación de Birulés sobre la relación entre el quién y la narración permite insistir en los límites que tenemos como individuos para dar cuenta de nuestra propia identidad. Según Birulés (1994), el significado de la acción permanece oculto para el agente y se revela al narrador. Lo que se revela al segundo está oculto para el primero mientras se realiza el acto o se encuentra atrapado en sus consecuencias pues, para el agente, la significación de lo que hace no se encuentra en la historia que le sigue.

Con esto, Birulés (1994) evidencia la posibilidad de una pluralidad de sentidos, así como el hecho de que estos se oculten al agente y se revelen al narrador. La pluralidad de sentidos e interpretaciones de la historia es consecuencia de la pluralidad de espectadores. El elemento de ocultamiento se comprende desde la ilimitación de la acción, que no permite prever las consecuencias de la iniciativa del agente, cuyo significado se revela cuando la historia ha terminado. La mirada retrospectiva del espectador permite mostrar el significado de la acción que emerge a través del relato y, a su vez, muestra el quién de dicha acción.

El descubrimiento del quién, en contraposición del qué (substancia) es alguien, se halla implícito en todo lo que alguien hace y dice. Pero se trata de una identidad frágil, puesto que depende de la autoexhibición y de la permanencia del acto de contar: la narración

amplia, en tanto refiere a la capacidad de desplazarse hacia otras perspectivas y tenerlas en cuenta para sopesar la validez del propio juicio.

nos identifica mediante el relato de nuestras propias acciones. Esto es, Arendt entiende que para el sujeto no hay conocimiento inmediato de sí, sino continuas re-apropiaciones por medio del relato (Birulés, 1994, p. 26).

La fragilidad del quién hace referencia a su contingencia. El quién se construye a través de las acciones que son percibidas por los otros. Por ello, la autoexhibición y la permanencia son sus condiciones. Lo que se muestra con cierta constancia es percibido por los otros, y por nosotros mismos, e incorporado a un mismo relato identitario. La identidad, entonces, está condicionada por la narración de aquellas acciones que son vistas por los demás y revelan aspectos que pueden permanecer ocultos al propio agente. Esto, visto desde el problema del reconocimiento, cuestiona la plausibilidad de la lucha por este. Luchar por ser reconocidos tal y como consideramos que somos podría ser legítimo, pero, desde la filosofía arendtiana, la forma de obtener dicho reconocimiento se lograría, quizá, manteniendo cierta apariencia para ser vistos como nos vemos a nosotros mismos.

Desde esta perspectiva, es el espectador quien está facultado para reconocer, puesto que, por definición, tiene la posición adecuada para percibir, juzgar y contar; es decir, es quien cumple con las condiciones para expresar su *dokei moi*. Todas las condiciones propuestas por Kant y retomadas por Arendt acaban apuntando a un sentido comunitario, al *sensus communis*, que reivindica el juzgar como miembros de una comunidad. La validez del juicio, siguiendo a Arendt (2007), no sería objetiva ni subjetiva sino intersubjetiva o representativa. Proponemos pensar el reconocimiento del mismo modo y ello implica que el criterio para reconocer o negar no puede estar separado del sentido moral. Camps (2006) propone que el sentido común es indistinguible del sentido moral.

El sentido común y el sentido moral están vinculados porque hay un criterio moral para elegir nuestras compañías. Es decir, del mismo modo en que procuramos mantener la armonía entre el dos-en-uno que constituye nuestra unicidad, asimismo debe haber cierta coherencia entre quién somos y con quién nos relacionamos. Por eso la comunidad a la que apelamos, los otros a los que tenemos en cuenta para sopesar nuestro juicio, no pueden ser personas que nuestro sentido moral rechace. En palabras de Arendt (2007, p. 150): “He tratado de mostrar

que nuestras decisiones acerca de lo correcto y lo incorrecto dependen de la compañía que elijamos, de aquellos con quienes deseemos pasar nuestra vida”.

Si bien juzgamos o, en este caso, reconocemos como miembros de una comunidad que en cierto sentido elegimos tal y como elegimos a nuestros amigos, ello no significa que dicha elección esté condicionada por una identidad de opiniones. No se trataría de seres idénticos que ven y juzgan los fenómenos de la misma manera sino de miembros que, pese a la pluralidad o gracias a esta, son capaces de dialogar y de entender el punto de vista del otro.¹² Arendt, siguiendo a Aristóteles, afirma: “El elemento político contenido en la amistad es que en el diálogo veraz cada uno de los amigos puede entender la verdad inherente a la opinión del otro. Más que como un amigo a nivel personal, el amigo comprende cómo y bajo qué articulación específica el mundo común se le presenta al otro, quien como persona permanece siempre desigual o distinto” (Arendt, 2016, p. 55).

Esta concepción de la amistad nos regresa a la cuestión con la que finalizamos el primer apartado y nos permite unir ambas perspectivas del reconocimiento, la *vita activa* y la *vita contemplativa*. El reconocimiento es una actividad política, entre otras razones, porque la cuestión de la posibilidad del conflicto invita a la solución mediante el diálogo y la persuasión en lugar de la violencia.¹³ Como en el caso de los juicios, podemos intentar que los otros cambien de opinión o ser persuadidos para cambiar la nuestra:

Arendt sabe que una opinión pública unánime tiende a eliminar a los discrepantes, pues la unanimidad de las masas jamás es el resultado de un acuerdo, sino una expresión de fanatismo. Compartir el juicio significa comunicación; apelar al asentimiento de los demás,

¹² “Las ventajas de pensar desde el lugar del otro son precisamente las variadas y diferentes perspectivas que nos brinda y la nueva materia que añade a nuestro pensar, ilustrándolo” (Sahuí Maldonado, 2002, p. 254).

¹³ Aun cuando Arendt tiene reflexiones muy agudas sobre la violencia: “La violencia puede siempre destruir al poder; del cañón de un arma brotan las órdenes más eficaces que determinan la más instantánea y perfecta obediencia. Lo que nunca podrá brotar de ahí es el poder” (Arendt, 2006, p. 73). Pareciera que lo que se intuye es que el violento no tiene poder y el poder es una forma concertada de acción.

persuadirlos, significa estar juntos en un mundo común (Birulés, 2007, p. 228).

4. Conclusión

Hemos presentado un acercamiento al problema del reconocimiento de la identidad visto desde la filosofía arendtiana tomando como referentes tanto la *vita activa* como la contemplativa. Retomamos la acción perteneciente a la *vita activa* para caracterizar el reconocimiento como político, imprevisible, frágil, que no pertenece a un solo agente ni está bajo su control, sino que surge en la trama de las relaciones humanas y las historias interpretadas.

Como la acción, el reconocimiento está condicionado por el espacio de aparición. De ahí la importancia de abordarlo desde la perspectiva de la *vita contemplativa*, entendiéndolo como un tipo de juicio y, por tanto, también condicionado por dicho espacio de aparición. La esfera pública, el mundo, posibilita que surja la acción y también es el espacio en el que el yo se revela y puede ser reconocido. La revelación de un fenómeno es algo que aparece ante los ojos del espectador. De ahí la relación con la *vita contemplativa*. La acción y el reconocimiento del agente solo son posibles en el mundo de las apariencias, pero adquieren su significado gracias a un ejercicio reflexivo del espectador y el uso de su facultad de juicio. Desde ambas perspectivas, el reconocimiento resulta ser de carácter fenoménico y recae sobre las facultades del espectador más que del agente.

Es cierto que de ninguno de nuestros dos referentes hemos tomado su elemento fundamental para caracterizar el reconocimiento. Por el lado de la acción, no hacemos referencia al elemento de la novedad, y por el lado del juicio, no nos referimos a la discriminación o distinción. No obstante, la acción y el juicio nos dan luces para comprender el reconocimiento como un fenómeno de carácter frágil e imprevisible que escapa al control del agente; es, más bien, una actividad en la que el espectador apela a su sentido común y, además, que en ocasiones puede requerir del tiempo y de la mirada retrospectiva.

El carácter condicional del espacio de aparición coincide como característica tanto desde la perspectiva de la *vita activa* como desde la contemplativa. Es esta condición la que permite que el quién del agente se revele mediante su palabra y acto, y sea visto por los espectadores para que estos comuniquen sus *dokei moi* y se mantengan en el ámbito

del discurso y la persuasión, mas no de la violencia. Sin embargo, cabe resaltar cómo, en la búsqueda de reconocimiento, la violencia puede trasladarse al discurso al devenir un ejercicio de humillación, degradación o intimidación más que de persuasión. En todo caso, esto será entendido como otras formas de violencia.

Esto significaría que, en tanto actividad del espectador, el reconocimiento —a pesar de las formas subjetivas de percibir un fenómeno— nunca es meramente subjetivo sino intersubjetivo, pues como en el caso del juicio, reconoce como miembro de una comunidad, teniendo en cuenta a los otros y siempre desde la posibilidad de cambiar o hacer cambiar de parecer. Son los espectadores los que, mediante sus facultades discursivas, finalmente otorgan o niegan el reconocimiento que los agentes reclaman.

Los límites de estos últimos, como vimos, refieren a una suerte de coherencia, de conducta consistente a partir de la autopresentación. El agente cuenta con la posibilidad de procurar aparecer siempre de determinada manera, de actuar y hablar tal y como se percibe a sí mismo, si es que es así que quiere ser reconocido. La única condición para esto sería la autoconciencia, pues para poder, en cierto sentido, controlar cómo queremos que nos vean primero nos debemos ver a nosotros mismos.

Referencias

- Arendt, H. (2002a). Comprensión y política (las dificultades de la comprensión). C. Sánchez Muñoz (trad.). *Daimon Revista Internacional de Filosofía*, 26, 17-30. <https://revistas.um.es/daimon/article/view/12041>
- Arendt, H. (2002b). *La vida del espíritu*. F. Birulés y C. Corral (trads.). Paidós.
- Arendt, H. (2003). *Conferencias sobre la filosofía política de Kant*. C. Corral (trad.). Paidós.
- Arendt, H. (2006a). *Hombres en tiempos de oscuridad*. C. Ferrari y A. Serrano de Haro (trads.). Gedisa
- Arendt, H. (2006b). *Sobre la violencia*. Santillana Ediciones Generales (trads.). Alianza.
- Arendt, H. (2007). *Responsabilidad y juicio*. M. Candel y F. Birulés (trads.). Paidós.
- Arendt, H. (2011). *La condición humana*. R. Gil Novales (trad.). Paidós.
- Arendt, H. (2016). *La promesa de la política*. E. Cañas (trad.). Booket.

- Arendt, H. (2021). La esfera pública y la privada. En H. Arendt, *La pluralidad del mundo* (pp. 281-239). R. Gil Novales (trad.). Taurus.
- Barrior, C. (2010). Una perspectiva de las funciones del relato en Hannah Arendt: verdad, significado y juicio. *Cuadernos del Sur—Filosofía*, 39, 31-45.
- Beiner, R. (2003). Ensayo interpretativo. Hannah Arendt y la facultad de juzgar. En H. Arendt, *Conferencias sobre la filosofía política de Kant* (pp. 157-270). R. Beiner (ed.). C. Corral (trad.). Paidós.
- Birulés, F. (1994). Hannah Arendt: modernidad, identidad y acción. En M. Vilanova (ed.), *Pensar las diferencias* (pp. 21-29). PPU.
- Birulés, F. (2007). *Hannah Arendt. El orgullo del pensar*. Herder.
- Butler, J. (2012). *Sujetos del deseo. Reflexiones hegelianas en la Francia del siglo XX*. E. Luján Odriozola (trad.). Amorrortu Editores.
- Camps, V. (2006). La moral como integridad. En M. Cruz (ed.), *El siglo de Hannah Arendt* (pp. 63-85). Paidós.
- Campillo, N. (2005). Mundo y pluralidad en Hannah Arendt. *Intersticios*, 10(22-23), 87-100. <https://roderic.uv.es/rest/api/core/bitstreams/18e26d9b-1848-42f1-85f8-bb43d63add56/content>
- Forti, S. (1999). Introduzione. En C. Lefort, M. Abensour, O. Guaraldo, J.-F. Lyotard, R. Esposito, M. Vatter, É. Tassin, J. Kohn, H. Höing, A. Cavarero, R. J. Bernstein y L. Savarino, *Hannah Arendt* (pp. i-xxxiii). Bruno Mondadori.
- Fraser, N. (2003). ¿Redistribución o reconocimiento? En N. Fraser y A. Honneth, *Redistribución o reconocimiento* (pp. 13-15). P. Manzano Bernádez (trad.). Morata.
- Garrido, F. (2020). *Los espectadores en la conformación del mundo: Hannah Arendt y la participación política en las sociedades contemporáneas*. [Tesis doctoral, Universidad Nacional del Sur]. Repositorio Institucional Digital de la Universidad Nacional del Sur. <https://repositoriodigital.uns.edu.ar/handle/123456789/5541>
- Honneth, A. (2014). *El derecho de la libertad*. G. Calderón (trad.). Katz Editores.
- Honneth, A. (1997). *La lucha por el reconocimiento*. M. Ballesteros (trad.). Crítica.
- Muñoz, M. T. (2020). *Hannah Arendt. Nuevas sendas para la política*. Bonilla Artiga Editores.
- Narváez Álvarez, C. A. (2020). *El mundo y su pérdida: un intento de comprensión desde Hannah Arendt*. [Tesis de doctorado, Universidad de Barcelona]. Tesis Doctorals en Xarxa. <https://hdl.handle.net/10803/670258>

- Quintana, L. (2010). Identidad sin sujeto: Arendt y el mutuo reconocimiento. *Ética y política*, 12(2), 430-448.
- Ricœur, P. (2006). *Caminos del reconocimiento. Tres estudios*. A. Neira (trad.). Fondo de Cultura Económica.
- Rodríguez Suárez, L. P. (2009). Acción y mundo en Hannah Arendt. En Á. Prior Olmos y Á. Rivero Rodríguez (eds.), *La filosofía de Ágnes Heller y su diálogo con Hannah Arendt. Congreso internacional: Murcia, del 13 al 15 de octubre de 2009* (p. 29). Servicio de Publicaciones de la Universidad de Murcia.
- Sahuí Maldonado, A. (2002). Hannah Arendt: Espacio público y juicio reflexivo. *Signos Filosóficos*, 4(8), 241-263.
- Sánchez, C. (2005). Hannah Arendt como pensadora de la pluralidad. *Intersticios*, 221-237.
- Sidorova, M. (2019). The Role of the Judging Spectator in Politics: Arendt's Approach. *The Russian Sociological Review*, 18(3) 245-259. <https://doi.org/10.17323/1728-192x-2019-3-245-259>
- Sorrentino, V. (2009). *Amor mundi* y política en Hannah Arendt. *Revista Laguna*, 25, 19-30. <http://riull.ull.es/xmlui/handle/915/13829>
- Taylor, C. (1993). *El multiculturalismo y la política del reconocimiento*. M. Utrilla de Neira (trad.). FCE.

<https://doi.org/10.21555/top.v760.3369>

La interpretación dialéctica de la hermenéutica

The Dialectical Interpretation of Hermeneutics

Luis Eduardo Gama Barbosa

Universidad Nacional de Colombia
Colombia

legamab@unal.edu.co

<https://orcid.org/0000-0002-0133-1966>

Recibido: 27 - 01 - 2025.

Aceptado: 20 - 07 - 2025.

Publicado en línea: 10 - 07 - 2026.

Cómo citar este artículo: Gama Barbosa, L. E. (2026). La interpretación dialéctica de la hermenéutica. *Tópicos, Revista de Filosofía*, 76, 373-407. <https://doi.org/10.21555/top.v760.3369>



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

Resumen

El auge del tema de la interpretación en la filosofía contemporánea ha suscitado una crítica a aquellas posturas que hacen de todas las realizaciones de la praxis humana una forma de interpretación, aclimatando un relativismo para el cual no existiría ninguna verdad objetiva. Este artículo sostiene que la hermenéutica de Hans-Georg Gadamer se libera de este reproche al elaborar una noción dialéctica de “interpretación” para la cual interpretar no es imponer un significado arbitrario, sino mediar una situación particular concreta con un espacio común de sentido. Aquí se examina la forma como, según Gadamer, se realiza esta mediación, tanto en la comprensión del mundo social como al nivel más fundamental del lenguaje vivo. Al final, se insinúan algunas vías de desarrollo que abre este examen, en particular la de una filosofía social hermenéutica capaz de dar orientaciones de sentido a las sociedades capitalistas contemporáneas.

Palabras clave: dialéctica; hermenéutica; interpretación; mediación; comprensión; fusión de horizontes; diálogo hermenéutico; dialéctica de pregunta y respuesta; horizonte de sentido; Hans-Georg Gadamer.

Abstract

The rise of the theme of interpretation in contemporary philosophy has sparked criticism of those positions that treat all human praxis as a form of interpretation, promoting a relativism for which no objective truth exists. This paper argues that Hans-Georg Gadamer's hermeneutics avoids this criticism by developing a dialectical notion of interpretation in which interpreting does not mean imposing an arbitrary meaning, but rather mediating a particular concrete situation through a common space of meaning. Here it is examined how, according to Gadamer, this mediation takes place both in the understanding of the social world and at the more fundamental level of living language. Finally, some avenues for further development opened up by this examination are suggested, in particular that of a hermeneutic social philosophy capable of providing guidance on meaning to contemporary capitalist societies.

Keywords: dialectics; hermeneutics; interpretation; mediation; understanding; fusion of horizons, hermeneutic dialogue; dialectic of question and answer; horizon of meaning; Hans-Georg Gadamer.

La cuestión de la interpretación constituye uno de los enclaves fundamentales del paisaje filosófico contemporáneo. Tematizada ya en Platón como la actividad mediadora del poeta en tanto “intérprete de los dioses” (*Ion*, 534e) e integrada luego a las formas clásicas de la hermenéutica en tanto doctrina de la interpretación de textos autoritativos, la interpretación ha ganado, por lo menos desde la segunda mitad del siglo pasado, una relevancia inusitada en el campo de la reflexión filosófica, en sus diversas corrientes y latitudes geográficas. En este sentido, se llega a considerar que una teoría de la interpretación “se ha convertido por sí misma en una teoría general del conocimiento y la realidad” (Margolis y Rockmore, 2000, p. 4) y que el concepto de interpretación es “quizás el concepto filosófico central de nuestro tiempo” (Shusterman, 1996, p. 14).¹

No cabe duda de que el auge de la noción está en relación con el esfuerzo por superar cierta tradición metafísica fundada en la convicción de que es posible aprehender la verdad del ser sin mediaciones, esto es, como una presencia que se manifiesta sin velo “en un lenguaje pleno y sin sintaxis” (Derrida, 2006, p. 310), en un lenguaje entonces “que ya no puede recibir más glosa, y no requiere interpretación” (Rorty, 1991, p. 89). El anhelo metafísico por un lenguaje así comienza a derruirse con la evidencia de que un conocimiento objetivo y neutral de la realidad resulta imposible, con la constatación del carácter esencialmente simbólico del mundo social, la cultura y la historia, o con la convicción de que el yo no descansa en una base trascendental, sino que se autoconstituye permanentemente en sus autointerpretaciones.

Todos estos temas, como se sabe, van ganando peso a partir de Kant y encuentran quizás en Nietzsche una primera elaboración general. En efecto, la “ontología” de la voluntad de poder descansa en el supuesto de una universalización de la interpretación que promueve el concepto a rasgo fundamental de la realidad, con lo que además la razón pura en sentido kantiano es desplazada por una razón interpretativa (cfr. Margolis y Rockmore, 2000, p. 1). Es, sin embargo, a partir de la segunda mitad del siglo XX que la cuestión de la interpretación se expande

¹ A menos que la bibliografía final indique lo contrario, las traducciones son de mi autoría.

de manera inédita por los más diversos territorios del pensamiento filosófico. Por supuesto es en las filosofías de Heidegger y Gadamer donde, prolongando su antiguo parentesco con la hermenéutica clásica, la noción se ha aclimatado de forma más natural. Pero más allá de este vínculo esperado, la problemática es también constitutiva de la llamada deconstrucción, de ciertos enfoques pragmatistas² o de algunos planteamientos de la filosofía analítica o postanalítica.³ Por lo demás, el auge de esta problemática en el mundo académico se alimenta también de las preocupaciones y crisis de las sociedades contemporáneas. Como se ha señalado, la crisis del sentido en el mundo actual tiene menos que ver con la presunta falta de significado de la existencia humana en las condiciones postcapitalistas e hipermodernas de nuestras sociedades que, por el contrario, con una sobreoferta de verdades y una disponibilidad infinita de información que desorienta y demanda una mejor comprensión y un ejercicio más consciente de nuestra capacidad de dar sentido, esto es, de interpretación (cfr. Berner y Thouard, 2015, p. 7; Berner y Morel, 2023, p. 9).

En medio de este panorama general ha hecho carrera la idea de que la interpretación es infinita o ilimitada. Esta formulación admite las más diversas declinaciones: desde la tesis de cuño nietzscheano que afirma el carácter perspectivístico e interpretativo de toda la praxis humana (Abel, 1995, p. 13), pasando por el enunciado hermenéutico que señala que el sentido de un texto es inagotable, hasta ciertas posturas constructivistas propias de los *cultural studies* que diluyen toda forma de realidad objetiva en un juego sin fin de significantes. Esta idea, que por un lado resulta bastante seductora como *plaidoyer* de la imaginación y la creatividad, y por su potencial emancipador frente al universalismo tirano de la racionalidad científica, ha encontrado sin embargo numerosos

² Por ejemplo en Richard Shusterman (1994; Shusterman *et al.*, 1991) y su “interpretative turn”, pero también en el enfoque hermenéutico-pragmático de corte más antropológico de Johann Michel (2017).

³ Piénsese, p. ej., en Donald Davidson (1984), que engrana su semántica y su teoría de la acción con la noción de “interpretación”, o en el proyecto de Nelson Goodman (1978), que entiende la filosofía como un análisis de los diversos procesos simbólicos e interpretativos con los que construimos el mundo. Del mismo modo, la filosofía interpretacionista de Günter Abel (1995 y 2004) combina recursos nietzscheanos y métodos analíticos para su propósito de entender toda forma de conocimiento y acción como una praxis de interpretación.

detractores, aun entre aquellos que están dispuestos a considerarse filósofos de la interpretación. Así, por ejemplo, el proyecto de Michel de extender el campo de la hermenéutica hacia objetos y prácticas (como la naturaleza o la inteligencia de la vida ordinaria) considerados hasta ahora por fuera del radio del interpretar se presenta al mismo tiempo como una “teoría restringida de la interpretación”.⁴ La apoteosis de la interpretación parece así encontrar sus límites, no en la nostalgia de un conocimiento metafísico infinito, libre de perspectiva y situación, sino desde el interior del propio campo interpretacionista donde no todos están dispuestos a entregarse a un despliegue sin restricciones de la interpretación que, en últimas, solo haría inservible el concepto y haría su propia posición vulnerable a un relativismo banal.

Esta tensión que recorre el campo actual de la filosofía de la interpretación sirve de trasfondo al problema que este artículo se propone abordar. Aquí me interesa analizar el concepto de “interpretación” propio de la hermenéutica de Hans-Georg Gadamer. A primera vista, Gadamer parecería ubicarse del lado de los defensores de una interpretación ilimitada y omnipresente; su planteamiento de la inseparabilidad entre comprensión e interpretación, aunado a la tesis del carácter universal del comprender, parecen en efecto apoyar esta

⁴ Michel hace eco aquí del propósito de Oliver Scholz (2008, p. 77) por restringir la interpretación a aquellos objetos o situaciones cuyo sentido no se comprende de manera espontánea y requieren entonces un esfuerzo del pensamiento para atribuirles sentido. En ese orden de ideas, los límites de la interpretación se hacen evidentes en hechos como el de que “no interpretamos permanentemente el sentido de nuestra identidad, el sentido de las cosas, el sentido de las reglas, el sentido de las intenciones del otro. Los comprendemos espontáneamente” (Michel, 2023, p. 108). La misma intención de limitar la omnipresencia de la interpretación impulsa el trabajo de Élisalde (2000), y también, en el ámbito anglosajón, la obra del ya citado Richard Shusterman. Este último se dirige contra la universalidad de la interpretación defendida por Gadamer, incluso compartiendo su espíritu antifundacionalista. Para Shusterman (2002, p. 32), si todo es interpretación entonces el concepto mismo pierde sentido, por lo que propone distinguir casos como la comprensión ininterrumpida e inmediata del lenguaje de aquellos donde la comprensión del sentido se pierde y se hace necesaria la mediación de la interpretación. Cabe recordar en todo caso que la distinción entre una comprensión inmediata y una mediada, que es propiamente la que se llamaría interpretación, ya está presente en Schleiermacher (1838), para quien solo cuando fracasa la comprensión se hace necesaria la interpretación.

idea. No obstante, la universalidad del fenómeno hermenéutico del comprender (o mejor, del comprender interpretante) solo con muchas restricciones puede expresarse con la fórmula “todo es interpretación”, y en todo caso no deriva en una teoría general que reduzca todas las expresiones de la praxis humana a una forma de interpretación, y menos en un relativismo que disuelve la realidad en un juego de significantes arbitrario.

El presente artículo se propone aclarar todo esto a partir de un análisis del planteamiento de Gadamer que entiende la interpretación como una forma de mediación dialéctica. Elaborada en este registro dialéctico, la interpretación hermenéutica hace evidente un compromiso con lo universal que la libra del reproche del relativismo. De este modo, este examen permite situar de manera más precisa la posición de la hermenéutica de Gadamer dentro del panorama actual de las filosofías de la interpretación, a la vez que ilumina posibles líneas de desarrollo de este enfoque. Nuestra exposición toma pie en el análisis existencial de Heidegger en *Ser y tiempo* sobre el comprender y la interpretación (sección 1), y pasa luego a elaborar el tema en la obra fundamental de Gadamer, *Verdad y método*: primero en la teoría de la interpretación histórica que se encuentra en la segunda parte de esta obra (sección 2), y luego al hilo de la ontología del lenguaje que se desarrolla en la parte tercera (sección 3). Una última sección trata finalmente de recoger de manera muy sucinta los resultados y conclusiones de este estudio.

1. Los fundamentos heideggerianos del problema: la inseparabilidad de comprender e interpretar

Contra la filosofía hermenéutica y su insistencia en hacer de la interpretación una actividad universal inherente a la totalidad de la experiencia humana se suelen presentar objeciones como la que se encuentra en un pasaje de la *Philosophische Grammatik* de Wittgenstein:

Por supuesto que yo a veces *interpreto* signos, le doy a los signos una interpretación; pero eso no ocurre siempre que yo comprendo un signo (si alguien me pregunta “¿Qué hora es?” no hay allí un proceso interno de interpretación laboriosa; yo simplemente reacciono a lo que veo y oigo. Si alguien agita un cuchillo delante de mí, yo no digo “interpreto eso como una amenaza”) (PG, II, § 9, p. 47; cfr. Descombes, 1991).

El planteamiento de Wittgenstein se apoya en experiencias tan evidentes que parece incontestable. Vale la pena, sin embargo, examinar sus presupuestos implícitos. Wittgenstein parte de que la interpretación es un proceso por el cual se hace comprensible un conjunto de signos cuyo sentido no es claro inicialmente. Con ello asume que una oración cuyo significado resulte evidente no requiere en absoluto ser interpretada. Como es sabido, para Wittgenstein, una frase no es comprensible como resultado de una aprehensión mental de un contenido teórico, sino más bien porque se sabe actuar conforme a ella, como solo comprende una orden quien sabe cumplir lo que ella en efecto demanda. Como sea, lo que quiero destacar de este pasaje es que representa una posición bien conocida para la cual la comprensión y la interpretación son dos cosas esencialmente distintas, y en buena medida independientes pues es posible comprender un enunciado sin necesidad de interpretarlo,⁵ así como también la interpretación solo tendría lugar justamente cuando la comprensión no existe. Desde esta perspectiva, comprender (*Verstehen*) no es tanto un ejercicio reflexivo de captación de un significado sino una familiaridad preteórica con este que se manifiesta ante todo en la acción. E interpretar (*Auslegen, Interpretieren*), por su parte, mienta algo así como un hacer explícito, esclarecer, descifrar o traducir en términos claros, signos cuyo significado resulta de manera inmediata oscuro o sujeto a controversia.

Pues bien, este planteamiento sería fundamentalmente antihermenéutico, pues la hermenéutica, o por lo menos la hermenéutica filosófica que se desarrolla en el siglo XX a partir del pensamiento de Heidegger y Gadamer, descansa sobre dos principios fundamentales: la inseparabilidad de la comprensión y la interpretación,⁶ y la universalidad de esta comprensión interpretante o interpretación compenso-

⁵ Por eso dice Wittgenstein para indicar un absurdo: "Si alguien dijera 'cada frase necesita una interpretación' es como si dijera que ninguna frase puede ser comprendida sin un añadido" (PG, II, § 9, p. 47).

⁶ Para Angehrn (2004, p. 141), ese entrelazamiento de comprender e interpretar determina el paradigma de la hermenéutica moderna, pero este mismo autor nos recuerda cómo en la antigua tradición hermenéutica se trazaba una nítida distinción entre estas dos funciones: entre el acto de comprender o captar el sentido del texto (*subtilitas intelligendi*) y la interpretación que lo profundiza y despliega (*subtilitas explicandi*). Estas dos funciones, por lo demás, se complementaban con el momento de la aplicación (*subtilitas applicandi*).

ra, es decir, el hecho de que esta interpretación es inherente de manera esencial a todas las esferas de la praxis humana. Si para Wittgenstein se puede comprender sin interpretar, y en muchas ocasiones no interpretamos en absoluto, para la hermenéutica la interpretación, vinculada internamente con la comprensión, acompaña de forma inevitable toda realización de la existencia humana.

Pero en realidad, la noción wittgensteniana de “interpretación” no discute la perspectiva hermenéutica sino que se apoya en presupuestos muy distintos. Por eso no tiene sentido querer ver en la filosofía de este último una refutación de la hermenéutica (cfr. Descombes, 1991). Para Wittgenstein y la tradición analítica de la filosofía del lenguaje, interpretar es en esencia un ejercicio de sustitución: sustitución de un enunciado por uno más claro que al explayar el sentido aún oscuro en el primero lo sustituye. En ese contexto, la traducción suele verse como un paradigma de este ejercicio interpretativo, pues en ella enunciados extraños se sustituyen por otros cuyo sentido es evidente. Esto vale también de manera particular para la traducción radical de Davidson que no sustituye directamente un enunciado por otro (“es regnet” por “llueve”) sino que aclara su significado a través de un acceso conductual a lo que los hablantes de un lenguaje asienten como verdadero.

Pues bien, la tesis que quiero sostener aquí es que para la hermenéutica interpretar no es sustitución sino mediación. Y que en esta mediación no se trata simplemente de reemplazar un enunciado por otro más explícito y comprensible que entonces ya no requeriría de nuevas interpretaciones, sino de determinar un sentido general en una situación específica, es decir de concretar en un contexto particular de la praxis humana una significación común y compartida. En ese orden de ideas, la interpretación hermenéutica no opera en el plano horizontal del lenguaje sustituyendo enunciados que se situarían al mismo nivel óntico, por decirlo de alguna forma, sino que encarna el viejo problema filosófico de la relación entre lo universal y lo particular: la cuestión de la mediación entre lo general y lo particular, entre lo abstracto y lo concreto, del que ya se ocupó de manera especial la dialéctica platónica. En lo que sigue quiero explicar cómo tiene lugar esta mediación de la interpretación hermenéutica y mostrar en qué sentido ella puede considerarse una forma de dialéctica. Para ello, conviene primero rastrear los antecedentes de la postura de Gadamer en la analítica existencial de Heidegger.

En el § 32 de *Sein und Zeit*,⁷ Heidegger introduce expresamente la noción de “interpretación” (*Auslegung*), que él entiende aquí como el “desarrollo del comprender”. En la interpretación, dice Heidegger:

[...] el comprender se apropia, comprendiendo, lo comprendido. En la interpretación no se vuelve el comprender otra cosa, sino él mismo. La interpretación se funda existencialmente en el comprender, en lugar de surgir este de ella. La interpretación no es el tomar conocimiento de lo comprendido, sino el desarrollo de las posibilidades proyectadas en el comprender (Heidegger, *SZ*, § 32).

El comprender al que se hace referencia aquí no consiste en un estado mental o una operación intelectual de la conciencia, sino señala el carácter específico de la existencia humana de estar inmersa siempre en entramados u horizontes de significación que constituyen las articulaciones de sentido que orientan su praxis y determinan en general su posición ante el mundo. Más que un hacer de la conciencia de un sujeto, el comprender indica entonces la situación de toda existencia humana concreta (*Dasein*) que solo se realiza en el marco de un mundo de significaciones desde donde se perfilan las posibilidades de realización que le están abiertas. Comprender señala por ello más un “estar comprendido”, esto es, estar abarcado o inserto en un espacio general de sentido del que se participa, que una actividad consciente y reflexiva.⁸ No se trata así de un rendimiento psicológico sino de una estructura ontológica del existir humano, o en la terminología de Heidegger, de un *existenciario*. Pero este comprender tampoco se deja inscribir en un marco

⁷ En adelante citado como *SZ*.

⁸ Deliberadamente he acentuado con esta caracterización del comprender el carácter “arrojado” (*Geworfenheit*) de la existencia más que su aspecto proyectante (*Entwurf*). No desconozco con ello que en el pasaje que citamos es este último aspecto el que prima sobre el arrojamiento del “estar comprendido” que solo se destacará más directamente en la exposición de la situación anímica (*Befindlichkeit*). Me propongo con esto sugerir un matiz que permita distinguir el comprender (como arrojamiento) del interpretar (como proyecto), sin negar, claro está, la inseparabilidad de los dos fenómenos. Para una presentación del nexo comprender/interpretar como proyectante, cfr. Demmerling (2001, pp. 101 y ss).

epistemológico, pues no se trata aquí del conocimiento de un hecho objetivo, sino de una familiaridad con un mundo denso de sentido que se deja abrir en múltiples posibilidades de significación.

Es este comprender ontológico el que se desarrolla, según Heidegger, como interpretación. Interpretar no significa entonces sino hacer real, en una situación concreta, una de las posibilidades de sentido que ya están abiertas en el comprender. O, como dice Heidegger, en la interpretación un sentido previamente comprendido, y esto quiere decir simplemente abierto pero no apropiado, se inserta en un plexo específico de acción y se realiza *como (als)* algo.⁹ Así, por ejemplo, solemos tener una comprensión general del objeto que llamamos martillo, pero solo la interpretación que se le da a este según un contexto práctico específico (como “herramienta”, como “arma” o como “símbolo de la clase trabajadora”) realiza esa comprensión y la hace concreta. Con esto ya se subvierte la perspectiva tradicional, aún vigente en Wittgenstein, para la cual la comprensión resulta de una interpretación lograda cuando esta ilumina el sentido ensombrecido. Desde el planteamiento ontológico existencial es más bien la interpretación lo que se apoya en el comprender, pues ella solo hace real un proyecto de verdad que ya estaba inscrito en el horizonte amplio de lo comprendido.

Pero de manera más profunda lo que se señala es que no se trata en realidad de dos operaciones separadas, sobre las que valiera la pena preguntarse cuál precede a cuál. Por ello se dice que en la interpretación el comprender “no se vuelve otra cosa” sino que más bien en ella este se consume verdaderamente.¹⁰ Y tampoco se trata de que solo la ausencia de comprensión solicite la interpretación, cuyo éxito entonces implicaría el acceso a una nueva comprensión que haría innecesario el interpretar,

⁹ “Das umsichtig auf sein Um-zu Auseinandergelegte als solches, das *ausdrücklich* Verstandene, hat die Struktur des *Etwas als Etwas*” (SZ, § 32). La estructura del “algo como algo” resulta determinante para la interpretación en este contexto. Con ella también se hace visible el nexo que une a la interpretación con la comprensión: “Las posibilidades que de un modo más o menos flotante se comprenden cuando algo comparece se vuelven ahora concretas bajo la forma de un tomar eso que aparece en un posible uso [...]. Ver algo en cuanto algo es explicitar las posibilidades de trato, esto es, explicitar la comprensión” (De Lara, 2015, p. 152).

¹⁰ “In der Auslegung wird das Verstehen nicht etwas anderes, sondern es selbst” (SZ, § 32).

como aún lo entiende Wittgenstein. Lo que ocurre más bien es que una verdadera comprensión solo se alcanza a través de la interpretación que la concreta y hace real, pues sin ella se trataría solo de una abstracta y vacía premonición de sentido, difusa, imprecisa y sin ninguna efectividad ni materialidad. Por su propia naturaleza, siguiendo su vocación más interna, por decirlo así, toda comprensión tiende a realizarse como interpretación, que no se independiza entonces como una actividad distinta, sino que representa su más acabada forma de ser. Por eso para Heidegger lo que tiene lugar en esta constelación del comprender y el interpretar debe verse como un “fenómeno unitario” (*ein einheitliches Phänomen*) (SZ, § 32, p. 151)¹¹ y no al modo de una relación que vincule actividades en principio autónomas y separadas.

En el párrafo al que hago referencia, la cuestión de la unidad entre comprender e interpretar sirve para introducir la célebre cuestión del círculo. El asunto se formula primero al modo de una objeción: si la interpretación se mueve dentro de la estructura de lo previamente comprendido, esto es, si ella tiene que haber comprendido ya lo que trata de interpretar, “¿cómo va a dar resultados científicos sin moverse en un círculo?” (SZ, § 32, p. 152). El peligro que se insinúa aquí es el de una circularidad viciosa: una especie de petición de principio en la que aquello que se busca comprender con la interpretación se encuentra ya de antemano comprendido. Heidegger muestra que esta objeción solo tiene sentido desde un modelo teórico del conocimiento cuyo ideal es aprehender al ente de manera pura, esto es, que hace del conocer la acción de una conciencia objetiva y libre de prejuicios sobre un mundo que le es inhóspito y desconocido. Su análisis fenomenológico desmonta este modelo al revelar la mutua pertenencia de la existencia humana y el mundo, al mostrar al Dasein como *ser-en-el-mundo*. Quien quiere interpretar ya está comprendiendo el mundo, y aquí no hay ningún vicio lógico sino la simple expresión del modo mismo de ser de la existencia humana: “Este círculo del comprender no es un círculo en el que se movería una cierta forma de conocimiento, sino que es la expresión de la existencial *estructura de lo previo* propia del Dasein” (SZ, § 32, p. 153). Al final, Heidegger no niega la presencia de un círculo,

¹¹ Que sea un fenómeno unitario no significa que sean lo mismo. Por eso en este contexto Heidegger adscribe a cada fenómeno una estructura distinta: el comprender descansa en la estructura de lo previo (*Vor-Struktur*) mientras que a la interpretación, como ya vimos, le corresponde la estructura del “como” (*als*).

pero este no señala un defecto lógico, sino que define la forma misma del ser del existir, la estructura ontológica de ese ente peculiar en el que comprender e interpretar ya están entrelazados en su modo de ser.

Todo este análisis resulta bastante conocido, y sin embargo es muy poco lo que Heidegger nos dice en este pasaje sobre esta circularidad “buena” que define la estructura del existir. Con toda evidencia se trata de un círculo entre el comprender y el interpretar: de un lado la interpretación viene determinada por un comprender previo, pero a la vez una interpretación temática puede “retornar” al nivel de una “comprensión no destacada” (SZ, § 32, p. 153), con lo que se cierra el círculo. Sin embargo, una descripción así nos parece pobre y sujeta a distorsiones. En ella parece tomarse el círculo como un ir y venir constante que oscila entre los extremos de la comprensión y la interpretación, con lo que se rompería la unidad del fenómeno que se ha señalado.¹² En efecto, lo que indica Heidegger aquí no es tanto el encadenamiento constante de dos procesos que pese a su remisión mutua siguen siendo independientes, sino la vinculación intrínseca que hace del comprender algo atravesado necesariamente por la interpretación, así como esta está traspasada de punta a punta por la comprensión.¹³ En este orden de ideas, así como la comprensión contiene en sí misma la interpretación que la consume, así también una interpretación lograda resulta ser

¹² En la literatura especializada se suele entender el círculo “bueno” como el círculo virtuoso que es atento a los prejuicios que guían la interpretación y procede a una evaluación crítica de estos. Algo así permitiría entonces distinguir entre buenas y malas interpretaciones (cfr. Mulhall, 1996, p. 87). Una lectura así no es errada y hasta el propio Heidegger la sugiere en este mismo § 32 cuando señala que lo que importa es “entrar correctamente al círculo”. Sin embargo, antes que exponer esa tarea crítica, aquí nos interesa determinar ese círculo más elemental que es propio del existir y que aparece como un fenómeno unitario. Aquí se trata, ese es nuestro punto, de un círculo que no se traza entre los polos en todo caso separados del comprender y el interpretar, sino de uno que los contiene y vincula.

¹³ Grondin (1996, p. 190) recuerda que el término alemán para interpretación (*Auslegung*) significa tanto “explicitación” como “interpretación” en el sentido de la posibilidad de la existencia humana de estar despierta, atenta o vigilante de su propia condición situada. Como explicitación, la interpretación se mantiene a distancia de la comprensión, si bien ambas están mutuamente remitidas como en un círculo. Solo en su segundo sentido, como condición despierta del Dasein, la interpretación y la comprensión constituyen un fenómeno unitario.

siempre una comprensión que se establece. Para Heidegger, se trata de una unidad de orden fenoménica, y por ello evidente a la experiencia. En efecto, cuando decimos que *comprendemos* no nos referimos con ello a una intelección abstracta y vagamente general, sino a un comprender algo de determinada forma, en una determinada constelación de referencias; y cuando digo que estoy *interpretando* algo no me encuentro tanteando el significado de ese algo en la oscuridad total, sino guiado ya por anticipaciones de sentido que brotan de una comprensión ya previa que me orienta.

La imagen del círculo que Heidegger introduce solo de manera muy elusiva en este contexto resulta confusa y conduce a malentendidos.¹⁴ Más que hablar entonces de circularidad, propongo entender la unidad fenoménica de comprender e interpretar como un proceso de mediación total entre un universal que solo gana realidad en lo particular, y un particular que solo se eleva a su verdad cuando reconoce su lugar dentro de lo universal. Se trata en el fondo de un viejo problema filosófico que ya ocupó a Platón en su esfuerzo por captar de qué manera las cosas del mundo perceptible participan de unos *eidos* universales que no son simples abstracciones sino realidades perfectamente intuibles en lo sensible; o a Hegel, cuya filosofía entera no es sino el esfuerzo por mostrar que en todo fenómeno concreto habita una razón espiritual que lo anima, pero que a la vez solo gana su ser concreto con la mediación de esta realidad efectiva. A esta mediación de lo universal en lo singular y de lo singular en lo particular se la nombra en esta tradición con la noción de “dialéctica”. Lo que propondré, pues, es entender la hermenéutica filosófica como una forma contemporánea de la dialéctica.

2. La dialéctica del comprender histórico en Gadamer

De todo lo expuesto parece claro que la ontología existencial de Heidegger no brinda una buena perspectiva para examinar la mediación dialéctica que aquí asumo como propia de la hermenéutica. Se podría incluso sospechar que el abandono del proyecto de *Ser y tiempo* por parte de Heidegger proviene justamente de que esta obra no logró desarrollar de manera cabal la mediación con el ser que debería ligar la existencia

¹⁴ Heidegger mismo se habría mostrado reticente a utilizar la noción de “círculo” para describir la estructura existencial del Dasein. En ese sentido, no es correcto hablar de una doctrina heideggeriana del círculo del comprender (cfr. Grondin, 2000, p. 126).

concreta del Dasein con un ámbito de trascendencia universal. Este, sin embargo, no es el lugar para evaluar esta conjetura. Por el contrario, los planteamientos posteriores de Hans-Georg Gadamer desarrollaron las observaciones de Heidegger sobre la comprensión y la interpretación en esta dirección, y por eso conviene detenerse en ellos.

Sería posible en efecto entender todo el pensamiento hermenéutico de Gadamer como un esfuerzo por introducir la cuestión de la mediación en el fenómeno del comprender interpretante de *Ser y tiempo*, aunque también en problemáticas heideggerianas posteriores, como la del lenguaje.¹⁵ En este sentido, Gadamer habría buscado una suerte de convergencia entre Hegel y Heidegger o, para decirlo en su propio lenguaje, una fusión de estos dos horizontes del pensar por la cual una mediación dialéctica transfigurada se pondría al servicio de la pregunta por el ser. No es el lugar para desarrollar esto a profundidad. Para nuestros intereses basta señalar algunos pasajes de *Verdad y método*¹⁶ donde este propósito se hace evidente. De manera destacada esto tiene lugar primero en la discusión sobre la comprensión propia de las ciencias históricas que domina los capítulos centrales de la obra. Para abordar este tema, y de paso superar la hermenéutica del XIX que hizo del comprender el método de las ciencias del espíritu, Gadamer hace propio el descubrimiento fenomenológico de Heidegger de la precomprensión, pero en lugar de insistir sobre su naturaleza ontológica, examina cuidadosamente la manera como esta comprensión se realiza en tanto interpretación comprensora.

¹⁵ Que la cuestión de la mediación yace en el corazón de la interpretación lo muestra la etimología de este último término: del latín *interpres* o *interpretis*, que significa “mediador”. En el término alemán *Auslegung* resulta menos fácil captar este matiz. Del mismo modo, parece evidente que toda interpretación realiza una mediación entre algo a lo que se le adscribe un sentido (un texto, una acción) y alguien a quien se le hace accesible ese sentido. Por ello es usual que el paradigma de la interpretación sea la traducción. Con esta asunción básica, sin embargo, no se ha aclarado en nada la naturaleza de esta mediación. En todo caso, algunos críticos de la universalidad de la interpretación han entendido claramente este punto y buscan poner un límite a la interpretación al fijar un límite para la mediación: “la tesis de una mediación sin límite es problemática: si no hay nada inmediato no se entiende cómo la mediación misma sería posible” (Élissalde, 2000, p. 181).

¹⁶ En adelante citado como *VM*.

Un primer momento de este examen consiste en reconocer el carácter esencialmente prejuicioso de toda comprensión. Con ello ya estamos más allá (o más acá) de Heidegger, pues lo que en este se mostraba como una estructura ontológica *a priori*, esto es, como la apertura de significatividad en la que por necesidad se encuentra comprendido todo Dasein, se presenta aquí en su dimensión histórica concreta.¹⁷ En efecto, el carácter de lo “previo” con el que Heidegger señala de manera aún formal la estructura del comprender toma en Gadamer el rostro de los prejuicios, es decir, de aquellos contenidos significativos que, surgiendo de las tradiciones a las que pertenecemos, orientan nuestra praxis y constituyen nuestro ser concreto. Lo previo del comprender ya no es entonces una indicación formal vacía, sino que se llena con los prejuicios específicos que conforman la tradición, si bien esta última hay que entenderla más como un horizonte móvil y compartido de sentido que al modo de una comunidad histórica específica.

El reconocimiento del carácter prejuicioso del comprender y de nuestra pertenencia a tradiciones no debe leerse entonces como expresión de un rancio conservadurismo de cuño reaccionario. Con estos elementos, Gadamer prolonga el análisis ontológico heideggeriano pero acentuando ahora aquellos aspectos de la comprensión que la fuerzan a desarrollarse, es decir a desplegarse y realizarse como interpretación. No es posible, en efecto, “mantener mucho tiempo una comprensión incorrecta sin que se destruya el sentido del conjunto” ni “mantener ciegamente los propios prejuicios” (Gadamer, *VM*, p. 335) frente a la alteridad insuperable de otra opinión. En estos momentos se hace evidente la unidad originaria del fenómeno pues aquí, por su propio impulso, la comprensión se vuelve interpretación; en esta última no

¹⁷ En ese sentido, “la contribución de Gadamer a la empresa de Heidegger consiste en agregar y elaborar el contexto histórico de lo que aparece en Heidegger como una descripción fenomenológica, más o menos estricta, de la estructura de lo previo” (Fehér, 2016, p. 283). No obstante, lo que aquí aparece como una “contribución” señala también uno de los puntos en que los dos enfoques en apariencia divergen, a saber, la cuestión de la tradición. En efecto, mientras que para Heidegger la tradición debe desmontarse con el fin de lograr una descripción más originaria de la vida fáctica, en Gadamer esta tiene un aspecto productivo como transmisora de los prejuicios que hacen posible la comprensión. Para un análisis matizado de estas presuntas diferencias, cfr. Olay (2013).

se pasa a algo distinto, sino que el que quiere comprender tiene que interpretar, y no entregarse al “azar de sus propias opiniones” ignorando obstinadamente la “opinión del texto” (Gadamer, *VM*, p. 335).

Esta última alusión al texto no debe dar lugar a confusiones. La comprensión interpretante que aquí se examina ya no se sitúa solamente en el marco de una hermenéutica filológica, sino que, determinada por el análisis ontológico de Heidegger, reviste un carácter universal y está por ello presente en la totalidad de la experiencia humana. En ese sentido, el segundo momento del análisis en el que Gadamer se ocupa de determinar más de cerca la realización de la interpretación, una vez reconocida la pertenencia a la tradición, no vale solamente para el trabajo de las ciencias modernas del espíritu, sino que busca esbozar un proceso propio de la praxis humana en general: no un procedimiento o método científico, sino la realización de la existencia humana interpretante que se puede rastrear en cualquier ámbito de su experiencia. Para nombrar este segundo momento, Gadamer emplea varias nociones, pero podemos concentrarnos aquí en la que parece recoger toda la complejidad del proceso: la *fusión de horizontes*.¹⁸

Esclarecer esta noción implica, en primer lugar, reconocer que la pertenencia a la tradición, como primer momento de la comprensión, no significa afirmar una unidad incuestionable con esta. En realidad, afirma Gadamer, en nuestra relación con la tradición, en tanto horizonte de sentido que nos sostiene, lo que se da es más bien una “verdadera polaridad de familiaridad y extrañeza” (Gadamer, *VM*, p. 365). En el desarrollo de este argumento, esta extrañeza se presenta primero al

¹⁸ Esto puede objetarse si se considera que la noción se introduce en *VM* para mostrar que la comprensión histórica se realiza como un proceso de fusión entre el horizonte del presente y el del pasado (cfr. Gadamer, *VM*, p. 376). En esa medida, parece tratarse de una noción metodológica que busca resolver, al modo de una “realización controlada”, los problemas peculiares de una comprensión “convertida en tarea científica” (*VM* p. 377). No obstante, como ocurre en Gadamer con todas las nociones asociadas con la comprensión, la “fusión” tiene un doble aspecto: ontológico y existencial, por una parte; metódico y científico. No hay que tomarla solo como un momento procedimental, sino de manera más fundamental, como algo que realiza la existencia en sus formas más inmediatas de praxis. Para Thaning (2015, p. 115), la fusión es una noción trascendental que está en obra cada vez que comprendemos, y Grondin (2000, p. 150) señala que con ella no se solucionan tareas metodológicas, sino que se desarrolla una forma de “estar despiertos” que es esencial para el comprender.

modo de una instancia de alteridad que hace frente a la comprensión hasta entonces asegurada y viene a perturbar la relación de familiaridad que se tenía con las cosas. Se trata del encuentro inesperado con lo otro, que nos incomoda pues pone en evidencia las limitaciones de algunos de los prejuicios que hasta entonces sostenían nuestro mundo entero de significaciones y orientaban nuestra acción en él. Una alteridad de este tipo puede manifestarse al modo de un acontecimiento inesperado que disloque las significaciones culturales habituales, en la emergencia de nuevos fenómenos sociales, ante la evidencia de un evento del pasado que nos resulte incomprensible desde el presente, en el encuentro con un grupo humano desconocido o simplemente en las fricciones que enturbian con frecuencia el trato con otra individualidad.

Estos ejemplos representan casos paradigmáticos de la comprensión social, la comprensión histórica, la comprensión multicultural o la comprensión del tú, que tienen lugar o bien de manera silvestre y espontánea, o bien en las formas metodológicamente protocolizadas del conocimiento científico. Con toda evidencia cada una de estas modalidades del otro reviste especificidades propias que determinan un tipo particular de realización de la comprensión interpretante. Pero más que poner de relieve las diferencias, me interesa mostrar que todas estas instancias de alteridad representan rupturas de una verdad establecida que interpelan a la comprensión y la invitan a interpretar, es decir a desplegar nuevas significaciones que restauren la trama de sentido y renueven el campo de lo comprendido. En último término, lo que en todas ellas se constituye es un horizonte de lo singular o de lo particular que hace frente al horizonte general de sentido, y obliga a movilizar la comprensión interpretante que cumple entonces su propósito cuando restaura el sentido en una fusión de los horizontes contrapuestos.

Todo lo anterior apunta a mostrar que esta fusión de horizontes, que encarna el fenómeno unitario de la comprensión interpretante, tiene la estructura de la mediación: mediación entre el horizonte de una verdad general compartida y el horizonte más o menos puntual de una alteridad que la disloca, entre la tradición y lo otro o, en términos más formales, entre lo universal y lo singular o particular. Así las cosas, la comprensión es interpretación porque en ella una verdad general logra mediar con una instancia particular ajena produciendo un proyecto de sentido más rico, o lo que es igual, porque lo que al principio se manifiesta como una singularidad impenetrable consigue recogerse en la generalidad de un nuevo campo de significado. La interpretación no

opera como la sustitución plana de un texto oscuro por otro que hace explícito lo que allí se dice, sino que tiene lugar porque una verdad solo general se media con lo concreto, o lo que es equivalente, porque un fenómeno en apariencia aislado se integra en un marco de significaciones compartidas. No es otra cosa lo que realiza la comprensión histórica cuando consigue mediar la aparente opacidad de un evento inesperado con la significatividad general que sustenta la tradición, y no porque esta última se imponga sobre el primero, sino porque de esta mediación surge una nueva verdad que revitaliza lo transmitido con la apertura de nuevos posibles. Y lo mismo pasa en la comprensión intercultural, como ha mostrado Taylor (2005, p. 211; 2009, p. 108), que tiene lugar como una verdadera mediación entre culturas en la que cada una va reconociendo su propia situación relativa a la otra en un juego de contrastes donde ninguna predomina de manera absoluta.

Parece claro que el tipo de mediación que se da en el caso de un acontecimiento singular inesperado (una revolución, una epidemia) no es el mismo que ocurre cuando se trata del encuentro de dos horizontes culturales distintos. En términos lógicos, el primero demanda la mediación de un universal con un singular, mientras en el segundo lo que se media son dos totalidades particulares. Pero esta circunstancia no refuta el carácter mediador de la interpretación, sino que solo muestra su complejidad. Ya Platón había advertido este hecho y había concebido su dialéctica como la ciencia suprema que realiza la mediación atendiendo a la diversidad de sus formas: como mediación de los *eidos* con las cosas particulares, como mediación de las ideas entre sí, como mediación entre ellas y los llamados géneros vocales, etc. E incluso en la maravillosa reflexión sobre la comunicación de los géneros que aparece en el *Sofista* (253d) contempla la posibilidad de que varios *eidos* permanezcan separados sin mediación posible entre ellos. Frente a esta profunda penetración en el problema de la mediación que realiza la dialéctica platónica, el método científico moderno, que solo conoce la mediación en la forma de la ley universal que determina de manera externa los casos concretos particulares, resulta pobre y unilateral.

Con esta referencia a Platón podemos retornar a la cuestión de la dialéctica. En efecto, también en la compleja mediación interpretativa que ocurre como fusión de horizontes Gadamer reconoce una forma peculiar de dialéctica que él llama *dialéctica de pregunta y respuesta* y que tiene como modelo precisamente al diálogo socrático. Lo que en esta referencia a Platón destaca Gadamer como dialéctico es, ante todo, la

vinculación entre pensar y dialogar. “Dialéctica” se entiende entonces como “el arte de llevar una auténtica conversación” (Gadamer, *VM*, p. 444) y esto es válido tanto para Platón como para la determinación hermenéutica de la comprensión interpretante. Esta referencia a la conversación señala, primero, que la verdadera comprensión no tiene lugar como una meditación en solitario o al modo de una argumentación en paralelo donde cada parte trata de refutar las posturas rivales para hacer prevalecer su propia posición. Pero lo importante en esto no es tanto el fijar el número de los que participan de la conversación, sino asegurar que la comprensión que en ella ocurre se abra hacia lo otro que la disloca y fractura. Lo que aquí se promueve, en últimas, no es precisamente un diálogo intersubjetivo, sino más bien un diálogo con las cosas mismas: uno que acoja los fenómenos en su alteridad real y que, en solitario o en compañía, sea capaz de reconocer los límites de su campo establecido de sentido ante la otredad que se manifiesta. En este contexto, el verdadero rasgo definitorio de este diálogo o conversación hermenéutica es que este está animado por un constante preguntar, pues una genuina pregunta ocurre precisamente como una ruptura de los significados ya fijados, que demanda una ampliación del ámbito de lo comprendido.¹⁹ El buen intérprete es, entonces, quien mejor dialoga, pero no porque sepa concertar acuerdos o alcanzar consensos sino porque posee el arte de preguntar, que es el arte de acoger la interpelación de lo otro con el que recomienza la interpretación.

Sin embargo, no se trata tan solo de que el comprender se realice desde la pregunta que lo desestabiliza y lo fuerza a devenir interpretación: más que un intercambio plano de preguntas y respuestas, el “arte de la conversación” ocurre como una progresiva penetración en el ámbito de sentido de lo que se busca comprender, y esta es la verdadera forma de la mediación. Que la interpretación es mediación significa entonces que no tiene lugar como una mera presentación de opiniones alternativas, pero tampoco como la simple subsunción de un singular en un concepto general. Es mediación en el sentido en que es mediación constante, fusión de lo singular que interrumpe con un horizonte general de

¹⁹ De aquí que no se trate de una dialéctica que culminaría, como en Hegel, en la parusía del absoluto. Gadamer mismo ha señalado que en el recurso al diálogo socrático-platónico intentó detectar “el fenómeno originario del lenguaje” (1992, p. 321) buscando una reorientación hermenéutica de la dialéctica especulativa del idealismo alemán.

sentido nunca del todo esclarecido y que solo se va desplegando con las posibilidades que se hacen visibles desde una alteridad concreta.²⁰ El arte de preguntar no solo es, pues, el arte de dejarse interpelar, sino el arte de mantener con las preguntas correctas el diálogo que penetra en la dimensión de sentido de lo investigado. Esta dialéctica de inspiración socrática constituye así la lógica de realización del comprender interpretante de la hermenéutica.

La interpretación, o comprensión interpretante, que revela la filosofía hermenéutica como rasgo esencial de la existencia humana es, pues, dialéctica. Se trata por lo pronto del sentido básico de la noción que la liga con la conversación y el diálogo. Ya desde Platón, sin embargo, este diálogo no es un mero intercambio de pareceres u opiniones personales, sino un ejercicio por el cual se va develando un campo de sentido común a los que conversan, y a la luz del cual estos van corrigiendo o afinando sus puntos de vista iniciales. Esta conversación dialéctica es una forma de mediación, pues allí se vinculan las opiniones solo individuales con un marco de significación más amplio que ya no es privilegio de uno solo sino un horizonte de verdad que cobija a todos. Por ello, contra toda apariencia, cuando, en los diálogos de Platón, los jóvenes atenienses deponen sus opiniones ante la fuerza de las preguntas de

²⁰ No es posible desplegar aquí, ni siquiera de manera general, la constelación de cuestiones y problemáticas que queda definida con la relación entre dialéctica y hermenéutica en el pensamiento de Gadamer. Aquí nos hemos limitado a poner el foco, y eso de manera aún introductoria, en el carácter dialéctico de la interpretación, para marcar una diferencia central entre Gadamer y otras filosofías de la interpretación. No obstante, somos conscientes de que, al entender la interpretación como una dialéctica de mediación, pisamos un terreno resbaladizo y abierto a la polémica. En un comentario ya clásico, Caputo (1987, pp. 81 y 113) coincide con nosotros en entender la fusión de horizontes como una versión contemporánea de la mediación dialéctica hegeliana, pero en ello ve un gesto metafísico que le impide a Gadamer desarrollar una genuina hermenéutica radical. Risser (2012, p. 70), por su parte, reconoce el carácter dialéctico de la interpretación en Gadamer, pero señala que esta dialéctica sigue el modelo dialógico del diálogo abierto Platón y no el de la mediación con una totalidad cerrada propio de Hegel. Sin desconocer las tensiones con Hegel, aquí sostenemos que la mediación de la interpretación en Gadamer ni presupone una totalidad metafísica ni se opone a la generación de múltiples proyectos de sentido. Para una elaboración amplia del carácter dialéctico de la hermenéutica de Gadamer, cfr. Swayne (2010).

Sócrates, no es este último en tanto individuo quien los deslumbra con su autoridad, sino el poder del *logos*, de la razón universal inscrita en las cosas, quien los convence de la unilateralidad de sus puntos de vista. Gadamer considera que una conversación de este tipo, que interpreta la realidad mediando lo singular que acontece con lo comprendido en la tradición, define el trabajo de las ciencias del espíritu. Estas interpretan el mundo social al modo de una dialéctica de pregunta y respuesta en una mediación siempre atenta a lo novedoso sin aspirar a una verdad absoluta que fuera invulnerable a la contingencia del acontecer.

3. La mediación total del lenguaje

Con todo, la caracterización hecha hasta ahora no recoge aún todo el sentido de la unidad originaria de la comprensión y la interpretación, ni da cuenta plenamente de su carácter dialéctico universal. Ella se enfoca en los rendimientos de una comprensión temática, es decir de una que se vuelca sobre un objeto (un texto, un evento del pasado, una cultura extraña, otra individualidad) y por ello puede transformarse potencialmente en una postura teórica objetivante. Vista más de cerca, esta comprensión temática solo tiene lugar como una sucesión de etapas en la que una primera comprensión alcanzada se enrarece, con lo que desencadena un momento interpretativo que restaura el sentido para alcanzar una nueva comprensión, y así sucesivamente. La dialéctica de pregunta y respuesta que estructura este proceso solo refuerza la impresión de que lo que aquí ocurre es una serie sucesiva de momentos de interpretación (preguntas) y momentos de comprensión (respuestas) que no hace justicia a la unidad del fenómeno. Dicho de otra forma, la comprensión no sería ella misma interpretación, sino solo tendría un vínculo necesario con esta última. Y por lo mismo la mediación entre una universalidad o generalidad de sentido y una singularidad concreta no sería completa, pues se trataría de instancias distintas repartidas en momentos distintos: la comprensión recogería siempre lo universal de un sentido compartido, mientras que solo la interpretación representaría el momento donde se presta la voz a lo singular.

Se trata entonces, para Gadamer, de hacer manifiesta esta unidad en la que se cumple una mediación plena entre el comprender y el interpretar, entre lo universal y lo singular concreto, de modo que estos ya no sean simples momentos sucesivos aunque necesariamente ligados. A mi parecer, el análisis del lenguaje que constituye el tema de la tercera parte de *Verdad y método* busca abrir el terreno donde esta unidad se

hace evidente. De manera explícita Gadamer (VM, p. 461) afirma que este paso al lenguaje tiene como objeto determinar el médium en que la experiencia de la comprensión tiene lugar, algo así como la condición ontológica que hace posible el comprender efectivo, histórico y temático examinado antes.²¹ Pero resulta claro que ese paso a lo ontológico no significa la erección de la comprensión sobre un fundamento incondicionado, sino más bien el acceso a un ámbito donde la unidad de lo universal y lo singular, del comprender y el interpretar, no se ha roto aún como ocurre ya en la sucesión dialéctica de pregunta y respuesta que determina toda comprensión temática. Con Hegel, se podría decir, *mutatis mutandis*, que el paso al lenguaje representaría el momento de la inmediatez que se rompe enseguida (o ya se ha roto desde siempre) en la comprensión histórica donde se distingue el sentido del sinsentido, la tradición del acontecimiento, lo familiar y compartido de la extrañeza. Como en Hegel también, esa inmediatez de la unidad del lenguaje es solo aparente, pues a ella le es inherente una forma particular de mediación que examinaremos enseguida. Como sea, lo que sí es claro es que el lenguaje no representa aquí una abstracta condición trascendental de la comprensión, sino un auténtico fenómeno, y por ello la unidad que le es esencial no es mero postulado sino algo de lo que podemos hacer experiencia.

Según esto, el análisis sobre el lenguaje de Gadamer no es una teoría sobre este, sino una auténtica fenomenología.²² La pregunta hermenéutica por el lenguaje no es entonces por su lógica operativa o por sus rasgos determinantes, sino por la experiencia más inmediata que

²¹ Este paso hacia el lenguaje sería entonces un paso atrás hacia la dimensión sobre la que descansa el comprender interpretante, y no la tematización de un nuevo objeto. Por eso afirma Gadamer que “el fenómeno hermenéutico se revela en su carácter lenguajístico” (*Sprachlichkeit*) y que el lenguaje, como la comprensión, no pueden tomarse como objeto porque estos “abarcan todo lo que de un modo u otro puede llegar a ser objeto” (VM, p. 485).

²² Una fenomenología que en la tradición de Heidegger hace del fenómeno algo que se oculta: “Lo que es el lenguaje forma parte de lo más oscuro que existe para la reflexión humana. La lenguajidad le es a nuestro pensamiento algo tan terriblemente cercano, y es en su realización algo tan poco objetivo, que por sí misma lo que hace es ocultar su verdadero ser” (Gadamer, VM, p. 457). Que el suyo sea un abordaje fenomenológico pone también en evidencia la distancia de Gadamer con respecto a los enfoques analíticos de la filosofía del lenguaje y el *linguistic turn* (cfr. Fehér, 2000, p. 192).

hacemos de este, por la manera en que este comparece primordialmente ante la experiencia humana antes de que esta, transformada en un comportamiento objetivante, haga de él un mero instrumento a su disposición. Pues bien, para esta fenomenología, antes de que el lenguaje sirva como herramienta para describir el mundo o los estados internos del sujeto, y por supuesto mucho antes de que se nos presente en su carácter de estructura o como un conjunto de reglas sintácticas o gramaticales, el lenguaje comparece ante nuestra experiencia como sentido, es decir como apertura de un espacio de significatividad que nos convoca y nos atrae. En efecto, la palabra que alguien nos dirige, o incluso aquella que simplemente escuchamos por azar, tiene algo de una fuerza gravitacional que nos arrastra e interpela. Incluso si se trata de una palabra quebradiza o difusa, y aun tratándose de una que proviene de una lengua ajena, la palabra que se emite o se escucha porta y abre consigo todo un horizonte de sentido que invita a ser compartido.²³

Como es claro ya, no se trata, en esta fenomenología, del lenguaje del historiador, el filólogo, el científico social o el dialéctico platónico que va comprendiendo e interpretando su objeto con nociones que, entre preguntas y respuestas, se van refinando, penetrando de a poco en el campo de sentido de lo investigado. Aquí se trata de la experiencia más elemental con el lenguaje, donde ya no se distinguen momentos sucesivos que se interrumpen, sino que tiene lugar como un *acontecer* de sentido que nos porta como un movimiento continuo donde el sentido es inseparable de su concreción y la comprensión no se separa de su objeto.²⁴ La experiencia primaria con el lenguaje es, pues, la de una

²³ Se trata de una apertura de sentido que, no obstante, es a la vez la experiencia de su límite y del encuentro con la alteridad: "Lo que se manifiesta en el lenguaje no es la mera fijación de un sentido pretendido, sino un intento en constante cambio, o más exactamente, una tentación reiterada de sumergirse en algo con alguien" (Gadamer, 1992, p. 324).

²⁴ Aquí no puedo explorar con detalle la noción de "acontecer", o "acontecer de sentido", que para Gadamer describe fenomenológicamente el ser del lenguaje. El tema recorre toda la tercera parte de *VM*: desde la caracterización de la palabra cristiana como acontecer de sentido (Gadamer, *VM*, p. 512), pasando por el acontecer de la tradición cuya palabra llega hasta nosotros para la comprensión (*VM*, p. 553), hasta el planteamiento último del lenguaje como un acontecer especulativo donde lo finito de lo dicho se liga con la infinitud del sentido (*VM*, pp. 565 y 567). No obstante, desde la caracterización de la interpretación como mediación que aquí elaboro se hace visible este carácter

apertura, como dice Gadamer, pero esta apertura no es la de un espacio iluminado donde se hubieran establecido posiciones de significación fijas para ser luego tematizadas y discutidas, sino como un salir a luz permanente de diversas configuraciones de significación a cuyo paso se va articulando un mundo que, en su diversidad de posibles, es común a los que comparten esa experiencia.

Este acontecer de sentido, como realización primaria del lenguaje, es el ámbito en el que se alcanza la mediación total entre la universalidad de un sentido compartido y lo singular concreto, y por ello en él se hace evidente la unidad originaria entre comprensión e interpretación. Como se ha dicho, no se trata aquí de un postulado *a priori*, sino de un fenómeno que se constata en la experiencia. Esa mediación se advierte, primero, si se atiende a la vida misma del lenguaje que hace surgir un sentido general y comunicable mediante el flujo de palabras individuales. Aquí la mediación es total porque el sentido general (el momento universal) solo surge ligado a lo singular de las palabras individuales, pero estas a su vez solo exhiben su verdad en el contexto de lo general que las soporta. No se distinguen aquí dos elementos separables, sino que lo general solo se da mezclado en la combinatoria de lo singular, y lo singular solo gana su ser en lo general. O, más precisamente, en el flujo del lenguaje ya no se puede hablar de una singularidad y de una universalidad puras que se pudieran aislar de manera independiente: solo hay singulares universalizados y un universal que penetra y envuelve a las palabras.²⁵ Paul Valéry ha ideado una bonita imagen para representar esta mediación total de la palabra y el sentido. Para

aconteciente, si bien no llegamos a su caracterización ontológica última. Para un análisis más cercano de este nexo entre lenguaje y acontecer, cfr. Esquisabel (2003, pp. 289 y ss.); para la relación de este acontecer del lenguaje con el carácter emergente y performativo de la interpretación, cfr. Davey (2006, pp. 234 y ss.).

²⁵ Hay que advertir, sin embargo, que el carácter acontecimental del lenguaje implica que la mediación total de la que hablamos aquí no debe entenderse como un acople o vinculación estática, sino como un acuerdo que se disloca permanentemente. Por ello hay un “límite implícito en toda experiencia de sentido” y lo “mentado en el lenguaje rebasa siempre aquello que se expresa” (Gadamer, 1992, p. 323). La interpretación es mediación, pero se trata de una “mediación nunca perfecta entre hombre y mundo”, una que ni es añoranza de una inmediatez perdida ni busca arribar a un sentido último; en esa medida, lo único “dado” es la mediación, “el que siempre comprendemos algo como algo” (Gadamer, 1992, p. 327).

él, el sentido se da siempre como en el flujo móvil de las palabras, pues solo se comprende si logramos sumirnos en este flujo sin detenernos en los pasos individuales de las palabras que lo constituyen. Solo hay comprensión gracias a “la velocidad de nuestro paso sobre las palabras” (Valéry, 1973, p. 400), pues cada palabra individual que en su fluidez va desplegando el sentido se parece a una de esas tablas que se ponen sobre una zanja o una grieta y que soportan el paso de un hombre en movimiento, pero que se tambalea o se rompe, hundiendo con ello el sentido, cuando este hombre se detiene y hace de ella un objeto de su atención.

Vista con detenimiento, se pueden identificar, en la imagen de Valéry, varias formas de mediación. En primer lugar, aquella que realiza el lenguaje en tanto acontecer donde se compenetran la palabra individual y el todo del sentido, pero luego este flujo del lenguaje no se autonomiza ni se hace objeto para una conciencia, sino que solo es manifestación de sentido para quien comprende, esto es, para quien se deja comprender por él, para quien participa y se integra a este movimiento y avanza a la “velocidad de las palabras”. Esta es la segunda mediación, a saber, la que unifica la comprensión y su objeto, pues el sentido no pasa al frente del comprender, sino que va aconteciendo al paso que este se produce. Pero en realidad cabría hablar aquí de una tercera mediación, porque ese sentido comprendido, o que se va comprendiendo, no requiere luego ser interpretado, es decir mediado con la propia situación de quien comprende, sino que ya es siempre sentido interpretado: es comprendido en tanto es interpretado, o sea concretado en una configuración particular de la significación y de la acción. Esta tercera mediación representa entonces, ahora sí, la verdadera unidad originaria del fenómeno del comprender interpretante, que no era evidente aún en las formas de comprensión temática de las ciencias sociales.

Ahora bien, en esta experiencia elemental del lenguaje que estamos analizando, esas tres mediaciones no se distinguen en realidad y todo tiene lugar como en un mismo impulso. En esta realización primaria del lenguaje, cuando yo hablo, o cuando escucho hablar, asisto a una apertura de un campo de sentido donde ya no capto la contribución, por decirlo así, de cada palabra a la erección de este; pero al mismo tiempo, el sentido que así comparece no es un objeto para la comprensión, sino que justamente se muestra como el espacio de significación en que ya estoy comprendido, y por último, este espacio es a la vez comprendido como algo que me dice algo a mí en mi situación, como algo que ya

se ha coloreado con rasgos significativos que provienen de mi contexto particular, y solo así es visible, es decir, se trata de una comprensión interpretada.

Se puede objetar que hay realizaciones del lenguaje donde estas formas de mediación total no tienen lugar. La poesía, p. ej., sería el caso de un campo de significación donde el peso específico de cada palabra individual no es despreciable; las ciencias sociales, por su parte, se ocupan de elaborar conceptos para configuraciones de sentido del mundo social que son opacas y que por ello tienden a objetivarse, y un caso como el de la traducción parece ser más bien el de un sentido que el traductor ha comprendido pero para el que busca aún una interpretación concreta en otra lengua. Todo esto es cierto, pero aquí se trataba de examinar una experiencia más básica, un rendimiento vital donde el lenguaje no se hace explícito, donde no llama la atención sobre sí ni como palabra poética ni como concepto ni como término por traducir, sino que se integra como flujo del sentido en el fluir mismo la existencia humana, comprensora e interpretante. Se trata de una experiencia más básica no por ser la más verdadera, sino porque en ella el lenguaje se manifiesta aún en la unidad primaria de lo universal y lo singular, del comprender e interpretar, que luego pierde constantemente en sus otras realizaciones.

Gadamer la llama “una realización vital” (VM, p. 463) e incluso llega a afirmar que en ella no se da aún ningún proceso interpretativo. Pero lo que quiere decir con esto es más bien que la interpretación no se destaca aquí: ocurre de forma espontánea y no como un hacer del individuo, sino como parte del movimiento de las cosas mismas. En esta vida espontánea del lenguaje comprender e interpretar se funden, pues el sentido que el lenguaje abre no flota en el aire, sino que de inmediato se particulariza y se vuelve concreto, no solo porque se encarna por necesidad en una lengua histórica específica, sino además porque cada hablante y cada oyente lo recoge en su propia situación y lo tiñe con las significaciones de su ahí particular. Así las cosas, todo acontecer del sentido en el lenguaje se transforma por sí mismo en una forma de traducción, no porque allí el sentido se desplace deliberadamente al horizonte de otra lengua, sino porque la comprensión ejecuta naturalmente un desplazamiento a cada situación individual en la que se realiza, de modo que se diversifica y dispersa, manteniendo no obstante su unidad fundamental. Toda palabra compartida es siempre ya una palabra traducida a nuestra situación, interpretada, y por eso

ella ya ha abierto, aún en la comunidad que instala, microdiferencias y desajustes que son el origen de la incomprensión, aunque en sus formas primeras ellas no enturbien la comunicación ni la vida práctica.

Las anteriores consideraciones llevan de vuelta al tema de la dialéctica. En efecto, este análisis fenomenológico ha fijado la atención en la realización más propia del lenguaje y ha puesto de manifiesto la unidad del comprender y el interpretar que allí tiene lugar, como una mediación total entre lo universal y lo singular concreto. Pues bien, lo que hay que destacar ahora es que todo ello lo hace el lenguaje por sí mismo y no tanto los individuos que hablan o escuchan. Es el lenguaje el que comparece ante nosotros como apertura de sentido que nos comprende y orienta; es el lenguaje el que realiza por sí mismo esa concreción particular que es la interpretación y por la cual gana significado nuestra situación propia. Claro que hay ejecuciones del lenguaje proyectadas por una conciencia subjetiva que toman al lenguaje como su herramienta: explicar, argumentar, explicitar, proyectar, y todas las formas de atribución de sentido que Wittgenstein llama “interpretar”. Pero este tipo de actividades son derivadas, pues toman pie en la apertura que ya ha instalado el lenguaje; dicho de otra forma, en ellas una conciencia hace del lenguaje un instrumento para su uso aprovechando la potencia de significación que le es inherente. Por su propia naturaleza, por su propio ser, como diría Gadamer, el lenguaje es el acontecer y deriva de un sentido que se realiza en trabazón con la comprensión que interpreta, y que por ello lo especifica en una situación concreta. Nosotros participamos de esta deriva y ayudamos a realizar el comprender, pero en esto no ocupamos nunca el centro que orienta o impulsa el movimiento, y en realidad participar aquí es integrarse al movimiento de las cosas mismas, ajustando nuestro paso a la velocidad de las palabras como dice Valéry.

Pues bien, esta participación en el ser de lo real o en el movimiento de las cosas mismas resulta ser, para Gadamer, el rasgo central de lo que desde Platón se llama “dialéctica”. La dialéctica puede entonces entenderse, primero, como un método de conocimiento filosófico: el arte de la conversación de Sócrates como arte de penetrar mediante preguntas en las articulaciones del *logos*, pero también el método hegeliano que no es un hacer de la conciencia subjetiva, sino un entregarse de esta al movimiento del absoluto que se manifiesta, son formas de esta dialéctica. También Gadamer, como vimos, estructura la comprensión interpretante de la historia o la filología al modo de una

lógica de pregunta y respuesta que es, en esencia, una figura de este método dialéctico.

Pero antes de ser un método, la dialéctica encarna toda una concepción acerca del ser de la realidad y del lugar que allí ocupa el ser humano. Para esta concepción, la realidad no está fijada al modo de constelaciones objetivas de hechos, sino que aparece como un horizonte articulado por significaciones y sentidos del cual participa el ser humano, no al modo de un sujeto frente a un objeto, sino como aquella instancia donde el sentido gana claridad sobre sí mismo. En ese orden de ideas, Gadamer reconoce la esencia de la dialéctica no en su carácter de método, sino como unidad entre ser y pensar, como “mediación total de pensamiento y ser”. Desde su enfoque hermenéutico esta mediación ya no queda asegurada en el elemento metafísico del *logos* como en Platón, o del *espíritu* como en Hegel, sino que tiene lugar en el lenguaje como el médium fenoménico donde se realiza la “trabazón de acontecer, comprensión e interpretación” (Gadamer, *VM*, p. 552), y por ello podemos tener experiencia de ella. La interpretación hermenéutica es entonces dialéctica, primero porque tiene lugar como un diálogo que se entabla con lo otro que descentra mi horizonte de significado y fuerza la comprensión a desplegarse, pero ante todo porque ella tiene lugar cuando el ser humano reconoce su participación en un horizonte general de sentido que él realiza en lo concreto, no de manera arbitraria y externa, sino siguiendo las posibilidades de significación inscritas en la realidad misma.

4. Resultados y conclusiones

Para concluir, quiero recoger en unas pocas tesis todos estos complejos planteamientos y mostrar de manera muy somera algunas vías de investigación que se desprenden de ellos. Aquí he tratado de elaborar, con atisbos de la hermenéutica de Gadamer, una concepción de la interpretación para la cual en toda interpretación tiene lugar una mediación total entre un sentido general y una situación particular que realiza y concreta este sentido. Esto significa entonces que:

(1) Interpretar no es el simple ejercicio por el cual un fenómeno cuya significación es oscura resulta esclarecido. No es sobre todo la sustitución plana de un texto por otro que lo explicita e ilumine, y que entonces ya no tendría necesidad de ser interpretado. En general, no se hace justicia al interpretar si se lo toma como la actividad por la cual un individuo da sentido a una realidad que no lo tiene. Tampoco Nietzsche,

con su llamado a interpretar desde múltiples voluntades de poder, está cerca de atisbar toda la complejidad del fenómeno.

(2) La mediación que es la interpretación se hace evidente en las distintas formas de conocimiento de las ciencias sociales, cuyo objeto (el texto, el pasado, lo social) no son hechos sólidos sino configuraciones significativas. En este contexto, comprender (es decir, interpretar) dicho “objeto” significa ser capaz de mediar entre un campo de sentido ya asegurado y las alteridades singulares que acaecen y que no parecen ser acogidas allí. Una mediación similar ocurre en el caso de la interpretación artística o jurídica, y tiene lugar también en aquellas formas de interacción dialógica que producen sentido y no se reducen a un mero intercambio de opiniones: el diálogo pedagógico, el diálogo entre generaciones o entre culturas, etc.

(3) Esta mediación no es una que imponga unilateralmente quien interpreta, sino que tiene lugar al interior del movimiento mismo del lenguaje que nos rebasa, y que como acontecer del sentido siempre supera las posibilidades de significación que el comprender interpretante va haciendo concretas. Por eso mismo, interpretar es mediar, pero la mediación no es un hacer caprichoso de una conciencia individual, sino que su curso tiene lugar en el campo abierto de sentido del lenguaje del que participa el intérprete. En tanto parte del movimiento de sentido del lenguaje, la interpretación es, para Gadamer, universal. Esto no significa que siempre estemos interpretando conscientemente, pero sí que en cualquier forma de nuestra experiencia de mundo se hace efectiva (aunque pase inadvertida) una posibilidad de significación inscrita en el lenguaje.

(4) Por tener lugar en el médium del lenguaje, la interpretación hermenéutica es dialéctica. Aquí “dialéctica” se toma en el sentido elemental del término, esto es, como la pertenencia del ser humano a un espacio de significación que lo condiciona y orienta, pero que él mismo contribuye a articular en lo concreto gracias a su comprensión. Este espacio de sentido no es otro que la vida del lenguaje: un campo móvil que no está articulado en esencias universales, como los *eidos* platónicos o las determinaciones del concepto en Hegel. Vista así, la interpretación pasa por detrás de la oposición entre objetivismo y relativismo: ni consiste en develar un sentido fijo escondido por detrás de los fenómenos, ni opera como un libre adjudicar significaciones a una realidad maleable y sin sustancia.

(5) La interpretación que es parte de la vida del lenguaje no se realiza como un movimiento que se impulse sin resistencias. Al contrario, está amenazada permanentemente por todos los intentos de reducir el lenguaje a unas significaciones fijas que se instrumentalizan por intereses particulares. Cuando esto ocurre, el lenguaje oculta la pertenencia común al sentido y deviene un campo de confrontación entre proyectos de sentido individuales que pugnan por imponerse. También se pervierte la mediación viva del lenguaje si eso común se reduce a una supuesta verdad universal que quiere imponerse sobre la amplia diversidad del mundo. La primera tarea de la interpretación hermenéutica es entonces estar en guardia contra las objetivaciones del lenguaje, su instrumentalización y control.

(6) De igual manera, la mediación no tiene lugar como una simple realización espontánea que ocurriera sin fricciones. Ella requiere de un esfuerzo sostenido que define las tareas propias de la comprensión interpretante. El fracaso o la ruptura de la mediación es lo que llamamos la “no comprensión” o la “no interpretación”. Estas tienen lugar cuando resulta imposible la mediación entre el universal y el particular: cuando, por ejemplo, disponemos de una intelección solo abstracta de una verdad general que resulta vacía y sin vida porque no logra encarnarse en realizaciones concretas ni traducirse en instituciones o principios efectivos. Esta forma de lo universal sin mediación es quizás lo que define el fenómeno del fanatismo. Pero también fracasa la comprensión interpretante en la otra dirección, cuando el singular no es capaz de superar su aislamiento porque no logra reconocer su pertenencia a formas de sentido compartidas que preceden la constitución de una subjetividad individual. Esta forma de lo singular sin mediación es quizás el rasgo más característico de las sociedades liberales y del capitalismo contemporáneo.

Estas sumarias tesis buscan sintetizar los resultados centrales del análisis desarrollado, pero permiten también determinar el lugar de la hermenéutica gadameriana en el panorama de la filosofía contemporánea de la interpretación, a la vez que subrayan su pertinencia a algunas de las problemáticas más urgentes por las que atraviesan las sociedades actuales.

Con relación a lo primero —tesis (1) y (2)—, debe resultar evidente que el planteamiento hermenéutico encarna una postura muy original dentro del amplio espectro de enfoques filosóficos en torno a la interpretación. Lejos de someter el concepto a un análisis meramente

lógico-lingüístico, como es común en el *linguistic turn* anglosajón, la interpretación es, para la hermenéutica, un rasgo ontológico constitutivo de la realidad y de la existencia humana. Esta ontologización de la noción no conduce sin embargo al extremo antirrealista, tan corriente en ciertas posturas posmodernas, para el cual toda presunta realidad no es más que un juego de significantes o la resultante de una pugna entre fuerzas interpretativas solo particulares. Más allá de la concepción lógica y de este último enfoque constructivista y relativista, la hermenéutica elabora una poderosa noción de “interpretación” como mediación que, a mi modo de ver, hace más justicia a la complejidad del asunto, y evita los *impasses* en que caen las otras posturas. Por lo demás, más que de la elaboración teórica del concepto, la hermenéutica opera aquí como una auténtica fenomenología que sustenta su análisis en las diversas experiencias de interpretación presentes en el ancho campo de la praxis humana: desde la conversación más trivial hasta la profunda experiencia del arte, pasando por supuesto por la experiencia de comprensión de las ciencias sociales e históricas.

Gran parte de la originalidad de este enfoque hermenéutico reside en la rehabilitación de la dialéctica que le acompaña. Como se señaló antes —tesis (3) y (4)—, la dialéctica de la hermenéutica quiere dejar atrás el compromiso metafísico de la dialéctica tradicional (Platón o Hegel) rescatando de allí, sin embargo, el contenido nuclear de la noción. Según este, más que una metodología de análisis filosófico, la dialéctica hace efectiva la mediación del ser humano, situado e histórico, con el orden de sentido superior al que este pertenece. Para la hermenéutica, ese horizonte abarcante de posibilidades de significación no es otro que el del lenguaje en su acontecer dinámico. La interpretación hermenéutica es dialéctica porque permite a los individuos introducirse en esta apertura móvil del lenguaje y hacer manifiesta la diversidad de sentidos que vertebran su situación. De esta manera, un genuino interpretar se convierte en la experiencia de algo común: no de una verdad idéntica para todos, sino de un marco de significaciones compartidas que articula nuestro presente concreto en la diversidad de sus posibles proyectos de verdad.

En esta valoración de una experiencia de lo común encuentro uno de los planteamientos más valiosos y actuales de la hermenéutica. Este reconocimiento de un horizonte abierto de sentido que rebasa toda particularidad puede desarrollarse en varias direcciones. Por ejemplo, hacia una ontología posmetafísica donde la oposición clásica entre lo

universal y lo particular, o entre verdad y apariencia, fuese superada en la evidencia de un universal que solo se realiza en la mediación con lo singular y concreto; o hacia una epistemología no fundacionalista que defienda la legitimidad de un conocimiento que no descansa en bases absolutas, sino que tiene lugar como la participación en un acontecer incesante de verdad. Aquí, sin embargo, me interesa mostrar que la experiencia de lo común puede ser la piedra angular de una filosofía social hermenéutica. En cierta forma, esta ha sido la motivación oculta que ha movido todo este análisis. Los resultados (5) y (6) ya insinuaron los ejes de este desarrollo; para concluir, quiero ampliar un poco este punto.

Usualmente considerada como una reflexión sobre la comprensión histórica o como una ontología especulativa del lenguaje, creo que no es advertido de manera suficiente el lugar que puede ocupar la hermenéutica en el contexto de la filosofía y crítica social contemporáneas. Los análisis anteriores sobre la interpretación dialéctica dan indicaciones en este sentido. En primer lugar, desde la estructura de la mediación que está a su base, la hermenéutica puede evaluar fenómenos sociales y políticos actuales e identificar su carácter problemático. Así, ella puede denunciar la injusticia de formas de estado totalitarias, y en general de todo intento por monopolizar la verdad, como quiebres de la mediación esencial de lo universal con lo concreto. Pero también puede denunciar al individualismo radical de nuestras sociedades de consumo como una empresa que absolutiza al singular y le oculta posibilidades de experiencias de lo común que son esenciales para una praxis genuinamente social.

En segundo lugar, como fenomenología del lenguaje, la interpretación hermenéutica toma la forma de una razón práctica cuya tarea es salvaguardar la vida del lenguaje en tanto acontecer abierto de sentido, develando todo intento de envarar su movimiento, de apropiárselo para intereses particulares o de deformarlo con usos y prácticas que desnaturalicen su capacidad de apertura de mundo. En el reino de las redes sociales, la comunicación inmediata y la inteligencia artificial, no es poco lo que hay que hacer en este sentido.

Referencias

- Abel, G. (1995). *Interpretationswelten. Gegenwartsphilosophie jenseits von Essentialismus und Relativismus*. Suhrkamp.
- Abel, G. (2004). *Zeichen der Wirklichkeit*. Suhrkamp.

- Angehrn, E. (2004). *Interpretation und Dekonstruktion. Untersuchungen zur Hermeneutik*. Velbrück.
- Berner, C. y Morel, C. (2023). *Philosophies du sens*. Septentrion. <https://doi.org/10.4000/books.septentrion.147865>
- Berner, C. y Thouard, D. (2015). *L'interprétation. Un dictionnaire philosophique*. Vrin.
- Caputo, J. (1987). *Radical Hermeneutics: Repetition, Deconstruction and the Hermeneutic Project*. Indiana University Press. <https://doi.org/10.2979/2735.0>
- Davey, N. (2006). *Unquiet Understanding. Gadamer's Philosophical Hermeneutics*. SUNY Press. <https://doi.org/10.1353/book5103>
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford University Press.
- De Lara, F. (2015). El ser-en como tal. En R. Rodríguez (ed.), *Ser y Tiempo de Martin Heidegger. Un comentario fenomenológico* (pp. 145-166). Tecnos.
- Demmerling, C. (2001). Hermeneutik der Alltäglichkeit und In-der-Welt-Sein. En Th. Rentsch (ed.), *Martin Heidegger. Sein und Zeit* (pp. 83-108). Akademie.
- Derrida, J. (2006). La mitología blanca". En J. Derrida, *Márgenes de la filosofía* (pp. 247-311). C. González Marín (trad.). Cátedra.
- Descombes, V. (1991). The Interpretative Text. En H. Silvermann (ed.), *Gadamer and Hermeneutics: Science, Culture, Literature* (pp. 247-268). Routledge.
- Élissalde, Y. (2000). *Critique de l'interprétation*. Vrin.
- Esquisabel, O. (2003). Sprache, Geschehen und Sein. Die Metaphysik der Sprache bei H.-G. Gadamer. En M. Wischke y M. Hofer (eds.), *Gadamer Verstehen / Understanding Gadamer* (pp. 282-301). Wissenschaftliche Buchgesellschaft.
- Fehér, I. (2000). Zum Sprachverständnis der Hermeneutik Gadamers. Wort, Gespräch und Sache. En G. Figal, J. Grondin y D. Schmidt (eds.), *Hermeneutische Wege* (pp. 191-205). Mohr Siebeck.
- Fehér, I. (2016). Pre-judice and Preunderstanding. En N. Keane y C. Lawn (eds.), *The Blackwell Companion to Hermeneutics* (pp. 280-288). Blackwell. <https://doi.org/10.1002/9781118529812.ch32>
- Gadamer, H.-G. (1996). [VM]. *Verdad y método*. A. Agud y R. de Agapito (trads.). Sígueme.
- Gadamer, H.-G. (1992). Texto e interpretación. En H.-G. Gadamer, *Verdad y método II* (pp. 319-347). M. Olasagasti (trad.). Sígueme.

- Goodman, N. (1978). *Ways of Worldmaking*. Hackett Publishing. <https://doi.org/10.5040/9781350928558>
- Grondin, J. (1996). L'herméneutique dans *Sein und Zeit*. En J.-F. Courtine (ed.), *Heidegger 1919-1929. De l'herméneutique de la facticité à la métaphysique du Dasein* (pp. 179-192). Vrin.
- Grondin, J. (2000). *Einführung zu Gadamer*. Mohr Siebeck.
- Heidegger, M. (1993). [SZ]. *Sein und Zeit*. Niemeyer.
- Margolis, J. y Rockmore, T. (2000). *The Philosophy of Interpretation*. Blackwell.
- Michel, J. (2017). *Homo Interpretans*. Hermann.
- Michel, J. (2023). *Qu'est-ce que l'herméneutique ?* PUF.
- Mulhall, S. (1996). *The Routledge Guidebook to Heidegger's Being and Time*. Routledge.
- Olay, C. (2013). Die Überlieferung der Gegenwart und die Gegenwart der Überlieferung. Heidegger und Gadamer über Tradition. En G. Figal (ed.), *International Yearbook for Hermeneutics* (pp. 196-219). Mohr Siebeck.
- Platón. (1985). Ion. En Platón, *Diálogos I* (pp. 243-270). E. Lledó (trad.). Gredos.
- Platón. (1988). Sofista. En Platón, *Diálogos V* (pp. 319-482). N. L. Cordero (trad.). Gredos.
- Risser, J. (2012). *The Life of Understanding: A Contemporary Hermeneutics*. Indiana University Press. <https://doi.org/10.2979/6588.0>
- Rorty, R. (1991). Deconstruction and circumvention. En R. Rorty, *Essays on Heidegger and Others: Philosophical Papers II* (pp. 85-106). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609039.006>
- Schleiermacher, F. (1838). *Hermeneutik und Kritik*. Reimer.
- Scholz, O. (2008). Compréhension, interprétation, herméneutique. En C. Berner y D. Thouard (eds.), *Sens et interprétation* (pp. 67-80). Septentrion. <https://doi.org/10.4000/books.septentrion.75173>
- Shusterman, R. (1994). *Sous l'interprétation*. J.-P. Cometti (trad.). Éditions de l'Éclat.
- Shusterman, R. (1996). *Von der Interpretation. Sprache und Erfahrung in Hermeneutik, Dekonstruktion und Pragmatismus*. B. Reiter (trad.). Passagen.
- Shusterman, R. (2002). Pragmatism and Criticism: A Response to Three Critics of Pragmatist Aesthetics. *The Journal of Speculative Philosophy*, 16(1), 26-38. <https://doi.org/10.1353/jsp.2002.0005>

- Shusterman, R., Bohman, J. F. y Hiley, D. R. (1991). *The Interpretive Turn: Philosophy, Science, Culture*. Cornell University Press.
- Swayne, L. (2010). *Gadamer's Dialectical Hermeneutics*. Lexington Books.
- Taylor, C. (2005). Comprensión y etnocentrismo. En C. Taylor, *La libertad de los modernos* (pp. 199-222). H. Pons (trad.). Amorrotu.
- Taylor, C. (2009). *El multiculturalismo y "la política del reconocimiento"*. M. Utrilla de Neira (trad.). FCE.
- Thaning, M. (2015). *The Problem of Objectivity in Gadamer's Hermeneutics in Light of McDowell's Empiricism*. Springer. <https://doi.org/10.1007/978-3-319-18648-1>
- Valéry, P. (1973). *Cahiers I*. Gallimard.
- Wittgenstein, L. (1969). [PG]. *Philosophische Grammatik*. Blackwell.

<https://doi.org/10.21555/top.v760.3327>

Interpretation of Space and Understanding of Self: Reading Foucault and Ricoeur in One Frame

Interpretación del espacio y comprensión del yo: una lectura de Foucault y Ricoeur en un mismo marco

Vidya Mary George

Goa University

India

philosophy.vidya@unigoa.ac.in

<https://orcid.org/0000-0002-9166-8211>

Recibido: 12 - 12 - 2024.

Aceptado: 16 - 03 - 2025.

Publicado en línea: 10 - 08 - 2026.

How to cite this paper: George, V. M. (2026). Interpretation of Space and Understanding of Self: Reading Foucault and Ricoeur in One Frame. *Tópicos, Revista de Filosofía*, 76, 409-434. <https://doi.org/10.21555/top.v760.3327>



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

Abstract

This paper explores how the interpretation of space opens possibilities for self-understanding in the later works of Michel Foucault and Paul Ricoeur. First, the thinkers' projects in post-structuralist hermeneutics of the subject/self are situated in the aftermath of the "overthrow of the Cogito" by the "masters of suspicion." Next, an analysis of space as a text-analogue is passed through to understand the self's exteriorisation of meaning in built space. Here, Foucault's history of the Western concept of space is unpacked through Stephan Günzel's exposition, and Foucault and Ricoeur are discussed as thinkers on space. Finally, "refiguration of space" is identified as the phase in which possibilities for the revealing and transforming of the self are opened. The paper shows that "spatiality of the self" offers another point of intersection to further imaginary conversations between Foucault and Ricoeur in the context of their shared concern for the self.

Keywords: Michel Foucault; Paul Ricoeur; hermeneutics of the subject/self; hermeneutics of suspicion; masters of suspicion; interpretation; meaning; self-understanding; spatial turn; spatiality; built space; architecture; refiguration of space.

Resumen

Este artículo explora cómo la interpretación del espacio abre posibilidades para la autocomprensión en las obras posteriores de Michel Foucault y Paul Ricoeur. En primer lugar, se sitúan los proyectos de ambos pensadores en la hermenéutica postestructuralista del sujeto/yo, tras el "derrocamiento del *cogito*" por los "maestros de la sospecha". A continuación, se analiza el espacio como análogo textual para comprender la exteriorización del significado del yo en el espacio construido. En este marco, la historia de Foucault sobre el concepto occidental de "espacio" se discute a partir de la exposición de Stephan Günzel, y Foucault y Ricoeur son analizados como pensadores del espacio. Por último, se identifica la "refiguración del espacio" como la fase en la que se abren posibilidades para la revelación y la transformación del yo. El artículo muestra que la "espacialidad del yo" ofrece otro punto de intersección para continuar las conversaciones imaginarias entre Foucault y Ricoeur en el contexto de su preocupación compartida por el yo.

Palabras clave: Michel Foucault; Paul Ricoeur; hermenéutica del sujeto/yo; hermenéutica de la sospecha; maestros de la sospecha; interpretación; significado; autocomprensión; giro sobre el espacio; espacialidad; espacio construido; arquitectura; refiguración del espacio.

1. Introduction¹

Despite being contemporaries within the camp of continental philosophy and having had several common concerns, the late twentieth-century French thinkers Michel Foucault and Paul Ricoeur had few interactions with each other during their lifetimes—"except at very localised points of intersection," as Ricoeur observed (Ricoeur, 1998, p. 79; see also Goetz, 2004; Castonguay, 2010, p. 83, n. 1). Recent studies have attempted to construct imaginary conversations and philosophical dialogues between the two thinkers regarding their shared concern for the self (Davidson & Michel, 2010, p. 8), specifically the constitution of the self by the self, the role of understanding and interpretation in the constitution of the subject, the conception of freedom as practices of the self, and the status of ethics (see Barthélémy, 2010; Castonguay, 2010). The search for self-understanding, imbued with an ethical concern for one's relations with oneself and others in institutions, has been identified as one of the common guiding threads that run through their works (Goetz, 2004), situating them within "a philosophical lineage that can be traced back to the Socratic injunction to 'know oneself,'" framed by the "care of the self" (Michel, 2015, p. 103; see Davidson, 2016, p. 159).

This paper attempts to show that "spatiality of the self" offers yet another point of intersection to further the imaginary conversations between Foucault and Ricoeur in the context of their shared concern for the self. First, the concern that Foucault and Ricoeur share for the self in their later works is identified as projects in post-structuralist hermeneutics of the subject/self, situated in the aftermath of the "overthrow of the Cogito" (Ricoeur, 1992, p. 5) by those whom Ricoeur designated as the

¹ This paper was written as part of my doctoral research in Philosophy at the School of Sanskrit, Philosophy and Indic Studies (SSPIS), Goa University, supported by the UGC Junior Research Fellowship (Government of India). I am indebted to Prof. Dr. Koshy Tharakan, my doctoral advisor (SSPIS, Goa University), for his thoughtful insights and mentorship in shaping my work. I am grateful to the anonymous referees for their generous engagement with the paper. My deep gratitude to Dr. Sophia Stone (Wisdom's Edge Foundation, Inc.; Lynn University, USA), Divya Mary George, Linta Mary Philip, Fr. Renny Markose (India), and Vidyasagar Sharma (Bielefeld University, Germany) for their steady support, guidance, and goodwill throughout this research journey.

“masters of suspicion”: Marx, Nietzsche, and Freud (Ricoeur, 1970, p. 33). Next, an analysis of space as a text-analogue is passed through to explore how the interpretation of space and self-understanding hang together, and to understand the self’s exteriorisation of meaning in built space. Here, Foucault’s history of the concept of space in the West is unpacked in light of Stephan Günzel’s exposition, and Foucault and Ricoeur are discussed as thinkers on space. Finally, “refiguration of space” is identified as the phase in which possibilities for the revealing and transforming of the self are opened.

2. A shared concern for the self: projects in post-structuralist hermeneutics

Foucault’s and Ricoeur’s concern for the self became more pronounced in their later works, which can be considered projects in post-structuralist hermeneutics of the subject/self. The highly eclectic concept of “post-structuralism” is difficult to define due to the ambiguities and disagreements surrounding it. Following Gary Gutting (1998a & 1998b), this paper uses the concept in the sense of philosophical reflections on or reactions to structuralism rather than as a simple rejection or extension of it. The implicit theoretical basis of structuralism was “a kind of Cartesianism without the subject”: it treated “distinctively human domains as formal structures in which meanings were constituted not by conscious subjects but by relations among the elements of a formal system” (Gutting, 1998b). Post-structuralist critiques of structuralism retain its elimination of the foundational role of the subject but deny the internal coherence of formal systems based on two major theses: (1) no system is autonomous in the way structuralism requires, and (2) dichotomies underlying structuralism do not hold up under scrutiny (Gutting, 1998a & 1998b). The retention of the term “post-structuralism” instead of “postmodernism,” which the current academic literature more typically invokes, seems more apt here to delineate Foucault’s and Ricoeur’s responses to the idea that “the subject [is] not the master of meaning” (Taylor, 2015, p. xii) and to suggest their engagement with tradition, rather than a “radical rupture” with it that the latter term implies (Taylor, 2015, p. xv). Meanwhile, Keane and Lawn (2016) observe that contemporary Western hermeneutics is “almost exclusively bound up with the activity of interpretation” (Keane & Lawn, 2016, p. 1),

compared to its more polysemic past. It involves the interplay of three interconnected yet distinct aspects:

- (1) [...] the concrete process of understanding as the task of explicating and assessing our interpretation of texts; (2) [...] the historical centrality of establishing a rule for the proper use of interpretation [...] ; and (3) [...] the philosophical theory and method with which we can fix or ascertain the nature, character, conditions, and limits of every possible act of understanding [...]
- (Keane & Lawn, 2016, p. 1).

Foucault is widely recognised as a post-structuralist, though not as often as a hermeneutician. Dreyfus and Rabinow (1983, p. 124) argue that Foucault employs an “interpretive analytic of our current situation” but not a hermeneutic method. However, Foucault is readily included as a “major figure” in a few key references on hermeneutics (for instance, see Foucault, 1990; Keane & Lawn, 2016), especially for his contributions to and expansion of the hermeneutics of suspicion by way of analysing “cultural meaning as a site in which competing fields of power configure and confront one another” (Davey, 2008, pp. 695-696). On the other hand, Ricoeur is widely recognised as a seminal figure in contemporary hermeneutics, though not as often as a post-structuralist. However, many recent studies, notably that by Johann Michel (2015), characterise Ricoeur as a post-structuralist, especially for his critique of the foundational subject, his attempt to surpass structuralism, and rereading Marx, Nietzsche, and Freud as hermeneuticians of suspicion (Michel, 2015, p. xxiv).²

Despite and due to their difficulties of “fully belonging” to both theoretical branches, this paper presupposes the later works of Foucault and Ricoeur as contributing “up to a point” to a “specific type” of post-structuralist hermeneutics in their distinct ways, in a manner similar to that proposed by Michel (2015, p. xxvi). Ricoeur (1998, p. 79) had observed that he and Foucault were on paths where they encountered each other infrequently. Even so, as Hans Joas (2019) notes, Michel (2015 & 2019) shows how the two thinkers “learned from each other and modified their respective philosophies accordingly” (Joas, 2019, p.

² See Michel (2015) for a list of studies that classify Ricoeur as a post-structuralist as well as those that discuss the difficulties with such a classification.

xiv), furthering the possibilities of reading them in one frame. As Joas summarises Michel's argument (and responds to Dreyfus and Rabinow):

Ricoeur integrated more and more aspects of a Nietzschean hermeneutics of suspicion into his approach—a process that had already begun with his interest in Freud and Marx or Karl Mannheim. Foucault, who had been labelled—by Hubert Dreyfus and Paul Rabinow—to be “interpretive but not hermeneutic,” developed in his later work a study of the genealogy of self-examination that can be called a “hermeneutics of the self.” [...] Both provide, as Johann Michel writes, an anthropological foundation for the “forms of auto-interpretation when a human agent is confronted with the problematicity regarding himself, when the meaning of his existence no longer makes sense” (Joas, 2019, p. xiv).

3. The quest for meaning: towards a hermeneutics of the subject/self

The self for which Foucault and Ricoeur share a concern in their later works emerges in the aftermath of the critique of the philosophies of the subject by Marx, Nietzsche, and Freud—whom both thinkers jointly treated for their contributions to the “hermeneutics of suspicion,” and whom Ricoeur famously designated as “masters of suspicion” (Ricoeur, 1970, p. 33; Foucault, 1990).³ The Cogito posited by Descartes in his search for a certain and indubitable foundation of all knowledge and meaning was a subject that answered to the question of what I am rather than who I am. The “exaltation of the Cogito” came at the price of its being reduced to a thinking thing bereft of understanding or agency (Ricoeur, 1992, pp. 5-11). When the masters of suspicion took up the problem of the Cartesian doubt in the nineteenth century, it was “to look upon the whole of consciousness primarily as ‘false’ consciousness.”

³ The primary references in this paper for the joint treatment of Marx, Nietzsche, and Freud are the published English translations of Foucault's “Nietzsche, Freud, Marx” (Foucault, 1990), presented and originally published in 1964, and Ricoeur's “Interpretation as Exercise of Suspicion” (Ricoeur, 1970), delivered in 1961-1962 and originally published in 1965.

They exercised suspicion—which until then had been directed only towards language and things—to consider consciousness itself as being something other than what it appears to itself (Ricoeur, 1970, p. 33; 1989, p. 148; Foucault, 1990, p. 60). They entered the problem of false consciousness to “make their ‘conscious’ methods of deciphering coincide with the ‘unconscious’ work of ciphering which they attributed to the will to power, to social being, to the unconscious psychism.” Nevertheless, all methods were united in their common intention: to reduce the illusions and lies of consciousness (Ricoeur, 1970, pp. 32-34).

When the interpreters began interpreting themselves per the techniques of interpretation, the heritage of the philosophies of the subject, “those from Descartes to Kant, then from Kant to Fichte, and finally to the Husserl of the *Cartesian Meditations*” (Ricoeur, 1992, p. 5), underwent a tremendous shift, inflicting “inexhaustible narcissistic wounds” on Western culture (Foucault, 1990, p. 61). It became uncertain if meaning and the consciousness of meaning coincided in consciousness at all. The new hermeneutics of suspicion posed interpretation as an exercise of suspicion against interpretation as a recollection of meaning and the reminiscence of being (Ricoeur, 1970, p. 35). It was no longer sufficient to seek meaning through recollection and verbalisation of the immediate consciousness of meaning. Rather, meaning necessitated mediation—an art of interpretation, a deciphering of its expressions (Ricoeur, 1970, p. 33; 1989, p. 149).

Ricoeur asserts that the contribution of the masters of suspicion to the heritage of the philosophies of the subject was not simply an exercise in scepticism, as is commonly misunderstood and dismissed. Their mastery lies in not only suspecting and demystifying consciousness but also extending it rather than detracting it (Ricoeur, 1970, p. 35). Indeed, the role of suspicion, as Taylor (2015) notes, the role of suspicion was to “open an alternative space for human vitality to flourish” (Taylor, 2015, p. xiv). As Foucault argues, the overthrow of the Cartesian subject “does not create a deficiency; it does not constitute a lacuna that must be filled. It is nothing more, and nothing less, than the unfolding of a space in which it is once more possible to think” (Foucault, 1970, p. 373). In the absence of the imposition of the Cartesian subject and the emptiness of reflexive self-certitude (Davidson, 2016, p. 160), many contemporary thinkers have superseded the philosophies of the subject. By being attentive to and assimilating the often-overlooked positive possibilities of the masters’ critiques (Ricoeur, 1970, p. 32), they have

reconceptualised the subject/self within newly unfolding interpretive spaces. And yet, to those who still do not consider the Cogito to have been overthrown, Foucault responds with a silent, philosophical laugh:

To all those who still wish to talk about man, about his reign or his liberation, to all those who still ask themselves questions about what man is in his essence, to all those who wish to take him as their starting-point in their attempts to reach the truth, to all those who, on the other hand, refer all knowledge back to the truths of man himself, to all those who refuse to formalise without anthropologising, who refuse to mythologise without demystifying, who refuse to think without immediately thinking that it is man who is thinking, to all these warped and twisted forms of reflection we can answer only with a philosophical laugh—which means, to a certain extent, a silent one (Foucault, 1970, p. 373).

For Foucault, the new relation between meaning and the consciousness of meaning that the hermeneutics of suspicion brought forth implied that the interpretation of consciousness became an essentially incomplete task. There is nothing to interpret, only interpretations of interpretations. Therefore, interpretation cannot but always correct itself. To impose one interpretation over others is to give up on the infinity of interpretations and to force the death of interpretation (Foucault, 1990, pp. 63-66). The rejection of *a priori* theories of the subject is what Foucault celebrates when he declares man to have disappeared and died (Foucault, 1970, p. 373), a statement often misunderstood and criticised as having given up on conceptualisations of the subject/self altogether (Allen, 2008, p. 60). On the contrary, he asserts that the new hermeneutic question turns to the “Who?”—the subject that poses the interpretation (Foucault, 1990, p. 66), a form, not a substance, which is “not primarily or always identical to itself” (Foucault, 1997, p. 290). The active constitution of the subject by itself in various forms through the practices of the self became the object of analysis in Foucault’s later investigations into the ethics of the care and concern of the self and others as a practice of freedom (Foucault, 1997, pp. 281-290).

For Ricoeur, to enforce a final interpretation or a last word on the conflict of interpretations, or the plurivocity of texts and actions, is to do violence (Ricoeur, 1981, p. 177). He, too, declares the Cogito to have

been shattered and his question of selfhood to have parted ways with the philosophies of the subject (Ricoeur, 1992, p. 4). He then turned his attention toward a hermeneutics of the self considered as a “triple dialectic of reflection and analysis, of selfhood and sameness, and of self and other” (Ricoeur, 1992, p. 21), which he placed “at an equal distance from the apology of the cogito and from its overthrow” (Ricoeur, 1992, p. 3). Ricoeurian hermeneutics is a “philosophy of detours”: “The recourse to analysis [...] is the price to pay for a hermeneutics characterised by the indirect manner of positing the self” (Ricoeur, 1992, p. 17). Ricoeur’s claim here, as Taylor explains, is that “meaning is understood not on the basis of intuition but in meaning’s externalisation in objects” (Taylor, 2015, p. xii). The self implied reflexively in the long route of analysis is rich, complex, and multidimensional, one that answers to the question “Who?”: “Who is speaking? Who is acting? Who is recounting about himself or herself? Who is the moral subject of imputation?” (Ricoeur, 1992, p. 16). The intimate degree to which the other is constitutive of selfhood (Ricoeur, 1992, p. 3) is such that “the autonomy of the self will appear then to be tightly bound up with solicitude for one’s neighbour and with justice for each individual” (Ricoeur, 1992, p. 18), thus binding his hermeneutics of the self with the practical wisdom of his “little ethics” in the quest for the “ethical intention,” i.e., the “aiming at the good life with and for others, in just institutions” (Ricoeur, 1992, p. 172).

4. Meaning’s exteriorisation in built space: an analysis of space

In Ricoeurian hermeneutics—the philosophy of detours—the paradigm for interpretation is provided by the model of the text. Understanding the text, however, is not an end in itself: “It mediates the relation to himself of a subject who, in the short circuit of immediate reflection, does not find the meaning of his own life. [...] The constitution of the self is contemporaneous with the constitution of meaning” (Ricoeur, 1981, p. 120). Self-understanding requires passing through “the detour of understanding the cultural signs in which the self documents and forms itself” (Ricoeur, 1981, p. 120), which include “those which are fixed by writing, including all the sorts of documents and monuments which entail a fixation similar to writing” (Ricoeur, 1981, p. 159). That is, Ricoeur construes built space as a text-analogue: he draws a direct and open analogy between architecture as inscription in stone or any hard material, and narrativity as inscription in language (Ricoeur, 2016, p. 32). He thus opens possibilities for the application and

expansion of hermeneutics to architecture and, more broadly, to space. In what follows, a detour of reflection by way of an analysis of space is passed through to explore how the interpretation of space and self-understanding hang together, and to understand self's exteriorisation of meaning in built space.

Space, for Foucault, is a historical entity subject to time and power. The history of space is a history of powers (Foucault, 1980b, p. 149). In his oft-cited article "Of Other Spaces," he traces a brief and rough history of the concept of space in the West: the contemporary space of "emplacement" has substituted the modern space of "extension," which itself had replaced the medieval space of "localisation" (Foucault, 2008, p. 14). Foucault's account has been criticised as being hasty and simplistic (for instance, see esp. Casey, 1997, p. 298); however, unpacking it in light of a more recent exposition of the conceptual history of space, such as that offered by Stephan Günzel (2009 & 2012), demonstrates its continuing relevance.

Günzel (2009 & 2012) identifies three epochs in the general history of the concept of space: the Aristotelian period, the Newtonian period, and the contemporary period wherein Newton's spatial conception was rethought, especially by Einstein. In each epoch, he notes, "the physical concept of space rested upon geometrical as well as cosmological concepts of spatiality, which furthermore was set in a cultural context by which certain understandings of space were encouraged and others ignored or suppressed" (Günzel, 2009, p. 1173).

During the Aristotelian period, the concept of space as *topos* or "place" was dominant. As Günzel (2009) notes, Aristotle considered place as "the surrounding border of any material object" (Günzel, 2009, p. 1173), tying spatiality to the locality and substance of an object. Space as place is non-geometrical and has no extension, is invariable and never void. Things can change their places, but place itself does not move. The cosmological background of Aristotelian physics—a finite universe with Earth as the centre and a first mover located at the outermost sphere who caused any physical process in the world—prevailed and persisted into the Middle Ages (Günzel, 2009, pp. 1173-1174). The concept of space that presupposes the cosmology Günzel outlines aligns with what Foucault calls the medieval space of "localisation," wherein space was seen as hierarchic and oppositional:

In the Middle Ages it was a hierarchic ensemble of places: sacred places and profane places, protected places and open spaces without defence, urban places and rural places (so far for the real life of humans). For cosmological theory, there were the super-celestial places as opposed to the celestial, and the celestial place was in turn opposed to the terrestrial place. There were the places where things found themselves placed because they had been violently displaced, and then, on the contrary, the places where things found their emplacement and natural rest. It was this complete hierarchy, this opposition, this intersection of places that constituted what could very roughly be called medieval space: the space of localisation (Foucault, 2008, p. 15).

Later, Aristotle's substantialism was strongly backed by Descartes, with his concept of physical space as *plenum*—"continuous and of a material character" (Günzel, 2009, p. 1175).

During the Newtonian period, a new concept of physical space as volume—"the (geometrical) product of a three-dimensional extension" (Günzel, 2009, p. 1175)—was established. Newtonian absolute space is the homogeneous, indivisible totality in which any object is located. Against the Aristotelian-Cartesian conceptions of space it replaced, absolute space was no longer tied to locality or substance and could be void. In the cosmological framework of Newtonian physics—an infinite universe with no centre—there was no place for a divine first mover, as matter (as mass) itself was considered the cause of movement (Günzel, 2009, pp. 1174-1175). The conception of space Günzel discusses aligns with what Foucault calls the space of "extension," wherein space was seen as "infinite and infinitely open" ever since Galileo, reducing medieval place to a point in the movement of a thing:

This space of localisation opened up with Galileo, for the real scandal of Galileo's work is not so much his discovery, or rediscovery, that the earth revolved around the sun, but his constitution of an infinite and infinitely open space. In such a space the place of the Middle Ages found itself dissolved as it were; the place of a thing was nothing but a point in its movement, just

as the rest of a thing was only its movement indefinitely slowed down. In other words, starting with Galileo, starting with the seventeenth century, extension supplanted localisation (Foucault, 2008, pp. 15-16).

Newton's spatial conception was contested and rethought in several ways. Leibniz refuted his concept of absolute space, conceiving space instead as the relations between *topoi*, the construction of which is even prior to the perception of objects.⁴ On the other hand, Kant considered geometric space as three-dimensional extension to be the very condition of perception and hence prior to and independent of relations, particular situations, and individual perceptions. In the early twentieth century, Einstein rethought Newtonian absolute space—which he called the container model of space—as relative. In his relativistic space-time theory, space is interweaved with time and no longer seen as infinite, as it has expanded since the Big Bang (Günzel, 2009, pp. 1175-1176).

In response to physical and mathematical concepts of space, phenomenologists argued for a description of spatiality from within lived bodily experience, against explanations of spatial structures according to the mechanics of a body's functions. In this "counterproposal," spatial dimensions arose from a "point zero" inherent in the subject (Günzel, 2012, p. 312), thereby incorporating the lived body into discussions on space (Tharakan & George, 2025, p. 147). Foucault considered phenomenological descriptions of space as primarily concerned with "inner space" and "fundamental for reflection in our time" (Foucault, 2008, p. 16). Reflecting on Bachelard's description of space as heterogeneous, he writes:

Bachelard's monumental work, the descriptions of phenomenologists have taught us that we do not live in a homogeneous and empty space, but on the contrary in a space thoroughly loaded with qualities and perhaps also haunted by fantasy. The space of our primary perception, the space of our dreams and that of

⁴ Günzel (2012) notes that Leibniz can be considered the forerunner of "topology"—a "non-materialistic and even non-extensional conception of space as the structure of relations (of spatial objects)" (Günzel, 2012, p. 307). This abstraction in geometry is said to have urged the contemporary understanding of space in sociocultural discourse (Günzel, 2012, p. 307).

our passions hold within themselves qualities that seem intrinsic: it is a light, ethereal, transparent space, or again it is a dark, rough, encumbered space; it is a space from above, a space of summits, or on the contrary a space from below, a space of mud; it is a space that can be flowing like lively water, or it is a space that is fixed, solidified, like stone or like crystal (Foucault, 2008, p. 16).

Phenomenological meditations on spatiality, such as those by Husserl, anticipated and urged a transition from the modern to the contemporary understanding of space (Tharakan & George, 2025, p. 147).

Meanwhile, by the twentieth century, the conception of space in sociocultural discourse had moved away from “a materialistic and deterministic view of (physical) space to an immaterial and possibilistic view of (cultural) space.” Cultural space is less concerned with the origin, essence, or ‘true’ concept of space than with spatial practices and the production of spatiality (Günzel, 2012, p. 307). In the contemporary period, space is understood as emergent rather than existing a priori, and composed of relations rather than structures (Warf, 2009, p. 75; Bachmann-Medick, 2016, p. 216), and is thus able to better account for the dynamicity and heterogeneity of the processes that shape space today.⁵ Contemporary society has the “network” as its organising principle, wherein distance is defined “functionally rather than in terms of physical length” (Warf, 2009, p. 70). The notion of space as relational and heterogeneous in the network era aligns with what Foucault calls the space of “emplacements” in the present epoch of space:

The present epoch would perhaps rather be the epoch of space. We are in the epoch of simultaneity; we are in the epoch of juxtaposition, the epoch of the near and the far, of the side-by-side, of the dispersed. We are at a moment, I believe, when our experience of the world is less that of a great life developing through time than

⁵ A few recent contributions to the relational understanding of space view space as the relations between things and actors that have to be constantly rebuilt by the latter (Peters & Kessl, 2009, pp. 25-26), even at the risk of overemphasising their creative possibilities (Schroer, 2019, p. 9).

that of a network that connects points and intersects with its own skein. [...] We are in an epoch in which space is given to us in the form of relations between emplacements. [...] We do not live in a kind of void, inside which we could place individuals and things. [...] We live inside a set of relations that delineate emplacements that cannot be equated or in any way superimposed (Foucault, 2008, pp. 14-16).

The understanding of space as relational and heterogeneous has been central to the spatial turn in social theory in the late twentieth century. With a strong theoretical critique provided by post-structuralism, this ongoing movement acknowledges and responds to “a longstanding ontological and epistemological bias that privileges time over space in all the human sciences, including spatial disciplines like geography and architecture” (Soja, 2008, p. 11). It therefore calls for a “reinsertion” or “reassertion” of space in the humanities and social sciences (Peters & Kessl, 2009, p. 20; Warf & Arias, 2009, p. 1). In the wake of its “long, slow, and painful undertaking,” it has come to be reasserted that space matters not only because all our thoughts and actions necessarily occur in space, but also because “where events unfold is integral to how they take shape” (Warf & Arias, 2009, pp. 3-10; see Vendra & Furia, 2021, p. 1). Today, space unfolds as “suspended between matter and meaning” (Warf & Arias, 2009, p. 10), a result of “social and political action that has its material correspondence in architecture [and] structures” (Bachmann-Medick, 2016, p. 229). The contemporary conception of spatiality helps us grasp “how space and social relations are made through each other” (Hubbard & Kitchin, 2011, p. 499). The spatial turn has also driven profound shifts in the understanding of identity and subjectivity. The fixed, unified, autonomous, and transcendental identity at the core of the Cogito has mostly made way to account for heterogeneous, shifting, and even contradictory aspects of the nature and constitution of selves “who lie at the changing intersections of different language games” and axes of identity (Warf & Arias, 2009, p. 6). The following two sections explore how Foucault and Ricoeur held space for the spatiality of the self in their distinct ways, for thinking spatially about the self, specifically in relation to built space or architecture, when the self is confronted with the problematicity regarding oneself.

5. Foucault and the spatiality of the self

Foucault is widely recognised as a “key thinker on space” (Hubbard & Kitchin, 2011), whose discussions on the significance of space for the human condition are credited for renewing a cultural approach to space (Günzel, 2012, p. 313) and leading to the spatial turn in the humanities and social sciences (Schroer, 2019, pp. 4-5). He recognised the “devaluation of space that has prevailed for generations” (Foucault, 1980a, p. 70), identifying the “discourse of philosophers” as one of the reasons for it (Foucault, 1980b, p. 149):

At the moment when a considered politics of spaces was starting to develop, at the end of the eighteenth century, the new achievements in theoretical and experimental physics dislodged philosophy from its ancient right to speak of the world, the cosmos, finite or infinite space. This double investment of space by political technology and scientific practice reduced philosophy to the field of a problematic of time. [...] Along with this goes a correlative devaluation of space [...] (Foucault, 1980b, pp. 149-150).

He observes that the longstanding neglect of space meant that space was treated as “the dead, the fixed, the undialectical, the immobile,” compared to time, which was seen as “richness, fecundity, life, dialectic” (Foucault, 1980a, p. 70). It also meant a delay in the emergence of space as a “historico-political problem” (Foucault, 1980b, p. 149): “Space used to be either dismissed as belonging to ‘nature’ — that is, the given, the basic conditions, ‘physical geography,’ in other words a sort of ‘prehistoric’ stratum; or else it was conceived as the residential site or field of expansion of peoples, of a culture, a language or a State” (Foucault, 1980b, p. 149). Foucault acknowledged the intersection of time and space, emphasising that “a whole history remains to be written of spaces — which would at the same time be the history of powers” (Foucault, 1980b, p. 149). He also observed that “the anxiety of today fundamentally concerns space, no doubt much more than time. Time probably appears only as one of the various possible operations of distribution between the elements that are spread out in space” (Foucault, 2008, p. 15). Although only a few of Foucault’s works discuss space explicitly — the most acclaimed

ones being the interviews “Questions on Geography” (Foucault, 1980a), “The Eye of Power” (Foucault, 1980b), and “Space, Knowledge, Power” (Foucault, 2000), and the article “Of Other Spaces” (Foucault, 2008)—he admitted to his spatial “obsessions” and preoccupations, conceding that “geography must indeed necessarily lie at the heart of [his] concerns” (Foucault, 1980a, p. 77). Indeed, most of his works are imbued with the theme of space, and not only use spatial, strategic metaphors but also study spaces as objects (Foucault, 1980a, p. 70; 2000, p. 362).

In Foucault’s works, space acts as the support, or the condition of possibility, for relating knowledge and power. In comparison to the “inner space” of phenomenologists, it was “outer space,” or sociocultural space, that he was attentive to (Foucault, 2008, p. 16). In his analysis of the genealogy of knowledge and the formation of discourses, he was concerned with the “politics of knowledge,” i.e., the tactics and strategies of power deployed through channels of geopolitics, by which “knowledge functions as a form of power and disseminates the effects of power” (Foucault, 1980a, pp. 69-77). He considered space as fundamental to any form of communal life and any exercise of power, and hence as having an effect on the constitution of personal identity (Foucault, 2000, p. 361). For Foucault, the individual “is not a pre-given entity which is seized on by the exercise of power. The individual, with his identity and characteristics, is the product of a relation of power exercised over bodies, multiplicities, movements, desires, forces” (Foucault, 1980a, pp. 73-74).

Foucault analyses architecture specifically for the “specific effects” it brings about in the field of social relations, rather than merely as an “element in space” (Foucault, 2000, p. 361). He admits that his analyses of architecture have been “vague,” and that, in them, architecture is “only taken as an element of support, to ensure a certain allocation of people in space, a canalisation of their circulation, as well as the coding of their reciprocal relations” (Foucault, 2000, p. 361). In other words, he was interested in architecture’s “interconnection” with social relations, and not “the primacy of this over that” (Foucault, 2000, p. 362). He observed that, if separated from one another, it would become impossible to understand “the effective practice of freedom by people, the practice of social relations, and the spatial distributions in which they find themselves” (Foucault, 2000, p. 356). For Foucault, liberty and resistance are relations realised and exercised in the actual practices of people, not that which can be given or possessed. Therefore, there is

nothing inherent or fundamental in the order of spaces as such—their structure, function, or the intention of the architect—that can either thwart or guarantee liberty to its inhabitants. Even in the most terrifying of systems, such as concentration camps, “aside from torture and execution which preclude any resistance [...] there always remain the possibilities of resistance, disobedience, and oppositional groupings” (Foucault, 2000, p. 354). Similarly, institutions and laws intended to guarantee liberty cannot automatically assure it merely by their establishment, and even retain the possibility of being turned around. Still, the exercise of freedom is not “completely indifferent to spatial distribution” (Foucault, 2000, p. 356)—his position is not that “one may as well leave people in slums, thinking that they can simply exercise their rights there” (Foucault, 2000, p. 355). Architecture can and does produce “positive effects when the liberating intentions of the architect coincide with the real practice of people in the exercise of their freedom” (Foucault, 2000, p. 355). For Foucault, an architect is not one among the people “through whom power passed or who are important in the fields of power relations” (Foucault, 2000, p. 356). However, the architect and her projects need to be taken into account so as to understand “a certain number of the techniques of power that are invested in architecture” (Foucault, 2000, pp. 354–357).

6. Ricoeur and the spatiality of the self

Ricoeur’s discussions on space are severely limited compared to those on time, so much so that recent scholarship observes that “his work could be (and in fact has been) considered a perfect example of the temporal bias” in philosophy and the humanities (Vendra & Furia, 2021, p. 2). Almost every aspect of Ricoeur’s philosophical thought, especially that concerning the self, “revolves around time rather than space” (Furia, 2022, p. 193), “even in contexts where spatial language might be more appropriate than temporal” (Gschwandtner, 2017, pp. 169-170). Ricoeur directly, though briefly, analyses space in the context of “the meaning and the task of human being” in only a handful of works—including the section “Inhabited Space” in *Memory, History, Forgetting* (2004) and the article “Architecture and Narrativity” (2016). Consequently, Ricoeur was rarely addressed as a thinker on space. However, recent scholarship also demonstrates that “the concept of space is profoundly intertwined with the recurring topics of his thought” (Vendra & Furia, 2021, pp. 2-3) and that “he offers decisive suggestions for how to conceive of

a hermeneutic account of human space” (Umbelino, 2017, p. 234), opening possibilities to envision a spatial turn for the philosopher of the temporal self.⁶ Ricoeur’s insights on space as a “polysemic concept” (Vendra & Furia, 2021, p. 1), though not explicitly developed by himself, are being recognised as having crucial implications for furthering our understanding of the spatiality of built space and space-language and self-environment relationships (Vendra & Furia, 2021, pp. 1-4; Furia, 2022). The growing interest in a “Ricoeurian project of a hermeneutics of space” (Umbelino, 2015, p. 199) is fragmentary so far, promising room for elaborations on the multiple dimensions along which he pursues the problem of space (Vendra & Furia, 2021, p. 3).

Ricoeur traces the discourse of space as a “dialectic of lived space, geometrical space, and inhabited space” (Ricoeur, 2004, p. 153). Lived space includes the living body surrounded by “places” of life (Ricoeur, 2016, p. 32). The corporeal space is “immediately linked with the surrounding space of the environment” in privately shared memory, from which “we pass by degrees to collective memory and its commemorations linked to places consecrated by tradition” (Ricoeur, 2004, p. 149)—its named, public spaces that invite ritualised gatherings (Ricoeur, 2004, pp. 149-151). The phenomenology of places is such that “beings of flesh and blood occupy, leave, lose, [and] rediscover” them (Ricoeur, 2004, p. 153):

My place is there where my body is. But placing and displacing oneself are primordial activities that make place something to be sought out. And it would be frightening not ever to find it. We ourselves would be devastated. The feeling of uneasiness—*Unheimlichkeit*—joined to the feeling of not being in one’s place, of not feeling at home, haunts us and this would be the realm of emptiness. But there is a question of place because space is not yet filled, not saturated. In truth, it is always possible, often urgent, to displace oneself, with

⁶ For instance, there is no mention of Ricoeur as a thinker on space in canonical texts such as *Key Thinkers on Space and Place* (Hubbard & Kitchin, 2011). However, the 2021 issue of the distinguished journal *Études Ricoeuriennes/Ricoeur Studies*—12(2)—is entirely dedicated to exploring “Ricoeur and the Problem of Space” (for a list of works that explore Ricoeur as a thinker on space, see Vendra & Furia, 2021, in the same issue).

the risk of becoming that passerby, that wanderer, that flâneur, that vagabond, stray dog that our fragmented contemporary culture both sets in motion and paralyses (Ricoeur, 2004, p. 149).

The placements and displacements of the body cannot be spoken of, thought, or experienced without some reference to “points, lines, surfaces, volumes, distances” inscribed on a geometrical space detached from it. Unlike the lived space of the lived body—the absolute here and “the landmark for any there”—three-dimensional geometrical space has no “privileged” places or “ethical overtones”; rather, it is measurable according to Cartesian coordinates, and each point in it can be “deduced from other points” (Ricoeur, 2004, pp. 149-153; see Ricoeur, 2016, p. 32). At the boundaries of lived space and geometrical space lies inhabited space or “site.” This “third space” or “mixed space” results from the superimposition of places on geometric space solely through an “act of construction” or architecture. Therefore, constructed space consists of “a system of sites for the major interactions of life”—be it “fixed space or space for dwelling” or “space to be traversed”—that are also “geometrical, measurable, and calculable” (Ricoeur, 2004, pp. 149-150; see Ricoeur, 2016, pp. 31-33).⁷

Ricoeur analyses architecture as a parallel to narrativity—thereby also investigating the entanglements of spatiality and temporality—by way of trying “to transpose to the architectural plane the categories linked to threefold mimesis”: prefiguration, configuration, and refiguration (Ricoeur, 2004, p. 527, n. 2). Rather than reproducing reality, Ricoeurian mimesis functions “in a dialectical manner, as being first imitation, next reconstruction, and, finally, the transformative capacity of experience” (Ricoeur, 1998, p. 83). The act of inhabiting is the “prefiguration of space” or the presupposition of the architectural act, “inasmuch as the need for shelter and circulation lays out the interior space of a dwelling and the given intervals to traverse” (Ricoeur, 2004, p. 527, n. 2; see Ricoeur, 2016, p. 33). The act of construction is the process of “configuration of space,” and it works simultaneously with the act of inhabiting, shaping one another at the same time:

⁷ A similarity, up to a point, can be observed between Ricoeur’s notions of places, geometrical space, and sites, and Foucault’s notions of spaces of localisation, extension, and emplacement, respectively.

The functions of inhabiting are continually ‘invented,’ in both senses of the word (to discover and to create), at the same time that the processes of construction are inscribed in the design of the architectural space. We might say that the act of inhabiting, and the building resulting from the construction, shape one another at the same time (Ricoeur, 2016, p. 36).

In the act of construction, inhabiting is redesigned so that “the last word is given to a thoughtful inhabiting.” The builder configures space according to three “rules of rationality”: (1) emplotment—the “spatial synthesis of the heterogeneous” or “the intention of internal coherence” between parts of a building; (2) intelligibility—the intention to clarify and make comprehensible the inextricable through inscription in the “enduringness of materials”; and (3) intertextuality—the intention to contextualise or confront the new building next to, against, and after the “network of already-there buildings” and ways of inhabiting (Ricoeur, 2004, p. 150; 2016, pp. 36-39). Due to the ideological presuppositions of the builder, the architectural project is inclined to disregard or outweigh the inhabitants’ needs and expectations stemming from the act of inhabiting, and “is too often the representation of the ‘competent authorities’ regarding the inhabiting needs of [...] [the] masses” (Ricoeur, 2004, p. 150; 2016, pp. 33-39). Even so, the building does not complete its journey in its configuration by the builder—it gives itself not only to be seen but also to be read (Ricoeur, 2004, p. 151):

[...] it will not suffice for an architectural project to be well thought-out, or even for it to be held to be rational for it to be understood and accepted. All planners ought to learn that an abyss can separate the rules of the rationality of a project [...] from the rules of acceptability to a public (Ricoeur, 2016, p. 39).

The inhabitants of the built space read and reread it from the point of view of their way of inhabiting, approaching it with their needs, expectations, resistance, and challenges, which, when confronted by the space’s propositions of meanings, elicit responses ranging “from passive reception, subdued, indifferent reception, to hostile and angry reception” (Ricoeur, 2004, p. 527, n. 2; see Ricoeur, 2016, p. 39). In the inhabitants’ inhabiting as a response or even an answer to building, the

architectural project undergoes a “refiguration of space.” Receptive and active inhabiting implies plural, polemical, and intertextual readings of space—“careful rereading of the urban environment, a continuous relearning of the juxtaposition of styles, and thus also of life stories of which the monuments and all the buildings carry the trace” (Ricoeur, 2016, p. 39).

7. Conclusion: self-understanding by way of interpretation of space

In Ricoeurian hermeneutics of space, it is in the mimetic phase of refiguration that the interpretation of space opens possibilities for the understanding of the self for inhabitants through the transformative capacity of experience (Umbelino, 2015, pp. 205-206; Ricoeur, 2016, p. 39). Ricoeur considers the model of reading excellent for reevaluating the act of inhabiting. In the act of reading a space, the space reveals and transforms the inhabitant while undergoing a transformation of its own, as a text does:

The text unfolds its capacity to illuminate or clarify the life of the reader; it has both the power of discovering, of revealing the hidden, the unsaid of a life shielded from Socratic scrutiny, and that of transforming the banal interpretation that the reader makes according to the bent of day-to-day life. To reveal [...], but also to transform, that is what carries the text outside itself (Ricoeur, 2016, p. 39).

The dialectic of revealing and transforming is such that the interpretation of space could culminate in the self-interpretation of the inhabitant. Then, the inhabitant “understands himself better, understands himself differently, or simply begins to understand himself” (Ricoeur, 1981, p. 120), constituting the self as meaning is constituted. An interplay of revealing and transforming is also consequential in Foucault’s analyses of the subject: revealing the genealogy of the formation of the subject and the possibilities for practices of self-transformation are crucial to the Foucauldian project of the hermeneutics of the subject (Foucault, 1986, p. 64; Koopman, 2013, p. 526).

Thus, as an object in which meaning is exteriorised, built space mediates self-understanding. Even so, as Taylor (2015) notes, Ricoeur

asserts that we are not the masters of meaning's exteriorisation in objects, since the exteriorisation "may be a result of an unconscious or subconscious intention, [...] may differ from the inspiring intention, and [...] may develop autonomy" (Taylor, 2015, pp. xii-xiii)—as in the case of a building's interpretation by its inhabitants. There is, therefore, a need to emphasise the positive "creativity of the externalisation" (Taylor, 2015, pp. xii-xiii). Simultaneously, there is a need to engage with tradition to incorporate the critique of ideology, or a hermeneutics of suspicion, within the hermeneutics of space, since built space, like other external products of life, can fall under the sway of false consciousness, games of power, or ideological presuppositions (Foucault, 2000, p. 363; Davidson, 2016, p. 160).

In conclusion, this paper has shown that the "spatiality of the self" offers yet another point of intersection to further imaginary conversations and philosophical dialogues between Foucault and Ricoeur in the context of their shared concern for the self. Despite their shared concerns and similarity of themes, Foucault and Ricoeur followed distinct paths in their search for self-understanding, differing in their methodologies, and use of concepts, and ideas on thinking about those concepts (Goetz, 2004). Then, reading the two thinkers in one frame requires the reader to hold space for their works in a manner similar to Gillian Rose's paradoxical spaces—"complex and contradictory spatialities" that speak of "power, resistance, and the acknowledgement of difference" (Rose, 1993, p. 141).

Bibliography

- Allen, A. (2008). The Impurity of Practical Reason: Power and Autonomy in Foucault. In A. Allen, *The Politics of Our Selves: Power, Autonomy, and Gender in Contemporary Critical Theory* (pp. 45-71). Columbia University Press. <https://doi.org/10.7312/alle13622-004>
- Bachmann-Medick, D. (2016). Spatial Turn. In D. Bachmann-Medick, *Cultural Turns: New Orientations in the Study of Culture* (pp. 211-244) A. Blauhut (trans.). De Gruyter. <https://doi.org/10.1515/9783110402988>
- Barthélémy, A. (2010). Herméneutiques croisées: conversation imaginaire entre Ricœur et Foucault. *Études Ricoeuriennes/Ricoeur Studies*, 1(1), 55-67. <https://doi.org/10.5195/errs.2010.7>
- Casey, E. S. (1997). *The Fate of Place: A Philosophical History*. University of California Press.

- Castonguay, S. (2010). Michel Foucault et Paul Ricœur, vers un dialogue possible. *Études Ricoeuriennes/Ricoeur Studies*, 1(1), 68-86. <https://doi.org/10.5195/errs.2010.22>
- Davey, N. (2008). Twentieth Century Hermeneutics. In D. Moran (ed.), *The Routledge Companion to Twentieth Century Philosophy* (pp. 693-735). Routledge.
- Davidson, S. (2016). Intersectional Hermeneutics. In S. Davidson & M.-A. Vallée (eds.), *Hermeneutics and Phenomenology in Paul Ricoeur: Between Text and Phenomenon* (pp. 159-174). Springer International Publishing. <https://doi.org/10.1007/978-3-319-33426-4>
- Davidson, S. & Michel, J. (2010). Foreword: Hermeneutics of the Self. *Études Ricoeuriennes/Ricoeur Studies*, 1(1), 5-8. <https://doi.org/10.5195/errs.2010.46>
- Dreyfus, H. & Rabinow, P. (1983). *Michel Foucault: Beyond Structuralism and Hermeneutics*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226154534.001.0001>
- Foucault, M. (1970). Man and His Doubles. In M. Foucault, *The Order of Things: An Archaeology of the Human Sciences* (pp. 330-374). Tavistock/Routledge (trans.). Routledge. <https://doi.org/10.4324/9781315660301-9>
- Foucault, M. (1980a). Questions on Geography. In C. Gordon (ed.), *Power/Knowledge: Selected Interviews and Other Writings 1972-1977* (pp. 63-77). C. Gordon, L. M. J. Mepham & K. Soper (trans.). Pantheon Books.
- Foucault, M. (1980b). The Eye of Power. In C. Gordon (ed.), *Power/Knowledge: Selected Interviews and Other Writings 1972-1977* (pp. 146-165). C. Gordon, L. M. J. Mepham & K. Soper (trans.). Pantheon Books.
- Foucault, M. (1986). *The History of Sexuality, Volume 3: The Care of the Self*. R. Hurley (trans.). Pantheon Books.
- Foucault, M. (1990). Nietzsche, Freud, Marx. In G. L. Ormiston & A. D. Schrift (eds.), *Transforming the Hermeneutic Context: From Nietzsche to Nancy* (pp. 59-68). A. D. Schrift (trans.). State University of New York Press.
- Foucault, M. (1997). The Ethics of the Concern of the Self as a Practice of Freedom. In P. Rabinow (ed.), *Ethics: Subjectivity and Truth – The Essential Works of Foucault, 1954-1984. Volume 1* (pp. 281-302). R. Hurley et al. (trans.). The New Press.
- Foucault, M. (2000). Space, Knowledge, Power. In J. D. Faubion (ed.), *Power: The Essential Works of Foucault, 1954-1984* (pp. 349-364). R. Hurley et al. (trans.). The New Press.

- Foucault, M. (2008). Of Other Spaces (1967). In M. Dehaene & L. De Cauter (eds.), *Heterotopia and the City* (pp. 13-30). M. Dehaene & L. De Cauter (trans.). Routledge.
- Furia, P. (2022). Understanding and Explanation: Paul Ricoeur and Human Geography. *Continental Philosophy Review*, 55(2), 193-214. <https://doi.org/10.1007/s11007-021-09554-9>
- Goetz, R. (2004). Paul Ricoeur et Michel Foucault. *Le Portique*, 13-14. <https://doi.org/10.4000/leportique.639>
- Gschwandtner, C. M. (2017). Space and Narrative: Ricoeur and a Hermeneutic Reading of Place. In B. B. Janz (ed.), *Place, Space and Hermeneutics* (pp. 169-181). Springer. <https://doi.org/10.1007/978-3-319-52214-2>
- Günzel, S. (2009). Space. In H. J. Birx (ed.), *Encyclopedia of Time: Science, Philosophy, Theology, & Culture*. 3 (pp. 1173-1177). Sage.
- Günzel, S. (2012). Space and Cultural Geography. In B. Neumann & A. Nünning (eds.), *Travelling Concepts for the Study of Culture* (pp. 307-321). De Gruyter. <https://doi.org/10.1515/9783110227628.307>
- Gutting, G. (1998a). Post-structuralism. *Routledge Encyclopedia of Philosophy* (Version 1). <https://doi.org/10.4324/9780415249126-N045-1>
- Gutting, G. (1998b). Post-structuralism in the Social Sciences. *Routledge Encyclopedia of Philosophy* (Version 1). <https://doi.org/10.4324/9780415249126-R018-1>
- Hubbard, P. & Kitchin, R. (2011). *Key Thinkers on Space and Place*. Sage.
- Joas, H. (2019). Preface. In J. Michel, *Homo Interpretans* (pp. xi-xiv). Rowman & Littlefield.
- Keane, N. & Lawn, C. (2016). *The Blackwell Companion to Hermeneutics*. Wiley-Blackwell. <https://doi.org/10.1002/9781118529812>
- Koopman, C. (2013). The Formation and Self-transformation of the Subject in Foucault's Ethics. In C. Falzon, T. O'Leary & J. Sawicki (eds.), *A Companion to Foucault* (pp. 526-543). Wiley-Blackwell. <https://doi.org/10.1002/9781118324905.ch27>
- Michel, J. (2015). *Ricoeur and the Post-structuralists: Bourdieu, Derrida, Deleuze, Foucault, Castoriadis*. S. Davidson (trans.). Rowman & Littlefield International. <https://doi.org/10.5771/9781783480968>
- Michel, J. (2019). *Homo Interpretans*. D. Pellauer (trans.). Rowman & Littlefield. <https://doi.org/10.5040/9798881811679>
- Peters, M. A. & Kessl, F. (2009). Space, Time, History: The Reassertion of Space in Social Theory. *Policy Futures in Education*, 7(1), 20-30. <https://doi.org/10.2304/pfie.2009.7.1.20>

- Ricoeur, P. (1970). Interpretation as Exercise of Suspicion. In P. Ricoeur, *Freud and Philosophy: An Essay on Interpretation* (pp. 32-36). D. Savage (trans.). Yale University Press.
- Ricoeur, P. (1981). *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. J. B. Thompson (ed. & trans.). Maison des Sciences de l'Homme-Cambridge University Press.
- Ricoeur, P. (1989). Psychoanalysis and the Movement of Contemporary Culture. In D. Ihde (ed.), *The Conflict of Interpretations* (pp. 121-159). W. Domingo (trans.). The Athlone Press.
- Ricoeur, P. (1992). *Oneself as Another*. K. Blamey (trans.). The University of Chicago Press.
- Ricoeur, P. (1998). *Critique and Conviction*. K. Blamey (trans.). Columbia University Press.
- Ricoeur, P. (2004). *Memory, History, Forgetting*. K. Blamey & D. Pellauer (trans.). The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226713465.001.0001>
- Ricoeur, P. (2016). Architecture and Narrativity. E. Brennan, R. Carney & S. Lelièvre (trans.). *Études Ricoeuriennes/Ricoeur Studies*, 7(2), 3-42. <https://doi.org/10.5195/errs.2016.378>
- Rose, G. (1993). A Politics of Paradoxical Space. In G. Rose, *Feminism and Geography: The Limits of Geographical Knowledge* (pp. 137-160). Polity Press. <https://doi.org/10.2307/3060126>
- Schroer, M. (2019). Spatial Theories/Social Construction of Spaces. In A. M. Orum (ed.), *The Wiley Blackwell Encyclopedia of Urban and Regional Studies* (pp. 1-11). Wiley-Blackwell. <https://doi.org/10.1002/9781118568446>
- Soja, E. (2008). Taking Space Personally. In B. Warf & S. Arias (eds.), *Spatial Turn: Interdisciplinary Perspectives* (pp. 11-35). Routledge.
- Taylor, G. H. (2015). Preface. In J. Michel, *Ricoeur and the Post-structuralists* (pp. xi-xv). Rowman & Littlefield International.
- Tharakan, K. & George, V. M. (2025). Husserl's Crisis Text and the Spatial Turn in Philosophy of Science. *Philosophia Scientiæ*, 29(1), 137-150. <https://doi.org/10.4000/13cnk>
- Umbelino, L. A. (2015). On the Ricœurian Project of a Hermeneutics of Space. In A. Wiercinski (ed.), *Hermeneutics-Ethics-Education* (pp. 199-206). Lit Verlag.
- Umbelino, L. A. (2017). On Paul Ricoeur's Unwritten Project of an Ontology of Place. *Critical Hermeneutics*, 1(1), 233-246. <https://doi.org/10.13125/CH/3150>

- Vendra, M. C. & Furia, P. (2021). Introduction: Ricoeur and the Problem of Space. Perspectives on a Ricoeurian 'Spatial Turn'. *Études Ricoeuriennes/ Ricoeur Studies*, 12(2), 1-7. <https://doi.org/10.5195/errs.2021.576>
- Warf, B. (2009). From Surfaces to Networks. In B. Warf & S. Arias (eds.), *Spatial Turn: Interdisciplinary Perspectives* (pp. 59-76). Routledge. <https://doi.org/10.4324/9780203891308>
- Warf, B. & Arias, S. (2009). Introduction: Reinsertion of Space in the Humanities and Social Sciences. In B. Warf & S. Arias (eds.), *Spatial Turn: Interdisciplinary Perspectives* (pp. 1-10). Routledge. <https://doi.org/10.4324/9780203891308>

<https://doi.org/10.21555/top.v760.3365>

La ausencia de la necesidad en Husserl como concepto explicativo del fenómeno del valor. Una propuesta desde Carl Menger y Ludwig Feuerbach

The Absence of Need in Husserl as an Explanatory Concept of the Phenomenon of Value: A Proposal from Carl Menger and Ludwig Feuerbach

Diego Colomé

Universidad de los Andes, Chile

dcolomes@uandes.cl

<https://orcid.org/0000-0002-8025-032X>

Recibido: 18 - 01 - 2025.

Aceptado: 20 - 05 - 2025.

Publicado en línea: 10 - 08 - 2026.

Cómo citar este artículo: Colomé, D. (2026). La ausencia de la necesidad en Husserl como concepto explicativo del fenómeno del valor. Una propuesta desde Carl Menger y Ludwig Feuerbach. *Tópicos, Revista de Filosofía*, 76, 435-474. <https://doi.org/10.21555/top.v760.3365>



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

Resumen

Este artículo explora algunos de los problemas que plantea el modelo explicativo husserliano de los actos afectivos en relación con el fenómeno del valor. En la primera parte se analiza la teoría del valor de Husserl desde el punto de vista de sus actos fundantes; se pone especial atención al problema de la relación entre actos objetivantes y no-objetivantes. En esta parte se concluye que el esquema intencional con el que Husserl analiza los actos afectivos presenta problemas que se derivan de la interpretación “visual” que él hace de este. En la segunda parte se presentan las ideas de Carl Menger y Ludwig Feuerbach sobre teoría del valor y acción, con base en las cuales se propone que el elemento que falta en el esquema husserliano es el concepto de “necesidad”. Con este concepto se propone un esquema intencional ampliado de los actos afectivos con una doble intencionalidad de la conciencia dirigida tanto al estado del sujeto como al objeto.

Palabras clave: Husserl; fenomenología; teoría del valor; actos afectivos; Menger; Feuerbach; necesidad; actos no objetivantes; intencionalidad.

Abstract

This paper explores some of the problems raised by the Husserlian explanatory model of affective acts in relation to the phenomenon of value. It is divided into two main parts. In the first, Husserl's theory of value is analyzed from the perspective of its foundational acts, paying special attention to the problem of the relation between objectifying and non-objectifying acts. This section concludes that the intentional schema with which Husserl analyzes affective acts presents problems that derive from his “visual” interpretation. In the second part, the converging ideas of Carl Menger and Ludwig Feuerbach on value theory and action are presented, based on which it is proposed that the missing element in the Husserlian schema is the concept of need. With this concept, an extended intentional scheme of affective acts is developed, featuring a double intentionality of consciousness directed to both the subject and the object.

Key words: Husserl; phenomenology, value theory, affective acts, Menger, Feuerbach, necessity; non-objectifying acts; intentionality.

1. Introducción¹

El proyecto filosófico de Husserl es el de una crítica fenomenológica de la razón. La razón, por su parte, se divide en especies, cada una de las cuales responde a un tipo de acto fundamental de conciencia (cfr. Melle, 1994). Estos actos son tres: intelectivos, valorativos y volitivos. Por lo tanto, tres también son las clases de razón correspondientes: razón lógico-cognoscitiva, razón axiológica y razón práctica (cfr. Melle, 1990 y 1994; Peucker, 2008a).

En este trabajo nos concentraremos en la razón axiológica y sus actos correspondientes de valoración y sentimiento, pues en este ámbito es donde encontramos, principalmente, la teoría husserliana del valor. Husserl desarrolla su teoría del valor dentro del marco de su proyecto de fundación de una ética fenomenológica, proyecto que se encuentra recopilado principalmente en los tomos XXVIII y XXXVII de Husserliana. En Husserliana XXVIII, *Vorlesungen über Ethik und Wertlehre 1908-1914*, Husserl busca instituir una axiología y una práctica formales como nuevas disciplinas filosóficas.² Para lo cual utiliza el método de analogía con la lógica formal, método que le sirve para mostrar que, así como los principios de la lógica formal gobiernan (es decir, dan racionalidad a) el uso lógico técnico de la razón cognoscente, así también tiene que haber una disciplina análoga que gobierne mediante principios formales y universales el uso práctico y técnico de la razón axiológica y práctica. De modo, entonces, que, así como en la lógica la razón prescribe principios universales (por ejemplo, de identidad, tercero excluido, contradicción, etc.), así también, análogamente, prescribe principios racionales universales a su uso axiológico y práctico (cfr. Crespo, 2012, pp. 31 y ss.). Por tanto, así como hay una lógica formal debe haber, análogamente, una axiología y una práctica formales (cfr. Hua XXVIII: 3 ss.).

En la axiología formal Husserl busca estudiar, pues, el elemento estructural racional de los actos afectivos o *Gemütsakte*. Estos actos son los responsables de la donación de valor, y los valores y valoraciones

¹ Este trabajo ha sido posible gracias al Programa de Postdoctorado FAI de la Universidad de los Andes (Chile).

² Estas dos disciplinas conforman el estudio de la ética formal. Vale aclarar que este trabajo no trata de la ética husserliana, sino específicamente de los actos afectivos y su rol en la constitución del valor.

se dan bajo normas racionales que les imprimen un orden racional *a priori*. De manera, entonces, que la razón axiológica da el orden formal mientras que los actos afectivos dan el contenido material de los valores. Este es el marco general dentro del cual se desarrolla la axiología formal husserliana.

El desarrollo de lo que sigue está dividido principalmente en dos partes: la primera se centra en analizar el concepto husserliano de “acto afectivo” (*Gemütsakt*) y su estructura intencional. A partir de este análisis surge la primera parte de la tesis que este trabajo defiende, a saber, que la manera que tiene Husserl de conceptualizar (y, por lo tanto, comprender) el fenómeno del valor (de la emergencia y vivencia del valor) presenta problemas estructurales que se derivan de una precomprensión (o sesgo) latente en el esquema interpretativo que él implícitamente utiliza, cuya característica principal es la de concebir la intencionalidad de la conciencia principalmente como intencionalidad “visual” o vidente. La segunda parte de nuestro análisis ofrece un modelo de comprensión ampliado que toma como base el esquema que usa Husserl, pero complementado con las ideas de dos autores heterogéneos que, a primera vista, parecen no tener relación con el problema en cuestión, como son el economista austríaco Carl Menger y el filósofo alemán Ludwig Feuerbach. Con base en las ideas de estos autores, introducimos en el modelo de comprensión intencional, como elemento clave, el concepto de “necesidad”, postulando con ello la existencia, en la experiencia del valor, de una doble intencionalidad de la conciencia que se dirige al mismo tiempo tanto hacia el objeto (que es la parte a la que Husserl dedica toda su atención) como hacia el estado del sujeto (que es el complemento que agregan estos dos autores).

2. Sobre los actos afectivos y las valoraciones en Husserl

2.1. Problema de la intencionalidad de los *Gemütsakte* como actos no-objetivantes

Desde el punto de vista de Husserl, cuando valoramos objetos del mundo lo que hacen los actos afectivo-valorativos es constituir una nueva capa sobre estos objetos y, con ello, una nueva región predicativa

respecto a ellos,³ pues entonces, además de ser objetos puramente naturales, pueden también ser juzgados como poseyendo índices axiológicos positivos o negativos: buenos o malos, bellos o feos, útiles o inútiles, etc. Con esta nueva capa, los objetos se transforman en objetos valiosos que pueden, entonces, ser objetos y motivadores de la voluntad. Peucker (2008a) resume de la siguiente manera este modelo de sedimentación:

En sus primeros años o al menos hasta *Ideas I* (1913), Husserl ha descrito los actos de voluntad⁴ como fenómenos doblemente fundados. Según esto, los actos de voluntad presuponen en su constitución en primer lugar un acto representante [*einen vorstellenden Akt*], gracias al cual algo querido en general puede ser representado, y contienen, en segundo lugar, siempre una valoración gracias a la cual lo representado puede transformarse en un objetivo de la voluntad [*Willensziel*] con valor positivo. El primer elemento de la fundamentación de los actos de voluntad es, por tanto, proporcionado por actos de conciencia representante [*vorstellende Bewusstseinsakte*], mientras que el componente valorativo surge a través de rendimientos de conciencia emotivos o relativos al sentimiento [...]. Según esto, la voluntad presupone necesariamente tanto un momento representante como uno valorante (Peucker, 2008a, p. 1).⁵

³ “Los bienes, obras, objetos de uso práctico, etc., reenvían hacia actos valorantes y prácticos en los cuales las ‘meras cosas’ reciben esta nueva capa de ser [*Seinsschicht*]. Si dejamos de considerar estas capas de ser, entonces nos vemos reconducidos a la ‘naturaleza’ como el ámbito de las meras cosas [*Sachen*]” (Hua IV: 214; cfr. también Hua XLIII/2: 32 ss).

⁴ Aquí se habla de actos de voluntad, pero vale exactamente lo mismo para los actos de valoración, porque estos están supuestos en los actos de voluntad.

⁵ Peucker hace esta descripción como correspondiente a la visión de Husserl en sus primeros años, hasta *Ideas I*. Sin embargo, el mismo Peucker sostiene que, a pesar de desarrollos y modificaciones posteriores, se trata de un modelo al que Husserl “nunca renunció completamente en sus años posteriores” (Peucker, 2008a, p. 1). Por lo demás, a menos que la bibliografía final indique lo contrario, las traducciones presentadas a lo largo del artículo son de mi autoría.

Desde el punto de vista de las valoraciones, los actos afectivos presuponen la donación previa del objeto que se ha de juzgar. Se tienen que fundar, por lo tanto, en otro tipo de actos, que son justamente aquellos que *dan* el objeto, razón por la cual son llamados por Husserl “actos objetivantes”.⁶ Los principales actos objetivantes son la representación (dentro de la cual encontramos la percepción y las distintas formas de presentificación) y el juicio. Por su parte, los actos afectivos son, por contraste, no-objetivantes, pues no dan el objeto: no dan la rosa o el atardecer; la rosa y el atardecer se dan a la conciencia principalmente gracias a la representación y no gracias a su valoración. Así resume Melle estas mismas ideas:

Los actos anímicos y volitivos valorantes son, siguiendo a Husserl, actos no-objetivantes. Tales actos son dependientes de los actos objetivantes, esto es, actos del representar [*Vorstellens*] en forma de percepciones y presentificaciones así como actos del juzgar, en dos sentidos: por una parte, están necesariamente fundados en actos objetivantes —primero se me tiene que dar un objeto o estado de cosas en forma de representación o de juicio antes de que me pueda yo referir a ello valorativa o volitivamente—; por otra parte, son necesarios los actos objetivantes para que el agrado valorante se transforme en un conocimiento de valor que pone y predica el valor de forma objetiva (Melle, 1988, pp. xxxv-xxxvi).

⁶ Los actos objetivantes son aquellos que dan el objeto de la experiencia, vale decir, la unidad que permanece siendo la misma a través de las diferentes manifestaciones y escorzos de esta. Objeto es aquello “idéntico consigo mismo”, esto es, “idéntico en la ‘conciencia de representación’ [*Vorstellungsbewusstsein*]” (Hua XLIII/2: 32). Por su parte, “la representación, en el más amplio sentido, es el acto que, como siempre, constituye la *unidad* intencional que es determinable en el pensamiento lógico como idéntica [a sí misma] y diferente frente a otras” (Hua XLIII/2: 32). El acto representante es, por lo tanto, el más importante de los actos objetivantes; está ligado a los actos lógicos del pensamiento y gracias a ello es capaz de dar unidad y, por lo tanto, sentido de objeto a la multiplicidad con la que este se manifiesta. “El representar convierte lo no-valor [*Nicht-Werte*] en objeto o es la fuente aperceptiva de objetos libres de valor [...]. El representar se dirige a las ‘meras cosas’, el afecto [*Gemüt*] se dirige al ser-valor de las cosas, a las cosas de valor [*Wertsachen*]” (Hua XLIII/2: 37).

Esta idea vuelve a aparecer en las *Vorlesungen zur Einleitung in die Ethik* de 1920 y 1924 (periodo de la segunda fase de la ética husserliana) en las que Husserl afirma: “los actos valorantes y volitivos se fundan en actos de conocimiento [...] [pues] lo que no me puedo mínimamente representar no lo puedo valorar en absoluto” (Hua XXXVII: 274).⁷ Esto, por definición, quiere decir que los actos afectivos no son *vorstellende Akte*, actos representantes (cfr. Peucker, 2008a). Por esta razón, para Husserl, los actos afectivos o valorativos son “por así decir mudos y, en cierto sentido, ciegos” (Hua XXVIII: 68), pues “la razón meramente valorante no ve, no capta, no explica, no predica” (Hua XXVIII: 69). Por lo tanto, tratamos aquí con un tipo de acto que depende completamente de los “ojos” de los actos objetivantes. “El entendimiento objetivante debe primero dar a los actos axiológicos y prácticos ‘los ojos del intelecto’ (Hua XXVIII: 68)” (Melle, 1990, p. 36).

Hay varios elementos de este modelo husserliano en los que nos interesa poner atención. Primero, ¿en qué sentido se puede hablar de una *wertende Vernunft*, de una razón valorante? ¿No es esta expresión en cierto sentido contradictoria? ¿Es capaz la “razón” de “valorar”? ¿Qué es realmente aquello que da valor a los objetos valorados? Ciertamente, como dice Melle, “sin el agrado valorante [*wertendes Gefallen*] no hay para nosotros ni objetos de valor ni determinaciones de valor” (1988, p. xxxvi), y este *Gefallen* implica fundamentalmente un acto de tipo afectivo, es una vivencia afectiva. Entonces, nuevamente, ¿qué facultad es la que propiamente da el valor? La idea de Husserl parece ser que los actos afectivos nos dan la “intuición” del valor (correlativa con los afectos como contenidos sensibles), mientras que la razón nos da su “conocimiento”. Pero ¿cómo pueden los actos afectivos darnos la intuición, la captación, de los valores si la “razón valorante” es ciega, no es representante?⁸

⁷ Esto muestra que Husserl no renuncia a su modelo estratificado cuando amplía su investigación hacia la pasividad y la fenomenología genética. El cambio de nomenclatura introducido en *Ideas I* (donde Husserl afirma que los actos valorativos y volitivos también son, a su manera, objetivantes; cfr. Hua III/1: 272) no modifica este esquema general, pues a pesar de que se le reconozca a los actos afectivos una forma peculiar de objetivación, siguen siendo entendidos como actos fundados (cfr. Parrella, 2025, p. 19). Sobre esto volveremos más adelante.

⁸ En Hua XLIII/2: 21 podemos leer: “Luego, es seguro afirmar: solo en actos afectivos [*Gemütsakten*] puede encontrarse el sentido del predicado de valor, pero los actos afectivos por sí solos no objetivan”.

A raíz de estos puntos nos preguntamos por la capacidad explicativa de los conceptos husserlianos respecto al fenómeno del valor. Husserl parece hacer uso de un modelo explicativo que, a nuestro juicio, presenta algunos problemas estructurales; más aún, es justamente su uso el causante de esos problemas.

En primer lugar, como adelantamos en la introducción, nos preguntamos por la posibilidad de que Husserl sea víctima de un sesgo en favor de la “visualidad” como criterio último de lo que él entiende por “evidencias” y que esta preferencia por lo visual sea causante, como criterio, de producir expectativas sobre un tipo de evidencias que el fenómeno del valor, en lo que tiene de afectivo, no puede cumplir a causa de su propia naturaleza. A raíz de lo cual surgirían las complicaciones explicativas. Entendemos este hipotético sesgo como un primado implícito de lo visual en el esquema analítico husserliano que, para efectos prácticos, llamaremos “criterio visual”. Se trata de un criterio que otorga primacía a los actos de percepción (*Wahrnehmung*) al considerarlos como los originaria y paradigmáticamente “videntes”, dadores originarios del objeto mismo, en referencia a los cuales serían también videntes solo aquellos actos derivados de estos (recordar, imaginar y en general todo presentificar). No cuestionamos que la percepción nos de experiencia directa de los objetos, sino que sea el *único* criterio de donación directa, frente al cual la afectividad emotiva tendría, por tanto, siempre un carácter derivado o secundario. Nos preguntamos, pues, si la afectividad del *Gemüt* no da también originariamente un elemento de la experiencia a través de su sola afectividad.

Recordemos que para Husserl la percepción es “el modo originario [*Urmodus*] de la *Anschauung*” (Hua VI: 107, citado en Breyer, 2010, p. 306) y por esta razón su objeto se da a la representación inmediata como “él mismo”. Y es esta cualidad que tiene la percepción la que da el criterio para el conocido “principio de todos los principios” de la fenomenología, que sostiene que el conocimiento fenomenológico se tiene que fundar en la “intuicionalidad y autodonación originaria de sus objetos [*Anschaulichkeit und originärer Selbstgegebenheit ihrer Objekte*]” (Breyer, 2010, p. 306).⁹ Esta intuicionalidad originaria lleva en sí el criterio de lo visual, es la que hace “evidente” al objeto mismo.

⁹ Cabe considerar, por ejemplo, cuando en *Ideas I* Husserl dice que la “intuición de esencia [*Wesenserchauung*] es un acto originariamente dador [*originär gebende Akt*] y como tal es el análogo de la percepción sensible y no de la imaginación”

Adelantamos, entonces, aquí la tesis de que esta primacía de la percepción como criterio de toda *Selbstgegebenheit* y, por tanto, de toda *Anschaulichkeit*, lleva implícito el criterio visual que acabamos de mencionar y que, creemos, es una posición que lleva inevitablemente a los problemas que presenta la relación intencional entre los actos objetivantes y los no-objetivantes, pues de él se configura un modelo que es adecuado, ciertamente, para el análisis de los objetos dados en la percepción externa, pero, creemos, no para el análisis de los actos afectivos. Así, si, como afirma Husserl, la razón valorante es ciega y muda, entonces, ¿qué es lo que *ven* los juicios de valor en los objetos puramente objetuales (sin valor) para asignarles un valor?, y ¿cómo hacen esto? ¿Cómo pueden asignar valor si la razón valorante que juzga en ellos es ciega?

La división entre actos videntes y no-videntes parece dejar fuera justamente el aspecto que la investigación sobre los actos de valor debe considerar, porque si los actos videntes (objetivantes) son los que ven (y dan) el objeto, entonces su visión se restringe justamente al puro campo de la objetualidad del objeto, a su solo aspecto representacional en tanto objeto (sin capa de valor).¹⁰ Por lo tanto, el yo, en la medida

(Hua III/1: 43; énfasis añadido). Con base en esto, a renglón seguido Husserl enuncia el principio de todos los principios: “que toda intuición originariamente dadora es una fuente de derecho del conocimiento, que todo lo que se nos presenta originariamente en la ‘intuición’ (en su realidad corporal, por así decir) sea simplemente tomado tal como se da, pero solo dentro de los límites en los que se da” (Hua III/1: 43-44). Amparados en este principio los análisis husserlianos tienden a explicar las valoraciones y actos afectivos con categorías visuales, es decir, tienden a llevar lo no visual a lo visual. De ahí que Husserl pueda afirmar, por ejemplo, que “el valorar es una conciencia exactamente [*genau*] en el sentido del percibir o representar” (Hua XXXI: 291). También es sintomático de esto mismo la valoración que él hace del matemático, por ejemplo, en la carta de 1919 dirigida a Arnold Metzger en la que dice lo siguiente: “Yo hablo de la fenomenología como el matemático de la matemática: ella es ciencia *propriadamente dicha* [echte], extraída a partir de evidencia pura, un campo de posibles proposiciones verdaderas y falsas. A pesar de todos los escépticos y filósofos confusos, el [matemático] habla así porque él ‘*ve*’ [denn er ‘sieht’]” (*Briefwechsel*, IV: 413 ss., citado en Marbach, 2009, p. xix; las cursivas son del propio Husserl). Por su parte, dice Husserl, “la mera razón valorante no ve”, sino que son “los actos lógicos” los que “iluminan” y “hacen visible” (Hua XXVIII: 69).

¹⁰ “Si hay algo ahí a la manera de la percepción o del recuerdo, si hay algo ante la *mirada* espiritual [*vor dem geistigen Blick*] a la manera del juicio, que

que *ve* solo con esta facultad representativa y objetivante no puede, por definición, percibir nada de valor en el objeto representado. De modo que esta facultad “visual” objetivante no puede ser la causante de la emergencia del valor y la valoración (lo cual pone en duda si estos actos son realmente “los ojos” de los actos afectivos). Tiene que entrar otro tipo de acto, justamente el “valorante”. Pero, de nuevo, este es “mudo y ciego” en la medida que no es un acto representante. ¿Cómo entonces puede valorar? Al decir que es la facultad “valorante” (*wertende*) se da el problema por resuelto sin haberlo resuelto.

Peucker (2008b) apunta a este mismo problema cuando sostiene que:

Esta diferenciación [entre actos objetivantes y no-objetivantes] produce problemas si Husserl también afirma que los sentimientos o las emociones nos dan un cierto acceso a propiedades específicas que los actos teóricos no pueden revelar, esto es, propiedades de valor. Aquí la pregunta que surge es cómo los sentimientos pueden hacer esto si no son actos objetivantes. ¿Cómo pueden abrir un aspecto del mundo si no son actos que hacen que algo sea objetivo? ¿Cómo puede ser entendible la donación de valor en la conciencia afectiva si esta no es llevada a cabo por actos objetivantes? Husserl se hace estas preguntas cruciales, pero admite que no pudo resolverlas con su teoría ética temprana (Peucker, 2008b, p. 318).¹¹

Peucker ve un problema en la posición de Husserl al distinguir los actos afectivos de los teóricos bajo la clasificación de no-objetivantes y objetivantes, respectivamente, pues “al hacer esto, distingue los actos emotivo y valorante del acto objetivante, pero también les adscribe a aquellos una función objetivante, es decir, dadora de valor” (2008b, p. 318, n. 29). Esta es la razón por la que, como mencionamos en nota

algo está hecho de tal y tal manera, entonces eso es conciencia objetivante” (Hua XXVIII: 338-339; énfasis añadido).

¹¹ Esto último lo dice Husserl en Hua XLIII/2: 39: “Me ha costado infinitos esfuerzos llevar a cabo esta concepción: desde los años noventa, en que reparé en el ‘aparecer-de’ de todos los actos, lo intenté constantemente, y dado que no vi ninguna posibilidad no pude arreglármelas [*konnte nicht fertig werden*] con este aparecer, con esta conciencia-de de todos los actos”.

precedente, luego Husserl cambiará su *nomenclatura* en *Ideas I* y dirá que todos los actos son, a su manera, objetivantes, es decir, constituidores originales de un tipo de objetualidad propia. Para ejemplificar esto toma, en ese pasaje de *Ideas I*, precisamente el caso de los actos valorantes:

[...] por ejemplo, la conciencia valorante constituye, en contraste con el mero mundo de cosas, una objetualidad [*Gegenständlichkeit*] de nuevo tipo, una objetualidad “axiológica”, un “ente” de una nueva región, justamente en la medida en que a través de la esencia de la conciencia valorante en general son prefiguradas tesis dóxicas actuales como posibilidades ideales que ponen de relieve objetualidades de un nuevo tipo de contenido —valores— como “mentadas” en la conciencia valorante. En los actos afectivos son afectivamente mentadas [...] (Hua III/1: 272).

El mero objeto del mundo es evidentemente un objeto, y lo que Husserl llama “valor”, vale decir, el objeto valioso, es también un objeto y como su ser total no es el ser mera cosa del mundo sino justamente un objeto valioso, su valor tiene que estar incluido en su objetualidad como propiedad suya. Por lo tanto, el acto de la conciencia que constituye el valor debe ser también, a su manera, un acto objetivante. Pero obsérvese que en el pasaje citado no hay mención sobre cómo llega la conciencia valorante a “mentar” (*meinen*) su objeto como poseyendo valor. La conciencia valorante o los actos afectivos ya son valorantes por el hecho de ser tales, pero ¿de dónde les viene su mentar las cosas valorativamente? ¿Cuál es el origen de esas tesis dóxicas valorativas? Se dirá: el origen es la facultad valorativa, la conciencia valorante o los actos afectivos, lo cual, claro, no nos ayuda a resolver el problema.

Por estas razones nos parece que las modificaciones conceptuales introducidas en *Ideas I* no van en la línea que nos permitiría avanzar en la solución de nuestro problema. El reconocimiento de la presencia de una propiedad objetivante en los actos afectivos no nos permite resolver el problema ni del origen ni de la intencionalidad de esos actos. Se reconoce y, en parte, se resuelve el problema señalado por Peucker, pero se asume que *de alguna forma* los actos afectivos constituyen el objeto valioso. El análisis parece partir asumiendo ya como dado el objeto valioso.

Pero quisiéramos señalar aún otro aspecto del problema relativo a la naturaleza de los actos valorantes y a su forma peculiar de intencionalidad. Quisiéramos volver al problema del elemento visual. En el § 37 de *Ideas I*, Husserl se refiere, de manera general, a la intencionalidad y al “estar-dirigido-a” del yo en el *cogito*. Ahí leemos que:

[...] al *cogito* mismo le pertenece una inmanente “mirada-hacia” [*Blick-auf*] el objeto que, por otro lado, brota desde el “yo”, de modo que nunca puede faltar. Este *Ichblick* [mirada del yo] hacia algo es, dependiendo del acto, mirada-hacia percipiente en la percepción, ficcionante en la ficción, agradable en el agrado, queriente en el querer (Hua III/1: 75).¹²

Este *im Blick Haben* (tener-en-la-mirada), “tener-en-los-ojos-del-espíritu [*im geistigen Auge Haben*]” (Hua III/1: 75), obsérvese bien, pertenece “a la *esencia* del *cogito*, del acto como tal” (Hua III/1: 75). Es decir, haber acto significa tener el objeto intencional en la mirada o frente a los ojos del espíritu. El acto en general es esencialmente, por tanto, *vidente*, su ser consiste en *ver*.¹³ Por lo tanto, los actos afectivos o valorantes, por el hecho de ser actos, deben ser necesariamente también videntes. Su forma de dirigirse hacia su objeto se comprende según el modelo de la percepción visual. Y por esta razón Husserl no tiene inconveniente en determinar (en la cita) tanto los actos representantes (percepción, ficción) como los afectivos (agrado, querer) según el mismo esquema: para él en esencia poseen *la misma estructura intencional*, determinada por la visión de la mirada del espíritu, porque la intencionalidad es para Husserl una sola, es siempre y solamente “conciencia-de” que, como tal, es siempre entendida por él según el esquema $C \rightarrow O$,¹⁴ donde C representa la conciencia y O su objeto. En esto quisiéramos hacer énfasis, porque lo que sostenemos es que, con esta manera de concebir

¹² Hemos preferido mantener homogéneas las formas de los participios para hacer más claro el estilo y el sentido de la oración original.

¹³ En Hua XXXVII: 74, Husserl dice, por ejemplo, que tanto las propiedades cóscicas como el valor del objeto son “vistos” (*erchauen*) y que tanto la percepción como la valicepción son actos que poseen una “dirección de la mirada sobre las cosas” (*Blickrichtung auf Sachen*).

¹⁴ Aquí el símbolo de la flecha no se usa con su sentido lógico, sino como representación de la dirección intencional de la conciencia.

la intencionalidad de los actos afectivos, según el modelo de los actos representantes y estos, a su vez, basados en el modelo visual, es que comienzan los problemas con la teoría de Husserl para describir los actos afectivos.

Por lo tanto, si este problema surge de la preferencia implícita por el modelo visual, entonces para intentar escapar de él podemos considerar otras formas sensibles de dárse nos los objetos y ver si nos pueden ayudar a pensar formas no visuales de donación. El tacto, por ejemplo, es una facultad no vidente que, sin embargo, no por ello está impedido de darnos un aspecto de los objetos o del mundo —y este aspecto a su vez (hay que remarcarlo) *no es visual*, es decir, no puede ser reducido a una comprensión visual—. ¹⁵ En este caso, la no videncia no significa ausencia de presencia.

Así, tomando esto en cuenta, podemos decir que la facultad valorativa es una facultad fundamentalmente *afectiva* y *no visual*, cuyo análogo *más adecuado* no es el sentido de la vista, sino el táctil. Supone que algo nos afecta, pero de manera que su presencia se hace manifiesta en nuestra conciencia mediante un sentimiento y no una representación. Representación y sentimiento son categorías epistémicamente excluyentes. *El valor no se ve, pero se siente*: este adagio representa en síntesis la tesis central de este trabajo. Y este sentir afectivo, tal como el sentir táctil del ciego, es, no obstante, una forma de donación del mundo por sí misma, que por sí misma da una presencia a la conciencia.

Pero más aún, también debemos decir, invirtiendo la perspectiva, que los actos objetivantes y sus correspondientes facultades, por videntes que sean, son, sin embargo, ciegos para la dimensión que sí percibe la facultad afectiva. Así, en general podríamos decir que lo que la conciencia “ve” en el objeto es algo más que la sola objetualidad de este, que la conciencia no solo ve con los ojos de las facultades representacionales, sino también con los ojos de la facultad afectiva, por así decir. Lo que percibe esta facultad no es algo solamente representacional, sino algo que *se siente*. Ahora bien, lo sentido en el

¹⁵ Por supuesto, no sostenemos que Husserl no tenga conciencia de que existen otras formas de donación de los objetos del mundo. Baste pensar en sus análisis sobre la música o los campos sensoriales. Lo que decimos es que Husserl da primacía a la facultad visual por sobre las demás y que esto afecta al modo de comprender la forma de donación de los objetos (sobre todo la donación de valores).

sentimiento no tiene figura ni forma representacional alguna; Ziri6n Quijano (2024, pp. 153 y ss.), siguiendo a Husserl, lo ha llamado, por ejemplo, el “colorido” de la vida (cfr. tambi6n Quepons, 2015), pero esto, claro, es solo una met6fora. La funci6n del sentimiento no es visual, sino, en t6rminos generales, valorativa, pues es gracias a ella que se puede asignar un signo axiol6gico (positivo/negativo) a lo representado. Lo que m6s adelante veremos es que el sentimiento con car6cter valorativo no se relaciona solo con el objeto representado, sino tambi6n y al mismo tiempo con el estado particular del sujeto valorante, por lo que el juicio de valor correspondiente no solo se remite al objeto expl6citamente valorado, sino tambi6n, y de manera necesaria, al estado del sujeto. Esta referencia al sujeto y su estado particular es lo que se deja afuera en el esquema intencional simple $C \rightarrow O$ en el que la conciencia es claramente analogada con el sentido de la vista. Adelant6ndonos a lo que diremos m6s adelante, lo que propondremos es un modelo ampliado en el que se destacará una doble direccionalidad de la conciencia. Modelo que, por ahora, podemos representar con el esquema: $S \leftarrow C \rightarrow O$.

Cuando, por ejemplo, Husserl toma como ejemplo de *Gemütsakt* el alegrarse y sostiene que “la alegría se dirige a la cosa que produce alegría [*auf die erfreuende Sache*]” (Hua XXVIII: 336), en este caso al buen clima, se puede ver claramente la forma binaria de abordar el problema de la intencionalidad valorativa, siempre solo con estos dos elementos: la conciencia (C) y su objeto (O), el alegrarse (del yo) y el buen tiempo.¹⁶ Por eso dice Husserl que la conciencia de alegría “no es conciencia

¹⁶ Se me ha se~alado con gran acierto que produce extra~eza hablar de un esquema intencional dual en Husserl cuando 6l expl6citamente sostiene que la intencionalidad tiene una estructura triple: *ego-cogito-cogitatum* (cfr. Hua I: 83 ss.; Hua Mb IV: 49). Esto me permite aclarar que mi esquema dual, $C \rightarrow O$, lleva ya impl6cita esa “triple” estructura, pues es simplemente una esquematizaci6n de la idea intencional fundamental de que toda conciencia es “conciencia-de”. El *ego*, mediante su *cogitatio*, tiene conciencia de un *cogitatum*. Por lo tanto, mi esquema $C \rightarrow O$ equivale a: [*ego-cogito*] $\rightarrow$ [*cogitatum*]. Ahora bien, si se acepta que el sentimiento sea tambi6n una *cogitatio*, entonces lo que se sostiene en este trabajo es que falta otro *cogitatum*, pero esta vez en la direcci6n del estado del sujeto. De modo que en los actos valorativos tendríamos: [*cogitatum*] $\leftarrow$ [*ego-cogito*] $\rightarrow$ [*cogitatum*]. Un *cogitatum* es correlato de un acto objetivante; el otro, de un acto no-objetivante, de un sentimiento, con su peculiar car6cter intencional. Haremos una articulaci6n m6s completa de este esquema (con otra nomenclatura) m6s adelante, en la secci6n dedicada a Menger.

que en sí misma traiga a manifestación [*Erscheinung*] cosa alguna” (Hua XXVIII: 339), y agrega: “ciertamente, en el sustrato fundante del acto valorante se lleva también a cabo un manifestar [*Ercheinen*] o un mentar [*Meinen*], pero solo en él [*eben in ihr*], y el valorar mismo no hace nada nuevo respecto a ello. Si este se suprimiese, la objetivación sería exactamente la misma [*genau dieselbe*]. El valorar no lleva a cabo ninguna otra objetivación” (Hua XXVIII: 339; énfasis añadido).¹⁷ Sin embargo, si el alegrarse es un acto afectivo, entonces se experimenta como un sentimiento (positivo) y, por tanto, no es el caso que este sentimiento no dé cuenta de nada: da cuenta del estado del propio sujeto, que ha pasado de una situación menos positiva a una más positiva, porque ahora se alegra del “buen” tiempo (siguiendo el ejemplo de Husserl). Pero ¿por qué se juzga el tiempo como bueno? Sostenemos que se lo juzga así *en relación* con el estado del sujeto: ese mismo sol que para mí es signo de buen tiempo para otro animal puede ser signo de algo negativo o incluso nefasto. No se trata, por lo tanto, de juicios que puedan ser comprendidos con la sola relación al objeto; no son juicios puramente “objetivos”, sin presencia del (estado del) sujeto. El juicio “la manzana es roja” es un juicio objetivo que, como tal, es independiente del estado anímico del sujeto. Y aquí sí funciona el esquema intencional simple $C \rightarrow O$. Pero el juicio “la manzana es buena” ya no puede no considerar al sujeto, porque el calificativo “buena” tiene una referencia esencial y necesaria a su estado y no solo al “valor” del objeto.¹⁸ Por ejemplo, después de comer cuarenta manzanas, la número cuarenta y uno puede matarme, y este hecho, esta posibilidad, está determinada por mi condición como ser vivo, *en relación con la cual* hay objetos que adquieren determinado signo de valor. La valoración, por lo tanto, está determinada por las

¹⁷ Con la nomenclatura de *Ideas I* se diría que el valorar sí lleva a cabo una nueva objetivación, pero de otra índole, sin realizar ninguna objetivación en el objeto como tal: desde el punto de vista de su ser como puro objeto, el objeto “libro” con valor es exactamente el mismo objeto “libro” sin valor.

¹⁸ Para Husserl, el correlato objetivo del acto intencional afectivo es “el valor” (cfr. Crespo, 2012, p. 44). Y por ello también cuando “coloreamos” el mundo producto de un estado afectivo, la causa o motivo de ello se encuentra también en el objeto. Así, por ejemplo, Ziriñ Quijano (2024, p. 162): “El valor en el objeto o del objeto [...] es el fundamento del resplandor, no el resplandor mismo”. La referencia y la determinación del valor parecen venir siempre y exclusivamente del lado del objeto.

condiciones sensibles del sujeto, y lo está *esencialmente*.¹⁹ No es lo mismo la manzana número dos que la número cuarenta, pero ¿qué cambió? No ha cambiado nada en las manzanas mismas, en el *objeto* intencional; lo que cambió es el estado del sujeto y con él sus actos valorativos sobre las manzanas.

Creemos, pues, que cuando Husserl dice que el acto axiológico no trae a manifestación cosa alguna se equivoca, y se equivoca porque considera la “manifestación” solo desde el punto de vista, de nuevo, de lo visible, del *Scheinen* de la *Erscheinung*. Pero el hecho es que el sentimiento implicado en el acto axiológico es también una manifestación, solo que no del objeto, sino del sujeto. Es un error sostener que el sentimiento no manifiesta nada por el hecho de no manifestar nada visible.

El valor, dice Husserl, pertenece al objeto; el valorar se relaciona con ciertos “momentos” del objeto, y el sentimiento pertenece a la vivencia del sujeto y su conciencia (cfr. Hua XLIII/2: 54). Nosotros diríamos, sin embargo, que el valorar es al mismo tiempo relación con el objeto y con el sujeto. Husserl parece centrarse solo en el hecho de que valorar es “valorar el objeto”. Ciertamente es valorar el objeto, pero esa dirección de la valoración está mediada por la valoración implícita del sujeto hacia su propio estado. La valoración del vaso de agua está mediada por la valoración implícita de la satisfacción de la sed. El padecer sed es un estado del sujeto (ya en sí mismo valorado por su sola presencia en el ánimo como padecimiento, como estado negativo) y su satisfacción es un nuevo estado del sujeto, valorado positivamente ya sea como expectativa (por lo tanto, como posible estado futuro) o como estado actual (por ser vivencia de satisfacción).

2.2. Los actos afectivos originarios también son fundados

Esto no quiere decir, sin embargo, que Husserl no haya sido consciente de algunos problemas implicados en su modelo explicativo.²⁰

¹⁹ Debido a su manera de comprender la intencionalidad de la conciencia valorativa, Husserl ($C \rightarrow O$) llega a hacer el siguiente tipo de afirmaciones: “[...] un valor es lo que es sea o no aprehendido valorativamente. Ser valor, ser bello o ser bueno no quiere decir que alguien considera algo como valioso” (Hua XXXI: 290). La subjetividad del sujeto está excluida, por lo tanto, para el ser del valor: el valor es un objeto frente al cual puede estar o no la conciencia.

²⁰ Partiendo por el “profundo abismo” (*tiefe Kluft*) que desde el comienzo él vio que separaba a los actos representantes de los valorantes y sus

Tenía conciencia, por ejemplo, de que en la experiencia real no hay tal cosa como un darse puro del solo objeto representacional sobre el cual se agregaría, luego, ese nuevo elemento, en sí independiente, que sería el valor del objeto. Sabía que en realidad no hay tal separación y que el objeto se da como una unidad en la que se entremezclan ambos elementos.²¹

Como vimos arriba en los pasajes citados de las lecciones de ética de los años veinte, para Husserl sigue siendo el caso que los actos valorativos y volitivos son “fundados” en los actos representantes (esta idea nunca se modifica; cfr. Husserl, 1939, pp. 1-123). Cuando ahí dice: “lo que no me puedo mínimamente representar no lo puedo valorar en absoluto” (Hua XXXVII: 274), afirma una idea que para él permanecerá siendo fundamental hasta el final.²²

correspondientes intencionalidades (cfr. Hua XXVIII: 337).

²¹ Sería necesario hacer un análisis sobre los avances y aportes de la fenomenología genética al problema del origen e intencionalidad de los actos afectivos, análisis que aquí, sin embargo, nos llevaría a extendernos demasiado (al respecto, cfr. Cabrera, 2021). Sobre fenomenología genética se puede consultar Hua I (109 ss., §§ 37-39), Bernet *et al.* (1993, pp. 195 y ss.), Lohmar (2017) y Rizo-Patrón Lerner (2024). Dicho a grandes rasgos, el análisis genético muestra que la constitución de la experiencia y sus objetos (incluido el objeto valioso) vienen ya pre-constituidos, por así decir, gracias a la pre-donación pasiva pre-predicativa del mundo entorno. Mundo concreto que, por tanto, está ya constituido por elementos práctico-emotivos, por una historia, por tradiciones y, en cada sujeto, por una biografía. De modo que el análisis genético nos advierte que el modelo de estratificaciones estáticas de la constitución es solo un dispositivo analítico, es decir, que no hay algo así como un darse puro del objeto sobre el cual, en segunda instancia, se constituiría su valor, como una capa agregada, sino que el objeto total viene ya constituido pasivamente. Sin embargo, nos parece que estos importantes avances en la teoría fenomenológica van en otra dirección, tienen otro foco de interés que el que atañe a nuestro problema.

²² La posición de Max Scheler sobre este problema es una variante interesante y más afín a nuestra manera de comprender el fenómeno en cuestión. A la inversa de la teoría de los actos representantes como los originariamente fundantes y, en consecuencia, los demás como actos necesariamente fundados, Scheler sostenía que *antes* del acto representante u objetivante tiene que estar ya operativo el acto valorante. Esto lo recuerda Lotz (2006) al tratar de la intencionalidad práctica en Husserl: “Los valores y los propósitos *hacen posible*, de acuerdo con Scheler, que las cosas sean dadas en ciertas gradaciones de valor, de las cuales tengo que ya tener conciencia para poder dirigir mi atención hacia ellas. Por ejemplo, Scheler argumentaría que yo debo ya ser consciente [*be aware*]

Más aún, en realidad la arqueología que hace Husserl de las fuentes originales de los actos que llevan ascendentemente primero a la constitución de la experiencia y, luego, al conocimiento de lo experimentado muestra la radicalidad de la tesis husserliana del carácter fundado de los actos valorativos. Pues, en efecto, para Husserl el proceso analítico de de-construcción de la experiencia tomada en general lleva “finalmente a los datos sensibles [*Empfindungsdaten*] como objetos originarios primitivos últimos que ya no son constituidos por ningún tipo de actividad del yo [*Ichaktivität*]” (Hua IV: 214). Pero ¿qué sucede en el caso de la experiencia afectiva? ¿Cuáles son sus últimos elementos constitutivos? “Si volvemos hasta el fundamento más profundo”, dice Husserl, “de los actos que se encuentran antes de todo pensamiento, llegamos aproximadamente a: 1) las sensaciones sensibles [*sinnliche Empfindungen*], sensaciones en sentido corriente; 2) los sentimientos sensibles más simples (sensaciones de sentimiento [*Gefühlsempfindungen*]). *Estas últimas se fundan en las primeras*” (Hua XLIII/2: 37; énfasis añadido).²³ Por lo tanto, los sentimientos más originarios ya son actos fundados, “fundados en las sensaciones [*Empfindungen*]” (Hua XLIII/2: 38). De donde se sigue que la sensación empírica, “sensible”, es el elemento más originario de la experiencia. “Las *Empfindungen* forman la ‘materia prima’ [*rohen Stoff*] de las apercepciones empíricas, sobre todo de las percepciones [*Wahrnehmungen*] empíricas, las sensibles [*sinnlichen*]. [...] Pero de tal modo que en estas apercepciones no entra ningún acto afectivo [*Gemütsakte*]. Ellas constituyen ser, no valor” (Hua XLIII/2: 38). Husserl separa aquí fuertemente el “grupo de apercepciones” que constituyen el ser y que se remiten correlativamente al entendimiento (*Verstand*) del grupo de apercepciones que constituyen valor y que se remiten al sentimiento. “Estas últimas entran en su forma más primitiva ya como sentimientos sensibles: no son meras sensaciones [...], sino ya conciencia-de, y ciertamente conciencia-afectiva-de, valoración del tipo más primitivo, vale decir, *valoraciones de contenidos sensibles* [*von Empfindungsinhalten*]” (Hua XLIII/2: 38; énfasis añadido).

Husserl mantiene, pues, el mismo esquema de fundación de los actos afectivos cuando busca en estos sus constituyentes más elementales; vale decir, así como en las valoraciones de objetos la valoración se funda

del valor positivo de la comida *antes* de que pueda ser atraído y ‘tirado’ por un objeto que tiene el aspecto de ‘comida’” (Lotz, 2006, p. 126).

²³ Al respecto, cfr. también Zirió Quijano (2024, pp. 160-161).

en la presencia del objeto, así también sigue ocurriendo en el nivel más primitivo: los sentimientos sensibles, como formas más primitivas de valoración, suponen la donación previa de su “objeto”: los contenidos sensibles. Estos datos y contenidos sensibles son, pues, puramente “empíricos”, sensibles; no pertenecen a la esfera del valor. Este “empirismo” de Husserl permea también su concepto de *Wertnehmung* (valicepción o percepción de valor), visible ya en la palabra misma, pues concibe esta función como un análogo de la percepción (cfr. Parrella, 2025, p. 13; Quepons, 2015)²⁴ y, por lo tanto, de nuevo tomando los actos representantes como criterio de donación. Esta *Wertnehmung* “se diferencia de la percepción por el hecho de que, al contrario que esta, es un acto del sujeto afectivo [*des fühlenden Subjekts*] fundado en aquella” (Fabbianelli, 2010, p. 311). Se funda en la percepción, pero, además, al ser entendida como análoga a esta también es analogado el modo de darse su objeto, es decir, así como en la percepción se da el objeto mismo a través de las apariciones subjetivas (cfr. Hua IV: 201 y 209), así también se da, según Husserl, el valor mismo a través de la vivencia de valor (cfr. Quepons, 2024, p. 168) sin que se confunda con esta. Nuevamente, no creemos incorrecta esta descripción de Husserl ni mucho menos: percibir una rosa como bella implica adjudicarle la cualidad de belleza, obviamente. Sin embargo, el problema que vemos en la analogía con la percepción es el convertir el valor (para la valicepción) en un análogo del objeto (para la percepción), con lo cual se corre el riesgo de autonomizar excesivamente el valor, de hacerlo tan autónomo como la existencia de la rosa (cfr. Hua XXXVII: 74).²⁵ Para todos es evidente que la rosa, como objeto, no es una creación o un producto del sujeto, pero no se puede adjudicar el mismo grado de independencia respecto al sujeto al *valor* de la rosa.

La primera conclusión, pues, que sacamos de estos análisis es que para Husserl lo originariamente primero²⁶ es el mundo natural, dado por los sentidos externos, por los datos sensibles. Por esta razón Asemissen lo ha criticado llamándolo “sensualista” (cfr. De Almeida, 1972, p. 40) y

²⁴ Luego la concibe, además, también como un análogo de la apercepción (cfr. Crespo, 2012, p. 46).

²⁵ Sobre el paralelismo entre actos del entendimiento y actos emotivos en Husserl, cfr. el pasaje de los Ms. A (VI 12 1/273a) citado por Crespo (2012, p. 44).

²⁶ Al menos desde el punto de vista analítico, considerando el modelo de fundación.

sosteniendo que el concepto de “sensación” (*Empfindung*) que utiliza Husserl para explicar su modelo de fundaciones no es realmente un concepto de sensación, pues “utiliza un concepto de sensación no comprobado [*unüberprüften*]”, por lo que “sobre las sensaciones mismas [...] no se hace cuestión en absoluto” (De Almeida, 1972, p. 40). Para algunos analistas, por tanto, la teoría husserliana de la sensibilidad es problemática, y para nosotros ello va unido al hecho de la primacía que Husserl da al objeto empírico (en última instancia, a los *Empfindungsinhalten*).

Hecho que se relaciona, a su vez, con nuestra segunda conclusión. Tanto con el análisis estático como, luego, con el genético, Husserl va en busca de la fundación más originaria del valor y, sin embargo, su análisis no se gira en ningún momento hacia el rol fundacional del sujeto (en tanto sujeto de necesidades). Es el objeto el que acapara la atención analítica: toda conciencia es conciencia de objeto; por lo tanto, también lo es la conciencia afectiva. El *objeto* valorado es el polo intencional, y esto es así incluso en las primitivas sensaciones de sentimientos, pues estas ya son actos dirigidos a un objeto (a contenidos sensibles) y, por lo tanto, desde su manifestación más originaria actos que son comprendidos por Husserl bajo el esquema intencional $C \rightarrow O$. Por esto se ha visto aquí una “debilidad referencial” del acto afectivo, pues su intencionalidad parece ser solo “secundaria” (Cabrera, 2021) con respecto a la intencionalidad primaria dirigida al objeto mismo. Ciertamente hay un lado subjetivo en la constitución del valor, pero al entenderlo bajo este esquema Husserl parece restringirlo solo a su rol secundario relativo a la constitución del objeto, de estar al servicio del objeto. En general, el problema que quiere resolver Husserl es cómo se constituyen *los objetos*, el natural y el cultural (de ahí el modelo intencional). Por lo tanto, la actividad afectiva es vista principalmente en función del objeto. Pero, como ya hemos dicho, el acto afectivo requiere de una relación, que es aún más fundamental, hacia el estado del sujeto, sin la cual es imposible que existan valores en el mundo. Al tener presente esta relación, en lo relativo al ámbito valorativo-afectivo, desde nuestro punto de vista es en realidad el objeto el que está en función de la actividad afectiva y no a la inversa.

Incluso la teoría de la *Erregung* (estimulación) está modelada con el objeto como elemento central: es el objeto el que estimula, el que llama

y motiva.²⁷ Cuando se refiere a la estimulación, Husserl utiliza el siguiente esquema para relacionar la valoración del sujeto y la estimulación que viene del objeto valioso (Ov):²⁸

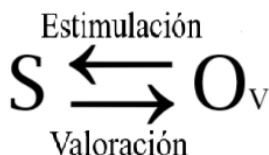


Figura 1. Esquema de la *Erregung* (estimulación) del objeto que motiva la valoración del sujeto

Si el objeto (en tanto que objeto valioso) es el que estimula, entonces Husserl se pregunta cómo es que adquiere esas propiedades “objetuales”. Incluso le adjudica al objeto (en tanto objeto valioso) ser *Wertquell* (fuente u origen del valor) (Hua XLIII/2: 55). La interpretación de Husserl parece, pues, partir del objeto valioso como hecho dado, como primer elemento en el proceso de valoración, a partir del cual emerge y se irradia la estimulación (por ejemplo, desde un “objeto agradable”, la estimulación del agrado) que, al llegar al sujeto, despierta en él el sentimiento correspondiente (por ejemplo, el agrado) gracias al cual este, el sujeto, se dirige valorativamente al objeto (por ejemplo, valorándolo como objeto “agradable”), todo lo cual presenta, como es evidente, varios problemas, como los de la precedencia y la procedencia. Nuevamente, no encontramos en el horizonte interpretativo husserliano

²⁷ Crespo (2012, p. 48) afirma, por ejemplo, que en el caso de la alegría “no se trata propiamente de una dirección de la conciencia al objeto ‘alegrante’. Lo que más bien sucede es que *la alegría procede de este objeto*. Esta alegría tiene su *origen último* no en el mero ser del objeto, sino en su carácter valioso” (énfasis añadido).

²⁸ En Hua XLIII/2: 56, a propósito de la vivencia de alegría o de tristeza, Husserl distingue “una valoración que apunta hacia el ‘objeto’, que se dirige hacia el ‘objeto’ [...] y la estimulación emotiva [*Gefühlserregung*] que se dirige hacia el sujeto, es decir, la alegría y la tristeza mismas que, por una parte, caracterizan al objeto como alegre o triste, pero al mismo tiempo conmueven (*angreift*) al sujeto como: estoy triste, estoy alegre”.

la fundamental *relación* entre el objeto valorado y el estado del sujeto como real principio del proceso de valoración, relación sin la cual ningún objeto podría aparecer como *Wertquell* ni adquirir la capacidad de estimular.

2.3. El problema del tránsito de la intencionalidad sin valor a la con valor

Pues bien, cuando Husserl nos dice que los actos no-objetivantes dan un elemento objetual (son *gegenstandgebende Akte*) a pesar de no dar el objeto (son *nicht objektivierende Akte*), quiere decir que tienen una intencionalidad, porque están dirigidos al objeto previamente dado, aunque no a la manera del objeto puramente teórico, sino en cuanto “valor” o en cuanto “bien”: “un tal acto del ánimo [*Gemütsakt*] posee en sí direccionalidad hacia un objeto, pues tiene conciencia de un objeto por el hecho de fundarse en un acto objetivante, objeto al cual puede luego dirigirse intencionalmente” (Melle, 1988, p. xxxvii).

Basados en lo que hemos visto previamente, esta descripción nos parece ahora ambigua y de dudoso valor explicativo del fenómeno del valor, porque si los actos valorativos se dirigen intencionalmente al objeto “previamente dado”, entonces solo podrían dirigirse a lo que es este objeto: una pura presentación (o representación) sin capas de valor. Pero si, como dice Husserl, se dirigen a este objeto en cuanto “valor” o “bien”, entonces esto implicaría que el objeto ya era valor o bien. La valoración no puede dirigirse sino hacia aquello que (ya) valora. Si nos representamos, nuevamente, el acto intencional con el esquema $C \rightarrow O$, tendríamos que decir que la conciencia valorante C_v no puede dirigirse al mero objeto O . Si ya hay dirección hacia el objeto por parte de C_v , entonces el objeto tiene que aparecer ya como objeto valorado O_v .²⁹

²⁹ Husserl problematiza este mismo punto en Hua XLIII/2: 55-56, constatando cierta tensión en sus explicaciones. Por ejemplo, cuando dice que en la vivencia de la alegría la valoración se dirige hacia el objeto se corrige en nota al pie diciendo: “Esto no es claro. El ‘objeto’ está simplemente ahí como valioso. La reflexión encuentra el valorar” (Hua XLIII/2: 56, n. 1). Por lo tanto, la frase “la valoración se dirige hacia el objeto” no es precisa, porque el objeto ya está ahí como valioso, y por ser ya valioso puede entonces estimular el sentimiento (por ejemplo, de alegría). Por tanto, la pregunta es: ¿cómo llega el objeto a adquirir este carácter de valioso, en primer lugar?

Tenemos, pues, el problema de explicar la relación y el tránsito entre estas dos relaciones intencionales:

a. $C \rightarrow O$

b. $C_v \rightarrow O_v$

Un acto valorante es un acto que está ya dando valor, no es un acto “en tránsito” de dar valor. Por lo tanto, implica que el objeto al que se dirige posee ya valor debido al propio acto valorante. Por eso nos parece que no puede decirse que el *acto valorativo* se dirige intencionalmente al objeto previamente dado en su solo carácter de objeto y, por lo tanto, que *no puede decirse que el objeto intencional del acto valorativo sea el objeto dado por el acto representante*, con lo cual el modelo de actos fundados y fundantes no parece de ayuda para esclarecer la intencionalidad afectiva.³⁰ Obviamente, cuando se juzga que el jardín es agradable, el jardín tiene que estar dado, tenemos que tener conciencia del jardín. El problema aparece con la separación analítica y clasificación de los actos involucrados.

La dificultad para explicar el tránsito entre *a* y *b* se transformó, entonces, en un problema para Husserl. Como afirma Melle (1994, p. 69): “La diferencia entre actos objetivantes y actos no-objetivantes permanece siendo para Husserl un problema fundamental de la descripción de la conciencia”.

Nuestra hipótesis de por qué ocurre esto es doble. En primer lugar, creemos que el esquema que utiliza Husserl para representarse la intencionalidad de los actos valorantes ($C_v \rightarrow O_v$), entendida esta

³⁰ La teoría de las *Empfindungen* referida en la sección anterior tiene también la consecuencia de negarle a los actos no-objetivantes una materia propia, y en consecuencia les obliga a buscar su materia en los actos objetivantes, teniendo que “fundarse” en ellos, de ahí la noción de “intencionalidad” debilitada o secundaria (cfr. Cabrera, 2021). Pero esto es justamente lo que estamos poniendo ahora en cuestión. Esta falta de materia propia de los actos no objetivantes es mencionada por Zhang y Yu: “el acto no-objetivante no tienen una materia propia, por lo tanto, *no tiene una relación directa con un objeto*. [...] Así, el acto no-objetivante debe adquirir materia y relación a un objeto gracias a un acto objetivante, para devenir con ello en un acto de conciencia” (2009, pp. 135-136). Nosotros sostenemos, en cambio, que, al ser actos no solo referidos a objetos sino también al sujeto, los actos valorantes adquieren su materia prima del estado del sujeto.

exclusivamente según un modelo *unidireccional* de la conciencia valorante dirigida hacia el *objeto* valorado, como si en el caso de los actos valorativos la conciencia solo tuviese en consideración el *objeto* que le interesa (por ejemplo, al valorar un libro solo considerase el libro); este esquema, decimos, no da cuenta de la totalidad de los elementos involucrados en el fenómeno en cuestión y, en segundo lugar, tal como lo hemos adelantado más arriba, lo que falta en este esquema es la igual participación del estado del *sujeto* valorante, pero no simplemente en tanto que “sujeto valorante” (porque esto no indica ni explica por qué el sujeto es valorante), sino en tanto que es fundamentalmente un sujeto de *necesidades*.³¹

En consecuencia, en lo que sigue propondremos un nuevo esquema, no binario (sujeto-objeto, conciencia-objeto), sino tripartito, que pone la *necesidad* del sujeto en la función de motor principal que desencadena el proceso de valoración y al yo (a la conciencia) como un mediador entre objeto y necesidades. Con la inclusión de este factor (la necesidad) nos preguntaremos, luego, por la intencionalidad de la conciencia que ya no solo tiene que considerar el objeto valorado sino al mismo tiempo también las necesidades del sujeto. Construiremos nuestro esquema basándonos en las ideas de dos autores muy heterogéneos y, probablemente, inesperados en la discusión de estos temas, a saber, por una parte, en la teoría del valor del economista austríaco Carl

³¹ Hay algunos textos en los que Husserl se refiere explícitamente a las necesidades, por ejemplo, Hua IV: 187-188; Hua XV: 593 ss.; Hua XXXIX: 375-8; Hua XLII: 69, 84-5, 95 y 108, etc. Es interesante notar que todos estos, con la sola excepción del breve pasaje de Hua IV, son textos de los años treinta, con lo cual se puede conjeturar que la relevancia de la relación entre necesidad y satisfacción se hizo más clara para Husserl en su última etapa. No obstante, la mayoría de estos textos están compuestos en gran medida por notas sin mayor desarrollo. Y, en general, en ellos la necesidad es siempre analizada dentro del estudio de los instintos e impulsos y prácticamente nunca está relacionada explícitamente con el problema del valor. Por otra parte, en la primera etapa del pensamiento husserliano hay referencias a la “satisfacción” (satisfacción del interés o de la intención), pero en el sentido de la psicología fenomenológica (por ejemplo, en Hua X: 140-146), que es muy distinto al que aquí nos interesa. Por último, es sintomático también el hecho de que en Hua XXVIII (el volumen dedicado a la ética y explícitamente a la teoría del valor) Husserl no haga referencia alguna a la necesidad (no hay referencias a *Bedürfniss*, *Bedarf* o *Not*, y *Notwendigkeit* solo se usa en su sentido lógico).

Menger (1840-1921) y, por otra, en la filosofía de la voluntad de Ludwig Feuerbach (1804-1872).

3. Naturaleza y relevancia del concepto de “necesidad”

3.1. Menger: necesidad y valoraciones subjetivas. Teoría proyectiva de la valoración

Mephistopheles: Wo fehlts nicht
irgendwo auf dieser Welt?

[Mefistófeles: ¿Quién no carece de algo
en este mundo?].

Goethe, *Fausto* (v. 4889)

En lo que respecta a la teoría del valor del economista austríaco, la siguiente cita nos instala *in medias res*: “el valor de los bienes se fundamenta en la *relación* de los bienes con nuestras necesidades” (Menger, 2016, p. 178; énfasis añadido). De entrada, nos encontramos, pues, con una aproximación distinta a la de Husserl. Este es el núcleo principal de la teoría del valor de Menger: el hacer depender el fenómeno esencialmente de la relación de los objetos valorados con nuestras necesidades personales. Este carácter relacional implica, por tanto, que el valor no tiene su fundamento “en los bienes mismos” (2016, p. 178), es decir, no está en los objetos (pues se trata de un fenómeno “vital”, 2016, pp. 173 y 176); no es, por tanto, *objetivo*, sino que “es por su propia naturaleza totalmente *subjetivo*” (2016, p. 179).³²

Así, pues, el valor no es algo inherente a los bienes, no es una cualidad intrínseca de los mismos, ni menos aún una cosa autónoma, independiente, asentada en sí misma. Es un *juicio* que se hacen los agentes

³² Nos reservamos la duda sobre si se puede afirmar que, aun siendo subjetivo, sea “totalmente” subjetivo (*durchaus subjektiv*). Podríamos interpretar la expresión mengeriana como queriendo decir “totalmente *personal*”, en el sentido de que el juicio de valor debe pasar por la conciencia de la persona y además adecuarse a sus necesidades.

económicos sobre la *significación* que tienen los bienes de que disponen para la conservación de su vida y de su bienestar y, por ende, *no existe fuera del ámbito de su conciencia* (2016, p. 178; énfasis añadido).³³

En Menger, por tanto, el valor no es un fenómeno autónomo, sino uno que surge siempre de forma relativa, por el encuentro entre el sujeto y el objeto, específicamente entre la interpretación que el sujeto hace del objeto en relación con sus propias necesidades. De modo que aquí lo relevante para el fenómeno de valor es el estado de *necesidad* del sujeto.

En su encuentro con el sujeto, según la teoría mengeriana, los objetos del mundo tomados en general pueden adquirir varios sentidos específicos, y en lo relativo al valor estos pasan primero por transformarse en “objetos útiles” (*Nützlichkeiten*) (Menger, 2016, p. 103) cuando *reconocemos* (o creemos reconocer) en ellos una capacidad para satisfacer nuestras necesidades (por lo tanto, una relación causal). Luego, estos objetos reconocidamente útiles se transforman en “bienes” cuando además de este reconocimiento podemos *disponer* de ellos voluntariamente, para llevarlos efectivamente a satisfacer nuestras necesidades. De aquí que Menger indique cuatro propiedades que necesariamente deben tener los bienes para ser tales, a saber: 1) lo más fundamental, que haya una necesidad; 2) que el bien tenga capacidad para dar satisfacción a la necesidad; 3) que el sujeto reconozca o sea consciente de esta capacidad y relación causal, y 4) que haya disponibilidad sobre dichos bienes (cfr. Menger, 2016, p. 104). Con esto tenemos que el objeto ha pasado por tres estadios: primero, como simple objeto del mundo, luego, como objeto útil y, finalmente, como bien. Todo ello determinado siempre en relación con el sujeto y sus necesidades.

Sin embargo, el bien por sí mismo no define aún si ha de tener o no valor. Para esto falta una última relación fundamental: la relación de cantidades, esto es, cantidad (o intensidad) de necesidad y cantidad (disponible) del bien. Las principales relaciones de cantidad son dos: 1) la necesidad es mayor que la cantidad de bienes disponibles ($N > B$) y 2)

³³ No debe distraernos el hecho de que Menger hable aquí de “agentes económicos”, como si el fenómeno del que habla solo incumbiese a las acciones económicas. No es posible dar aquí una explicación completa, pero la teoría mengeriana es una teoría general y no puramente “económica”, en el sentido común de la palabra.

la necesidad es menor que la cantidad de bienes disponibles ($N < B$). De esta manera, si tenemos un tipo de necesidad N y disponemos de un tipo de bienes que la satisfacen B , puede ocurrir que la cantidad de bienes sea mayor que lo que requiere la satisfacción de la necesidad y, por lo tanto, todas las necesidades N quedan con ellos satisfechas, o puede ocurrir lo inverso, que la cantidad de B disponible no sea suficiente y, por tanto, parte de la necesidad N no quede satisfecha. Menger sostiene que es solo en este último caso que aparece el fenómeno del valor; solo aquí B adquiere valor. Y a esta relación de cantidad ($N > B$) la llama “relación económica” y a los bienes que entran en ella, “bienes económicos” (Menger, 2016, p. 150).³⁴

El valor, por tanto, emerge ahí donde hay una necesidad que satisfacer³⁵ y donde los bienes disponibles para satisfacerla no son suficientes (o son escasos). Si una específica necesidad N estuviese cubierta con bienes de sobra y se tuviese conciencia de que seguirá cubierta sin necesidad de realizar ninguna actividad específica para ello (como ocurre, en la gran mayoría de los casos, con el aire que respiramos a diario), entonces ese bien no tendría ningún interés práctico para nosotros (relacionado con la necesidad específica en cuestión: en el caso del aire, con la necesidad de respirar; podría adquirir, por supuesto, otro tipo de interés, por ejemplo, científico) y, por tanto, en relación específica con la necesidad N no tendría valor para nosotros. Pero en el momento que cambie la relación cuantitativa, pasando de $N < B$ a $N > B$, inmediatamente ese bien atrae nuestro interés y adquiere valor (en el Everest o en el fondo del mar, el oxígeno es escaso y, por lo tanto, adquiere valor y se le presta atención y cuidado).

³⁴ Me es imposible aquí explicar en qué medida estos bienes económicos no son solo “económicos” en el sentido usual de la palabra y cómo estas relaciones entre bienes y necesidades determinan un campo de experiencia mucho más amplio que el reducido ámbito de las acciones comúnmente comprendidas como económicas.

³⁵ La palabra “necesidad” (*Bedarf/Bedürfnis*) es ambigua y produce fácilmente muchos malentendidos. Aquí está entendida de manera práctica y en un sentido formal y, por lo tanto, general. Necesidad es todo aquello que implique algún grado de insatisfacción y que, por lo tanto, busque ser satisfecho. No está determinada ni preferentemente limitada por la dimensión material o biológica de la persona.

Si observamos, por lo tanto, la relación fundamental que nos presenta Menger, vemos en ella la participación de tres elementos absolutamente esenciales: un sujeto (con su conciencia), sus necesidades y una cantidad de bienes que pueden satisfacer parte o la totalidad de esas necesidades. Los sujetos hacen un “juicio” sobre la significación (importancia) que los objetos o bienes en cuestión tienen o tendrían para la satisfacción de sus necesidades. Los objetos son vistos a través del lente de las necesidades, y por eso adquieren valor (cuando estas no están cubiertas). El sujeto, el yo, funciona, pues, como una suerte de mediador que conecta objetos y necesidades o, lo que es lo mismo, insatisfacción y potenciales satisfactores.

Ahora bien, ¿qué significa exactamente que los agentes se hagan un *juicio*, como sostiene Menger? Ese juicio es un juicio de relevancia, de apreciación, y como tal ocurre solo en el ámbito de la conciencia del agente. Hay, sin embargo, que tener cuidado aquí de no dejarse llevar por los usos acostumbrados que suelen relacionar fuertemente conciencia y juicio con la actividad teórica. El tomar conciencia de la *importancia* de algo no pasa por un juicio teórico, por un juicio de tipo “vidente”. Se trata más bien de un juicio afectivo que, como tal, no es únicamente un juicio (“B es valioso”, “B es muy importante”, etc.), sino que implica necesariamente el concurso y mediación de afectos. Si nos quemamos la mano con la llama de una vela, el dolor *sentido* no surge por un juicio, sino que es el sentimiento mismo, experimentado de manera directa. El juicio que dice: “la vela es peligrosa” es posterior a la experiencia. Por tanto, en el juicio la parte que corresponde a la “significación” que los bienes disponibles tienen para las propias necesidades es la valoración y, como tal, esta *no es vidente*, o si se quiere, es vidente de otra forma. Adaptando nuestro adagio, diríamos: *se ve el objeto, pero se siente su valor*, porque el sentimiento es el sentido que tenemos para percibir nuestro propio estado interno, por así decir.

Se ve el objeto, pero se siente su valor: se ve el vaso de agua, pero se siente su valor. ¿Cuándo se siente su valor? Cuando tenemos sed. ¿Por qué, entonces, se *siente* su valor? Porque el sentimiento asociado a la valoración aparece en nuestra conciencia cuando hay una necesidad insatisfecha y la necesidad *solo se puede manifestar en nuestra experiencia como sensación o sentimiento*. La necesidad no es una realidad representable, no aparece bajo una forma o figura visual (no hay ni una *Vorstellung* de ella); no se percibe, por lo tanto, como se percibe el vaso. La necesidad es fundamentalmente un sentimiento negativo que busca

satisfacción (justamente por ser negativo), que tiene, por lo tanto, una *fuerza motriz* que se manifiesta a través de la sensación o el sentimiento que provoca. El valor del objeto, por ende, se siente porque en primer lugar se siente la necesidad que hemos asociado a ese objeto como potencial satisfactor de esta. Siento la sed, luego el agua adquiere valor (cuando no es abundante).³⁶ Si no existiese la sed en el mundo, el agua no tendría ningún valor como satisfactor de esta. Sería solo un extraño y abundante elemento químico. Dicho aún en otro giro, la necesidad es lo que *motiva* a la conciencia a hacer la conexión causal con los objetos y, así, a darles valor (cuando, a partir de ahí, toma conciencia o cree tomar conciencia de las condiciones dadas por Menger).

Como hemos dicho más arriba, cuando se trata de la experiencia de afectos, estamos frente a un tipo de conciencia distinto al de la conciencia vidente teórica (contrariamente a lo que sostenía Husserl).³⁷ El elemento que determina la relevancia de algo en un juicio de relevancia es la función afectiva y esta es, comparada con la función teórica, la que determina, por tanto, el valor.

Entonces, tener conciencia de una necesidad *es sinónimo de sentir necesidad*, y este sentimiento tiene un carácter negativo, pues indica un estado de insatisfacción. Por consiguiente, cuando Menger nos dice que el valor surge del juicio que el agente se hace sobre la importancia de los objetos en cuestión (bienes) para su bienestar o satisfacción, el elemento

³⁶ En estricto rigor, no es “el” agua (como un todo) lo que adquiere valor, sino específicamente la porción de agua involucrada en el momento práctico y decisonal de la persona. Pero tampoco es aquella porción de agua sin más, sino solo en relación con la específica necesidad de la persona con la que esta la ha asociado. Además, según lo dicho, la sed, o en general la necesidad, es una condición necesaria pero no suficiente para que emerja el sentimiento de valor: tienen que cumplirse también las condiciones mengerianas. Si además de tener sed tengo certeza de que “jamás se agotarán las cantidades de bienes [que tengo disponibles] para aquella necesidad” (Menger, 2016, p. 153), entonces una porción cualquiera de ese bien (de agua, en este caso) no tendrá valor para mí. Si por alguna razón yo perdiese esa porción de agua (porque se me cae, por ejemplo) no sentiré que he perdido nada valioso, porque tengo conciencia de que tengo a disposición las cantidades de agua que desee para reemplazarlo. Al respecto, ver el análisis de los cántaros de agua que hace Menger (2016, pp. 153 y ss.).

³⁷ “[...] el valorar es una conciencia exactamente [*genau*] en el sentido del percibir o representar, y esto es algo notable” (Hua XXXI: 291).

conector *activo*, por así decir, entre sujeto y objeto es este sentimiento y el motor de ello, como origen, es el propio estado de insatisfacción.

Todo esto nos lo podemos representar sucintamente con el siguiente esquema:

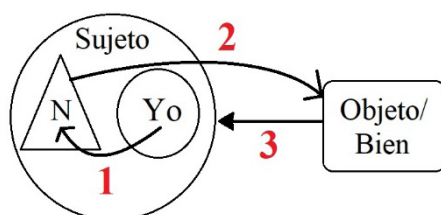


Figura 2. Teoría proyectiva del valor de Menger. El objeto adquiere valor cuando el sujeto proyecta expectativas de satisfacción sobre él a partir de la percepción de sus necesidades

Podemos llamar “proyectiva” a la teoría mengeriana, en el sentido de que lo que hace el sujeto, si nos representamos el proceso de manera diacrónica, es, primero, sentir su necesidad (1) y, luego, *proyectarla* sobre el objeto (2), a raíz de lo cual el objeto adquiere (cumpliendo, además, con las otras condiciones mengerianas) cualidad de bien valioso (3). A todo lo cual hay que agregar un cuarto momento que sería el de la acción del sujeto motivada por el bien. Lo que se proyecta sobre el objeto son, por tanto, las *expectativas de satisfacción*, razón por la cual el estado de necesidad tiene un carácter de *fuerza* tendencial o vectorial. Es decir, la manera en que el sujeto interpreta valorativamente un objeto está pre-dirigida por el propio estado de insatisfacción (tanto en su dirección como en su intensidad o fuerza). De modo, entonces, que la coloración o aura que adquiere el objeto cuando se convierte en un bien valioso se debe a esta proyección de expectativas de satisfacción: el vaso de agua en el desierto *aparece* como teniendo un “aura” enorme, una importancia superlativa, en contraste con el agua que bebemos a diario.

En síntesis, la forma de aparecer el objeto, en cuanto bien valioso, a la conciencia perceptiva o vidente no es el resultado del trabajo directo de este tipo de conciencia, sino el resultado de la mediación previa y necesaria de la afección y del sentimiento de insatisfacción (necesidad).

Esta mediación es *sine qua non* en el fenómeno de transformación de las cosas del mundo en bienes valiosos.

3.2. Feuerbach: necesidad y voluntad individual. Dialéctica del sentimiento

Si Carl Menger tiene el mérito de haber destacado el concepto de “necesidad” en la teoría del valor, Ludwig Feuerbach puede ser considerado como aquel que se introduce en la lógica interna de este concepto. Sus ideas relativas a esto fueron desarrolladas en los escritos sobre ética de su última etapa filosófica.³⁸ En ellos desarrolla una teoría ética y de la acción a partir del concepto de “felicidad” (*Glückseligkeit*) como fin último de la voluntad en general. Para él, voluntad significa esencialmente impulso hacia la felicidad, y felicidad, a su vez, se define no en sí misma sino en relación con su opuesto: el padecimiento, el dolor o la desdicha. Impulso a la felicidad significa, pues, buscar liberarse de un padecimiento, y el dejar de padecer es sinónimo de satisfacción. Esta idea general tiene consecuencias para nuestro tema sobre rol de la necesidad en los actos valorativos.

Partamos, entonces, analizando el siguiente pasaje en el que nuestro filósofo afirma lo siguiente:

[...] el hambre, a causa de una previa abstención o privación de alimento, me tiene en su poder; me fuerza a pensar solo en el estómago. Pero si el hambre es aplacada y, por consiguiente, se vuelve cosa del pasado, entonces tengo ahora tiempo y libertad para pensar en mi cabeza en vez de en mi estómago: *primum vivere deinde philosophari* (primero vivir y luego pensar o filosofar) [...]. [De esta forma], todo tiempo presente de la voluntad [“yo quiero *x*”] presupone un tiempo perfecto de su contrario o de un otro en general [-*x*]. Quiero trabajar o moverme, después de que y porque descansé; quiero descansar, después de que y porque he trabajado o me he movido, esto es, porque no he descansado. Sin este No del pasado no

³⁸ Nosotros nos remitiremos específicamente a su texto titulado “Sobre espiritualismo y materialismo, particularmente en relación con el libre arbitrio” (Feuerbach, 1960, pp. 91-229), escrito entre 1863 y 1866.

tengo fundamento, motivo, impulso para el querer (Feuerbach, 1960, p. 103).

Observemos, pues, que, según esta idea, el querer no se puede definir en sí mismo, sino solo en relación con un estado contrario. Quiero *A* porque *no* he tenido *A*, es decir, he estado viviendo con *-A*.³⁹ La falta o carencia de *A* es condición necesaria de querer *A*. Pero esta carencia no es un hecho teórico, lógico, sino un hecho real que se manifiesta en el ánimo a través de y gracias a un sentimiento. Esto implica una dinámica constante de balances y desbalances de sentimientos opuestos.

En el ánimo, determinado por la percepción sensible-afectiva, hay, por lo tanto, una dinámica de tensión y distensión percibida como una dinámica de sensaciones codeterminantes negativas y positivas. Las vivencias afectivas son las responsables de que se perciban ciertas situaciones como negativas y otras como positivas; son, por así decir, el *sensorium* mediante el cual el ánimo percibe el mundo entorno como pintado con una amplia paleta de tonalidades que llevan en sí el signo positivo o negativo (y sus grados), o —usando otra analogía— el *sensorium* mediante el cual el ánimo percibe el mundo como una topografía axiológica con valles y cimas de diferentes alturas.

Ahora bien, si el determinante de la voluntad y de la valoración es un estado negativo y este es percibido sensiblemente, entonces no puede darse sino bajo las específicas circunstancias de la persona concreta y su contexto específico. Estas circunstancias tienen que interactuar con su sistema afectivo específico. O como dice Feuerbach: “En pocas palabras, el querer se encuentra bajo todas las condiciones y formas de la finitud y de la temporalidad. Al ‘yo quiero’ está inseparablemente unida la pregunta ‘¿qué?’”. Una voluntad separada de la materia de la voluntad es un absurdo” (1960, pp. 110-111). Es decir, no existe *la* voluntad en general, sola, separada, en sí.⁴⁰ Esa expresión es solo un espejismo.

³⁹ Se entiende que a partir del hecho de que sea el caso que *-A* no se deriva necesariamente que entonces se deba querer *A*. Hay muchas cosas que no poseemos y que, sin embargo, no buscamos poseer o directamente no queremos. La deducción funciona solo a la inversa: si quiero *A*, ello significa que no poseo *A*, y este no poseerlo es condición necesaria de mi quererlo.

⁴⁰ Así, de nuevo, Feuerbach (1960, p. 124): “La voluntad real es siempre solo la voluntad de un ser determinado. Lleva una cantidad cualquiera de personas a una librería y déjales escoger un libro a voluntad. Cada una dirá: ‘yo quiero’; pero una dirá: yo quiero esta novela; otra dirá: yo quiero este libro de

Voluntad es impulso-hacia y el fin de ese “hacia”, como impulso de todos los impulsos, es la *Glückseligkeit*, la “felicidad”, es decir, la satisfacción.

Pero entonces ¿qué es lo que quiero? No otra cosa que el fin de lo que me produce aversión [*Widerwilligkeit*], el fin de un mal [*eines Übels*] —donde no hay ningún mal, no hay ninguna voluntad— el fin de un padecimiento [*eines Leidens*] —querer significa *no querer padecer* [*Wollen heisst nicht leiden wollen*]. En fin, no quiero otra cosa sino el no-ser de mi no-ser [...]. Mi voluntad solo es mi ser [*Wesen*] que se opone y contrapone a aquello que me oprime y me hiere, que me hace infeliz, que me arroja al suelo, que me quiere aniquilar. Querer es bienquerer [*Wohltwollen*] (Feuerbach, 1960, p. 111).

Querer es necesariamente querer un bien, mi bien, porque busca eliminar el estado de insatisfacción que el sistema afectivo personal ha detectado y por el cual ha sido motivado. Por lo tanto, el hecho de que padezcamos necesidades, o estados negativos en general, lleva en sí mismo, sostiene Feuerbach, una dirección teleológica natural, el motivo de un impulso con una finalidad específica: padecer (sentir) necesidades o insatisfacciones es el mecanismo por el cual la propia naturaleza de la que estamos constituidos (con todos sus condicionantes y limitaciones) nos indica la dirección en la que podemos encontrar nuestra propia felicidad. Dice Feuerbach al respecto:

“Yo quiero”, no significa otra cosa sino: yo quiero deshacerme en general de un tormento, de una carga, de un mal; quiero ser libre, feliz, aunque sea por un momento. La sed me fuerza [*nötigt mich*] a beber el agua envenenada; pero esta coerción [*Nötigung*] es, al mismo tiempo, solo la voluntad de hacerme feliz, de liberarme de la penuria de la sed. Contra voluntad padezco la sed, pero con voluntad la aplaco: sed es necesidad, pero beber es libertad; padecer sed es el infierno, pero aplacar la sed es el reino de los cielos. El infierno de la naturaleza [...] es enemigo del tormento en la medida

viajes; una tercera dirá: yo quiero esta obra de filosofía; cada una, sin embargo, a través de la diferencia de su voluntad, manifiesta la diferencia de su ser”.

en que este es enemigo del hombre: él quiere terminar, quiere su opuesto, quiere el reino de los cielos; pues la sed misma solo es voluntad de beber y no de padecer sed [*nicht zu dursten*]. [...] La naturaleza no solo me ha agobiado con la necesidad de la sed, sino que también me ha concedido, en las herramientas y medios para aplacarla, el poder liberarme de ella, y con ello la libertad para poder pensar en otras cosas muy distintas de la sed, pues aquel que no tiene agua en el cuerpo tiene solo agua en la cabeza —el sediento solo sueña con beber (Feuerbach, 1960, p. 123).

Se tiene conciencia del padecer sed porque se siente, y se siente con un signo axiológico negativo. Bajo estas circunstancias, debido a la negatividad de este estado, el sistema afectivo se ve estimulado con una *fuerza* que, a su vez, lleva en sí una *dirección*: no es cualquier cosa del mundo lo que calma la sed. Esta fuerza vectorial es, por lo tanto, teleológica: lleva prefigurada en sí misma su finalidad. Y como estamos naturalmente hechos para rechazar lo negativo, lo negativo surge para ser rechazado por nosotros, y esto es lo que mueve a la voluntad. Por lo tanto, en la medida que está implicado un principio teleológico, no se trata tan solo de rechazar lo negativo, sino al mismo tiempo de un esperar y buscar un estado mejor futuro, al menos en comparación con el estado inicial. Ambos aspectos, positivo y negativo, van, pues, esencialmente coordinados.

En conclusión, el sistema afectivo, justamente por ser el único capaz de detectar los estados positivos y negativos mediante el sentimiento, lleva inscrito en sí mismo un sistema dinámico y teleológico natural en la medida en que lo negativo por su propia naturaleza *es* deseo de transformación en su opuesto, en lo positivo y, por lo tanto, quiere pasar, quiere dejar de ser lo que es. Razón por la cual se transforma en la fuente principal de la valoración y en el motor de la acción al indicarle a esta una finalidad y al imprimirle a la voluntad una determinada cantidad de fuerza motriz.

Conclusión. La intencionalidad de la conciencia valorativa

En lo que respecta a la interpretación que hemos ofrecido de la teoría husserliana de los actos afectivos y valorativos cabe recordar que en este trabajo no hemos incluido los análisis de la fenomenología genética,

que buscar pensar a la persona de manera más concreta, como una que vive en un mundo histórico y cultural y que posee herencias, instintos, impulsos, ideas recibidas, hábitos, etc. Tampoco hemos considerado las ideas del último Husserl, el de los años treinta, sobre todo en lo que respecta a las referencias explícitas que aparecen en esta época sobre necesidades y satisfacción (lo cual sería un interesante tema de investigación por sí solo). Por lo cual, ciertamente estamos abiertos a que pueda haber otras interpretaciones distintas de la que hemos dado aquí.

Ahora bien, considerando, no obstante, el análisis que hemos hecho de la teoría husserliana y agregando a ello las ideas de Menger y Feuerbach, podemos extraer las conclusiones que siguen a continuación.

Los actos de valoración surgen a partir de un estado de necesidad que implica necesariamente su manifestación (en el sujeto) a través de un sentimiento, y uno con carácter negativo, respecto del cual lo valorado aparece con un signo afectivamente positivo porque se proyecta en ello la expectativa de satisfacción, es decir, de transformación de aquel sentimiento negativo en uno positivo. De esta manera el fenómeno del valor aparece bajo dos aspectos: 1) como un fenómeno relacional, que requiere necesariamente tanto del estado del sujeto como del objeto, y 2) como un fenómeno fundamentalmente vital, ligado a la condición humana finita, temporal y en constante estado de necesidad (vale decir, no es un fenómeno solo de una esfera particular de la praxis humana).

En un sentido, podemos decir que la necesidad constitutiva del sujeto posee *a priori* una relación valorativa con el mundo. El padecer sed implica *a priori* una relación valorativa con el agua que ofrece el mundo. Y esta conexión vital con el mundo, anterior a toda teoría y, como acabamos de decir, constitutiva del sujeto, implica una disposición (*Einstellung*) fundamental que está constantemente presente y que podemos llamar “necesitativa”. En consecuencia, podemos decir que los *Gemütsakte* son tales *porque son en primer lugar actos necesitativos*.

Desde nuestro punto de vista, por tanto, la conciencia valorante no opera de la misma forma que la conciencia teórica. La teórica, como la perceptiva, se dirige al objeto según el esquema: $C \rightarrow O$. La conciencia, en cambio, que al ver el agua siente deseos de beberla (la valora) y determina en esa dirección la acción, lleva en sí, explícito o implícito, el juicio “tengo sed”, es decir, “estoy en un determinado estado de necesidad”. Esto se expresa o se manifiesta en la realidad mediante el sentimiento, y este, a su vez, está determinado por el menor

o mayor estado de satisfacción del sujeto. Este estado previo tiene que llegar a la conciencia, pero no llega en la forma (privilegiada en los análisis de Husserl) del *ver*, sino del *sentir*. En la dimensión axiológica, la conciencia se dirige al objeto *porque* se encuentra ya en un estado afectivo determinado,⁴¹ ya es, *en su sentimiento*, conciencia de ese estado, aunque no lo sea explícitamente en el sentido de la conciencia visual temática. Dado que el sentimiento no es visual el carácter que imprime en el ánimo con el índice de “deseable” o “agradable” (o su contrario) no se da mediante la forma de la intencionalidad de la conciencia visual. El signo de valor positivo de un objeto aparece cuando proyectamos en él lo que nos indica nuestro estado afectivo al convertirlo en promesa de satisfacción. El valor del objeto es correlativo a mi experimentar (o sentirme) en carencia.⁴²

“Conciencia afectiva” es, pues, una expresión ambigua porque, al ser conciencia, parece exigir visualización del sentimiento (lo que es contradictorio),⁴³ como si se tratase de una conciencia representante con su intencionalidad unidireccional $C \rightarrow O$. Si consideramos un objeto visible, aquel libro que veo sobre la mesa, por ejemplo, ciertamente tengo conciencia de él a través de mi percepción sensible. Pero la conciencia representante no ve lo *interesante* que es el libro. Si, por lo tanto, distinguimos el tipo de conciencia que llamamos “visual” (o “representante”) del tipo de conciencia que llamamos “afectiva”, entonces tenemos que distinguir, correlativamente, una intencionalidad visual y una intencionalidad afectiva. Esta última se diferencia de la visual porque requiere o, más bien, supone aquella relación previa que mencionamos arriba de la conciencia con el propio estado necesitativo o carencial del sujeto.

Por lo tanto, el principal resultado que quisiéramos destacar de nuestro análisis es el hecho de que en los actos valorativos nos encontramos en presencia de dos instancias afectivas y, en consecuencia,

⁴¹ Idea similar a la tesis de Scheler que hemos referido en nota anterior, en la sección 2.2, al menos tal como la presenta Lotz (2006).

⁴² “Carencia” es también aquí (solo) un término relativo: es un menos respecto a un más.

⁴³ Para tratar de dar satisfacción a nuestro deseo de describir visualmente (mediante un objeto de la percepción) el sentimiento utilizamos metáforas como “colorido”, “resplandor”, “iluminación”, “oscuridad”, etc., pero no hay que perder de vista que estas son justamente metáforas.

de una doble intencionalidad: una dirigida a la necesidad (instancia negativa) y otra, basada en esta, dirigida al objeto que ha de adquirir valor (instancia de expectativa positiva o de satisfacción). Según esto, por tanto, el origen de la valoración no es solo el objeto valioso o el valor del objeto (como creemos que afirma la tesis husserliana), sino la combinación mutuamente requerida del objeto con sus características y la necesidad del sujeto. Lo que *ve*, por tanto, la conciencia “dóxica” (por ejemplo, perceptiva) en el objeto son sus características capaces de satisfacer la necesidad sentida. Al ver el agua, veo que sus características, diferentes de las de las piedras o los árboles, prometen ser efectivas en aplacar mi sed. Por esta razón el agua puede entonces ejercer una *Erregung*, una estimulación; no la ejerce por sí misma, desde su sola objetividad, sino por su relación con mi estado de necesidad de ella. Nuevamente, pues: el acto valorativo posee estos dos momentos afectivos: hacia a la necesidad como hacia el objeto.

Es a causa de esta dualidad afectiva que el esquema intencional $C \rightarrow O$ nos parece que no logra describir, en la esfera axiológica, el fenómeno afectivo y, por lo tanto, de valor, porque el sujeto está constantemente mediado por su propia constitución vital (en todos sus niveles: físico, biológico, anímico, espiritual) en tanto que ser de necesidades y, en consecuencia, también lo está su intencionalidad. Valoramos el agua porque tenemos necesidad de ella y así también valoramos estéticamente una obra de arte porque tenemos necesidad de la belleza. En su dimensión general, ambos fenómenos comparten la misma estructura formal.

Por último, creemos que la introducción del concepto de “necesidad” en las reflexiones axiológicas permite adoptar nuevos puntos de vista analíticos para repensar la naturaleza fenoménica de los actos afectivos y su forma peculiar de intencionalidad “invisible”.

Referencias

- Bernet, R., Kern, I. y Marbach, E. (1993). *An Introduction to Husserlian Phenomenology*. Northwestern University Press.
- Breyer, T. (2010). Wahrnehmung. En H.-H. Gander (ed.), *Husserl-Lexikon* (pp. 306-308). Wissenschaftliche Buchgesellschaft.
- Cabrera, C. (2021). La constitución de la conciencia volitiva y sus implicancias para la ética fenomenológica en Husserl. *Anuario Filosófico*, 54(2), 347-375. <https://doi.org/10.15581/009.54.2.005>

- Crespo, M. (2012). *El valor ético de la afectividad. Estudios de ética fenomenológica*. Ediciones Universidad Católica de Chile.
- De Almeida, G. A. (1972). *Sinn und Inhalt in der genetische Phänomenologie E. Husserls*. Martinus Nijhoff. <https://doi.org/10.1007/978-94-017-6538-1>
- Fabbianelli, F. (2010). Wert. En H.-H. Gander (ed.), *Husserl-Lexikon* (pp. 311-313). Wissenschaftliche Buchgesellschaft.
- Feuerbach, L. (1960). *Schriften zur Ethik und nachgelassene Aphorismen*. F. Jodl (ed.). Frommann Verlag.
- Husserl, E. (1939). *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*. Academia Verlagbuchhandlung.
- Husserl, E. (1952). [Hua IV]. *Husserliana. Gesammelte Werke. Band IV. Ideen zur einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch: Phänomenologische Untersuchungen zur Konstitution*. M. Biemel (ed.). Martinus Nijhoff.
- Husserl, E. (1966). [Hua X]. *Husserliana. Gesammelte Werke. Band 10. Zur Phänomenologie der Intersubjektivität. Erster Teil: 1905-1920*. I. Kern (ed.). Martinus Nijhof.
- Husserl, E. (1973). [Hua XV]. *Husserliana. Gesammelte Werke. Band 15. Zur Phänomenologie des inneren Zeitbewusstseins (1893-1917)*. R. Boehm (ed.). Martinus Nijhof.
- Husserl, E. (1976). [Hua III/1]. *Husserliana. Gesammelte Werke. Band III/1. Ideen zur einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie*. K. Schuhmann (ed.). Martinus Nijhoff.
- Husserl, E. (1988). [Hua XXVIII]. *Husserliana. Gesammelte Werke. Band XXVIII. Vorlesungen über Ethik und Wertlehre 1908-1914*. U. Melle (ed.). Kluwer Academic Publishers.
- Husserl, E. (1991). [Hua 1]. *Husserliana. Gesammelte Werke. Band 1. Cartesianische Meditationes und Pariser Vorträge*. S. Strasser (ed.). Kluwer Academic Publishers.
- Husserl, E. (1996). [Hua XXX]. *Husserliana. Gesammelte Werke. Band XXX. Logik und Allgemeine Wissenschaftstheorie. Vorlesungen 1917/18. Mit ergänzenden Texten aus der ersten Fassung von 1910/11*. U. Panzer (ed.). Kluwer Academic Publishers.
- Husserl, E. (2000). [Hua XXXI]. *Husserliana. Gesammelte Werke. Band XXXI. Aktive Synthesen. Aus der Vorlesung „Transzendente Logik“ 1920/1921. Ergänzungsband zu „Analysen zur passiven Synthesis“*. R. Breeur (ed.). Springer.

- Husserl, E. (2002). [Hua Mb IV]. *Husserliana Materialien. Band IV. Natur und Geist. Vorlesungen Sommersemester 1919*. M. Weiler (ed.). Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-010-0429-9>
- Husserl, E. (2004). [Hua XXXVII]. *Husserliana. Gesammelte Werke. Band XXXVII. Einleitung in die Ethik. Vorlesungen Sommersemester 1920 und 1924*. H. Peucker (ed.). Springer.
- Husserl, E. (2008). [Hua XXXIX]. *Husserliana. Gesammelte Werke. Band XXXIX. Die Lebenswelt. Auslegungen der vorgegebenen Welt und ihrer Konstitution*. R. Sowa (ed.). Springer.
- Husserl, E. (2013). [Hua XLII]. *Husserliana. Gesammelte Werke. Band XLII. Grenzprobleme der Phänomenologie*. R. Sowa y T. Vongehr (eds.). Springer.
- Husserl, E. (2020). [Hua XLIII/2]. *Husserliana. Gesammelte Werke. Band XLIII/2. Studien zur Struktur des Bewusstseins. Teilband II: Gefühl und Wert*. U. Melle y T. Vongehr (eds.). Springer. <https://doi.org/10.1007/978-3-030-35926-3>
- Lohmar, D. (2017). Genetische Phänomenologie. En S. Luft y M. Wehrle (eds.), *Husserl-Handbuch. Leben-Werk-Wirkung* (pp. 149-157). J. B. Metzler. <https://doi.org/10.1007/978-3-476-05417-3>
- Lotz, C. (2006). Action: Phenomenology of Wishing and Willing in Husserl and Heidegger. *Husserl Studies*, 22(2), 121-135. <https://doi.org/10.1007/s10743-006-9006-7>
- Marbach, E. (2009). Einleitung. En E. Husserl, *Philosophie als strenge Wissenschaft*. Felix Meiner Verlag.
- Melle, U. (1988). Einleitung. En E. Husserl, *Husserliana. Gesammelte Werke. Band XXVIII. Vorlesungen über Ethik und Wertlehre 1908-1914* (pp. xiii-xlix). Kluwer Academic Publishers.
- Melle, U. (1990). Objektivierende und nicht-objektivierende Akte. En S. Ijsseling (ed.), *Husserl-Aufgabe und Husserl-Forschung* (pp. 35-49). Kluwer Academic Publisher. https://doi.org/10.1007/978-94-009-2427-7_3
- Melle, U. (1994). La fenomenología de la voluntad de Husserl. C. Maldonado (trad.). *Ideas y Valores*, 95, 65-84.
- Menger, C. (2016). *Principios de economía política*. M. Villanueva Salas (trad.). Unión Editorial.
- Parrella, A. (2025). Introduzione. En E. Husserl, *Le emozioni e i valori* (pp. 5-30). A. Parrella (trad.). Morcelliana.
- Peucker, H. (2008a). *Die Grundlagen der praktischen Intentionalität*. [Conferencia]. Lebenswelt und Wissenschaft. XXI. Deutscher

- Kongreß für Philosophie, 15.-19. September 2008 an der Universität Duisburg-Essen, Alemania. http://www.dgphil2008.de/fileadmin/download/Sektionsbeitraege/13-2_Peucker.pdf
- Peucker, H. (2008b). From Logic to the Person: An Introduction to Edmund Husserl's Ethics. *The Review of Metaphysics*, 62(2), 307-325.
- Quepons, I. (2015). Apercepción de valor y tonalidad afectiva: problemas de la fenomenología husserliana de los sentimientos. *Investigaciones Fenomenológicas*, 12, 157-183. <https://doi.org/10.5944/rif.12.2015.29594>
- Quepons, I. (2024). Valor y generatividad: una contribución al estudio de la estructura de la conciencia del valor y las fuentes de la normatividad en la fenomenología de Husserl. *Investigaciones Fenomenológicas. Anuario de la Sociedad Española de Fenomenología*, serie monográfica: Anthony J. Steinbock. Fenomenología y generatividad (9), 167-185. <https://doi.org/10.5944/rif.9.2024.41544>
- Rizo-Patrón Lerner, R. (2024). Emotions, Science and Generativity: A Husserlian Perspective. *Investigaciones Fenomenológicas*, 9, 187-225. <https://doi.org/10.5944/rif.9.2024.41576>
- Zhang, W. y Yu, X. (2009). The Foundation of Phenomenological Ethics: Intentional Feelings. *Frontiers of Philosophy in China*, 4(1), 132-142. <https://doi.org/10.1007/s11466-009-0008-3>
- Zirión Quijano, A. (2024). *Sobre el colorido de la vida y la fenomenología de lo inefable*. Sb Editorial.

<https://doi.org/10.21555/top.v760.3332>

Descartes, un comienzo absoluto. La génesis del pensamiento moderno en la lectura heideggeriana de las *Regulae* de Descartes

Descartes, an Absolute Beginning: The Genesis of Modern Thought in the Heideggerian Reading of Descartes' *Regulae*

José Francisco Ramírez Olvera
Benemérita Universidad Autónoma de Puebla
México
jose.ramirezolv@alumno.buap.mx
<https://orcid.org/0009-0003-6482-7330>

Recibido: 18 - 12 - 2024.

Aceptado: 10 - 04 - 2025.

Publicado en línea: 10 - 08 - 2026.

Cómo citar este artículo: Ramírez Olvera, J. F. (2026). Descartes, un comienzo absoluto. La génesis del pensamiento moderno en la lectura heideggeriana de las *Regulae* de Descartes. *Tópicos, Revista de Filosofía*, 76, 475-509. <https://doi.org/10.21555/top.v760.3332>



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

Resumen

Aunque la investigación heideggeriana ha estudiado ampliamente la influencia de Aristóteles, la figura de Descartes no ha recibido una atención equivalente, pese a la publicación casi completa de la *Gesamtausgabe* y a la abundante documentación disponible. Desde sus primeras lecciones en Friburgo y Marburgo, Heidegger sostiene una confrontación interpretativa con el pensamiento de Descartes, la cual se prolonga incluso hasta los *Seminarios de Zollikon* en los años cincuenta y sesenta. A diferencia de Aristóteles, cuya presencia es afirmativa y formativa, Descartes aparece como una figura contrastante, lo que quizá explique el relativo desinterés académico por este vínculo. Este artículo examina el inicio del pensamiento moderno a partir de la interpretación heideggeriana de las *Reglas para la dirección del espíritu*, un texto que atraviesa buena parte de sus reflexiones sobre Descartes. Esto contribuye a un estudio más preciso de su relación y ofrece nuevos impulsos para la investigación especializada.

Palabras clave: Heidegger; Descartes; metafísica; ontología; modernidad; subjetividad; *mathesis universalis*; *Regulae*; método; confrontación.

Abstract

Although Heideggerian research has extensively studied the influence of Aristotle, the figure of Descartes has not received equivalent attention, despite the near-complete publication of the *Gesamtausgabe* and the abundance of available documentation. From his earliest lectures in Freiburg and Marburg, Heidegger maintains an interpretative confrontation with the thought of Descartes, which extends even to the *Zollikon Seminars* of the 1950s and 1960s. Unlike Aristotle, whose presence is affirmative and formative, Descartes appears as a contrasting figure, which may explain the relative academic neglect of this connection. This paper examines the beginning of modern thought through Heidegger's interpretation of the *Rules for the Direction of the Mind*, a text that permeates a significant portion of his reflections on Descartes. In doing so, it contributes to a more precise study of their relationship and offers new impulses for specialized research.

Keywords: Heidegger; Descartes; metaphysics; ontology; modernity; subjectivity; *mathesis universalis*; *Regulae*; method; confrontation.

Toda filosofía esencial sigue su destinación obedeciendo a la ley de un designio que la compone y ensambla: por ejemplo la *Crítica de la razón pura* de Kant, los tratados breves de Leibniz y sus cartas y compilaciones en parágrafos, las *Regulae* de Descartes, los *Diálogos* de Platón.

Heidegger (GA 95, p. 389; 2017, p. 328)¹

No es la victoria de la ciencia lo que caracteriza a nuestro siglo XIX, sino la victoria del método científico sobre la ciencia.

Nietzsche (FP IV, 15 [51]; 2008, p. 646)

1. Introducción

No sorprende que Heidegger jamás haya visto en Descartes un aliado en su pensamiento; de hecho, la interpretación más común y extendida en la literatura especializada tiende a presentarlo como un pensador anticartesiano, o incluso a caracterizar la relación entre ambos como una suerte de enemistad irreconciliable. Sin embargo, lo interesante a destacar aquí es que, pese a esa supuesta “enemistad” u “hostilidad”,² la

¹ Algunas aclaraciones bibliográficas. La *Obra integral* (*Gesamtausgabe*) de Martin Heidegger será citada en el presente artículo siguiendo el formato canónico: se utilizará la sigla GA (*Gesamtausgabe*), acompañada del número del tomo y la página correspondiente en la versión original en alemán. Si existe una traducción al español, se indicará, tras un punto y coma, el año de la edición y el número de página de la versión traducida. De manera similar, y solamente en el caso de las *Regulae*, se empleará la abreviatura AT, correspondiente a la edición completa de la obra de René Descartes, seguida del número de tomo y la página específica. Si hay traducción al español, se señalará de la misma manera: punto y coma, año de edición y número de página. El epígrafe de Friedrich Nietzsche seguirá la misma convención: se usará la abreviatura FP (*Fragmentos póstumos*), seguida del número de volumen y, posteriormente, el número del fragmento, según la edición crítica dirigida por Diego Sánchez Meca. Cuando no existan ediciones al español, las traducciones correrán por nuestra cuenta.

² Emmanuel Faye (2005) sostiene que la actitud de Heidegger hacia Descartes es abiertamente hostil, y que dicha hostilidad desfigura

figura de Descartes reaparece en distintos —y decisivos— momentos de la obra heideggeriana: a partir de los seminarios y lecciones tempranas, especialmente durante los años de Friburgo y Marburgo, pasando por el proyecto de la ontología fundamental, hasta el desarrollo del pensar histórico del *Ereignis* (es decir, el pensamiento *ontohistórico* posterior a la *Kehre*). Esta presencia se extiende incluso a sus reflexiones tardías, particularmente en torno a la superación de la metafísica, la pregunta por la técnica y la cuestión del *Gestell*. En suma, Descartes acompaña a Heidegger casi en todo su camino filosófico.

Tal vez no sea un dato tan conocido que, durante su periodo como profesor asociado (*Privatdozent*) en la cátedra I de Filosofía en la Universidad de Friburgo, entre 1919 y 1923, Heidegger llevó a cabo un par de seminarios dedicados a Descartes: en el semestre de verano de 1919 (*Einführung in die Phänomenologie im Anschluss an Descartes Meditationes*) y en el semestre de invierno de 1920/21 (*Phänomenologische Übungen für Anfänger im Anschluss an Descartes Meditationes*). Posteriormente, ya como profesor en la Universidad de Marburgo, ofreció un tercer seminario similar en el semestre de verano de 1925 (*Anfangsübungen im Anschluss an Descartes Meditationes*), igualmente centrado en el estudio de las *Meditaciones* (cfr. Xolocotzi Yáñez, 2011, pp. 60, 64, 67 y 105). En su regreso a Friburgo, durante el semestre de invierno de 1929/30, dirigió un cuarto seminario (*Über Gewissheit und Wahrheit im Anschluss an Descartes und Leibniz*), esta vez en torno a la certeza y la verdad según Descartes y Leibniz.

Se trata, así, de cuatro seminarios dedicados expresamente al pensador francés, además de las lecciones, conferencias, anotaciones y otros ejercicios que han sido conservados y publicados. Hasta donde tengo conocimiento, dichos seminarios no figuran en la *Gesamtausgabe*, por lo que carecemos de registros precisos sobre su contenido. De

deliberadamente su interpretación. En sus palabras: “La presentación de los primeros párrafos del curso del semestre de verano de 1933 nos permite concluir que Heidegger, al atacar a Descartes, no se entrega a una reflexión filosófica seria, sino que manifiesta únicamente una hostilidad radical y de principio, la cual se arraiga en el hecho de que, al apoyarse en el yo y en la conciencia, Descartes no puede sino obstaculizar la “cuestión fundamental” de la “filosofía” según Heidegger, que concierne al pueblo alemán en su destino y su “decisión” históricos, en el “movimiento” que lo conduce a crear el nuevo Estado nacional-socialista alemán” (p. 284).

acuerdo con el testimonio de Emmanuel Faye, quien, como es sabido, dista mucho de ser un simpatizante de Heidegger:

Quisimos saber si la edición de estos seminarios estaba prevista dentro de la llamada edición “integral” (*Gesamtausgabe*). Consultando al respecto, el profesor F.-W. von Herrmann nos respondió que, lamentablemente, los manuscritos de los seminarios sobre Descartes faltaban (carta al autor del 2 de mayo de 2002). Sin embargo, pudimos constatar, en agosto de 2002, durante investigaciones en el *Deutsches Literaturarchiv de Marbach* (donde se encuentra el archivo Heidegger), que existía una transcripción manuscrita de al menos dos de estos seminarios, pero que ambos manuscritos estaban *gesperrt*, es decir, no consultables antes del año 2026, salvo con una autorización expresa de Hermann Heidegger. El hecho de que no se haya programado la publicación de estos dos seminarios sobre Descartes en la edición llamada, erróneamente, integral, no se debe, por tanto, a una imposibilidad material, sino a una exclusión deliberada (Faye, 2005, p. 282).

Sea como fuere, lo verdaderamente significativo es que, contra las apariencias, Descartes ocupa un lugar notable en el itinerario filosófico de Heidegger. No hay otro pensador, que no sea alemán o griego, al que haya dedicado tantas menciones y espacios de reflexión. Como observa Christophe Perrin (2010, p. 141), Descartes está presente en el pensamiento de Heidegger desde los primeros seminarios hasta sus últimos escritos, y constituye una excepción en su *corpus*, al ser el único filósofo francés que figura como objeto central de múltiples cursos y seminarios. Con todo, el filósofo de La Flèche nunca llega a desempeñar, pese a toda esta atención, el papel de interlocutor afín para el pensador de Friburgo.

Este trato contrasta drásticamente con el lugar que ocupa Aristóteles en el pensamiento heideggeriano, por no hablar de la abundante producción especializada que ha suscitado dicha relación. Hoy tenemos una amplia documentación que permite afirmar que su figura es indispensable para comprender el despliegue del pensamiento heideggeriano, especialmente en la década de 1920. No se trata sólo del número de páginas que Heidegger le dedica —pues Descartes

bien podría ocupar, si no un lugar equivalente en cantidad, sí uno significativamente más destacado del que habitualmente se le reconoce—, sino de la influencia de su pensamiento, metodológico y temático, para la gestación, elaboración y consolidación del proyecto de la ontología fundamental. En Aristóteles, Heidegger encuentra un verdadero interlocutor filosófico, cuya obra es objeto de una apropiación creativa, y en ocasiones incluso violenta, que resulta harto evidente. Basta señalar, con la documentación que hoy día disponemos, que la pregunta por el sentido del ser, directriz de *Ser y tiempo* en 1927, prácticamente hunde sus raíces en los primeros ejercicios de Heidegger sobre el *De anima* de Aristóteles en 1921.³ Por esto, señalamos, Aristóteles se presenta como un verdadero aliado para el maestro de Meßkirch.

Con Descartes, la relación es muy distinta. Aunque ocupa un lugar en muchos momentos cruciales de la obra de Heidegger, su presencia está caracterizada más bien por la *confrontación* (*Auseinandersetzung*).⁴

³ “No es difícil de hallar la razón: *De anima* es el texto aristotélico dedicado al fenómeno de la vida [...] y lo que Heidegger buscaba en Aristóteles a principios de la década de 1920 era precisamente una *ontología de la vida*. Esto también explica por qué el seminario de 1921 estudia a *De anima* junto con una lectura de los libros centrales de la *Metafísica* sobre la οὐσία. En efecto, en esta lectura de la *Metafísica* junto con *De anima*, podemos ver una anticipación de la metodología misma de *Ser y tiempo*: abordar la cuestión del ser a través de un análisis del ser, no de la vida en general, sino de la vida humana” (Gonzales, 2024, p. 26). Otro trabajo importante del mismo autor en el que se aborda justamente el origen de *Ser y tiempo* a partir de la lectura del *De anima* es “El nacimiento de *Ser y tiempo*: la lectura crucial de Heidegger de 1921 del *De anima* de Aristóteles” (cfr. González, 2022, pp. 543-574). En los últimos años, la literatura especializada ha renovado el interés por el desarrollo filosófico temprano de Heidegger, utilizando como clave interpretativa los ejercicios iniciales sobre el *De anima* de Aristóteles. Estos estudios se han visto enriquecidos gracias al hallazgo de apuntes inéditos correspondientes al seminario de 1921, elaborados por dos de sus alumnos, Helene Weiss y Oskar Becker. Este descubrimiento ha permitido reconstruir con mayor precisión el camino que llevó a la formulación del proyecto de la ontología fundamental en *Ser y tiempo* de 1927. Entre las contribuciones destacadas que abordan este periodo temprano del pensamiento de Heidegger, cabe mencionar los trabajos de De Brasi (2024, pp. 141-159), Shiffman (2024, pp. 162-181) y Jimenez (2024, pp. 183-211).

⁴ El término alemán *Auseinandersetzung* puede traducirse como “enfrentamiento”, “disputa”, “forcejeo”, “polémica”, “altercado” o incluso “pleito”. Sin embargo, el sentido de esta caracterización entre ambos autores no es trivial ni

Para él, Descartes representa ciertamente un punto de inflexión en la historia del pensamiento occidental, un alejamiento de la tradición griega y escolástica hacia una orientación que coloca al hombre —ahora entendido como sujeto (*subiectum*)—, en el centro de la totalidad de lo ente en tanto que *fundamentum absolutum et inconcussum*,⁵ marcando así el inicio de un determinado modo de pensar que, a ojos de Heidegger, culminará irremediabilmente en la técnica contemporánea que impera en nuestros días, cuya muestra más evidente la encontramos en los distintos mecanismos de dominio total sobre lo ente en su conjunto.

En efecto, con la consolidación de la subjetividad, es decir, cuando el hombre asume por vez primera el rol de sujeto en sentido pleno, todo lo ente que aparece a su alrededor queda asegurado y determinado en su ser por referencia al propio sujeto o, tal vez mejor dicho, por su refracción hacia el sujeto. De este modo, el ente se convierte, por primera vez, en objeto en la re-presentación (*Vorstellung*) del sujeto; no en el sentido de una mera producción mental, sino que “representar quiere decir traer ante sí eso que está ahí delante en tanto que algo situado frente a nosotros, referirlo a sí mismo, al que lo representa y, en esta relación consigo, obligarlo a retornar a sí como ámbito que impone las normas” (GA 5, p. 84; Heidegger, 2010, p. 75). Este cambio otorga al hombre

meramente sugerente, sino que posee también un valor metódico en el modo en que Heidegger accede a lo pensado por Descartes. En efecto, la relación es ciertamente confrontativa, en el sentido de una disputa o lucha —sobre todo si se la compara con la relación más “sana” o cordial que mantiene con Aristóteles. No obstante, la confrontación (y más precisamente: la confrontación histórica) constituye para Heidegger la estrategia hermenéutica adecuada para abordar a los pensadores de la tradición metafísica, y en este caso particular, a Descartes como figura inaugural de la modernidad. Por esto, el término “confrontación” se entiende aquí en un doble sentido: por un lado, como un estado de enfrentamiento o lucha entre los pensadores; por otro, como una estrategia hermenéutica para acceder a lo pensado en la tradición filosófica.

⁵ “Verdad no significa aquí, pues, patencia de lo directamente presente; la verdad es caracterizada como aquello que puede ser constatado de forma clara, indudablemente segura, y esto quiere decir certera, para un yo que representa. El criterio de esta verdad en cuanto *certeza* es aquella evidencia que alcanzamos cuando, después de eliminar todo lo de alguna forma dubitable, nos topamos con aquello indubitable que debe ser reconocido como *fundamentum absolutum et inconcussum*, como fundamento absoluto e inquebrantable” (Heidegger, 2013b, p. 173).

moderno no sólo la certeza y seguridad necesarias para emprender el nuevo proyecto científico, sino que además establece las bases de un nuevo paradigma, en el cual la matematización de la naturaleza queda determinada por la supremacía del método, entendido este último como un proceder ordenado, regulado y universal que antecede y, por ello, condiciona tanto la aparición como la comprensión de todos los objetos de estudio: *mathesis universalis*.

Para Heidegger, tanto la ciencia como la técnica moderna encuentran su origen en esta reestructuración, cuyas consecuencias o, menos alarmante, resultados yacen allí donde la totalidad de lo ente se redefine como aquello que puede ser calculado, manipulado y medido, reduciendo al mundo, y al propio hombre, a un conjunto de recursos disponibles o meras existencias (*Bestände*), como leemos en la famosa conferencia “La pregunta por la técnica” de 1949 (cfr. GA 7, p. 28; Heidegger, 1992, p. 26). Este viraje no sólo implica una reorganización de la relación entre el hombre y el mundo, sino también una transformación esencial en la concepción de la verdad; se produce el paso de la *αλήθεια* (desocultamiento) a la *certitudo* (certeza), viraje que remite a la instauración cartesiana del método como medida de lo verdadero. En este sentido, Heidegger concibe a Descartes como un *comienzo absoluto* que traza los cimientos del pensamiento moderno. Justamente por este carácter fundacional, la figura cartesiana representa para él el polo necesario de toda confrontación filosófica.

No obstante, la *scholarship* heideggeriana suele centrarse, las más de las veces, en los efectos (o consecuencias) del pensamiento moderno inaugurado por Descartes según la lectura de Heidegger, como los recién comentados, antes que del estudio cuidadoso de su presencia en la obra heideggeriana. Trayendo a cuentas otra vez la comparación con Aristóteles, lo que hace falta —o al menos hasta donde tengo conocimiento—⁶ es una reflexión tal vez más rigurosa, y hasta

⁶ *Peccata minuta*: al momento de redactar este artículo, el único estudio relevante que pude encontrar dentro de esta línea de investigación —y que, sin duda, ha influido en las reflexiones aquí desarrolladas, a las que remitiré en más de una ocasión— fue el de Mathias Flatscher (2005). No obstante, gracias a las observaciones de los evaluadores de este texto, tuve conocimiento de cierta obra de Christophe Perrin (2013a) derivada de su tesis doctoral en la Universidad Paris-Sorbonne. Esta obra ofrece una interpretación exhaustiva y brillante de la recepción cartesiana en Heidegger, ¡precisamente aquello que buscaba y cuya

sistemática, y por qué no escolástica, sobre el papel que ocupa Descartes en el camino del pensamiento del maestro de la Selva Negra, así como se ha hecho, y se continúa investigando actualmente, destacadísimo en el caso del Estagirita.

ausencia lamentaba! Al seguir investigando, me encontré también con otros dos artículos del mismo autor que enriquecen la comprensión sobre el lugar determinante, aunque ciertamente reservado y cargado de prejuicios por parte de los especialistas, que ocupa Descartes en la obra de Heidegger. Por ejemplo, Perrin (2013b) aborda la lectura heideggeriana del *cogito* cartesiano reconociendo, además, el hecho de que Heidegger nunca dejó de reflexionar a lo largo de su obra sobre la célebre proposición *cogito, ergo sum*. Esto revela no solo su interés, sino también la dedicación sostenida que prestó al pensamiento del filósofo francés. Más cercano aún a lo que aquí se pretende es el artículo de Perrin (2010) donde se profundiza en el enfrentamiento filosófico entre Heidegger y Descartes. En dicho texto, Perrin examina la supuesta postura “anticartesiana” de Heidegger y la noción de *Auseinandersetzung* en su relación con Descartes y otros pensadores de la tradición, es decir, como una forma esencial de confrontación o encuentro crítico con el pensamiento del otro. El autor desmonta las lecturas tradicionales que han reducido la relación Heidegger–Descartes a un simple conflicto irreconciliable, razón por la cual Heidegger ha sido con frecuencia caracterizado como un pensador intrínsecamente anticartesiano. Perrin considera esta interpretación como un claro prejuicio. Esta es, de hecho —como también en nuestra lectura—, una de las razones principales por las que la investigación especializada ha tendido a desestimar el papel de Descartes en el pensamiento heideggeriano, incluso en contra de la abundante documentación disponible. Como señala el autor, “las frecuentes referencias que se le hacen —su nombre aparece nada menos que 57 veces en *Sein und Zeit* y está presente en 69 de los 79 tomos publicados hasta la fecha de la *Gesamtausgabe*— dan testimonio indiscutible de su presencia en el *Denkweg* del maestro de Messkirch y de su importancia en la historia de la filosofía o, si se prefiere, en la historia del ser” (2010, p. 141). Por mi parte, una obra hallada y digna de mención es la de la filósofa Sidonie Kellerer (2013), en la que se ofrece una rigurosa investigación sobre la recepción de Descartes en Alemania centrada en tres corrientes principales: el neokantianismo de Marburgo, la fenomenología de Husserl y el pensamiento de Heidegger. Y lejos de contradecir el reclamo que aquí sostengo —y me permito señalarlo en primera persona—, estas investigaciones, particularmente las de Perrin, no refutan la falta de atención a la relación Heidegger–Descartes, sino que más bien confirman la necesidad de profundizar aún más en el estudio en torno a la presencia del filósofo de La Flèche en el camino heideggeriano. Por ello, a lo largo del texto incluiré algunas referencias a estos trabajos, con el objetivo de contribuir al enriquecimiento y profundización de lo aquí pensado.

Mas la razón de esta especie de desdén o, mejor dicho, poca atención al problema quizás provenga del mismo Heidegger. En diversas ocasiones, este tacha de ingenuas o ficticias (*scheinbar*) las consideraciones que ven en la imagen de Descartes el comienzo de la modernidad, ignorando las fuentes medievales y escolásticas de las que Descartes, sabemos bien, tanto se nutre en su filosofía. En el curso de verano de 1933 escribe:

Este supuesto nuevo comienzo de la filosofía moderna con Descartes no solo no existe, sino que en verdad marca el inicio de una caída esencial aún más profunda de la filosofía. Descartes no devuelve a la filosofía a sí misma ni a su fundamento, sino que la aleja aún más de la pregunta fundamental (Heidegger, GA 36/37, p. 39).

Esta denuncia se expresa en contra de la imagen habitual de Descartes como aquel pensador que un día, al despertar temprano por la mañana, empezó a dudar de todo y fundó el comienzo del pensamiento moderno, desprendiéndose de las ataduras de la tradición medieval y escolástica (y griega, desde luego). Leemos en el mismo curso: “El radicalismo de la duda cartesiana y el rigor de la nueva fundamentación de la filosofía y del conocimiento en general es una farsa y, por tanto, la fuente de engaños fatales que aún hoy son difíciles de erradicar” (GA 36/37, p. 39). De este modo, lo ingenuo es, en todo caso, creerse la “mala novela” (*schlechter Roman*) —como describe Heidegger en otro momento—⁷ de que el comienzo de la modernidad tiene lugar con el radicalismo cartesiano, esto es, con el simple hecho de ponerse a dudar de todo hasta no poder dudar de quien duda, el yo o el *ego cogito*, desembocar por tanto en el sujeto, volverse subjetivista, y a partir de entonces instituir un nuevo paradigma de pensamiento, y tras ello la modernidad toda. Por desgracia, la historia es harto más compleja.

Según nuestro modo de ver, la clave o, si se quiere, la intención principal de la confrontación heideggeriana con Descartes yace fundamentalmente en la preocupación por comprender con mayor cabalidad el comienzo del pensamiento moderno. Justamente, situar

⁷ “Esta historia de Descartes de que llegó, dudó y con ello se convirtió en un subjetivista y así fundó la teoría del conocimiento moderna, se corresponde, es verdad, con la imagen usual. Sin embargo, no es más que, como mucho, una mala novela —pero en ningún caso una historia en la que se haga visible el movimiento del ser” (GA 41, p. 100; Heidegger, 2009b, pp. 129-130).

históricamente a “*Descartes en tanto que un comienzo absoluto*” (Heidegger, GA 76, p. 197; énfasis de Heidegger).

Ahora bien, y en contraposición a la visión acostumbrada al respecto, Heidegger curiosamente identifica este comienzo no en las obras de mayor renombre como lo son el *Discurso del método* o las *Meditaciones metafísicas*, sino en la breve e inconclusa (y por un tiempo hasta perdida) obra de juventud *Reglas para la dirección del espíritu* (*Regulae ad directionem ingenii*)⁸ de 1628, la cual, vale decir, las más de las veces se subestima o subsume bajo la lectura de las demás obras de madurez, es decir, leyendo las *Regulae* desde las *Meditaciones* o el *Discurso*, cuando —y esto ya ha sido explorado por varios pensadores—⁹ debería ser todo lo contrario,

⁸ A partir de este punto, nos referiremos a ella simplemente como *Regulae*.

⁹ Jean-Luc Marion (2008) es, quizás, el autor más conocido que persigue esta lectura de Descartes en su célebre obra *Sobre la ontología gris de Descartes: ciencia cartesiana y saber aristotélico en las Regulae*. En ella señala: “Una tentativa seductora ha ocupado a gran número de intérpretes: leer las *Regulae* a partir del *Discurso del método*. Si se supone de entrada que la ‘entera filosofía de Descartes se encuentra contenida como en compendio en el *Discurso del método*’, no queda más que concluir, lógica pero anacrónicamente, como hace Hamelin, que ‘la segunda obra (el *Discurso*) es el texto, la primera (las *Regulae*) el comentario del método cartesiano’; con Gilson: que ‘en definitiva las *Regulae* no contienen nada más que los cuatro preceptos del *Discurso*, y que lo que le añaden no consiste en preceptos suplementarios, sino reglas prácticas destinadas a facilitar su aplicación’; con Milhaud: que ‘las *Regulae ad directionem Ingenii* nos hacen conocer el detalle de éste (del método),... el *Discurso* desarrolla lo esencial en unas pocas reglas bien conocidas’; con N. K. Smith: que ‘son útiles, por provisional que sea su enseñanza’. A parte del desprecio por la lectura cronológica, en este modo de pensar se vuelve problemático todo lo que las *Regulae* ‘añaden’ por su parte con respecto al *Discurso* [...]. El *Discurso* sólo da cuenta de lo que le llega de las *Regulae*, no de éstas en cuanto tales. De ahí el ensayo de leer, al contrario, el *Método* a partir de las *Regulae* [...] entonces hay que pedirle a las *Regulae* que nos introduzcan en lo que para Descartes pueda querer decir método” (Marion, 2008, pp. 25-27). Sin embargo, resulta por lo menos curioso que esta estrategia de lectura haya sido propuesta mucho tiempo antes por el neokantiano Paul Natorp (1882). En el apartado “Die Grundabsicht der Metaphysik Descartes’ im Verhältniss zu den ‘Regeln’” (algo así como “La intención fundamental de la metafísica de Descartes en relación con las ‘Reglas’”) leemos justamente lo siguiente: “De ello se deduce que Descartes, cuando emprendió la empresa de su metafísica, registró conscientemente los rasgos fundamentales del método presentado en las *Reglas*. Por lo tanto, podemos esperar con certeza encontrar

aquellos más bien entendidos a partir de lo establecido en la obra de juventud. Para Heidegger, como se desprende del epígrafe que abre este artículo, la filosofía esencial de Descartes debe someterse a la ley del designio que la configura; tal designio se identifica, inauguralmente, con la palabra contenida en las *Regulae*.

En este panorama, el objetivo de este trabajo es dilucidar la génesis del pensamiento moderno a partir de la lectura que Heidegger realiza de las *Regulae* de Descartes. Primero, se analizará la confrontación entre ambos pensadores para establecer un marco interpretativo que permita situar históricamente la perspectiva sobre Descartes. Posteriormente, se presentarán los elementos fundamentales para comprender el inicio del pensamiento moderno en las *Regulae*, centrándonos especialmente en la idea de la *mathesis universalis*. Esto implicará examinar cómo se configura la necesidad del método matemático como medio para la unificación de las ciencias y el aseguramiento del camino hacia la búsqueda de la verdad. A partir de esta línea de análisis, se espera ofrecer algunos impulsos y consideraciones que, cuando menos, resulten pertinentes para el estudio de la recepción heideggeriana del filósofo de La Flèche.

2. Heidegger y Descartes: el sentido de la *confrontación*

Es quizá una opinión compartida que las interpretaciones heideggerianas de la tradición filosófica, particularmente de ciertos autores, suelen ser calificadas de forzadas o incluso violentas. En ocasiones, se omiten aspectos relevantes del pensamiento original, mientras que se incorporan —a veces sin explicitarlo del todo— elementos que se ajustan a la propia reflexión de Heidegger. Un ejemplo paradigmático de esta estrategia hermenéutica es su lectura de Aristóteles, como hemos señalado; y en esta misma línea podrían incluirse sus interpretaciones de Kant, Nietzsche o incluso Hölderlin. Sin duda, una evaluación superficial del método interpretativo de Heidegger podría sostener que sus lecturas de la tradición no solo son controversiales, sino también filológica y etimológicamente cuestionables. En distintos grados, los comentarios provenientes de círculos académicos más

que los puntos de vista sobre la esencia del conocimiento en los que se basa este método continúan en la metafísica de Descartes” (Natorp, 1882, p. 29). Sin entrar en mayores polémicas, resulta por lo menos relevante destacar aquí la atención que varios autores de renombre han dedicado al breve tratado de juventud del filósofo francés. Cfr. también al respecto Natorp (1896).

tradicionales y apegados al canon suelen mirar con cierto desdén tales interpretaciones, sin tener realmente un conocimiento profundo de los métodos o, mejor dicho, de las estrategias utilizadas por Heidegger para interpretar a los autores de la tradición.¹⁰

Debido a esto, su lectura de Descartes no constituye una excepción a este juicio; menos aún si se considera que el francés nunca figuró, para Heidegger, como un aliado. Por poner un ejemplo, Emmanuel Faye llega incluso a afirmar que “hoy en día es necesario liberar el pensamiento de la influencia destructiva de la ‘hermenéutica’ heideggeriana, para que sea posible volver a meditar libremente sobre la fecundidad de la contribución de Descartes a las futuras investigaciones filosóficas” (Faye, 2005, p. 301). Y aunque ciertamente la relación de Heidegger con Descartes es, en lo esencial, confrontativa —en marcado contraste con su diálogo más fértil con Aristóteles—, dicha confrontación no se reduce a una mera oposición o conflicto, sino que constituye una *estrategia hermenéutica* fundamental en su abordaje de la tradición metafísica.

Heidegger reflexiona en varios momentos sobre el sentido de la confrontación (*Auseinandersetzung*) como modo de acceso a la tradición. Mas no nos detendremos aquí en un comentario exhaustivo al respecto, sino que nos limitaremos, en las siguientes líneas, a ofrecer algunas indicaciones generales.

En el texto *Meditación*, por ejemplo, leemos que la confrontación es el modo en que el pensar debe medirse con la historia de la filosofía, con aquello legado por la tradición y, especialmente, con las posiciones fundamentales de los pensadores esenciales (cfr. GA 66, p. 68-80; Heidegger, 2006b, pp. 70-78). Entre ellos se encuentra, de manera destacada, la figura de Descartes. El objetivo de este enfrentamiento no consiste tan solo en oponerse a dichas posiciones —como suele entenderse en el uso habitual del término “confrontación”, o sea, como disputa u oposición sistemática—, ni tampoco en aceptarlas sin más. Pues, como afirma Heidegger, “ninguna determinación esencial de la verdad está ‘contra’ la otra en el sentido del mero rechazo; por ello, tampoco a favor de ella en el sentido de la mera asunción, sea en parte, sea completamente” (GA 66, p. 75; 2006b, p. 75). En cambio, el enfrentamiento con la tradición exige “volver siempre de nuevo a tales

¹⁰ Para una reflexión sobre el modo en que Heidegger aborda la tradición filosófica, particularmente en el contexto de sus lecciones, cfr. Xolocotzi Yáñez (2018b).

posiciones fundamentales, para, a través de la confrontación con ellas, ser impelidos a la singularidad de un preguntar originario" (GA 66, p. 75; 2006b, p. 75). Siendo así, la confrontación no consiste ni en revivir sin más ideas o interpretaciones del pasado, ni en desecharlas de forma inmediata, ni mucho menos en ejercer una crítica despectiva o en entrar en conflicto simplemente por disentir con tal o cual postura, sino que se despliega, más bien, como "un diálogo con los pensadores. Ahí yace su carácter histórico que no debe entenderse sólo respecto del pasado. El pensar para Heidegger es un recordar adveniente que acontece en la tradición misma" (Xolocotzi Yáñez, 2018a, p. 133).

Este diálogo —o, como también reflexiona Perrin, lucha sin conflicto— compartido en torno a una misma pregunta adquiere la forma de una crítica esencial, que reconoce la dignidad y profundidad del pensamiento del otro. O sea, no se trata de querer corregir, perfeccionar o, peor aún, superar las doctrinas y posiciones fundamentales de los grandes pensadores, ni de ofrecer la "verdaderísima explicación" de los clásicos de la filosofía. Esa tarea correspondería, más bien, al ámbito de la investigación historiográfica o, en su defecto, al saber escolar y erudito. En palabras de Heidegger, la confrontación no puede convertirse en un "cómputo de corrección e incorrección de teorías y opiniones; la representación del maestro de escuela, de que los pensadores 'hicieran' sin embargo respectivamente 'errores', que tendrían que ser eliminados, tiene su lugar en la 'escuela', pero nunca [...] en el diálogo con los pensadores" (GA 66, p. 69; 2006b, p. 71). No puede ser el intento de refutación historiográfica que, mediante el registro sistemático de doctrinas y el estudio erudito de las obras, aspire a interpretar más exactamente el pensamiento de un autor. Más aún: "Todo pensador esencial es irrefutable" (GA 66, p. 75; 2006b, p. 74), y no porque ya no pueda decirse nada sobre él, sino porque ha conquistado una posición fundamental originaria, a tal grado que lo torna singular y, por ende, insuperable. De hecho, para Heidegger carece de sentido contraponer argumentos en busca de contradicción o falsedad, pues en este terreno no está en juego ningún tipo de mejoramiento, ni de progreso, ni siquiera de depuración de doctrinas filosóficas (cf. GA 66, pp. 75-77; 2006b, pp. 74-76). Y esta es, justamente, su postura en el enfrentamiento con Descartes.

La confrontación, en este sentido, no debe orientarse hacia la veracidad doctrinal, sino por el carácter histórico del pensar, por mantener abierta la cuestionabilidad en torno a los pensadores, siempre

al hilo de un preguntar más originario. Se busca “liberar la historia del pensar a su futuro” (GA 66, p. 77; 2006b, p. 76) y, con ello, “poner así *ante el camino de lo futuro* las posiciones esenciales, insuperables, sólo equiparables a través de una singularidad del pensar inicialmente interrogador” (GA 66, p. 77; 2006b, pp. 76-77). Frente a la rigurosidad de la investigación historiográfica, Heidegger propone así una meditación histórica; o sea, un replanteamiento o una vuelta a preguntar en torno al pensamiento de los grandes filósofos, no en función de su época, sino desde el horizonte del acontecer histórico del ser.

Existe, por ello, una estrecha vinculación entre el sentido de la confrontación heideggeriana y su concepción de la historia (*Geschichte*), en la que el pasado no se comprende como algo irrevocablemente concluido, sino como un “pasado proyectivo” o adveniente. En este marco, “lo sido” conserva en sí el destino (*Geschick*) tanto de lo presente como de lo porvenir. Por esto, la forma en que Heidegger concibe la tradición no se ajusta a la noción de *Historie* (historiografía), que enfatiza un registro descriptivo del pasado, sino al sentido de la historia expresado en el término *Geschichte* (historia como acontecer). Esta distinción, aunque sutil, es fundamental, ya que permite pensar el pasado no como algo estático, petrificado e inmodificable, sino como un fenómeno dinámico que incide de manera constante en el presente y el futuro. Mientras la historiografía aspira a conocer archivísticamente el pasado —es decir, a documentarlo—, la meditación histórica busca penetrar en el propio acontecer de la historia: en aquello que, en tal acaecer, sigue teniendo efecto hoy. Como escribe Heidegger en el curso de 1936/37:

Lo que nosotros llamamos la *meditación historiográfica* es algo esencialmente diferente a la consideración historiográfica [...]. La palabra “histórico” [*geschichtlich*] quiere decir el acaecer [*Geschehen*], la historia [*Geschichte*] misma en cuanto ente. La palabra “historiográfico” quiere decir un modo de conocer. No hablamos de consideración histórica, sino que hablamos de meditación. Meditación [*Be-sinnung*] significa: penetrar en el sentido de lo acaecido, de la historia (GA 45, pp. 35-36; Heidegger, 2008, pp. 36-37).

De allí que, líneas después, señale: “Lo histórico no es lo pasado, tampoco lo presente, sino lo futuro, aquello que está puesto en la voluntad,

en la expectativa, en la preocupación. Esto que no se deja considerar, sino que debemos meditar al respecto" (GA 45, p. 40; Heidegger, 2008, p. 40). La historia no se presenta, entonces, como un archivo estático de ideas, sino como un ámbito de confrontación filosófica esencial que interpela directamente al pensamiento, tanto al presente como, sobre todo, al futuro. En este sentido, es decir, en tanto meditación histórica, la confrontación constituye un ámbito de responsabilidad frente a la historia, un modo de co-responder activamente a las preguntas que nos lega el pensamiento filosófico.

La confrontación no es otra cosa que la estrategia interpretativa en la que se considera la figura de un pensador. En las lecciones sobre Nietzsche leemos: "Confrontación es auténtica crítica. Es el modo más elevado y la única manera de apreciar verdaderamente a un pensador, pues asume la tarea de continuar pensando su pensamiento y de seguir su fuerza productiva y no sus debilidades" (GA 6.1, p. 3; Heidegger, 2013a, p. 19). Por esto, la confrontación se muestra como un compromiso genuino con aquello que ha sido pensado por un pensador, y no tan solo como un recuento externo o una explicación ingenuamente neutral de su filosofía.

Ahora bien, si la historia se entiende como *acontecer*, debemos reconocer entonces que "Estamos inmersos en gran medida en las tendencias interpretativas de la metafísica moderna, que guían nuestra comprensión del ser y, por tanto, la autocomprensión del hombre y el horizonte de interpretación para lo ente" (Flatscher, 2005, p. 104). Esta ha sido, en términos generales, la actitud predominante hasta nuestros días respecto de la tradición filosófica que nos ha sido legada, y en particular frente al pensamiento moderno inaugurado por Descartes. Y aunque el objetivo de Heidegger no es personificar el origen de la modernidad en la figura individual del filósofo francés —como si pudiera aislarse de su propio contexto histórico, él mismo heredero de una tradición filosófica previa—, lo cierto es que: "Toda la metafísica moderna, incluido Nietzsche, se mantendrá dentro de la interpretación de lo ente y la verdad iniciada por Descartes" (GA 5, p. 80; Heidegger, 2010, p. 72). Desde la perspectiva heideggeriana, este es el rasgo distintivo de lo que significa que Descartes sea, para la filosofía occidental, un comienzo absoluto; en él se despliegan las directrices del pensamiento moderno que se prolongan hasta nuestra época. Él las destina.

No obstante, lo verdaderamente importante en la confrontación con Descartes no es la figura personal del filósofo —no se trata de una

preocupación por las biografías filosóficas (cfr. Flatscher, 2005, p. 103)—, sino más bien de lo pensado históricamente por él, en el sentido que hemos delineado: como un compromiso con la historia que acontece en los albores del pensamiento moderno. En ello radica, sin duda, la grandeza de Descartes: en haber sido ese comienzo decisivo. O, para decirlo con una bella y poco común expresión de Heidegger sobre el filósofo francés: “Descartes, como un meteoro, apareció de la nada. Luz contra la oscuridad y la barbarie” (Heidegger, GA 23, p. 108).

En la confrontación con Descartes se juega, pues, el intento por penetrar en aquello que determina y distingue lo que hoy entendemos, con propiedad, justamente como pensamiento moderno. Tal confrontación no implica un rechazo, o un conflicto visceral, sino que se configura como una vía de acceso a la historia de la filosofía. Al enfrentarse a Descartes, Heidegger saca a la luz los presupuestos ontológicos que, hasta hoy, siguen configurando el horizonte de la época técnica contemporánea.

O para plantearlo en una pregunta: ¿qué es aquello que, en el fondo, distingue al pensamiento cartesiano —que además se presenta como un comienzo absoluto— de la filosofía griega, medieval y escolástica? Ante esto, y si bien “[la] posición fundamental de Descartes está soportada históricamente por la metafísica platónica-aristotélica y se mueve, a pesar del nuevo comienzo, dentro del planteamiento: ¿qué es lo ente?” (GA 5, p. 91; Heidegger, 2010, p. 80), lo que aparece o se trastoca por vez primera y que configura la posición fundamental de Descartes es “el trato de trabajo con las cosas y la proyección metafísica de la coseidad de la cosa” (GA 41, p. 68; Heidegger, 2009b, p. 96), y que rompe, en definitiva, con el proceder aristotélico según el cual las ciencias deben conformarse en relación con la cosa. O sea, que la ciencia se moldea por mor de la cosa en cuestión, por aquello que se propone investigar —así la medicina se ocupa de la salud; la retórica del discurso, la gramática de las palabras, la historia de los hechos, o la filosofía primera de lo ente en cuanto ente, etc. En todos estos casos, lo que impera es la cosa o, por decir así, el objeto de estudio.

Jean-Luc Marion habla aquí de una primacía de la cosa, en la que “la esencia de cada cosa gobierna la ciencia correspondiente, y no el espíritu que produce la ciencia” (Marion, 2008, p. 38). Con Descartes, sin embargo, tiene lugar por primera vez una inversión de este modo de relación con la cosa. La diferencia decisiva reside en el carácter anticipador del método cartesiano, que redefine el modo en que la ciencia trata con los entes. A partir de ahora, es la ciencia la que debe decidir

o prefigurar de antemano qué es la cosa y cómo ha de presentarse ante quien la produce: el sujeto, el *ego cogito*. Por ello, “la nueva pretensión del saber es la pretensión de saber *matemática*” (GA 41, p. 68; Heidegger, 2009b, pp. 96-97), pretensión que otorga a la ciencia certeza y seguridad. No se trata, sin embargo, de una forma cualquiera de saber matemático, sino de aquello que el propio Descartes denomina en las *Regulae* como *mathesis universalis* (cfr. AT X, p. 378; Descartes, 2018, p. 104).

A partir de este momento se establece el comienzo absoluto del pensamiento moderno, que, según hemos esbozado, no se encontraría en los grandes tratados cartesianos, sino en aquel breve escrito de juventud. De ahí que la confrontación de Heidegger con Descartes otorgue una importancia singular a las *Regulae*, concentrándose en diversos aspectos que, como veremos en las páginas siguientes, emergen por primera vez en el desarrollo del pensamiento occidental.

3. El proyecto (moderno, demasiado moderno) de las *Regulae*

La historia de las *Regulae ad directionem ingenii* es, como notablemente ya hicieron ver numerosos estudiosos de la obra de Descartes,¹¹ cuanto

¹¹ Marion hace un excelente resumen de la engorrosa historia del texto cartesiano: “No se pretende formular con ello ninguna paradoja, pues el manuscrito original que se encontró entre los papeles de Descartes, en Estocolmo, desaparece sin dejar huella, al término de una auténtica odisea que lo llevó, tras un naufragio, en París, en el fondo del Sena, a las manos de Chanut, Clerselier, J.-Baptiste Legrand y finalmente Marmion. Después, ni rastro del manuscrito. Esta pérdida no constituiría una desgracia irreparable si, al menos, una única tradición textual lo hubiese reproducido *ne varietur*. Imposible de recuperar él mismo, deja en cambio versiones, fragmentos y traducciones, cuya verdad no puede ya, sin embargo, autenticarse ni probarse. Curiosamente, las dificultades filológicas actuales proceden de la sobreabundancia de testimonios y de tradiciones, que la ausencia de una norma hace difícil de contrastar. Anótese en primer lugar el manuscrito de Ámsterdam (A), editado en esta ciudad en 1701, bajo el título de *Opuscula postuma, physica et Mathematica...* (J. y P. Blaeu), que sirve de principal autoridad para la edición de Adam y Tannery [...]. A continuación viene el manuscrito, o más bien, como el precedente, la copia del original, comprado por Leibniz a Schuller también en Ámsterdam, a cargo de la Biblioteca de Hanover (H); fue corregido por el propio Leibniz (de ahí la notación L) [...]. Por último, debemos a G. Crapulli la exhumación y reedición de una traducción neerlandesa, publicada en 1684 por J. H. Glazemaker, bajo el título *R. Descartes Regulen van de bestieringe des verstants*, en el volumen III de *Alle de Werken van Renatus Descartes*, Ámsterdam [...]. Finalmente y sobre todo, el

menos harto escabrosa y compleja, digna incluso, y sin exagerar, de alguna adaptación cinematográfica. Además, contamos hoy día con varias versiones del escrito, algunas perdidas y recuperadas por el mismísimo Leibniz, así como distintos títulos, y todavía es motivo de discusión sobre el orden en el que deben situarse algunas de las reglas, especialmente la importantísima *regula* IV.¹² Marion emplea aquí la atinada expresión “Texto sin texto” (Marion, 2008, p. 21), pues en su opinión los escritos previos de Descartes fracasan en el intento por comprender el proyecto de las *Regulae*. Es más, para él, así como en cierta medida también para Heidegger, son más bien las *Regulae* las que en todo caso posibilitan un mejor entendimiento tanto de la obra anterior como la posterior. Por esto, escribe, “[l]as *Regulae* no tienen genealogía en el pensamiento cartesiano porque, según estimamos, producen su génesis” (Marion, 2008, p. 24). Sin embargo, en el presente escrito no nos detendremos a presentar los detalles relativos a la complejidad del texto cartesiano ni a toda la historia que lo precede. Baste aquí pensar lo siguiente: ¿cómo confrontarse entonces con este texto sin texto? Marión sigue la clave interpretativa de leerlo en contraposición con lo que este intenta romper, dialogar con lo que intenta superar, a saber, la filosofía aristotélica. Heidegger, por su lado, busca situarlo históricamente dentro del despliegue del pensamiento occidental o, si se prefiere, en el marco epocal de la historia del ser; y según los términos descritos, esto no implica simplemente archivarlo en un período de la historia y luego olvidarnos de él, sino penetrar en aquello que las *Regulae* inauguran y

inventario de Estocolmo da el título extenso del manuscrito original: *Tratado de reglas útiles y claras para la dirección del Espíritu en la búsqueda de la Verdad*” (2008, pp. 21-23).

¹² “La regla IV está dividida en dos partes, atendiendo no sólo a la fecha de su redacción, sino también con respecto al tema fundamental de que se ocupa. Partes denominadas por J. P. Weber [...] IV-A, que se extiende desde 371 hasta 374, línea 5, y IV-B, desde 374, línea 16, hasta el final de la regla. El motivo inicial que pudo tener Weber para esta división probablemente fuese el hecho, como señala J. A. Schuster [...] de que en el manuscrito de Hannover IV-B estaba desplazada al final, después de la Regla XXI. Pero ni Adam-Tannery, ni Crapulli, ni Marion la han remitido, en sus ediciones, al final. Ambas partes son de época distinta. Pero parece evidente que IV-B es anterior a IV-A” (Juan Manuel Navarro Cordón, en Descartes, 2018, p. 95).

que, en tanto tal, marca y orienta la tradición filosófica hasta nuestro mundo contemporáneo.¹³

De este modo, si lo moderno (*modernus*) remite principalmente a aquello que se hace “al modo de hoy” —de acuerdo con su etimología—, ¿en qué consiste esencialmente el proyecto con el cual, por primera vez, Descartes establece el modo de pensar que define la modernidad?

3.1. La unidad de la ciencia y la exigencia de certeza

Veamos la primera regla: “El fin de los estudios debe ser la dirección del espíritu para que emita juicios sólidos y verdaderos de todo lo que se le presente” (AT X, p. 359; Descartes, 2018, p. 75). Para esto, Descartes considera un desvarío dedicarse, cada uno por su cuenta, al estudio de ciencias particulares (como en el modelo aristotélico), cuando en realidad, y con una mirada quizás más atenta al respecto, descubrimos que aquello subyacente en cada una de ellas no es otra cosa sino “la sabiduría humana, que permanece una y la misma, aunque aplicada a diferentes objetos, y no recibiendo de ellos mayor diferenciación que los que recibe la luz del sol de la variedad de las cosas que ilumina” (AT X, p. 360; 2018, p. 77). Pero esto no implica en modo alguno rechazar o desestimar a las demás ciencias; Descartes no propone que, de ahora en adelante, los hombres tienen que dejar a un lado el estudio de la astronomía, la física, la gramática, la matemática, etc. Sino que su sorpresa, y posterior reparo, radica más bien en el hecho de que las más de las veces los hombres estudian tanto estas ciencias “mientras que

¹³ Por este motivo, hemos decidido dejar de lado deliberadamente la lectura que Heidegger realiza de las *Regulae* en su primer curso en Marburgo durante el semestre de invierno de 1923/24, *Introducción a la investigación fenomenológica* (2008a), donde aborda la preocupación por el conocimiento reconocido en el pensamiento cartesiano (cfr. GA 17, pp. 195-228; Heidegger, 2008a, pp. 127-226). Asimismo, omitimos ciertas reflexiones de *Ser y tiempo* (2012) que, aunque no mencionan explícitamente a las *Regulae*, dejan entrever una referencia a ellas, particularmente en el apartado b) del § 69, donde se trata la génesis ontológica del comportamiento teórico (cfr. GA 2, § 69, pp. 472-481; Heidegger, 2012, pp. 371-378). Si bien en este periodo Heidegger reconoce el inicio del pensamiento moderno en Descartes y reflexiona sobre la ciencia y el comportamiento teórico, su intención es distinta. Hemos optado por excluir estas consideraciones en la presente lectura, que juega más con la estrategia heideggeriana de la narrativa histórica del ser. Esto refuerza, además, la necesidad de un trabajo sistemático que reconstruya el lugar que ocupa Descartes en la obra heideggeriana.

casi nadie se preocupa del buen sentido o de esta sabiduría universal, cuando, sin embargo, todas las otras deben ser apreciadas no tanto por sí mismas cuanto porque aportan algo a ésta" (AT X, p. 360; 2018, pp. 78-79). Descartes aboga por una sabiduría universal; de hecho, todo el trabajo del espíritu debe encaminarse justamente hacia esta dirección, pues, al fin y al cabo, todas las ciencias remiten inexorablemente a la única sabiduría universal; todas están entrelazadas entre sí y se dirigen hacia el mismo espíritu (cfr. AT X, p. 361; 2018, p. 80).

Hay un giro de timón o, en la expresión de Marion, un traslado del centro de gravedad en la propuesta cartesiana. Tradicionalmente, la división de las ciencias era por mor del objeto de estudio en cuestión, porque cada una se preocupaba de distintas cosas, y estas, las cosas o los objetos, dictaban las directrices de la ciencia. Ahora, el *dictum* proviene de la propia sabiduría humana, que no es más que una y la misma; por esto Descartes habla aquí de una sabiduría universal. Escribe Marion: "el centro de gravedad de la ciencia no reside en lo que se conoce, cuanto en aquél que conoce; no tanto en la cosa misma, cuanto en aquél que la aprehende [...] [:] el conocimiento del espíritu que las constituye en objetos" (Marion, 2008, p. 39). De este modo, la primera regla establece el fin último para con todo el proyecto de las *Regulae* —por esto es la primera de todas las reglas—, a saber: la *unidad de la ciencia* como el fundamento del nuevo paradigma del pensamiento en cuanto tal.

En la segunda regla leemos: "Conviene ocuparse tan sólo de aquellos objetos sobre los que nuestros espíritus parezcan ser suficientes para obtener un conocimiento cierto e indudable" (AT X, p. 362; 2018, p. 81). En efecto, dado que la intención de Descartes es alcanzar la unidad de la ciencia, *universalis Sapientia*, resulta indispensable que esta proporcione un conocimiento absolutamente seguro y libre de dudas. Descartes incluso rechaza por completo cualquier conocimiento que sea solo probable, por más que su estudio esté arraigado en la propia tradición, o simplemente porque muchos hombres doctísimos investiguen al respecto, pues, dice, "no se debe dar asentimiento sino a los perfectamente conocidos y de los que no puede dudarse" (AT X, p. 362; 2018, p. 82). La exigencia de certeza se dicta categóricamente desde las primeras palabras al comentario de esta regla: "Toda ciencia es un conocimiento cierto y evidente" (AT X, 362; 2018, p. 81). Aparece entonces la certeza en el proyecto cartesiano.

Sin embargo, y es algo que el mismo Heidegger reconoce, la idea de certeza no es, en primera instancia, exclusiva del pensamiento

moderno, ni con Descartes se habla por vez primera de la certeza en relación con el conocimiento (verdadero) de las cosas del mundo. Esto es harto conocido desde la determinación esencial de la verdad formulada por Platón y Aristóteles, y el pensamiento griego en general, que se extiende a lo largo del cristianismo, consolidándose durante la época medieval y particularmente en la filosofía escolástica, con santo Tomás de Aquino a la cabeza. Por esto, “no se trata de que Descartes considere la certeza como modo de saber el saber y su objeto sabido. Esto sucedía en lo griego, y sobre todo en la estoa y con los escépticos” (GA 88, p. 87; Heidegger, 2011, p. 123), lo cual también, señala en seguida, “sucedio en el cristianismo: aquí es decisiva y esencial la salvación del alma propia, si y cómo cada individuo puede estar seguro de ella” (GA 88, p. 87; 2011, p. 123).

El fenómeno de la certeza no es nada nuevo. Mas el giro que introduce Descartes sí lo es. En la tradición, como es bien sabido, la verdad es adecuación; en términos habituales, se trata siempre de la justeza o rectitud de la cosa con el intelecto, o del intelecto con la cosa. La verdad, entendida como *veritas*, precisa secundariamente de certeza; es decir, la certeza juega aquí un papel derivado, por decirlo de este modo, en el que sólo confiere seguridad o rigor a lo que de entrada se presenta como verdadero (*verum*). En cambio, lo que ocurre ahora es que:

[...] desde la certitudo y por medio de ella hay que determinar primero qué sea y tenga que ser lo *verum* y, por tanto, qué sea y tenga que ser el *ens* qua *ens*. Ahora la certitudo no es sólo un método de protección de lo verdadero (correcto), sino que pasa a ser una prefiguración de la esencia de la verdad. Se produce la transformación de la rectitud en certitudo. Por esto la certitudo [...] es sobre todo y propiamente *seguridad* de la *ratio* como tal en cuanto a su ser sí misma y a la facultad de sí misma. Se erige su dominio autónomo (GA 88, p. 87; 2011, p. 124).

Descartes quiebra, aquí, con la doctrina escolástica de la salvación. Esta brindaba igualmente seguridad al conocimiento del hombre, pero no a través de su propio intelecto (*ratio*), sino especialmente por medio de la *fides*; el hombre, como *ens creatum*, estaba irremediablemente sostenido por la omnipresencia de Dios, por lo que toda verdad se encontraba, por decir así, cubierta por el velo de la verdad revelada. Por esto, la

certeza de la verdad “siempre permaneció sostenida y asegurada por el *haber sido creada*, por su procedencia de Dios y su *inclusión* en la fe en cuanto conocimiento supremo y auténtico” (GA 88, p. 87; 2011, p. 124). Liberándose de la doctrina de la salvación, de la verdad revelada, el hombre debe buscar el modo de asegurarse del conocimiento de todo lo ente sin recurrir al camino de la *fides*. La segunda regla es explícita: se trata de centrarnos únicamente, de manera individual y con nuestro propio espíritu, en aquellos objetos que permitan alcanzar un conocimiento cierto e indudable. La exigencia de certeza no es un aspecto accesorio ni añadido a lo verdadero, sino un trastocamiento profundo en el corazón de la esencia de la verdad. En una palabra, la verdad deviene en certeza.

Por esto, hacia el final de esta regla se nos dice qué tipo de objetos deben ser estos, a saber, aquellos en los que podamos tener certeza y seguridad como los que proporciona justamente la aritmética y la geometría. Explica Descartes: “porque sólo ellas se ocupan de un objeto de tal modo puro y simple que no suponen absolutamente nada que la experiencia haya mostrado incierto [...]. Son las más fáciles y transparentes de todas, y tienen un objeto tal como el que requerimos” (AT X, p. 365; Descartes, 2018, p. 87). Sin embargo, y como el propio autor acota, esto no quiere decir que la ciencia universal o este nuevo pensamiento tenga por tanto que ocuparse solamente de cuestiones relativas a estas dos disciplinas (cfr. AT X, p. 366; 2018, p. 88), o que la filosofía deba convertirse en un simple matematismo, como algunos creyeron. Lo que se indica más bien es que el pensamiento debe asegurar la certeza de lo que conoce de tal modo como lo hacen la aritmética y la geometría —en suma: de la matemática en general, en donde la certeza, sabemos, se consigue por anticipado en las distintas operaciones, pues los resultados de estas no son otra cosa sino una confirmación de lo previamente conocido—. Con esto, “[la] certeza inmediata de sí misma es la forma *absoluta* del conocer —lo que *así* se conoce es” (Heidegger, GA 76, p. 203). Y este es precisamente el dominio autónomo de la seguridad de la *ratio* o el espíritu que se gesta con Descartes. No se persigue ya la simple adecuación o correspondencia entre el intelecto y la cosa, o viceversa; más bien, se busca una certeza que, de antemano, decida sobre el carácter de verdad o validez del objeto conocido. Es decir, la *certitudo* se erige como el criterio previo que establece qué será considerado verdadero o válido, así como lo que constituirá, o no, el ente o la cosa que se conoce.

Heidegger escribe en este punto: “Así pues, la ciencia surge cuando se lleva a cabo esta *regula* II, cuando se investigan los objetos en su sentido (cuando el conocimiento determina lo que es conocible y, sobre todo, lo que es digno de ser conocido)” (GA 76, p. 200). Para esto, desde luego, resulta crucial trazar un nuevo camino, distinto al que ofrecía la tradición, capaz de satisfacer plenamente las exigencias de la ciencia que Descartes tiene en mente. Como señala Heidegger en otro momento: “Mientras que en el mundo medieval el camino de la salvación y el modo de transmisión de la verdad (*doctrina*) estaban precisamente fijados, ahora se vuelve decisiva la *búsqueda* de nuevos caminos” (GA 6.2, p. 133; Heidegger, 2013a, p. 629). Esto implica, necesariamente, la emergencia de la pregunta por el método, una cuestión determinante para esta nueva ciencia, que se diferencia de la tradición previa al instaurar “la pregunta acerca de cómo conquistar y fundamentar una seguridad fijada por el hombre mismo” (GA 6.2, p. 133; 2013a, p. 629). Con esto se dilucida, además, que en la posición metafísica fundamental de Descartes no figure como tal la pregunta conductora de la metafísica tradicional, a saber, “¿qué es el ente?”, o cualquier otra formulación sobre el ente en general. En su lugar, se establece la pregunta por el método. Y para el de Friburgo, esta transformación “es el comienzo de un nuevo pensar por el que la época se vuelve una época nueva y la edad que le sigue se vuelve edad moderna” (GA 6.2, p. 142; 2013a, p. 635). *Incipit modernum*.

3.2. *Mathesis universalis*

Llegamos finalmente a la importantísima regla cuatro. Descartes la presenta del siguiente modo: “El método es necesario para la investigación de la verdad de las cosas” (AT X, p. 371; Descartes, 2018, p. 95). No es sorprendente, para cualquier estudioso de la obra de Descartes o incluso por gracia de la *bona mens* intrínseca al ser humano, la cual, según afirma el filósofo francés en el *Discurso*, “es la cosa mejor repartida del mundo” (Descartes, 2010, p. 39), que el interés por el método ocupe un lugar central en su pensamiento —de allí que en el título mismo de varias de sus principales obras haya una preminencia del método (cfr. GA 6.2, p. 134; Heidegger, 2013a, p. 629)—. Este énfasis metódico muestra la preocupación cartesiana por establecer, mediante el espíritu, fundamentos sólidos y universales para el conocimiento de las cosas, alejándose así de la dispersión y la poca seguridad que caracterizaban gran parte de la tradición filosófica, más allá de lo que la *fides* podría ofrecer. Sin embargo, y esta es una de las mayores apuestas

de la lectura heideggeriana, la cual sin duda podemos ir comprobando, la idea moderna del método no se gesta en la obra de madurez de Descartes, sino en estas pocas páginas de las *Regulae*; aquí se juega el destino todo del pensamiento occidental.

Por lo que preguntamos: ¿qué significa aquí la *necesidad* respecto al método en el proceso de investigación de las cosas, tal como lo leemos en las palabras de Descartes? Heidegger escribe al respecto en el curso de invierno de 1935:

Esta regla no quiere referir el lugar común de que una ciencia también ha de tener su método, sino que el procedimiento, es decir, el modo en que estamos en general detrás de las cosas (μέθοδος), decide ya de antemano lo que aprehendemos de verdad de ellas. El método no es un elemento entre otros del equipamiento de la ciencia, sino la instancia fundamental desde la que se determina por vez primera qué puede convertirse en objeto y cómo se convierte en objeto [*Gegenstand*] (GA 41, p. 103; Heidegger, 2009b, p. 132).

En concordancia con la primera regla, la unidad de la ciencia exige que el método no sea cualquiera, ni mucho menos uno distinto para cada ciencia, como ocurría en el modelo científico aristotélico. Tampoco el método debe responder a la particularidad de la cosa en cuestión, sino que debe establecer previamente la verdad de dicha cosa y, con ello, decidir desde ahora cómo debe presentarse dentro de la totalidad de lo ente, es decir, convertirla por primera vez en objeto. Se entiende ahora con mayor claridad que, para esta nueva ciencia universal, el método debe ser, igualmente, el uno y el mismo; o sea, el método buscado debe ser universal. Este garantizaría la certeza al modo matemático, con este grado de verdad, pero no para transformar el conocimiento de las cosas, o todo lo ente, en lo que tradicionalmente podría pensarse, esto es, como un resultado numérico. Leemos en la misma regla cuatro: “en absoluto pienso aquí en la Matemática corriente, sino que expongo cierta disciplina distinta [...]. Pues ésta debe contener los principios rudimentarios de la razón humana y desplegarse para hacer salir de sí verdades respecto de cualquier asunto” (AT X, p. 374; Descartes, 2018, p. 100). Para Descartes, la matemática de su tiempo es, cuanto menos, insuficiente para la tarea pretendida en el proyecto de la ciencia universal, pues esta se dedica al estudio de asuntos particulares, y con sorprendente efectividad, sin

duda, pero poco sirven para las pretensiones de universalidad. En las *Regulae* se busca ciertamente otra cosa, pues “no tendría en mucho estas reglas si no sirvieran más que para resolver vanos problemas, en los que los calculistas y geómetras ociosos acostumbraron a distraerse” (AT X, p. 373; 2018, pp. 99-100).

De este modo, y contra algunas lecturas tradicionales, el método universal buscado no es el de la matemática, sino lo que Descartes nombra *mathesis universalis*, “ya que en ésta se contiene todo aquello por lo que las otras ciencias son llamadas partes de la matemática” (AT X, p. 378; 2018, p. 104). ¿Y qué quiere decir concretamente esta expresión: *mathesis universalis*?

3.2.1. El sentido de la *mathesis* y lo “matemático”

Como bien señala Matthias Flatscher, el término *mathesis universalis* aparece únicamente en dos ocasiones a lo largo de las *Regulae*. Por esta razón, para comprender adecuadamente su significado dentro de la obra, es necesario precisar qué se entiende por *mathesis* y por lo “matemático” en este contexto (cfr. Flatscher, 2005, p. 108). Estas nociones, como seguro que ya puede entreverse, no solo remiten a lo técnico o a la disciplina de las matemáticas, sino que, en el marco cartesiano, adquieren un sentido más amplio vinculado al conocimiento universal y al método como base unificadora de todas las ciencias. Heidegger se ocupa de esto en el curso de verano de 1933, particularmente en el segundo capítulo titulado “Das System der neuzeitlichen Metaphysik und ihr erster Hauptbestimmungsgrund: das Mathematische” (“El sistema de la metafísica moderna y el primero de sus fundamentos determinantes: la matemática”). En el § 8 leemos: “Lo matemático debe entenderse aquí en un sentido más amplio y fundamental, no como una metodología específica de algún campo particular de las matemáticas” (Heidegger, GA 36/37, p. 30). Es decir, debe comprenderse en su sentido originario, que Descartes conoce muy bien de la tradición griega, a saber, como *mathesis* en su acepción más apegada a la etimología de la palabra.

Desde los griegos, lo matemático no se limita exclusivamente a aquello relacionado con números, operaciones, funciones, puntos, líneas o planos, etc., sino que “[l]o matemático [es], entendido como μάθημα, lo enseñable por excelencia, aquello que en un sentido eminente puede ser aprendido; μάθησις, el aprendizaje, y μανθάνειν, aprender” (GA 36/37, p. 31). De este modo, la *mathesis* alude tanto al proceso de aprendizaje o enseñanza como a aquello que puede ser aprendido

mediante dicho proceso. En este sentido, Van de Pitte coincide con la descripción que ofrece Heidegger, pues en uno de sus artículos señala: “Se está de acuerdo en que (etimológicamente) ‘*mathesis*’ y ‘*disciplina*’ significan lo mismo. Ambas palabras significan que algo es objeto de aprendizaje, o quizás indican el proceso por medio del cual [este algo] es aprendido” (Van de Pitte, 2016, p. 17). En las lecciones de invierno de 1935, el filósofo alemán prácticamente retoma esta discusión y la redacta sin menores modificaciones: “Lo matemático viene, según la acuñación de la palabra, del griego τὰ μαθήματα, lo que puede aprenderse y por ello, a su vez, lo enseñable; μανθάνειν significa aprender, μάθησις la enseñanza [...] como acudir a la enseñanza y aprender y [...] enseñanza como aquello enseñado” (GA 41, p. 70; Heidegger, 2009b, p. 98). Por esto, Heidegger usa en el mismo sentido *mathesis* y “matemático”. De hecho, las más de las veces emplea más bien este segundo término en sus reflexiones sobre el pensamiento moderno y, en este caso, sobre Descartes —cabe entonces la aclaración para evitar confusiones—.

Lo matemático (o la *mathesis*) implica esencialmente una especie de familiaridad con el trato de las cosas. “Τὰ μαθήματα significa para los griegos aquello que el hombre ya conoce por adelantado cuando contempla lo ente o entra en trato con las cosas” (GA 5, p. 71; Heidegger, 2010, p. 65). Los números son, sin duda, el caso insigne por excelencia. Heidegger pone el ejemplo de que cuando vemos tres manzanas sobre la mesa, o las que sean, nos percatamos de que son efectivamente tres porque antes conocemos el número tres, es decir, tenemos por anticipado el conocimiento de lo que implica la triplicidad de algún ente en el mundo; de no ser así, sería imposible identificar que son tres manzanas las que yacen sobre la mesa. Y este es el motivo, señala también Heidegger, por el que, en la tradición, y en el entendimiento cotidiano, lo matemático haya sido reducido únicamente a los números, cuando lo que determina en verdad la esencia de lo matemático, y de la *mathesis* como tal, es más bien el carácter *anticipador* no solo del conocimiento de los números, a través del cual podemos realizar operaciones matemáticas, sino de todo el conocimiento. Leemos en la lección de 1935:

Lo matemático es aquello abierto en las cosas en lo que ya nos movemos siempre y a partir del cual tenemos experiencia de ellas en general como cosas y como tales cosas. Lo matemático es aquella posición fundamental ante las cosas en la que nosotros tomamos previamente

las cosas en relación con cómo nos son, necesitan ser y deben ser ya dadas. Lo matemático es por esto el presupuesto fundamental del saber de las cosas (GA 41, p. 76; Heidegger, 2009b, p. 104).

De este modo, lo matemático o, si se prefiere aquí, la *mathesis* de la nueva ciencia, en tanto que método para el conocimiento de la verdad de las cosas, es nombrada justamente *mathesis universalis* en la medida en que esta funge ahora como el presupuesto fundamental, o condición de posibilidad (a la Kant, con sus salvedades), para con el conocimiento de la totalidad de lo ente —o sea, no para el conocimiento de este o aquel ente, no un método más entre otros disponibles, sino *el* método para el conocimiento de la realidad toda, por esto que se presuma aquí como el método que determina y destina el pensamiento moderno—. O en palabras de Heidegger: “La ciencia fundamental más general recibe de Descartes el nombre de *mathesis universalis* [...]. Esta ciencia fundamental y universal representa, al mismo tiempo, el modelo de toda ciencia y de toda cientificidad” (Heidegger, GA 36/37, pp. 34-35).

3.2.2. La primacía *anticipadora* del método moderno como disposición (orden y medida) de la totalidad de lo ente

Matthias Flatscher también identifica en este punto que, quizás desgraciadamente, Descartes no explicita con mayor detalle lo que quiere decir o más bien cómo es que funciona el método de la *mathesis universalis* (cfr. Flatscher, 2005, p. 112). Esto ha conducido a algunos investigadores a sostener que esta idea de juventud queda por tanto superada en los desarrollos posteriores, y que, en pocas palabras, Descartes termina abandonando el camino metodológico de la *mathesis universalis* (cfr. Van de Pitte, 2016, p. 7). Heidegger no es de esta opinión. Para él, en cambio, “las ciencias naturales modernas, la matemática moderna y la metafísica moderna surgen de la misma raíz de lo matemático [es decir, de la *mathesis*] en sentido amplio” (GA 41, p. 98; Heidegger, 2009b, p. 127). En este contexto, Heidegger identifica la *mathesis* con “lo matemático” y, simultáneamente, con la *mathesis universalis*. No establece una distinción estricta entre ambos términos, sino que los considera equivalentes en su reflexión. Así, para Heidegger, la *mathesis* y la *mathesis universalis* se integran en un único concepto, que prefiere designar simplemente como “lo matemático”. Y en el mismo curso páginas después escribe lo siguiente: “Lo decisivo sigue siendo el modo en que esta meditación acerca

de lo matemático se efectuó en la confrontación [*Auseinandersetzung*] con la metafísica (*prima philosophia*) heredada, y cómo desde ahí se determinó el destino posterior y la figura de la filosofía moderna" (GA 41, p. 104; 2009b, p. 133). Por esto, el método cartesiano de las *Regulae* entendido justamente como "método matemático" (o *mathesis universalis*, o simplemente lo matemático) es lo que, por un lado, quiebra el proceder legado por la tradición, la aristotélica especialmente, y lo que, por el otro lado, consecuentemente determina e instaura el destino del pensamiento moderno.

Hemos indicado dos cuestiones en torno al sentido del método, sobre todo que este debe preconfigurar cómo lo ente debe presentarse, logrando con esto la certeza deseada para lo conocido. ¿Cuál es esta nueva manera en que lo ente o, si se quiere, las cosas tienen que presentarse por anticipado en el mundo a partir del método? Veamos lo dicho en la regla cinco: "Todo el método consiste en el orden y disposición de aquellas cosas a las que se ha de dirigir la mirada de la mente a fin de que descubramos alguna verdad" (AT X, p. 379; Descartes, 2018, p. 106). En un primer momento, podría parecer que el método cartesiano pretende organizar y disponer las cosas después de haber decidido estudiarlas. Sin embargo, esta interpretación es incorrecta. Descartes no pretende, con el método, una suerte de "arreglo" o "acomodo" secundario de los entes, como si estos estuvieran de antemano en una especie de "desorden" en el mundo y necesitaran ser organizados. Por el contrario, el método establece, de manera previa, el modo en que los entes, en su totalidad, deben manifestarse en el mundo. Este modo no responde a cómo las cosas se presentan en el mundo por sí mismas, sino a cómo el espíritu decide de antemano sobre ellas; qué deben ser o no ser, qué deben valer o no, y qué debe considerarse o no. Solo desde esta preconfiguración el método garantiza la certeza requerida, cuya validez se pretende universal. En este sentido, el método no es "lo que >también< se añade como un orden que recae sobre algo ya presente en sí, sino que en él y a través de él >se convierte< el objeto por vez primera como tal" (Heidegger, GA 76, p. 197).

La ciencia universal buscada por Descartes en la figura de la *mathesis universalis* no tiene por añadidura también un método que le prescribe los pasos que deben seguirse en el camino de la investigación, o en el camino hacia la verdad de las cosas. Pues "[n]o hay que comprender aquí 'método' en sentido 'metodológico', como modo de exploración e investigación, sino en sentido metafísico, como camino hacia una

determinación esencial de la verdad" (GA 6.2, p. 133; Heidegger, 2013a, p. 629). El método actúa como un pro-ceder (*Vor-gehen*) anticipador que organiza de antemano aquello que debe ser investigado, determinando cómo ha de ser comprendido tal o cual objeto. Se trata de una decisión previa ante lo ente que establece precisamente qué debe (o no debe) ser reconocido como válido o, mejor aún, como verdadero en el mundo y, con ello, garantizar la certeza de lo que ahora comparece en el mundo por medio del método, o sea, según la disposición ordenada de la totalidad de lo ente en aquello que, tal como se estipulará en el pensamiento posterior de Descartes, deviene por primera vez objeto para el sujeto, o el *ego cogito*. En las lecciones de 1940 leemos:

Lo verdadero es sólo lo asegurado, lo cierto. Verdad es certeza, certeza para la cual resulta decisivo que el hombre como sujeto esté en ella en cada caso seguro y cierto de sí mismo. Por ello, para el aseguramiento de la verdad en cuanto certeza es necesario en un sentido esencial el pro-ceder [*Vor-gehen*], el asegurar-de-antemano. El 'método' adquiere ahora un peso metafísico [...]. 'Método' ya no es simplemente la secuencia de algún modo ordenada de los diferentes pasos en los que se mueve el examen, la demostración, la exposición y la correlación de conocimientos y fragmentos doctrinales en el modo de una *summa* escolástica [...]. 'Método' es ahora el nombre del proceder asegurador y conquistador frente al ente para ponerlo seguro como objeto para el sujeto. En este sentido metafísico se entiende *methodus* cuando Descartes, en su importante tratado *Regulae ad directionem ingenii*, aparecido sólo después de su muerte, establece, como *regula* IV (GA 6.2, p. 170; 2013a, p. 657).

Por esta razón, el método es en sí y por sí mismo aquello que, desde entonces, es decir, desde las *Regulae*, caracteriza esencialmente tanto la cientificidad moderna como el pensamiento moderno en general; o, como señala en otro momento el filósofo de Friburgo: "El método mismo en su autodesarrollo y por tanto (idealísticamente) la fuente (*fons*) de todas las ciencias" (Heidegger, GA 76, p. 201). Y solo a partir de esta instauración de la primacía del método —entendido como disposición ordenada de lo ente en su totalidad— "se ha de pasar a las cosas mismas, que tan

sólo deben ser consideradas en la medida en que tienen relación con el entendimiento” (AT X, p. 399; Descartes, 2018, p. 128).

4. Conclusión

En los *Seminarios de Zollikon* (Heidegger, 2013b), hacia 1969, ya en una etapa tardía de su pensamiento, Heidegger vuelve a reflexionar brevemente en algunas sesiones en torno a las *Regulae*. Y una vez más, localiza en esta obra de juventud de Descartes el inicio del pensamiento moderno, sobre todo —como hemos trabajado— en lo que concierne a la instauración del nuevo método científico, en tanto proceder anticipador de orden y medida de todo lo ente en su conjunto (cfr. Heidegger, 2013b, pp. 166 y 171). Este es el comienzo absoluto cartesiano, en el que aún no pinta explícitamente el sujeto o el *ego cogito*, que se considera habitualmente como el primer paso o el rasgo característico de la época moderna. Es, desde luego, un elemento importantísimo, aunque no el fundamental según nuestra lectura.

Como señala Nietzsche en el segundo epígrafe del escrito, no es la ciencia o la científicidad lo que determina la modernidad, pues ciencia ha existido desde los griegos (en un sentido distinto, por supuesto), sino más bien el triunfo del método sobre la ciencia. La prefiguración de la realidad a través del método que inaugura Descartes en las *Regulae* es lo que ha permitido, como tal, el modo de actuar del hombre hasta nuestros días, o sea, creyéndose amo y señor de la naturaleza, o en la formulación posterior en el *Discurso*: “hacernos maestros y poseedores de la naturaleza” (Descartes, 2010, p. 50). Este modo de proceder destina el modo de comportamiento y trato para con las cosas, lo ente en su conjunto, el cual Heidegger describe en los términos de un pasar por encima o un saltar por encima de las cosas (*über die Dinge hinwegspringen*), pues se trata de la proyección en la cual por primera vez “se pone aquello por lo que propiamente deben ser tomadas las cosas, cómo y como qué deben ser apreciadas las cosas” (GA 41, p. 92; Heidegger, 2009b, pp. 121-122). Dicho sobrepasamiento, en el que anticipadamente se dispone ordenadamente lo ente, devendrá en el actual dominio calculador de la técnica contemporánea, según adelantábamos.

De allí que Heidegger escriba: “la ciencia así fijada, esto es, este método, es el ataque más monstruoso del ser humano a la naturaleza” (Heidegger, 2013b, p. 174), pues desestima a las cosas mismas, casi como en una suerte de desinterés o indiferencia, con la pretensión ciega, y hasta monstruosa, de sobreponerse y desde el método dictar el qué y

cómo es que las cosas deben ahora comparecer en el mundo. Es, pues, el dominio sobre la totalidad de lo ente. Por ello, y a ojos de Heidegger, Descartes piensa de antemano el fundamento metafísico del proceder moderno en su totalidad, un fundamento que —como sabemos, y que ahora quizá se vuelve un tanto más claro— se extiende hasta el mundo contemporáneo. Pero no se trata de un simple invento o de una ocurrencia: no porque un día, de la nada, decidiera intrépida y radicalmente dudar de todo. Por el contrario, Descartes representa un “comienzo absoluto” porque lo pensado en las *Regulae*, breve e inconclusa obra de juventud, “quedó como fundamento para lo que vino después” (GA 6.2, p. 147; Heidegger, 2013a, p. 639).

Estas consideraciones sobre la lectura heideggeriana de las *Regulae* han permitido, en primer lugar, esclarecer la génesis del pensamiento moderno más allá de los lugares comunes heredados por la tradición, al centrarse en esta temprana obra del filósofo francés, en la que —como ahora podemos sostener— se encuentra el verdadero punto de partida. La concepción del método, entendida como *mathesis universalis*, constituye la piedra angular del desarrollo moderno y contemporáneo en los ámbitos de la ciencia, la tecnología y la filosofía. En segundo lugar, estas reflexiones han posibilitado una comprensión más profunda del alcance de las palabras de Heidegger en su lección de 1941: “Sólo quien haya pensado efectiva y largamente este escrito sobrio, sin consideraciones, hasta sus más escondidos recovecos, está en disposición de tener alguna idea de lo que ocurre en la ciencia moderna” (GA 41, p. 102; Heidegger, 2009b, pp. 131-132). En tercer lugar, se ha mostrado que la presencia de Descartes en el *Denkweg* heideggeriano dista mucho de ser marginal. Si bien este artículo no tuvo como fin estudiar, sistematizar y evaluar todo lo que dicha presencia implica —pues, ciertamente, no habría páginas suficientes—, algo se ha ganado: al menos algunas indicaciones que permiten poner sobre la mesa que la confrontación entre Heidegger y Descartes no ha recibido hasta ahora la atención que merece, y que, por ello, sigue siendo una tarea pendiente.

En definitiva —y eso espero—, lo planteado en estas páginas pretende servir como estímulo para futuros estudios heideggerianos dedicados al filósofo de La Flèche.

Referencias

- Descartes, R. (1908). [AT X]. *Œuvres : Vol. X. Physico Mathematica, Compendium Musicae, Regulæ ad directionem ingenii, Recherche de la vérité, Supplément à la correspondance*. C. Adam y P. Tannery (eds.). Léopold Cerf.
- Descartes, R. (2010). *Discurso del método. Meditaciones metafísicas*. M. García Morente (trad.). Austral.
- Descartes, R. (2018). *Reglas para la dirección del espíritu*. J. M. Navarro Cordón (trad.). Alianza Editorial.
- De Brasi, D. (2024). El seminario de Heidegger de 1921 sobre *De anima*: algunas observaciones filológicas. En J. Orejarena Torres y Á. Xolocotzi Yáñez (eds.), *Heidegger, lector de Aristóteles: En torno al origen de su interpretación fenomenológica* (pp. 139-159). SB Editorial.
- Faye, E. (2005). La pensée métaphysique de Descartes et son «interprétation» par Heidegger. En Y.-C. Zarka y B. Pinchard (eds.), *Y a-t-il une histoire de la métaphysique?* (pp. 263-301). PUF.
- Flatscher, M. (2005). Heideggers »Aus-ein-andersetzung« mit dem cartesianischen Methodenideal einer »Mathesis Universalis«. En M. Blamauer, W. Fasching y M. Flatscher (eds.), *Phänomenologische Aufbrüche* (pp. 102-119). Peter Lang Verlag.
- Gonzalez, F. J. (2022). El nacimiento de *Ser y tiempo*: La lectura crucial de Heidegger de 1921 del *De anima* de Aristóteles. En Á. Xolocotzi Yáñez, R. Gibu Shimabukuro y J. Orejarena Torres (eds.), *Aristóteles y la fenomenología del siglo XX: estudios en torno a la presencia de Aristóteles en la obra de Heidegger y Husserl* (pp. 543-574). Editorial Biblos.
- Gonzalez, F. J. (2024). Introducción al seminario de Heidegger del semestre de verano de 1921 sobre *De anima* de Aristóteles según los apuntes manuscritos de Helene Weiss y Oskar Becker. En J. Orejarena Torres y Á. Xolocotzi Yáñez (Eds.), *Heidegger, lector de Aristóteles: En torno al origen de su interpretación fenomenológica* (pp. 25-28). SB Editorial.
- Heidegger, M. (1992). [GA 7]. *Conferencias y artículos*. E. Barjau (trad.). Ediciones del Serbal.
- Heidegger, M. (2001). [GA 36/37]. *Gesamtausgabe. II. Abteilung: Vorlesungen (1919-1944). Band 36/37. Sein und Wahrheit*. Vittorio Klostermann.

- Heidegger, M. (2006a). [GA 23]. *Gesamtausgabe. II. Abteilung: Vorlesungen (1919-1944). Band 23. Geschichte der Philosophie von Thomas von Aquin bis Kant*. Vittorio Klostermann.
- Heidegger, M. (2006b). [GA 66]. *Meditación*. D. V. Picotti (trad.). Editorial Biblos.
- Heidegger, M. (2008). [GA 45]. *Preguntas fundamentales de la filosofía*. Á. Xolocotzi Yáñez (trad.). Comares.
- Heidegger, M. (2009a). [GA 76]. *Gesamtausgabe. III. Abteilung: Unveröffentlichte Abhandlungen / Vorträge-Gedachtes. Band 76. Leitgedanken zur Entstehung der Metaphysik, der neuzeitlichen Wissenschaft und der modernen Technik*. Vittorio Klostermann.
- Heidegger, M. (2009b). [GA 41]. *La pregunta por la cosa. Sobre la doctrina de los principios trascendentales de Kant*. J. García Gómez de la Valle (trad.). Palamedes Editorial.
- Heidegger, M. (2010). [GA 5]. *Caminos de bosque*. A. Leyte Coello y H. Cortés Gabaudan (trads.). Alianza Editorial.
- Heidegger, M. (2011). [GA 88]. *Posiciones metafísicas fundamentales del pensamiento occidental. Ejercicios en el semestre de invierno de 1937-1938*. A. Ciria (trad.). Herder.
- Heidegger, M. (2013a). [GA 6. 1 y 6.2] *Nietzsche*. J. L. Vermal (trad.). Editorial Ariel.
- Heidegger, M. (2013b). *Seminarios de Zollikon*. Á. Xolocotzi Yáñez (trad.). Herder.
- Heidegger, M. (2017). [GA 95]. *Reflexiones VII-XI. Cuadernos negros (1938-1939)*. A. Ciria (trad.). Trotta.
- Jimenez, E. R. (2024). Comentario sobre las notas de Helene Weiss en torno al seminario de verano sobre Aristóteles de Heidegger de 1921. En J. Orejarena Torres y Á. Xolocotzi Yáñez (eds.), *Heidegger, lector de Aristóteles: En torno al origen de su interpretación fenomenológica* (pp. 183-211). SB Editorial.
- Kellerer, S. (2013). *Zerrissene Moderne. Descartes bei den Neukantianern, Husserl und Heidegger*. Konstanz University Press. <https://doi.org/10.46500/83539031>
- Marion, J.-L. (2008). *Sobre la ontología gris de Descartes. Ciencia cartesiana y saber aristotélico en las Regulae*. A. García Mayo (trad.). Escolar y Mayo Editores.
- Natorp, P. (1882). *Descartes' Erkenntnistheorie. Eine Studie Zur Vorgeschichte Des Kriticismus*. Elwert'sche Verlagsbuchhandlung.

- Natorp, P. (1896). Le développement de la pensée de Descartes depuis les 'Regulæ' jusqu'aux 'Méditations'. *Revue de Métaphysique et de Morale*, 4(4), 416-432.
- Nietzsche, F. (2008). [FP IV]. *Fragmentos Póstumos. Volumen IV (1885-1889)*. J. L. Vermal y J. B. Llinares Chover (trads.). Tecnos.
- Orejarena Torres, J. y Xolocotzi Yáñez, Á. (2024). *Heidegger, lector de Aristóteles: En torno al origen de su interpretación fenomenológica*. SB Editorial.
- Perrin, C. (2010). Sur l'anticartésianisme prétendu de Heidegger: Le sens d'(une) Auseinandersetzung. *Daimon: Revista Internacional de Filosofía*, 51, 137-149.
- Perrin, C. (2013a). *Entendre la métaphysique. Les significations de la pensée de Descartes dans l'œuvre de Heidegger*. Peeters.
- Perrin, C. (2013b). L'impensé du je pense: L'explication heideggerienne du principe cartésien. *Cahiers de Philosophie de l'Université de Caen*, 50, 81-105. <https://doi.org/10.4000/cpuc.703>
- Shiffman, M. (2024). Comentarios al seminario sobre *De anima* de Heidegger en 1921. En J. Orejarena Torres y Á. Xolocotzi Yáñez (eds.), *Heidegger, lector de Aristóteles: En torno al origen de su interpretación fenomenológica* (pp. 161-181). SB Editorial.
- Van de Pitte, F. P. (2016). La *mathesis universalis* de Descartes. H. H. Benítez (trad.). *Revista de Filosofía*, 51, 7-26. <https://doi.org/10.5354/0718-4360.1998.43570>
- Xolocotzi Yáñez, Á. (2011). *Una crónica de Ser y tiempo de Martin Heidegger*. Editorial Ítaca-BUAP.
- Xolocotzi Yáñez, Á. (2018a). *Heidegger. Lenguaje y escritura*. Fontamara.
- Xolocotzi Yáñez, Á. (2018b). "Mis lecciones han sido caminos de bosque". Aspectos de la apropiación heideggeriana de la filosofía occidental. En Á. Xolocotzi Yáñez, *Heidegger. Lenguaje y escritura* (pp. 125-137). Fontamara.

<https://doi.org/10.21555/top.v760.3339>

La interrupción en el sistema de la absoluta objetividad de Friedrich Heinrich Jacobi

Interruption in Friedrich Heinrich Jacobi's system of absolute objectivity

Federico Vicum

Consejo Nacional de Investigaciones Científicas y Técnicas
Argentina

federicovicum@gmail.com

<https://orcid.org/0009-0002-2928-9840>

Recibido: 30 - 12 - 2024.

Aceptado: 10 - 07 - 2025.

Publicado en línea: 10 - 08 - 2026.

Cómo citar este artículo: Vicum, F. (2026). La interrupción en el sistema de la absoluta objetividad de Friedrich Heinrich Jacobi. *Tópicos, Revista de Filosofía*, 76, 511-548. <https://doi.org/10.21555/top.v760.3339>



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

Resumen

El realismo de Friedrich Heinrich Jacobi es inseparable de las polémicas en las que fue puesto a prueba. La más famosa de ellas involucra a la “cosa en sí” kantiana. Sin embargo, el realismo de Jacobi tiene que ser interpretado también en virtud de su relación con la filosofía de Spinoza. Partiendo de un pasaje de la novela *Allwill*, poco estudiado y sin embargo altamente valorado por el propio Jacobi, proponemos acercarnos al realismo jacobiano (el sistema de la absoluta objetividad) de la mano de la noción de “interrupción”. Esta noción no solo nos permite descifrar aspectos conceptuales del pensamiento de Jacobi, sino también comprender el modo en que este trabaja concretamente con la cultura escrita. La interrupción del puro vacío del pensamiento es, en Jacobi, el comienzo de la vida (como actuar libre) expresado desde las condiciones del discurso conceptual, poco propicias para esas expresiones.

Palabras clave: interrupción; sistema de la absoluta objetividad; realismo; cultura escrita; idealismo; sensación; spinozismo.

Abstract

Friedrich Heinrich Jacobi's realism is inseparable from the philosophical controversies in which he was involved. The most famous of them involves the Kantian “thing-in-itself”. However, Jacobi's realism has to be interpreted also according to its relation to Spinoza's philosophy. Starting from a passage of the novel *Allwill*, rarely studied and yet highly valued by Jacobi himself, we propose to explore Jacobi's realism (the system of absolute objectivity) using the notion of “interruption” as a reference point. This notion allows us not only to decipher conceptual aspects of Jacobi's thought, but also to understand the way in which he works concretely with written culture. The interruption of the pure emptiness of thought is, for Jacobi, the beginning of life (as free acting) expressed from the conditions of conceptual discourse, which are not really appropriate for such expressions.

Keywords: interruption; system of absolute objectivity; realism; written culture; idealism; sensation; Spinozism.

1. Introducción

La más célebre de las intervenciones polémicas de Friedrich Heinrich Jacobi probablemente sea aquella que señala las dificultades que, a su entender, tiene la *Crítica de la razón pura* para lidiar con las impresiones sensibles. En esta gran obra, serían inconciliables, por un lado, la “presuposición” de que los objetos “provocan impresiones en los sentidos” y, por el otro, el resto del sistema kantiano (JWA 2,1, p. 109).¹ Jacobi observa que “la palabra ‘sensibilidad’ no tiene significado si no se entiende por ella un medio distinto y real entre lo real y lo real, un medio efectivo de *algo* hacia *algo* [...]” (JWA 2,1, p. 109). A ese rasgo constitutivo de la sensibilidad no podría hacerle justicia, según Jacobi, la *Crítica* kantiana, a pesar de darlo por supuesto. De allí, entonces, el famoso *dictum* del “Apéndice sobre el idealismo trascendental”: “*sin* aquella presuposición no podría entrar en el sistema, y *con* aquella presuposición no podía permanecer en él” (JWA 2,1, p. 109).²

Es sabido que, en lugar del “idealismo” kantiano y postkantiano, Jacobi abraza el “realismo” (cfr. NA I, p. 239). Sin embargo, los alcances del realismo jacobiano no son fácilmente determinables, como observaremos al repasar algunas de las interpretaciones a las que ha dado lugar. Nuestro recorrido por la bibliografía especializada no será exhaustivo, pero sí nos permitirá ingresar plenamente en la discusión que nos concierne. Dado que nos proponemos centrarnos en un pasaje de la obra de Jacobi descuidado por los intérpretes, juzgamos necesario mostrar con la mayor claridad posible el modo en que nuestra propuesta se relaciona con otras concepciones sobre la obra jacobiana. El objetivo de este recorrido es presentar las líneas de investigación ya acreditadas en los estudios jacobianos con las cuales dialoga nuestra propuesta.

Birgit Sandkaulen sostiene que la filosofía jacobiana es una “metafísica de la acción” en lugar de una “metafísica de la fundamentación”

¹ Citamos todas las obras de Jacobi de acuerdo con las siglas convencionales en cada caso: el primer tomo del *Auserlesener Briefwechsel* es citado como AB I; el segundo, como AB II. El *Nachlaß* es citado como NA I. Finalmente, la *Gesamtausgabe* es citada como JWA, seguido de número de serie y tomo. Por lo demás, a menos que la bibliografía final indique lo contrario, las traducciones son nuestras.

² Cfr. Guzmán Grez (2024) para una crítica a la interpretación de Jacobi.

(Sandkaulen, 2000, p. 220). La certeza jacobiana tiene carácter práctico y no se genera en un acto cognitivo, sino en la experiencia de la praxis vital (cfr. Sandkaulen, 2019, pp. 42-43 y 48). En tal marco, el verdadero concepto de lo incondicionado es la libertad (Sandkaulen, 2000, p. 261). Al mismo tiempo, Sandkaulen propone una interesante reconsideración del nihilismo en el pensamiento de Jacobi, lo que también implica reconsiderar su relación con Spinoza. Apartándose de la tendencia usual, Sandkaulen advierte que no hay que trasladar hacia Spinoza la crítica del nihilismo que Jacobi dirige, en verdad, contra la filosofía kantiana y poskantiana (cfr. JWA 2,1, p. 215). Para Jacobi —según Sandkaulen—, el autor de la *Ética* no sería nihilista, precisamente porque el principio de la filosofía spinozista es el del “ser en toda existencia” (*Seyn in allem Daseyn*). Esto implica que es en las mismas cosas finitas, en su realidad no sustancial, donde se manifiesta la realidad de la sustancia (JWA 1,1, p. 39; cfr. Sandkaulen, 2023, pp. 44-45). El principio spinozista del “ser en toda existencia” coincidiría, según Sandkaulen, con la noción de la existencia como posición que Jacobi recibe del escrito precrítico de Kant *El único argumento posible para una demostración de la existencia de Dios*. Jacobi habría “transferido el lugar que, en Kant, ocupa la ‘existencia’ [*Dasein*] hacia la posición del ‘ser’, y [habría] determinado la ‘existencia’ [*Dasein*] expresamente como la esfera de lo finito” (Sandkaulen, 2024, § 6). De allí se sigue el principio que Jacobi le atribuye a Spinoza: el “ser en toda existencia”. En Jacobi, entonces, habría una noción de “ser” realista, proveniente de la filosofía spinozista y confirmada por el Kant precrítico.

Brady Bowman ha propuesto otra manera (compatible con la de Sandkaulen) de comprender el carácter spinozista de la filosofía de Jacobi. Según Bowman (2004, p. 161), Jacobi toma ideas de Spinoza para pensar contra el propio Spinoza. Si bien Bowman decide no indagar cuán coherente sea tal procedimiento, considera que la filosofía de Spinoza le permite a Jacobi formular un “empirismo especulativamente fundado” con el que se puede dejar atrás el idealismo escéptico de Hume (Bowman, 2004, p. 176). El empirismo jacobiano se fundaría en las “nociones comunes” de Spinoza, que provienen de los puntos comunes en que nuestro cuerpo coincide con otros cuerpos. Además, como anota Bowman, Jacobi concibe su concepto de “creencia” según el paradigma de la idea adecuada en Spinoza. En efecto, Jacobi mismo observa, en carta a Herder, que la creencia que él ha formulado en sus *Cartas sobre la doctrina de Spinoza* proviene expresamente de la

Ética del filósofo de Ámsterdam:³ la creencia no sería sino la certeza defendida por Spinoza, entendida como un “tener por verdadero que no proviene de fundamentos” (AB I, p. 390). A esta estructura de la creencia jacobiana se refiere Bowman al señalar su origen spinozista. La creencia jacobiana es asimismo afirmación de la existencia de la cosa que se presenta en esa creencia. Ahora bien, si Bowman, en 2004, limitaba esta característica al carácter spinozista de la creencia, en 2023 la lleva también hacia la creencia humeana. Según Bowman (2023, p. 59), Jacobi pudo compatibilizar la certeza spinozista con la creencia humeana en la medida en que las dos hacen inseparables la creencia/certeza y el contenido de la percepción. En efecto, en Hume, “la creencia no consiste en [...] el acto —por caso— de atribuir un predicado de existencia a una representación singular [...]. Más bien, está inseparablemente fundida con el mismo contenido de la propia percepción” (p. 55). A los fines de nuestra propia interpretación, el realismo spinozista del ser —siguiendo a Sandkaulen— y el realismo de la certeza/creencia —siguiendo a Bowman— nos permitirán echar luz sobre un pasaje crucial de la novela *Allwill*, de Jacobi, extrañamente poco valorado por la bibliografía especializada. En la sección 2 y especialmente en la 2.1, intentaremos demostrar la pertinencia de interpretar a Jacobi atendiendo no solo a aquellos puntos en que se distancia de Spinoza (que son los aspectos usualmente más destacados al estudiar a Jacobi), sino también aquellos en los que se acerca a él.

La estrategia de Bowman, que se apoya en sostener que Jacobi toma ideas de Spinoza para pensar contra Spinoza, tiene un notable antecedente en la interpretación de Hermann Timm. Para Timm, el conocido “salto” de Jacobi va desde Spinoza a Spinoza; es decir, el filósofo de la Ética se hace presente tanto en la “mediación absoluta del saber” como en la “inmediatez de la verdad que se autocerciora” (Timm, 1971, p. 57). Con todo, si por “realismo” se entiende la primacía del objeto sobre el sujeto, Timm no entiende el spinozismo jacobiano como una toma de posición realista:

³ Nos referimos a la conocida definición jacobiana de “creencia”: “una certeza inmediata, que no solo no necesita de ningún fundamento, sino que directamente excluye todos los fundamentos, y simplemente es la *representación misma que coincide con la cosa representada*” (JWA 1,1, p. 115).

[En Jacobi] es afirmada la identidad de la *cogitatio* y la *extensio*, lo ideal y lo real, el yo y el tú, pero no es explícitamente designada. Jacobi enseña su cooriginariedad, sin tematizar el origen como tal. Niega, con Spinoza, la representación de un primado del sujeto sobre el objeto (idealismo), o del objeto sobre el sujeto (realismo). No puede hacerse comprensible ni la cosa a partir del yo, ni el yo a partir de la cosa. Su duplicidad es irreductible. Sin embargo, es indudable, del mismo modo, el *factum* de su coexistencia en la conciencia. [...] Incapaz de explicar por sí misma cómo llega ella a la realidad, o cómo la realidad llega a ella, [la conciencia] se ve en la necesidad de valerse en cada uno de sus actos de una creencia trascendente de unidad (Timm, 1974, p. 209).

Esta creencia, en tanto “experiencia irreflexiva del origen” (Timm, 1974, p. 213), en verdad habría de expresarse no con un “yo creo”, sino con un “se cree en mí, [que] es una autorreferencia anónima, pues ha de ser un verse-a-sí-mismo de Dios” (Timm, 1974, p. 215). Esa experiencia natural, originaria y anónima es rota, con todo, por el trabajo intelectual (cfr. Timm, 1974, p. 146). Advertimos aquí que también esta idea de una ruptura de la experiencia originaria y anónima será importante para nosotros.

Según Timm, los distintos términos (“corazón”, “sentimiento”, “creencia”) con que Jacobi designa la originaria presencia de lo incondicionado en nosotros no son sino “distintos nombres para lo uno Innombrable de la verdad de Dios” (Timm, 1974, p. 215). El riesgo de Jacobi al querer exponer este núcleo interior de la creencia (el “se cree en mí” mencionado arriba) sería el de caer inevitablemente cerca de la inmanencia de la que Jacobi pretendía ser declarado enemigo, “la inmanencia de la mismidad cíclica” (Timm, 1974, p. 139). Este problema, a nuestro entender, es estructuralmente equivalente a lo que veíamos con Bowman y el origen spinozista-humeano de la creencia en Jacobi: el supuesto realismo de Jacobi se apoyaría en bases filosóficas tomadas de autores con intenciones intelectuales muy distintas de las de Jacobi, por no decir opuestas.

Este curioso rasgo del pensamiento jacobiano nos lleva a preguntarnos cómo evita Jacobi esa peligrosa cercanía respecto de las filosofías con las

cuales discute. Para empezar a esbozar nuestra propia respuesta, será clave que reparemos en las ideas de Marco Olivetti. Olivetti destaca (como veremos en breve) la misma sección de la novela *Allwill* en la que nos centraremos nosotros y, así, pone de relieve la obra que nosotros queremos estudiar. Además, su interpretación general de la filosofía de Jacobi nos aportará una buena clave de lectura que buscaremos profundizar. Según Olivetti, es central en Jacobi la distinción entre lenguaje y concepto, porque, con ella, se hace viable una concepción metafísica del primero (cfr. Olivetti, 2013, p. 85; 1971, p. 92). Tal concepción metafísica del lenguaje (proveniente de Johann Georg Hamann) implica verlo como un contenido en sí mismo. De lo que se trata, según Olivetti, es de ver al universo, en cuanto “cosmos”, como un lenguaje divino. Sobre esa concepción del lenguaje como contenido metafísico, Jacobi añadiría otra concepción del lenguaje, ahora entendido como el instrumento de la abstracción conceptual. En tanto esta abstracción conceptual opera sobre el cosmos/lenguaje divino, la actividad de abstracción no degenera en formalizaciones vacías, y el entendimiento, agente de esa abstracción, colabora con la razón en su sentido superior (cfr. Olivetti, 2013, p. 120). La relación humana con el cosmos entendido como lenguaje divino (o como libro divino) ya no pasaría por la percepción, que Olivetti parece confinar al Jacobi de la década de 1780, sino por la comprensión o la hermenéutica, que cobraría vigencia a partir de la década de 1790. Entre el ser humano y el cosmos, Dios sería el tercer elemento que permitiría que se desarrollara la actividad interpretativa (Olivetti, 2013, p. 95; 1971, pp. 88-89). Para nosotros, es crucial que Olivetti encuentre como uno de los pasajes centrales en la formulación de estas ideas las siguientes líneas de la novela *Allwill*, que provienen, en verdad, de la pluma del mismo Hamann (con pequeñas modificaciones realizadas por Jacobi):

Todos los fenómenos de la naturaleza son sueños, aspectos, enigmas, que tienen su significado, su secreto sentido. El libro de la naturaleza y [el] de la historia no son sino cifras, signos ocultos, que exigen una clave, de la cual necesitan también aquellos que creen en una revelación para la interpretación de la misma, [clave] que podría ser ella misma la intención, la única intención de una revelación y la prueba de su inspiración (JWA 6,1, p. 167).

Más adelante ahondaremos en nuestra posición respecto de la sección de *Allwill* de la cual Olivetti extrae únicamente este pasaje. Aquí solo diremos que, a nuestro entender, esa concepción del libro de la naturaleza y de la historia como lenguaje cósmico-divino tiene que ser entendida a la luz de lo que hemos dicho sobre la certeza/creencia de raíz spinozista en Jacobi, puntualmente a la luz de la ruptura de la experiencia original a-subjetiva tal como la entiende Timm. El libro divino del que habla Olivetti es, en nuestra propia interpretación, aquella experiencia originaria que, según Timm, se rompe mediante el trabajo intelectual. La ruptura que señala Timm es, para nosotros, la ruptura del libro divino propuesto por Olivetti. La objetividad metafísica que Olivetti cifra en el cosmos entendido como lenguaje es rota por la actividad intelectual abstrayente (la actividad del entendimiento). Por otro lado, si bien la abstracción no es, desde luego, la creencia/certeza (más bien es aquello que trabaja sobre el fondo dado por la creencia), sí se ve paradójicamente motorizada por esa creencia/certeza, como tendremos ocasión de constatar en la sección 2.2. Nosotros sostenemos que la interrupción originaria en el realismo jacobiano es la que corta el libro divino, es decir, la que corta esa experiencia anónima y a-subjetiva. Aunque es cierto que, en una visión superior, el entendimiento no es rechazado por la razón (como sostiene adecuadamente Olivetti), el entendimiento no esclarecido es víctima de un malentendido que, como veremos, habrá que desandar (sección 2.2). Ese malentendido nace de la confusión entre el sentimiento de lo incondicionado (dado en la creencia original) y la conceptuabilidad de todo lo sensible (necesaria para la abstracción intelectual). Mostraremos esa interrupción originaria en la sección 2 mediante el análisis del uso que Jacobi hace de un breve pasaje de la novela *Tristram Shandy*, de Laurence Sterne. No casualmente, cuando cita el *Tristram Shandy*, Jacobi refiere a un “malentendido” que es, precisamente, el que nosotros intentaremos descifrar.

Para cerrar la introducción y el estado de la cuestión, destacaremos el aporte de dos investigaciones muy apropiadas para justificar nuestro acercamiento a un problema y a un pasaje muy específicos de la obra jacobiana.⁴ Nos referimos a los trabajos de Cornelia Ortlieb y Carmen Götz. No es infrecuente entender a Jacobi como un filósofo que

⁴ Por razones expositivas, no podemos mencionar aquí a Knode (2024) y a Villacañas (1989), a los cuales volveremos en notas *infra*.

seculariza contenidos cristianos. Así lo hace, entre otros,⁵ Carmen Götz: Jacobi sería un ilustrado en el sentido de que comparte los intentos por inmanentizar la trascendencia de la posición de Dios y por convertir en instancia intramundana la promesa de salvación religiosa (Götz, 2008, p. 2). El objetivo ilustrado sería encontrar estructuras mundanas que, a su vez, permitan trascender los límites espaciotemporales. Por ello, Götz considera que es clave el proceso por el cual la cultura moderna se convirtió, en el siglo XVIII, en una cultura principalmente escrita. La escritura permite trascender los límites espaciotemporales, pero además otorga un gran lugar a la producción humana en esa búsqueda de trascendencia. La cultura mediatizada, según Götz, produce un mundo propio, lo que concuerda con el afán creativo ilustrado (cfr. Götz, 2008, pp. 55-56). Lo propio de este afán creativo, sin embargo, no es el empeño por trascender límites (cosa, que según Götz, siempre estuvo presente en la historia humana), sino el intento de producir esa ilimitación de un modo secular e intramundano (cfr. Götz, 2008, p. 476). Jacobi participa de ese intento, al que también critica, sin embargo, principalmente en dos aspectos: la concepción de la razón y la de la naturaleza (cfr. Götz, 2008, p. 479). En la sección 2 se verá que Jacobi elige como momento central de su obra un pasaje de la novela *Allwill* en el que, no azarosamente, se teje un vínculo con otra novela (*Tristram Shandy*), en función precisamente de lo que uno de los personajes de *Tristram Shandy* realiza sobre otra obra literaria, en este caso de Erasmo de Rotterdam. La importancia de la cultura escrita queda de manifiesto, entonces, en tres niveles: Jacobi citando al *Tristram Shandy*, y Sterne citando a Erasmo, todo ello clave, como veremos, para entender el pensamiento de Jacobi.

La transformación de la cultura en cultura principalmente escrita, que Götz caracteriza como un proceso típico del siglo XVIII, puede vincularse con la hipótesis de Cornelia Ortlieb. Para Ortlieb (2010, p. 12), el corazón de la filosofía jacobiana está marcado por la búsqueda de la certeza o de la evidencia en el contexto de la filosofía moderna. A diferencia, por ejemplo, de Klaus Hammacher —que también parte del problema de la certeza como motor de las inquietudes jacobianas (cfr. Hammacher, 1969, p. 10) y que cifra el realismo de Jacobi en un marco dialógico, asentado en la relación yo-tú (Hammacher, 1969, p. 126)—,

⁵ Cfr. Bollnow (1933, p. 235), Verra (1963, pp. xxiii-xxiv y 8) y Olivetti (2013, p. 78); en contra de la utilidad de la categoría de la secularización, cfr. Brunel (2008, p. 572).

Ortlieb considera que la respuesta jacobiana a la búsqueda de certeza es una “nueva forma de certeza filológica” (Ortlieb, 2010, p. 27), de la cual Jacobi sería pionero (Ortlieb, 2010, p. 417). Para Jacobi, según Ortlieb, la filosofía no es un sistema, sino un modo de escritura (cfr. Ortlieb, 2010, p. 426). Esta búsqueda de una certeza filológica se expresaría en el uso constante y peculiar que Jacobi realiza de sus propios textos y, sobre todo, de los textos ajenos. El proceder filológico jacobiano exige incluso tomarse en serio el montaje de citas y fragmentos que Jacobi introduce en sus textos (cfr. Ortlieb, 2010, p. 162). La filosofía, en Jacobi, no podría desprenderse de su “constitución moderna como crítica textual y ciencia de libros” (Ortlieb, 2010, p. 49). Para nuestra interpretación, la posición de Ortlieb es relevante en la medida en que mostraremos que el uso jacobiano de un texto ajeno —puntualmente, la novela *Tristram Shandy*, de Laurence Sterne— es absolutamente pertinente para comprender el pensamiento de Jacobi. Con Ortlieb, creemos que este uso de textos ajenos, fragmentariamente incorporados a la propia argumentación, es fundamental en el proceder jacobiano. El desarrollo de los temas mencionados hasta aquí —el spinozismo de Jacobi y su uso de *Tristram Shandy*— se encontrará en la sección 2.

Si nuestra pregunta se dirigía al modo en que Jacobi puede evitar los riesgos de tomar conceptos elaborados por autores con los cuales él mayormente discrepa, los aportes de Götz y de Ortlieb nos llevan a considerar que los conceptos con los que Jacobi se opone a sus propias fuentes no podrán carecer de un vínculo concreto con la cultura escrita de su propia época. Nuestra hipótesis es que uno de los conceptos con los que Jacobi se diferencia de los autores con los cuales dialoga es el de “interrupción”, que quedará delineado en su rasgo fundamental en la sección 2. Veremos luego, como ejemplo, el modo en que la “interrupción” pretende ser piedra de toque de las insuficiencias de la filosofía kantiana, precisamente una de las más admiradas y a la vez criticadas por Jacobi (sección 3). De ese modo, veremos a la interrupción operando concretamente en el análisis filosófico jacobiano. Cerraremos nuestro trabajo con una conclusión.

2. El realismo spinozista: de la absoluta objetividad al Libro de la Naturaleza

Ya hemos citado un pasaje de la novela *Allwill* que Marco Olivetti considera esencial para comprender el viraje de Jacobi hacia

una concepción hermenéutica del mundo (cfr. JWA 6,1, p. 167). El texto original proviene de Hamann, y sobre él Jacobi realiza leves modificaciones. Olivetti sostiene que, al adoptar el pensamiento de Hamann, Jacobi incorpora a su filosofía la idea de un lenguaje divino configurador del mundo. Nosotros creemos que a este pasaje de la novela *Allwill*, sobre el que se apoya Olivetti, hay que complementarlo (o, más bien, contrarrestarlo) con otro que se encuentra a pocas líneas de él. Este otro pasaje es, a nuestro entender, mucho más importante que el elegido por Olivetti, porque el propio Jacobi lo destaca como el momento en que más acertadamente expresó su propia filosofía. Con todo, tendremos que dar un breve rodeo para llegar a ese pasaje crucial de *Allwill*, rodeo que nos permitirá disponer de un panorama general sobre ciertos términos utilizados por Jacobi que, de no ser explicitados, serían muy oscuros.

Comencemos el rodeo. En el “Prefacio y también introducción a los escritos filosóficos reunidos del autor” (1815), Jacobi realiza una de las últimas miradas de conjunto sobre su trayectoria filosófica, siempre signada por la polémica. Allí escribe:

[...] el camino de la doctrina kantiana necesariamente conduce a un sistema de la absoluta subjetividad, y por ello agrada al entendimiento explicador (que se llama “filosofante”) y que finalmente, sin embargo, no explica, sino que solamente elimina. [La doctrina kantiana] tiene contra sí solo a la razón que se aleja de este camino, la razón que no explica, sino que revela positivamente e incondicionalmente decide, la razón que es la *creencia racional natural*. El camino de la doctrina jacobiana, en tanto lleva con igual necesidad a un sistema de la absoluta objetividad, desagrada al entendimiento que se atiene únicamente a lo conceptual (se llama también la *razón filosofante*), y tiene para sí solo a la razón que no explica, que inmediatamente revela, o la *creencia racional natural* (JWA 2,1, p. 391).

Aquí Jacobi distingue su propia doctrina de la kantiana. Jacobi se presenta como filósofo del “sistema de la absoluta objetividad”, en contraposición con la línea kantiana, que culmina en el “sistema de la absoluta subjetividad”. Más adelante (sección 3) tendremos oportunidad de decir algo más sobre la relación de Jacobi con la filosofía kantiana.

También más adelante nos explayaremos sobre la relación entre Jacobi y Spinoza, pero no es imprudente adelantar que en 1811, cuatro años antes de la publicación del texto recién citado, Jacobi había utilizado la expresión “sistema de la absoluta objetividad” para calificar, no a su propia filosofía, sino a la de Spinoza (cfr. JWA 3, p. 121). Ahora, con todo, nos interesa la nota al pie que Jacobi añade al pasaje recién traducido. En ella, Jacobi refiere a la novela *Allwill*, a las líneas que, para nosotros —y siguiendo en esto al propio Jacobi—, son claves para comprender su filosofía. Veamos en primer lugar la mencionada nota al pie del escrito de 1815:

O bien todos los conocimientos son, en último término, objetivos, es decir, representaciones de algo existente independientemente del sujeto que tiene representaciones, con lo cual ellos también tienen que ser encontrados en el entendimiento divino, pero no de un modo limitado, finito, sino de un modo infinito, que abarque al mismo tiempo todas las relaciones; o bien no hay en absoluto ningún conocimiento objetivo, ningún mundo, ningún Dios. Véase la parte I de la carta 15 de *Allwill*, donde quizás más intuitiva y conceptualmente que en ningún otro lugar de mis escritos se presenta lo que significa para mí la “absoluta objetividad”. Remito especialmente a las páginas 134 y 135 (JWA 2,1, p. 391, n.).

Con esta nota al pie, Jacobi reformula la diferencia entre su doctrina y la kantiana del siguiente modo: su doctrina, la jacobiana (el “sistema de la absoluta objetividad”), sostiene que todos los conocimientos son objetivos, lo cual equivale a decir que, de algún modo, también se encuentran en el entendimiento de Dios; la doctrina kantiana, en cambio, en tanto “sistema de la absoluta subjetividad”, tiene que concluir afirmando que no hay verdadero Dios, ni verdadero mundo, ni verdadero conocimiento. Para aclarar su propia posición, Jacobi remite a las páginas 134-135 de la novela *Allwill*, correspondientes al primer tomo de las obras completas editadas en vida de Jacobi.⁶ Hemos llegado,

⁶ Cfr. JWA 2,2, p. 758, donde los responsables de la edición crítica de este tomo confirman que Jacobi se está refiriendo a las páginas que nosotros destacaremos. Dadas las numerosas ediciones de *Allwill*, podría haber alguna

después del rodeo, al pasaje que nosotros consideramos crucial, en línea con el mismo Jacobi. En él, leemos lo siguiente:

Parecemos ser un soplo, a menudo solo una sombra de un soplo, o como dijo un viejo poeta: el sueño de una sombra. Pero un ser que no fuera más que sombra, un ser que fuera solo sueño, sería un absurdo. Nosotros *somos*, nosotros *vivimos*, y es imposible que haya una especie de vida y de existencia que no sea una especie de vida y existencia del mismísimo y *más alto ser*. Sonidos, colores y todo lo que nosotros podríamos considerar, de otra manera, como mero juego de los sentidos e ilusiones insustanciales [*wesenslose Täuschungen*] surgirá nuevamente, alguna vez, como intuición de lo verdadero, a partir de un contexto más grande [*wird einmal als Anschauung des Wahren aus einem größeren Zusammenhange neu hervorgehen*], y nos dejará conocer el fundamento del malentendido [*Mißverständes*] que tan indescriptiblemente nos inclina a querer *introducir raspando* [*hineinradieren*] en el Libro de la Naturaleza, siempre, un sentido [*Sinn*] mejor (JWA 6,1, p. 168).

Para comprender estas líneas, ciertamente poco diáfanas, hay que tener en cuenta el “sistema de la absoluta objetividad” defendido por Jacobi. El sistema de la absoluta objetividad nos decía que todos los conocimientos tienen que encontrarse en el entendimiento divino en tanto representaciones de algo que existe con independencia de nuestras representaciones. Los conocimientos se encuentran en Dios “de un modo que abarque al mismo tiempo todas las relaciones”, un modo que no es comparable con la finitud de nuestro conocimiento. Ahora, en el pasaje recién citado de la novela *Allwill*, Jacobi nos dice que “es imposible que haya una especie de vida y de existencia que no sea una especie de vida y existencia del mismísimo y *más alto ser*”. Todo lo que es vida, todo lo que tiene ser, se encuentra en el más alto ser, que es el propio Dios. Se trata aquí del mismo principio del sistema de la absoluta objetividad:

confusión sobre cuáles son las páginas apuntadas por Jacobi. Con todo, la referencia está confirmada porque ya en carta a Jean Paul del 16 de marzo de 1800 (cfr. NA I, p. 239) Jacobi señala el mismo pasaje, pero con la paginación correspondiente a la edición de *Allwill* de 1792.

todo se encuentra en Dios. Incluso los sonidos y los colores, usualmente desacreditados como ilusiones, han de ser considerados, en un “contexto más grande”, como “intuición de lo verdadero”.⁷ A la vez, percatarse de que todo, incluidas las mal llamadas ilusiones sensoriales, ha de ser visto como “intuición de lo verdadero” —más allá de que no podamos conceptualizar ese “contexto más grande” (el entendimiento divino) en el que todo es verdadero— nos permitirá comprender el origen del malentendido que nos lleva a distorsionar el Libro de la Naturaleza por querer introducir en él un supuesto sentido mejor (tal como se lee en el pasaje de la novela *Allwill*).

Articularemos en tres apartados (2.1, 2.2 y 2.3) las preguntas que nos depara el breve texto de *Allwill* en el que, según Jacobi, está expuesta con precisión la “absoluta objetividad”. Descifrar las oscuridades de este texto es clave para entender, entonces, el realismo jacobiano: 1) ¿afirmar que todos los conocimientos se encuentran en el entendimiento divino, y afirmar que toda vida es una especie de vida del más alto ser, no equivale acaso a convalidar posiciones filosóficas spinozistas o panteístas, contra las cuales siempre combatió Jacobi? 2) ¿De qué se trata ese malentendido que, según Jacobi, nos lleva a querer introducir algo —un sentido mejor— en el Libro de la Naturaleza? 3) Si Jacobi encuentra en este párrafo su “sistema de la absoluta objetividad”, ¿podemos

⁷ Este carácter verdadero de las “ilusiones” sensibles no es tenido en cuenta por Knode (2024), quien ofrece una excelente interpretación de las novelas jacobianas desde el punto de vista de las escenas “idílico-sentimentales” de descripción de la naturaleza. Según Knode, en las versiones tardías de las novelas (correspondientes a la década de 1790) Jacobi utiliza la descripción idílico-sentimental de la naturaleza para exponer una conciencia extremadamente subjetivista, casi solipsista, que incluso parece perder la capacidad de distinguir entre la percepción efectiva y la imaginación (cfr. Knode, 2024, pp. 298, 313 y 439). Así, las escenas idílicas de descripción de la naturaleza mostrarían una subjetividad hundida en la ilusión sensible y abrumada por el exceso de estímulos sensoriales. De ese modo, Jacobi criticaría —según Knode— al subjetivismo de la conciencia. Con esto, Knode señala un punto absolutamente correcto. Sin embargo, no toma en cuenta que Jacobi mismo elige el pasaje de *Allwill* que estamos analizando como la mejor expresión del “sistema de la absoluta objetividad”. En otras palabras, Knode no toma en cuenta el carácter verdadero de la ilusión sensible. La ilusión sensible no puede ser la última palabra de la filosofía de Jacobi, pero sí ha de ser más que mera ocasión de perdición para una conciencia hipersubjetivista.

extraer de allí alguna conclusión relativa al método de su pensamiento? ¿Hay algo en este pasaje que nos conduzca al modo en que Jacobi, según Ortlieb y Götz, construye su filosofía a partir del vínculo con la cultura escrita?

2.1. Jacobi y Spinoza

La cercanía de Jacobi con Spinoza ya ha sido corroborada en nuestra introducción. Allí observamos que Sandkaulen considera que, para Jacobi, Spinoza es un filósofo realista, en tanto el principio de su filosofía es el del “ser en toda existencia”. También tomamos en cuenta la interpretación de Bowman, que encuentra el origen de la creencia jacobiana en la certeza spinozista, en tanto esta es inseparable del contenido mismo que se afirma como existente en esa certeza. También Timm creía encontrar en Jacobi un principio spinozista puesto en juego contra otro factor spinozista, a saber, la inmediatez de la intuición del final de la *Ética* contra la mediación intelectual con que se inicia la obra mayor de Spinoza. La creencia jacobiana, para Timm, carece, en verdad, de sujeto: no se expresa con un “yo creo”, sino con un “se cree en mí”. Tal unidad originaria sería rota, según Timm, por la actividad intelectual.

Gracias a focalizarnos en el pasaje de la novela *Allwill* que el propio Jacobi destaca como clave de su filosofía, podemos proponer una perspectiva sobre la relación entre Jacobi y el spinozismo que los autores mencionados no toman en cuenta.⁸ Este pasaje —creemos nosotros— se conecta directamente con la “polémica del spinozismo”, originada por la publicación de las *Cartas sobre la doctrina de Spinoza al señor*

⁸ Hasta donde sabemos, la bibliografía especializada ha dedicado escasísima atención a las breves líneas de *Allwill* que aquí nos interesan. Una excepción es Wilde (1894), que se limita a advertir que en ellas solo se hace presente una “facultad de creencia que testifica la existencia de una realidad que encuentra expresión en las apariencias de los sentidos” (p. 68). Por su parte, Villacañas (1989, pp. 439-440) también cita el pasaje de *Allwill* que nosotros hemos destacado. Lo interpreta en función de su concepción general sobre la filosofía de Jacobi como una teoría alérgica a lo sensible, ascética y nihilista (negadora de la realidad material e histórica). Villacañas (1989, p. 440) parece leer el pasaje en cuestión en clave cristiana y encuentra en Jacobi una filosofía fuertemente reacia a lo sensible. No coincidimos con él sobre este último punto; nos inclinamos por las interpretaciones que ven en Jacobi cierta cercanía con el empirismo (cfr. Stiening, 2021; Bowman, 2023). Por último, un breve comentario sobre el mencionado pasaje de *Allwill* puede hallarse en Ortlieb (2010, pp. 205-206).

Moses Mendelssohn, en el año 1785. En ellas, Jacobi cuenta que, en un diálogo privado, Lessing le confesó que adhería a la filosofía spinozista. Dejaremos de lado el detalle de esa conversación. Nos remitiremos directamente a lo que puede aportar luz a nuestro problema, el de saber cómo hasta las ilusiones sensibles (según el pasaje de *Allwill*) pueden ser verdaderas en un contexto más amplio, que sería el pensamiento divino. Moses Mendelssohn, amigo de Lessing y eminente representante de la Ilustración alemana, es el destinatario de las cartas en las que Jacobi expone la filosofía de Spinoza. En el intercambio epistolar, Mendelssohn ya advierte algo que actualmente los investigadores confirman: Jacobi sostiene ideas spinozistas. Pero Mendelssohn va aún más lejos y, sin rodeos, le atribuye a Jacobi la que posiblemente fuera la única idea spinozista que este rechazaría con total convicción. Veamos *in extenso* este momento de la carta de Mendelssohn, que Jacobi incluye en su propio libro:

En la página 26⁹ me encuentro con un pasaje sencillamente inentendible para mí. “El pensamiento —dice usted— no es la fuente de la sustancia, sino que la sustancia es la fuente del pensamiento. Así, hay que asumir algo no pensante como lo primero, antes del pensar; algo que, si bien no en el orden de la realidad,¹⁰ sí tiene que ser pensado como lo primero y anterior en la representación, según el ser y la naturaleza interna”. Me parece que usted quiere pensar aquí, con nuestro amigo, algo que no es un pensamiento; quiere hacer un salto al vacío, hacia donde la razón no puede seguirnos. Quiere pensar algo que antecede a todo pensar, y que por tanto no puede ser pensable ni siquiera para el entendimiento más perfecto. Estimo que la fuente de todos estos aparentes conceptos es que usted considera que la extensión y el movimiento son la única materia y los únicos objetos del pensamiento, e incluso que estos lo son solo en tanto existen efectivamente. No sé con qué

⁹ Mendelssohn, que le escribe a Jacobi, se refiere a la primera edición de las *Cartas sobre la doctrina de Spinoza*.

¹⁰ Mendelssohn escribe allí “posibilidad”, pero Jacobi aclara que esto debe haber sido un error de copiado de Mendelssohn, porque su libro dice *Wirklichkeit*. Traducimos aquí siguiendo directamente la aclaración de Jacobi.

fundamento supone esto como dado. ¿No puede el ser pensante ser materia y objeto para sí mismo? Sabemos cómo es para nuestro ánimo padecer dolor, hambre, sed, frío o calor, o temer, esperar, amar, detestar, etc. Llame usted a estos pensamientos “conceptos” o “sensaciones” o “afecciones” del alma; alcanza con [ver] que el alma, en todas estas afecciones, no tiene ni extensión ni movimiento como objetos. Incluso en las mismas sensaciones sensibles: ¿qué tienen en común el sonido, el olor, el color o el gusto corporal con la extensión y el movimiento? Bien sé que Locke ha acostumbrado a los sabios a tomar la extensión, la impenetrabilidad y el movimiento como *qualitates primitivas*, y a reducir a ellas las apariencias de los restantes sentidos, como *qualitates secundarias*. ¿Pero qué fundamento tiene el spinozista para dar esto por válido? ¿No puede haber, en fin, un espíritu que piense como meramente posibles la extensión y el movimiento, aun si no existieran efectivamente? Para Spinoza, que tomó a la extensión como una propiedad de la sustancia única infinita, esto tiene que ser tanto más aceptable (JWA 1,1, pp. 177-178).

Mendelssohn le atribuye a Jacobi aquel principio spinozista que Jacobi rechaza con absoluta decisión: la hipótesis de que la sustancia es la fuente del pensamiento y de que, entonces, el pensamiento no es lo primero. De allí, Mendelssohn desprende la consecuencia de que Jacobi no aceptaría otro objeto del pensamiento que la extensión y el movimiento. ¿Pero acaso —se pregunta Mendelssohn— el ser pensante no puede pensarse a sí mismo? ¿Acaso no puede pensar algo que no tiene correlato ni en la extensión ni en el movimiento? ¿Acaso no somos conscientes de tal posibilidad en las más sencillas sensaciones visuales, auditivas, olfativas, etc.? Un spinozista, sostiene Mendelssohn —y Jacobi es un spinozista, según Mendelssohn—, no tiene por qué aceptar la distinción entre cualidades primarias y cualidades secundarias, distinción que supone que aquellas refieren a algo real o existente, mientras que estas no son más que sensaciones subjetivas. En otras palabras, un spinozista, en tanto sostiene la primacía de la sustancia única, no tiene fundamento alguno para afirmar que hay algo fuera de la sustancia. Un spinozista no tiene margen para pensar que haya

pensamientos que no son parte del pensamiento divino.¹¹ El desafío que Mendelssohn le lanza a Jacobi podría formularse así: ¿por qué usted, Jacobi, siendo spinozista —es decir, aceptando que la sustancia es lo primero y que el pensamiento es por sí mismo una expresión de la sustancia—, no acepta que el pensamiento puede pensarse a sí mismo, sin tener por objeto la extensión? Para un spinozista tendría que ser aceptable que el pensamiento se pensara a sí mismo, porque el pensamiento spinozista tiene un orden propio, independiente de la extensión. Allí están los colores, los sonidos, los sabores, para demostrar que el pensamiento puede pensar sin tener algo extenso como correlato directo de ese pensamiento en tanto tal. Con las más simples “cualidades secundarias” se advierte que el pensamiento puede pensarse a sí mismo y que, por lo tanto, no depende de la sustancia (ni de la extensión, ni del movimiento).

Jacobi se espanta de que Mendelssohn le atribuya la idea de que la sustancia es anterior al pensamiento (cfr. JWA 1,1, p. 53, n.). Sostener que la sustancia es anterior al pensamiento es precisamente la base del fatalismo, aquello contra lo cual Jacobi no se cansa de polemizar. Nada más lejos de él que esa premisa spinozista que le atribuye Mendelssohn. Sin embargo, creemos que la estrategia de Jacobi no es el mero rechazo de lo que le imputa su corresponsal. Por el contrario, Jacobi acepta aquello que hemos llamado el “desafío” de Mendelssohn, el reto de pensar spinozísticamente que todos los pensamientos, en tanto pertenecientes a una entidad divina, son verdaderos. Eso es precisamente lo que Jacobi hace en el pasaje de *Allwill*, del año 1792, apenas unos años después del intercambio con Mendelssohn. Allí, como ya vimos, Jacobi dice que incluso los colores, los sonidos y las ilusiones sensoriales en general han de ser verdaderas en un “contexto más grande”. En 1815, como vimos, Jacobi agrega que precisamente eso es el “sistema de la absoluta objetividad”: sostener que todos los conocimientos son verdaderos en el entendimiento divino, que comprende todas las cosas en todas sus relaciones. El sistema de la absoluta objetividad es el sistema que permite afirmar que incluso las ilusiones sensoriales son verdaderas, siempre y cuando las remitamos a un “contexto más grande”, que no es sino el pensamiento divino capaz de abarcar infinitamente todas las relaciones.

¹¹ Para el modo en que Jacobi interpreta el pensamiento y el entendimiento en Spinoza, cfr. JWA 1,1, pp. 101 y ss.

El sistema de la absoluta objetividad jacobiana afirma positivamente aquello que Mendelssohn le reclamaba a Jacobi: el pensamiento puede pensarse a sí mismo, pues no depende de una sustancia anterior a él. De esta manera, Jacobi asume la posición spinozista que le adjudicaba Mendelssohn ("spinozista" al menos en el sentido que le estamos dando a esa denominación en este contexto) y sutilmente, sin embargo, va alejándose del spinozismo. Pone a pensar a Spinoza contra Spinoza.

Veamos la articulación de este doble movimiento. 1) Jacobi toma la posición spinozista en la medida en que, según Spinoza, "todas las ideas son en Dios [...] y, en cuanto referidas a Dios, son verdaderas [...] y adecuadas" (Spinoza, *Ética*, II, prop. XXXVI; 2009, p. 158). Esto mismo es lo que Jacobi llama, en 1815, "sistema de la absoluta objetividad"; por lo tanto, no ha de sorprendernos observar que, ya en 1811, hubiera utilizado exactamente esa misma denominación para calificar a la filosofía de Spinoza (cfr. JWA 3, p. 121). El sistema de la absoluta objetividad es el de Spinoza en tanto las ideas, en su condición de pertenencia al pensamiento divino, son todas verdaderas. De allí que Jacobi acepte incluso aquello que parecería ser una consecuencia extrema del desafío de Mendelssohn: defender una posición según la cual hasta las ilusiones sensoriales tienen su cuota de verdad al ser consideradas dentro del pensamiento divino. 2) Con todo, Jacobi no puede darse por satisfecho con este realismo spinozista. Jacobi efectivamente cree que el "sistema de la absoluta objetividad" spinozista extrae su contenido del ámbito de la extensión. Por ello, Jacobi repite, en 1811, los términos de la discusión con Mendelssohn, pero esta vez haciendo propia la advertencia de su antiguo adversario: "Para Spinoza, el ser extenso de todo lo que es y opera [*Seyende, Wesende und Wirkende*] objetivamente (en su lenguaje, formalmente) es lo auténticamente real; el ser pensante, por el contrario, solo tiene representaciones de acuerdo con aquel" (JWA 3, p. 121). El sistema spinozista de la absoluta objetividad no puede salir de la esfera del pensamiento que afirma la existencia de la extensión. Es cierto que el spinozismo permite sostener que todas las ideas son verdaderas en el "contexto más grande" del pensamiento divino, pero también es cierto que todas las ideas posibles dentro del marco del spinozismo son —al menos según Jacobi— ideas cuyo contenido proviene de la extensión. El spinozismo no puede terminar de afirmar la anterioridad del pensamiento, precisamente porque el pensamiento spinozista es solo un atributo de la sustancia y, por eso mismo, puede afirmar la existencia real de las cosas externas, pero no ir más allá de estas. Spinoza no pudo

mostrar, como tampoco pudo Leibniz, cómo es que “el espíritu [puede existir y actuar] antes de la materia, el pensamiento antes del objeto” (JWA 1,1, p. 27). En consecuencia, el realismo spinozista necesita ser interrumpido porque debemos afirmar no solo la verdad interna del pensamiento, sino también la primacía del pensamiento.

Apropiándonos de los términos de Timm (según los vimos arriba), la solución jacobiana radica en la ruptura de la experiencia spinozista anónima —en la que son cooriginarios pensamiento y extensión— por medio de la actividad pensante. Así, esta ya no queda reducida a la condición de mero atributo de la sustancia, sino que es reconocida como lo primero. Precisamente la afirmación de la primacía del pensamiento es lo que Jacobi le reconoce al idealismo kantiano-fichteano: “La filosofía crítica tiene de bueno y saludable que ella ha mostrado [...] que la razón es lo primero, que tenemos que partir de la razón y que el espíritu tiene que ser pensado necesariamente antes del cuerpo [...]” (JWA 2,2, p. 538). El realismo, entonces, ha de ser interrumpido o “roto” por el idealismo. El realismo spinozista no puede darnos la captación de la primacía del espíritu que sí nos da el idealismo, porque aquel llega a afirmar la verdad de las ideas en tanto ideas, pero no la anterioridad del pensamiento. En otras palabras, llega a captar que el pensamiento puede pensarse a sí mismo, pero no hace de ese autopenamiento el punto de inicio de la filosofía y de la realidad en tanto pensada por nosotros. El ingreso al reino del pensamiento como lo primero estaba reservado a la filosofía crítica, pues son Kant y Fichte quienes “han descubierto la más alta mecánica del espíritu humano” (JWA 2,1, p. 207). Veamos, entonces, qué es esta “mecánica”, solo descubierta, no inventada, por Kant y Fichte. Esa mecánica es la que rige, desde siempre, toda actividad intelectual humana, impelida a comprender la naturaleza a partir de criterios conceptuales. Con esto, nos aproximaremos también al “malentendido” que, según el crucial pasaje de *Allwill*, nos lleva a querer introducir un mejor sentido en el Libro de la Naturaleza.

2.2. El malentendido de la mecánica del concepto

La actividad conceptual tiene un procedimiento determinado (la construcción del objeto que ha de ser conocido) y una finalidad concreta (el ordenamiento y la estabilización de nuestra relación con la naturaleza en función de nuestras necesidades; en otros términos, la autoconservación). Los seres humanos tienen que construir el objeto del conocimiento porque el mundo objetivo, librado a sí mismo, no mediado

por nuestros conceptos, les es hostil. En la *Carta a Fichte*, Jacobi explica la actividad conceptual —que la filosofía fichteana lleva a su máxima cumbre (JWA 2,1, p. 201 y p. 207)— como un continuo “desprenderse” respecto de una realidad que busca “devorar” al espíritu humano (JWA 2,1, p. 198). El conocimiento produce un efecto de distancia respecto de una realidad amenazante, y lo hace destruyendo el objeto tal como este viene dado naturalmente, para reconstruirlo de acuerdo con nuestros conceptos (cfr. JWA 2,1, p. 203). Tal es la “mecánica del espíritu humano” en lo que concierne al conocimiento conceptual de la naturaleza. No hay otro modo de acceder a este tipo de conocimiento, pues es “imposible que comprendamos aquello que no estamos en condiciones de construir” (JWA 1,1, p. 258).

Precisamente esta “mecánica del espíritu humano”, insuperable en el conocimiento de la naturaleza, produce el “malentendido” que nos lleva a entrar en relaciones equívocas con ella. En carta a Johann Franz Laharpe, escrita en francés, Jacobi recurre de nuevo al término “malentendido”. Allí explica que las ciencias naturales, por el éxito que han tenido en su tarea de conocer la naturaleza, han ocasionado el “malentendido” (*malentendu*) del que es víctima la filosofía europea, a saber, el de creer que hay una “manera fija de ser gobernado por la sola razón”, como si se pudiera trasladar el método constructivo de las ciencias naturales a todas las áreas de la vida humana (JWA 5,1, p. 175). En algún momento se verá —así lo cree Jacobi— que es inviable querer “construir el fondo con la forma, los deseos con la voluntad, el ser razonable con su concepción abstracta, la virtud con su precepto, todas las causas con los efectos de estas causas” (JWA 5,1, p. 175). La actividad constructiva del entendimiento pretende llegar al punto cero, lógicamente incondicionado, desde el cual fuera posible construir conceptualmente todo lo que hace a la vida del ser humano en su relación con el mundo natural y cultural. El problema del afán omniconstructivo del entendimiento es que él mismo (el afán) nace del “sentimiento racional” (esto es, no propio del mero entendimiento) de lo incondicionado: “sin el sentimiento racional positivo de algo más alto que el mundo sensible, el entendimiento nunca habría salido del ámbito de lo condicionado, y no habría llegado ni siquiera al concepto negativo de lo incondicionado” (JWA 2,1, pp. 412-413; las itálicas son de Jacobi). El concepto negativo de lo incondicionado es el concepto vacío, indeterminado, a partir del cual el entendimiento quiere lanzarse a construir su mundo. Sin el sentimiento de lo incondicionado positivo, el entendimiento no habría

llegado jamás a ese concepto. Con esto, tenemos las herramientas para detectar el origen del “malentendido”. El “malentendido” producido por el conocimiento natural se debe a la conjunción de dos factores bien distintos: el sentimiento positivo de lo incondicionado (irreducible al concepto) y la urgencia de nuestra actividad conceptual para poder vivir en un mundo reactivo a nuestra autoconservación. El malentendido se debe a la unión del sentimiento inmediato de inconceptuabilidad y la conceptuabilidad de todo lo sensible: cuanto más construye el entendimiento, más incondicionado debe ser su punto de partida (para abarcar así todas las posibles realidades que luego quiera construir); pero para poder construir a partir de eso incondicionado, más vacío deberá pensarlo. De tal manera, el entendimiento transforma lo incondicionado (cuyo origen es racional-espiritual) en mero caos indeterminado. Reduce toda realidad a un fundamento amorfo para desde allí legitimar su tarea constructiva. Por ello, tiene que negar el sentimiento positivo de lo incondicionado, que fue, en verdad, lo que lo llevó hacia el “concepto negativo de lo incondicionado”. Jacobi identifica el sentimiento positivo con la libertad. De allí se sigue que el entendimiento, que termina por negar ese sentimiento, derive en el fatalismo.

2.3. Jacobi como escritor lector

El breve pasaje de *Allwill* que Jacobi considera como la mejor expresión del “sistema de la absoluta objetividad” nos ha servido de puerta de entrada para varias de las discusiones y de los conceptos presentes a lo largo de su producción filosófica. Su utilidad, con todo, no se queda allí. También nos servirá para comprobar y profundizar las hipótesis de Götz y Ortlieb: Jacobi produce su filosofía en laboriosa relación con la cultura escrita. A su vez, podremos vincular esta hipótesis, concerniente al método jacobiano de composición textual y argumentación filosófica, con nuestra propia hipótesis respecto de la filosofía jacobiana, a saber: Jacobi hace de la interrupción una figura adecuada para señalar los límites de los sistemas filosóficos con los cuales discute.

El malentendido producido por el extraño maridaje entre sentimiento racional y entendimiento “nos inclina a querer *introducir raspando* [*hineinradieren*] en el Libro de la Naturaleza, siempre, un sentido [*Sinn*] mejor” (JWA 6,1, p. 168). Como aclara Jacobi en nota al pie, “estas últimas palabras parecen aludir a un pasaje del *Tristram Shandy*, parte III, cap. 37” (JWA 6,1, p. 168). Jacobi se refiere, desde luego, a la novela de Laurence Sterne. Tenemos aquí, por lo tanto, en un pasaje que el

propio Jacobi señala como clave en su filosofía, la referencia explícita a la obra de un novelista del siglo XVIII. Podemos comprobar, como decíamos, que Jacobi recurre a citas de obras ajenas como medio de expresión discursiva de su propio pensamiento. Analicemos con mayor detenimiento, por ello, el momento de *Tristram Shandy* utilizado por Jacobi.

En el capítulo 37 del tercer volumen de *Tristram Shandy*, el padre de Tristram, Walter Shandy (un hombre apasionado por el trabajo intelectual), intenta descubrir el sentido oculto de unas breves palabras de los *Colloquia* de Erasmo de Rotterdam, dedicadas a las narices (cfr. Erasmo de Rotterdam, 1677, pp. 24-27). Tras una frustrante primera lectura, según nos cuenta Tristram, Walter vuelve a la carga, acompañado por su hermano Toby:

Como el diálogo era de Erasmo, mi padre [...] volvió a leerlo y releerlo con gran aplicación [...] pero tampoco de esa manera logró sacar nada en limpio. –Tal vez haya otro significado, aparte del expreso, dijo mi padre; –los sabios, hermano Toby, no escriben diálogos sobre las narices largas porque sí. –Voy a estudiar los sentidos místico y alegórico; –todavía hay aquí material suficiente para volver sobre ello, hermano. Y mi padre siguió leyendo. –Bien; me parece que ahora es necesario que les comunique a sus reverencias y señorías que, aparte de los numerosos usos de tipo náutico que Erasmo atribuye a las narices largas, el dialoguista afirma que también tienen sus ventajas de carácter doméstico; pues en caso de apuro, –y falta de fuelle, servirá estupendamente *ad excitandum focum* (para avivar el fuego). La naturaleza, al otorgarle sus dones a mi padre, se había mostrado desmedidamente pródiga, y había sembrado las semillas de la crítica verbal, en su interior, a tanta profundidad como las de los demás conocimientos de que era poseedor; –de modo que cogió su cortaplumas y se puso a experimentar con la frase a ver si podía arañarle algún otro sentido más satisfactorio [*scratch some better sense into it*]. – – ¡Hermano Toby!, gritó mi padre: ¡He logrado aprehender el sentido místico de Erasmo a falta de una

sola letra, que no acaba de encajar! – – Pues hay que reconocer en toda justicia, hermano, respondió mi tío, que estás ya bastante cerca. – – ¡Bah!, exclamó mi padre al tiempo que se ponía de nuevo a raspar [*scratching on*]: – lo mismo daría que estuviese aún a siete millas. – ¡Lo he conseguido!, dijo mi padre haciendo chasquear los dedos: – ¡Mira cómo he corregido el sentido [*sense*], hermano Toby! – – Pero has desfigurado una palabra, replicó mi tío Toby. – Mi padre se caló los lentes, – se mordió el labio inferior – y, en un acceso de ira, arrancó la hoja del libro (Sterne, 1980, p. 167; 1978, pp. 200-201).

Javier Marías y Howard Anderson, en sus respectivas ediciones, indican en nota la razón por la cual Walter Shandy no queda conforme con el nuevo sentido que extrajo del libro de Erasmo (o, mejor dicho, que introdujo en él). Sterne hace una pequeña modificación en el texto de Erasmo, y en lugar de decir que las narices sirven para avivar el brasero (*foculo*) –tal como se lee efectivamente en el coloquio erasmiano– le adjudica al humanista la siguiente expresión: *ad excitandum focum* (“para avivar el fuego”). Sobre esta palabra, *focum*, trabaja con su cortaplumas el padre de Tristram, literalmente borrando una parte de la palabra. El resultado es el siguiente, tal como aclara Javier Marías: “*Focum* puede haber quedado convertido en *ficum* (higo, es decir, el sexo femenino) o en *locum* (que, aparte de su significado de lugar, también tenía en latín el de sexo de la mujer)” (Marías, en Sterne, 1978, p. 661, n. 103). En cualquiera de los dos casos, la expresión obtenida por Walter sería algo así como “para avivar el órgano sexual femenino”. Con su cortaplumas y su procedimiento de raspar una parte del texto erasmiano, lejos estuvo Walter de extraer un sentido “místico” o “alegórico”. Podemos ver en juego aquí un problema filosófico recurrente en Jacobi (el progresivo avance del naturalismo en la filosofía moderna) ligado al uso concreto de un texto literario. Veamos cómo se da esa unión.

Jacobi parece citar al mismo tiempo dos traducciones alemanas distintas de la novela de Sterne: de una de ellas parece tomar el verbo *hineinradieren* (en lugar de *herauskratzen*); de la otra, el sustantivo *Sinn* (“sentido”), en lugar de *Verstand* (“entendimiento”).¹² Creemos que Jacobi

¹² Una de las traducciones es la siguiente: “So, daß er sein Federmesser herauszog, und mit der Phrasis eine Erfahrung anstellte, ob er nicht einen bessern

eligió deliberadamente el verbo *hineinradieren* para indicar el intento de embutir un mejor sentido en la naturaleza. En inglés, la expresión en cuestión es *scratch some better sense into it*, lo cual indica que, al “raspar” o “arañar” la palabra, Walter Shandy pretende introducir algo en el texto (*into it*), como si suprimiendo algo de lo impreso en el libro pudiera ingresar en él un significado mejor. Esa acción de borrar y al mismo tiempo introducir algo mediante el mismo borrado está expresada en la palabra *hineinradieren*: *hinein* indica dirección “hacia adentro”, y *radieren* quiere decir, extrañamente, dos cosas casi opuestas: “borrar” y “grabar”, en el sentido artístico de “grabado”. Así, Walter Shandy borra una parte del texto impreso y, en el mismo acto de borrar, consigue “grabar” en él un nuevo sentido que, pese a su intención, no es ni mejor, ni místico, ni alegórico. Por el contrario, el sentido que Walter introdujo en el texto es absolutamente natural, en su acepción de “animal”. Exactamente lo mismo les ocurre a la ciencia naturalista y a la filosofía que explica su mecánica, es decir, al idealismo: queriendo encontrar el sentido de la naturaleza, no obtiene más que mera naturaleza, mera animalidad que no sabe de Dios — “[...] poseer razón y saber de Dios es una sola cosa, así como una sola cosa es no saber de Dios y ser animal” (JWA 3, p. 104) —. El apego a la letra, como el que se ve en Walter Shandy, termina no en la muerte del espíritu (pues el espíritu es divino; cfr. JWA 3, p. 104), sino en su distorsión, porque lo hace coincidir palmo a palmo con la naturaleza.

En la historia de la filosofía, Jacobi ve esa progresión (o regresión) constante desde el racionalismo mecanicista a la identidad entre espíritu y naturaleza en el largo periplo que abarca a Spinoza, Kant, Fichte y Schelling. Con Schelling, en tanto culminación y reunión de las filosofías de los otros tres autores, se lleva al punto máximo la “identidad del ser y la conciencia” (JWA 3, p. 120), identidad que no es sino lo propio del animal, en el que “ser y conciencia coinciden perfectísimamente” (JWA 3, p. 102). Así, la preferencia de la palabra *Sinn* (“sentido”) en lugar de *Verstand* (“entendimiento”) también puede vincularse con la polémica de Jacobi contra la filosofía racionalista-idealista: lo que nos lleva a “raspar” el Libro de la Naturaleza no es meramente el entendimiento, sino la

Verstand hinein radiren könnte” (Sterne, 1774, p. 179). La otra traducción reza: “Er nahm also das Federmesser heraus, und versuchte das Experiment, ob er nicht einen bessern Sinn aus der Sentenz heraus kratzen könnte” (Sterne, 1770, p. 317).

intención de ver el “sentido” de toda la naturaleza; es decir, la intención de darles “sentido” incluso a esas sensaciones que — como los colores o los sonidos — parecen no tenerlo y se nos muestran superficialmente como meros engaños. El origen de la tendencia hacia lo incondicionado, hacia el “sentido”, no viene del mismo entendimiento, sino del sentimiento no intelectual que tenemos de lo incondicionado. Como tenemos este sentimiento, el entendimiento no puede quedarse en lo condicionado, y busca el “sentido” del Libro de la Naturaleza; pero como, al mismo tiempo, el entendimiento no puede construir el sentido incondicionado, opta por destruir todo aquello que, en su incomprensibilidad, le haga recordar a lo incondicionado-inconstruible. Las ilusiones sensoriales, en tanto cualidades, no pueden ser construidas (cfr. JWA 1,1, p. 258); tal como Mendelssohn le recordaba a Jacobi, en esas sensaciones tenemos la prueba de un pensamiento que no se deja reconectar con la extensión y el movimiento. El entendimiento, entonces, opta por intentar reducir ese peculiar modo de pensar que se da en las cualidades secundarias: en lugar de remitirlo de un modo realista-spinozista a un entendimiento divino, lo encierra en sí mismo, en una conciencia “pura”. Con la conciencia pura, el entendimiento destruye — o reduce a cero — aquellas sensaciones. Justamente esto es lo que Kant realizará, según Jacobi, en la sección de las “Anticipaciones de la percepción” de la *Crítica de la razón pura*. Este argumento nos ocupará en la sección 3 de nuestro trabajo. En el caso del *Tristram Shandy*, la destrucción del Libro de la Naturaleza por parte del entendimiento está simbolizada en el momento en que Walter, ya frustrado, arranca la hoja del libro que está leyendo. Walter (es decir, el laborioso entendimiento) buscó el “sentido místico” del libro de Erasmo (que vendría a ser, para nosotros, el libro divino) para terminar destruyendo el libro (la interrupción del orden de sus páginas).

En lugar de la reducción idealista a la conciencia “pura”, Jacobi acepta, como momento spinozista de su filosofía, la verdad de esas sensaciones en tanto ideas pertenecientes al entendimiento divino y correspondientes a algo que existe con independencia de nuestras representaciones. Sin embargo, Jacobi no pretende simplemente rechazar el rol del entendimiento. Mediante la inclusión del pasaje de *Tristram Shandy*, Jacobi vuelve contra sí misma la actividad del entendimiento. Este, originariamente, toma por verdaderas todas sus sensaciones. Impulsado por el sentimiento racional positivo, busca el sentido incondicionado del Libro de la Naturaleza. Como no puede encontrar eso incondicionado en la naturaleza (sino que solo encuentra más y más naturaleza),

termina arrancando la hoja del Libro, como lo hizo Walter Shandy. El entendimiento, entonces, es la primera interrupción, en sentido lógico: es el que interrumpe esa verdad inmediata de todas las sensaciones para tratar de ordenar el Libro de la Naturaleza. Al mismo tiempo, no puede sino interrumpir este mismo Libro, pues no interrumpirlo (no arrancar la hoja del libro) equivaldría a poder construir lo incondicionado o el “sentido místico” (cosa imposible, según Jacobi). Encontrar ese límite del entendimiento es lo que Jacobi ya indicaba como objetivo de su tarea filosófica en la conversación con Lessing: en lugar de negar lo incondicionado mediante el afán explicativo-constructivo, Jacobi busca “saber dónde comienza el límite” de toda explicación (JWA 1,1, p. 29). Retomando la interpretación de Olivetti, podemos concluir que si bien es cierto que en Jacobi aparece la metáfora del Libro de la Naturaleza, también es cierto que el entendimiento le arranca una hoja a ese Libro. La interrupción que el entendimiento realiza en el Libro de la Naturaleza es necesaria e inevitable: sin ella, no podríamos siquiera responder a nuestras necesidades naturales. Al mismo tiempo, el entendimiento, al arrancar la hoja del Libro, muestra sus propios límites.

La interrupción es, además, un principio estructural de la narración de *Tristram Shandy*. La novela de Sterne se caracteriza por las múltiples veces en que el relato es interrumpido, recommenzado, suspendido, desviado hacia temas aledaños, etc. No es casual que Jacobi, después de citar las palabras de Sterne, introduzca una interrupción de la narración. Las palabras de Sterne están, en la novela *Allwill*, en boca precisamente del personaje Allwill, en conversación con la joven Cläre. Tras ser pronunciadas, la conversación se interrumpe: “Fuimos interrumpidos por el anuncio de que la cena estaba servida” (JWA 6,1, p. 168). Es evidente que Jacobi no escribe novelas con la técnica literaria de Sterne (y, por eso mismo, los alcances de la interrupción en Jacobi y Sterne difieren significativamente), pero creemos poder poner en relación la noción de “interrupción” de ambos. Gregori (2013, p. 230) entiende que la interrupción, en *Tristram Shandy*, opera como un principio voluntario que crea vida. Este matiz voluntario y creador de la interrupción es el que Jacobi pretende contraponer a la continuidad de la “cadena ininterrumpida” de la naturaleza (JWA 3, p. 117). Sobre el carácter vital de la interrupción —que, en rigor, exigiría ampliar los objetivos de este artículo— diremos unas breves palabras en la conclusión. Pasemos ahora a la ya anunciada sección sobre Kant.

3. Kant, o la interrupción imposible

3.1. La imposible interrupción de la pureza

El famoso pasaje del “Apéndice sobre el idealismo trascendental”, recordado al comienzo de este artículo, puede ser complementado con otro de 1815, en el que Jacobi sostiene que la razón kantiana no muestra su “facultad más alta” “al final” del sistema, mediante la admisión de la libertad práctica, sino justamente “al comienzo”, en la “presuposición absoluta de una *cosa* en sí, que no se revela a la facultad de conocimiento ni en los fenómenos ni *a través* de ellos, sino únicamente con ellos, de un modo *completamente positivo o místico*, incomprensible para los sentidos y el entendimiento” (JWA 2,1, p. 384). La presuposición de la cosa en sí, que el sistema kantiano necesita pero no puede terminar de afirmar, es el punto más elevado del sistema.

Al dejar en una condición ambigua la existencia de la cosa en sí, Kant no puede terminar de salir del ámbito de las representaciones de la conciencia. Sí puede, efectivamente, deslindar las condiciones del conocimiento, pero no puede salir de la pureza de esas condiciones (a menos que presuponga aquello que no puede presuponer cabalmente). El desafío de Jacobi a los filósofos kantianos es el de mostrar cómo se interrumpe la pureza de la unidad de las condiciones subjetivas de conocimiento, de modo tal que se vea la conexión del conocimiento con algo verdadero. Así, leemos en *Sobre el intento del criticismo de llevar la razón al entendimiento y darle a la filosofía en general una nueva perspectiva* (1802):

Seriamente les pregunto, ¿qué es aquello que [logra] llevar finitud a aquellas tres infinitudes [el espacio, el tiempo y la conciencia], a las dos de la receptividad y a la otra de la espontaneidad? ¿Qué es aquello que fecunda al espacio y al tiempo *a priori* con el número y la medida, y los transforma en un múltiple puro? ¿Qué es aquello que pone en oscilación a la espontaneidad pura, y [nos] lleva de la conciencia *a priori* a la conciencia? ¿Cómo llega a ser su pura vocal una consonante o, más bien, cómo se suspende su soplo insonoro ininterrumpido [*ununterbrochenes*] para al menos ganar,

interrumpiéndose a sí mismo [*sich selbst unterbrechend*],
una especie de vocal, un *acento*? (JWA 2,1, p. 290).

Según Jacobi, el criticismo debería mostrar, si es que quiere ser coherente consigo mismo, cómo surge lo finito y lo determinado desde la indeterminación de las condiciones del conocimiento. Tendría que realizar semejante demostración *a priori*, pues de lo contrario estaría ya dando por válida la exterioridad de lo sensible; esto es, estaría presuponiendo lo que no puede presuponer coherentemente dentro de su sistema. Se ve aquí “el carácter contenidístico [*contentutistico*] y directamente metafísico, en lugar de trascendental, de la interpretación que Jacobi hace del *apriori* kantiano” (Verra, 1963, p. 253). Jacobi se inquieta por el “contenido” de las condiciones del conocimiento. Es ese contenido el que el criticismo no puede explicar. Así, interpelando a los filósofos kantianos, Jacobi escribe estas palabras: “Mi dificultad en relación con sus tres fundamentos fluyentes de unidad, o sus tres continuidades distintas, es la misma: a saber, la interrupción de las mismas *a priori*; su tránsito hacia la multiplicidad a través de una finitización determinante [...]” (JWA 2,1, p. 303). Jacobi no pretende hacer aquello que, según él, el kantismo no puede ofrecer. Solo procura sacar a la luz aquello que opera como presupuesto en otras teorías que, sin embargo, tienden a negar ese presupuesto. En el caso de la filosofía kantiana, Jacobi asume sus conceptos y, desde ellos (esto es, desde la “mecánica del espíritu humano” descubierta por Kant y pulida por Fichte), se pregunta cómo podría explicarse el contenido finito y determinado de nuestras representaciones. El concepto para expresar esta pregunta desde el punto de vista de la pureza continua de las condiciones de conocimiento es el de “interrupción”: ¿cómo se interrumpe esa pureza, de modo tal que conozcamos un mundo efectivo, no puro? Según Jacobi, esa pregunta no tiene respuesta, porque la percepción de objetos realmente existentes fuera del yo es “un milagro incomprensible” (JWA 2,1, p. 390). En la filosofía de Jacobi, hay dos verdades irreductibles, que hay que aceptar en su irreductibilidad: “yo soy y hay cosas fuera de mí” (JWA 2,1, p. 194). Así, si en la sección anterior veíamos que la primacía del pensamiento quebraba la cooriginariedad de pensamiento y extensión, ahora es la extensión la que, veladamente, viene a interrumpir la pureza del pensamiento. En efecto, al afirmar resueltamente que “hay cosas fuera de mí”, Jacobi adhiere al principio más elemental del materialismo, según Olivier Bloch: “Actitud natural: creencia y adhesión

a la inmediatez de la vida concreta — actitud de pensamiento: certeza de la existencia de las cosas [...]” (Bloch, 1985, p. 33). En consecuencia, la pregunta de Jacobi por la “interrupción” de la pureza kantiana no es trascendental, pues va en la dirección contraria a la que Kant tiene en mente cuando define el adjetivo “trascendental” (cfr. KrV A11-B25; Kant, 2009, p. 79). Jacobi se preocupa, a diferencia de Kant, más por los objetos que por nuestro modo de conocer. La tarea del investigador es, según él, “descubrir y revelar la *existencia*... La explicación es para él [*i. e.*, el investigador] un medio, un camino hacia la meta; es su fin próximo, pero nunca el último. Su último fin es aquello que no se deja explicar: lo insoluble, lo inmediato, lo simple” (JWA 1,1, p. 29).

La revelación de la existencia —en la medida en que esto le quepa al investigador— pasa, entonces, por el reconocimiento de lo inexplicable. Jacobi encuentra en Kant la afirmación de lo inexplicable (la cosa en sí), pero solo a medias, pues el propio compromiso trascendental le impide a Kant realizar plenamente esa afirmación. La afirmación jacobiana de la existencia revelada proviene de la certeza spinozista, que es afirmación de sí misma y de su contenido. Esta certeza realista, como vimos en nuestra propuesta de análisis del pasaje de la novela *Allwill* (pasaje curiosa y altamente desestimado por los intérpretes), es asumida por Jacobi con total consecuencia: incluso las ilusiones sensoriales han de ser verdaderas, más allá de que su verdad sea inexplicable para nosotros (por depender de la infinitud del entendimiento divino). El realismo jacobiano, en tanto tiene que hacer de la sensibilidad un “medio distinto y real entre lo real y lo real” (JWA 2,1, p. 109), tiene que darles verdad incluso a las ilusiones sensoriales. Esa verdad es localizada en el entendimiento divino, siguiendo la línea spinozista. En Kant, Jacobi encontrará precisamente lo opuesto: la reducción de la sensación a la conciencia “pura” y humana. Podremos observarlo en la *Epístola sobre la filosofía kantiana* (1791), donde Jacobi señala, como una de las claves del kantismo, la sección de las “Anticipaciones de la percepción” de la *Crítica de la razón pura*.

3.2. La sensación suprimida

En la *Epístola sobre la filosofía kantiana*, Jacobi explica de esta manera las anticipaciones de la percepción:

La segunda categoría [Jacobi se refiere a las categorías del entendimiento según Kant] es la categoría de

cualidad, y comprende bajo sí a la realidad, la negación y la limitación. [...] La cualidad no es sino la conciencia pura afectada de tal o cual manera; es decir, [la cualidad] es la sensación [*Empfindung*]. Por lo tanto, el concepto originario de la cualidad sería la sensación pura. Pero como la sensación es pasividad y completamente material, no puede tener lugar, como tal, entre las facultades puramente activas y meramente formales. Por eso, ella tiene su lugar entre tales [facultades] meramente en la medida en que cada sensación tiene que tener un grado, que es sintéticamente y, en consecuencia, activamente conocido a través de una acción pura del entendimiento (categoría). Vea usted, sobre esto y sobre cómo genera esta categoría el principio de las anticipaciones de la percepción, la página 62 del *Manual* señalado. NB. Como la generación de este principio es uno de los trucos principales [*Hauptkunststücke*] de la filosofía kantiana, valdría la pena que usted viera al mismo maestro hacerlo. Véase *Crítica de la razón pura*, 2ª edición, pp. 207-218 (JWA 2,1, pp. 143-144).

Jacobi le sugiere al destinatario de su carta que lea el *Manual* de Carl Christian Erhard Schmid sobre la *Crítica de la razón pura* (el libro *Critik der reinen Vernunft im Grundrisse*). Siguiendo su consejo, veamos cómo presenta Schmid la categoría de cualidad y las anticipaciones de la percepción kantianas:

§. 133. Principio de la cualidad. En todos los fenómenos, la sensación —y lo real que a ella le corresponde en el objeto (*realitas phaenomenon*)— tiene una cantidad intensiva, es decir, un grado. Este principio subsume lo real del fenómeno, a saber, la sensación, en la medida en que ella es comparada con la no-sensación [*Nichtempfindung*] y es representado un tránsito hacia la misma, bajo el concepto de una cantidad.

§. 134. Prueba. La sensación es auténticamente lo distintivo y la auténtica materia del conocimiento empírico. Sin embargo, se puede conocer algo *a priori* en cada sensación, sin que sea dada una sensación

particular; ciertamente no su cualidad [...] pero sí la cantidad de la síntesis empírica, a través de la cual ella llega a efectivizarse.

§. 135. La sensación no tiene, por cierto, ninguna cantidad extensiva, pues su aprehensión no es una síntesis sucesiva, sino que colma un instante.

§. 136. No obstante, cada sensación (por ejemplo, un color, el calor, etc.) es susceptible de disminución, hasta desaparecer. Entre cada sensación (realidad) y la no-sensación (negación) hay un conjunto posible continuo de sensaciones intermedias [*Zwischenempfindungen*], que se acercan cada vez más al cero. En consecuencia, la sensación tiene una cantidad, que, empero, no puede ser aprehendida como multiplicidad, sino solo representada a través de su aproximación a la negación; una cantidad intensiva, un grado (Schmid, 1788, pp. 62-63).

Schmid es muy claro sobre el modo en que él entiende las anticipaciones de la percepción: se trata de comparar la sensación con la no-sensación y de advertir, por esa diferencia de toda sensación respecto del “cero” de sensación, que toda sensación tiene un grado, captable en un “instante”. Schmid incluye dentro de las sensaciones justamente las cualidades secundarias (los colores, por ejemplo), aquellas que Mendelssohn había considerado que, para un spinozista, no podían ser menos reales que la extensión y el movimiento. Por nuestra parte, hemos visto que Jacobi acepta el reto que le lanza Mendelssohn: spinozísticamente, Jacobi hace de las ilusiones sensoriales ideas verdaderas en el entendimiento divino. Jacobi, por lo tanto, lleva las ilusiones sensoriales hacia el entendimiento divino, garantizando su realismo. En sí mismas, son verdaderas; esto es, hay una serie de relaciones reales (independientemente de que nosotros no podamos agotarlas) que hace verdaderas a las sensaciones. En Kant, Jacobi encuentra —mediante Schmid— lo contrario: las sensaciones son reconducidas no al realismo del entendimiento divino, sino a la no-sensación, a la negación de la realidad. En otras palabras, Kant reduce las sensaciones a la conciencia “pura”:

[...] de la conciencia empírica a la pura es posible una alteración gradual, en la cual lo real de ella desaparezca enteramente, quedando una conciencia meramente formal (*apriori*) de lo múltiple en el espacio y en el tiempo; y por consiguiente [es posible] también una síntesis de la generación de la cantidad de una sensación, desde su comienzo, la intuición pura=0, hasta una cantidad cualquiera de ella (KrV B 208; Kant, 2009, p. 262).

Para Jacobi, justamente lo que no puede dar el criticismo es ese comienzo de la sensación al que se refiere Kant; es decir, no puede mostrar cómo se interrumpe la “intuición pura=0” para dar pie a una sensación distinta de cero. Así como Walter Shandy no podía descubrir el “sentido místico” y, enfurecido, destruía el libro de Erasmo, así Kant no puede construir las sensaciones y, resignado, las niega en el “cero” de la conciencia pura. La *Mathesis intensorum* que Schmid (1788, p. 65) atribuye a las anticipaciones de la percepción no puede explicar el comienzo de la sensación, que implicaría salir de la continuidad pura del espacio, del tiempo y de la conciencia. Solo puede, en verdad, retrotraer la sensación hacia su posible negación. El criterio de la sensación es, en el criticismo, no su realidad, sino su diferencia con la negación. De allí el “truco” de las anticipaciones de la percepción: Kant deja en pie solo una posible negación gradual (en dirección hacia la conciencia pura) para aquello mismo que era afirmado como verdad en el Dios impersonal de Spinoza y que Jacobi retoma, con Spinoza y contra Spinoza, para su propia noción de “Dios”.

En efecto, la unión de todos los temas trabajados hasta aquí tiene su punto culminante en el Dios personal por el que, como es sabido, Jacobi ha abogado insistentemente. No es casual que las cuestiones elaboradas en este trabajo aparezcan juntas en un solo escrito de Jacobi, una carta del 29 de diciembre de 1790 (AB II, pp. 47-51), evidente matriz del texto de la novela *Allwill*. Allí es afirmado el Dios personal y creador (AB II, pp. 48-49); allí es citado, con mayor extensión, el pasaje de *Tristram Shandy* sobre Walter y su cortaplumas (AB II, pp. 49-50); allí, por último, Kant aparece como el filósofo que llevó el idealismo de la filosofía moderna, negador de la verdad de las percepciones sensibles, a su punto más alto (AB II, pp. 50-51). La verdad de lo sensible, en Jacobi, va asociada a la verdad del Dios creador. Esta verdad de lo sensible, vista desde la “pureza” de nuestro pensamiento, se muestra como una inexplicable interrupción

de la pureza. Por ello, la necesidad de alejarse de la conciencia “pura” de Kant no puede sino leerse a la luz del rechazo que Jacobi siente por el Dios impersonal de Spinoza. No por azar Jacobi intenta explicar el pensamiento spinozista (es decir, el atributo de la sustancia llamado “pensamiento”) mediante la “conciencia pura” kantiana (cfr. JWA 1,1, p. 105). Evitar la conciencia pura kantiana es también un modo de evitar la impersonalidad del Dios spinozista, que no puede hacer del pensamiento lo primero.

4. Conclusión

Hemos intentado mostrar que uno de los conceptos con los que Jacobi expone el límite del entendimiento es el de “interrupción” (lo hemos visto ejemplificado con el kantismo y la imposibilidad de interrumpir la “pureza”). A la vez, este concepto nos permite comprobar que la elaboración filosófica de Jacobi va de la mano con un trabajo de lectura que alimenta de modos muy concretos su propia escritura: la referencia al *Tristram Shandy* es una pieza fundamental del pasaje que el propio Jacobi considera como la formulación más ajustada del “sistema de la absoluta objetividad”. Así, hemos expuesto que la interrupción es una figura central del “sistema de la absoluta objetividad”, pues aparece en la más elevada discusión filosófica y en la más material configuración del saber en tanto texto (escrito y leído).

Para finalizar, quisiéramos señalar que el concepto de “interrupción” no obtiene su relevancia meramente de su mayor o menor fecundidad para marcar los límites del kantismo o para mostrar el sutil entramado textual del pensamiento de Jacobi. Con él, se abre una nueva puerta de investigación, con vistas a un camino que aquí no recorreremos, pero que merece ser apuntado. Nos referimos al aspecto vital de la interrupción, mencionado al final de la sección 2. De acuerdo con la “metafísica de la acción” que Sandkaulen señala como lo característico del pensamiento de Jacobi, consideramos que para nuestro autor la interrupción es necesaria para la vida. Así lo dice el propio Jacobi:

Tenemos [...] la más profunda conciencia de algo extratemporal. En nosotros, lo llamamos el yo [*das Selbst*]; fuera de nosotros, Dios. A lo que está en el tiempo [lo llamamos] *Naturaleza* o lo transitorio. Todo lo transitorio está dado en algo no transitorio, y lo presupone; si no, no sería tal. Vivimos en tanto

incesantemente interrumpimos la eternidad y ponemos un comienzo y un fin [*Wir leben dadurch, daß wir unaufhörlich die Ewigkeit unterbrechen, einen Anfang und ein Ende setzen*] (JWA 5,1, p. 406).

Es curioso que lo que nos haga vivir sea la interrupción de la eternidad, dado que —en el contexto de ese breve parágrafo— lo eterno parece ser lo divino. Creemos, por ello, que hay que buscar una explicación más auspiciosa para estas palabras, y sugerimos pensar que la eternidad que debe ser interrumpida es la eternidad de lo transitorio, la eternidad del “tiempo eterno” que Jacobi denuncia en Spinoza (JWA 1,1, p. 251). Al “tiempo eterno” de la sustancia spinozista, que no es más que el reverso simétrico de la conciencia pura del idealismo (cfr. JWA 2,1, p. 195), hay que interrumpirlo con la libertad irreductible al concepto. Si en la sustancia de Spinoza no hay “comienzo o fin [*Anfang oder Ende*]” (JWA 1,1, p. 20), la interrupción que pone un comienzo y un fin se muestra como lo contrario del fatalismo y el comienzo de la acción libre.

Referencias

- Bloch, O. (1985). *Le matérialisme*. Presses Universitaires de France.
- Bollnow, O. (1933). *Die Lebensphilosophie F. H. Jacobis*. W. Kohlhammer Verlag.
- Bowman, B. (2004). *Notiones Communes und Common Sense*. Zu den Spinozanischen Voraussetzungen von Jacobis Rezeption der Philosophie Thomas Reids. En W. Jaeschke y B. Sandkaulen (eds.), *Friedrich Heinrich Jacobi. Ein Wendepunkt der geistigen Bildung der Zeit* (pp. 159-176). Felix Meiner.
- Bowman, B. (2023). Jacobi on the Nature of Mind and Intuitive Certainty. En A. J. B. Hampton (ed.), *Friedrich Heinrich Jacobi and the Ends of the Enlightenment: Religion, Philosophy, and Reason at the Crux of Modernity* (pp. 49-65). Cambridge University Press. <https://doi.org/10.1017/9781009244978.006>
- Brunel, P.-J. (2008). De Protée à Polyphème. La crise de la rationalité chez F. H. Jacobi. *Études Germaniques*, 63(3), 571-600. <https://doi.org/10.3917/eger.251.0571>
- Erasmus de Róterdam. (1677). *Colloquia: nunc emendatiora. Cum omnium Notis. Typis Danielis Elzevirii*. <https://books.google.com.ar/books?id=H-c1n7MK7X4C&printsec=frontcover&hl=es#v=onepage&q&f=false>

- Götz, C. (2008). *Friedrich Heinrich Jacobi im Kontext der Aufklärung. Diskurse zwischen Philosophie, Medizin und Literatur*. Felix Meiner.
- Gregori, F. (2013). *Homunculus ab ovo: Beginning as Continuity and Discontinuity in Tristram Shandy*. *Études Anglaises*, 66(2), 214-233. <https://doi.org/10.3917/etan.662.0214>
- Guzmán Grez, N. (2024). F. H. Jacobi y el problema de la afección en Kant: una solución metodológica. *Tópicos, Revista de Filosofía*, 69, 31-53. <https://doi.org/10.21555/top.v690.2521>
- Hammacher, K. (1969). *Kritik und Leben II: Die Philosophie Friedrich Heinrich Jacobis*. Wilhelm Fink.
- Jacobi, F. H. (1825). [AB I]. *Auserlesener Briefwechsel. Erster Band*. Gerhard Fleischer. <https://www.digitale-sammlungen.de/view/bsb10403083?page=%2C1>
- Jacobi, F. H. (1827). [AB II]. *Auserlesener Briefwechsel. Zweyter Band*. Gerhard Fleischer. <https://www.digitale-sammlungen.de/view/bsb11415034?page=%2C1>
- Jacobi, F. H. (1869). [NA I]. *Aus F. H. Jacobis Nachlaß. Erster Band. Ungedruckte Briefe von und an Jacobi. Nebst ungedruckten Gedichten von Goethe und Lenz*. R. Zoeppritz (ed.). Wilhelm Engelmann. <https://www.digitale-sammlungen.de/view/bsb11415035?page=%2C1>
- Jacobi, F. H. (1998). [JWA 1,1]. *Werke. Gesamtausgabe. Band 1,1. Schriften zum Spinozastreit*. K. Hammacher e I.-M. Piske (eds.). Felix Meiner.
- Jacobi, F. H. (2000). [JWA 3]. *Werke. Gesamtausgabe. Band 3. Schriften zum Streit um die göttlichen Dinge und ihre Offenbarung*. W. Jaeschke (ed.). Felix Meiner.
- Jacobi, F. H. (2004a). [JWA 2,1]. *Werke. Gesamtausgabe. Band 2,1. Schriften zum transzendentalen Idealismus*. W. Jaeschke e I.-M. Piske (eds.). Felix Meiner.
- Jacobi, F.H. (2004b). [JWA 2,2]. *Werke. Gesamtausgabe. Band 2,2. Schriften zum transzendentalen Idealismus. Anhang*. W. Jaeschke e I.-M. Piske (eds.). Felix Meiner.
- Jacobi, F. H. (2006). [JWA 6,1]. *Werke. Gesamtausgabe. Band 6,1. Romane I. Eduard Allwill*. C. Götz y W. Jaeschke (eds.). Felix Meiner.
- Jacobi, F. H. (2007). [JWA 5,1]. *Werke. Gesamtausgabe. Band 5,1. Kleine Schriften II (1787-1817)*. C. Goretzki y W. Jaeschke (eds.). Felix Meiner.
- Kant, I. (2009). *Crítica de la razón pura*. M. Caimi (trad.). Colihue.
- Knöde, F. (2024). *Literarische Anthropologie und empfindsame Idyllik. Zum Verhältnis von Literatur und Philosophie im Erzählwerk Friedrich Heinrich Jacobis*. Transcript Verlag. <https://doi.org/10.1515/9783839468586>

- Olivetti, M. M. (1971). Der Einfluß Hamanns auf die Religionsphilosophie Jacobis. En K. Hammacher (ed.), *Friedrich Heinrich Jacobi. Philosoph und Literat der Goethezeit* (pp. 85-117). Vittorio Klostermann.
- Olivetti, M. M. (2013). *L'ésito teologico della filosofia del linguaggio di Jacobi*. En M. M. Olivetti, *Opere II*. Fabrizio Serra Editore.
- Ortlieb, C. (2010). *Friedrich Heinrich Jacobi und die Philosophie als Schreibart*. Brill-Wilhelm Fink. <https://doi.org/10.30965/9783846749388>
- Sandkaulen, B. (2000). *Grund und Ursache. Die Vernunftkritik Jacobis*. Wilhelm Fink.
- Sandkaulen, B. (2019). *Jacobis Philosophie. Über den Widerspruch zwischen System und Freiheit*. Felix Meiner. <https://doi.org/10.28937/978-3-7873-3629-6>
- Sandkaulen, B. (2023). Jacobi and Spinoza. En A. J. B. Hampton (ed.), *Friedrich Heinrich Jacobi and the Ends of the Enlightenment: Religion, Philosophy, and Reason at the Crux of Modernity* (pp. 34-48). Cambridge University Press. <https://doi.org/10.1017/9781009244978.005>
- Sandkaulen, B. (2024). Sein in allem Dasein. En B. Sandkaulen, S. Schick y O. Koch (eds.), *Jacobi-Wörterbuch Online* [Versión v6]. <https://jwo.saw-leipzig.de/articles/v6/bacb2871>
- Schmid, C. C. E. (1788). *Critik der reinen Vernunft im Grundrisse zu Vorlesungen nebst einem Wörterbuche zum leichtern Gebrauch der Kantischen Schriften*. Crökerschen Buchhandlung. <https://www.digitale-sammlungen.de/view/bsb10044342?page=1>
- Spinoza, B. (2009). *Ética demostrada según el orden geométrico*. V. Peña (trad.). Alianza.
- Sterne, L. (1770). *Das Leben und die Meynungen des Herrn Tristram Shandy. Dritter Theil*. Gottlieb August Langen. <https://www.digitale-sammlungen.de/view/bsb10750311?page=%2C1>
- Sterne, L. (1774). *Tristram Shandis Leben und Meynungen. Dritter Theil*. J. J. C. Bode (trad.). Hamburgo. <https://www.digitale-sammlungen.de/view/bsb10753449?page=%2C1>
- Sterne, L. (1978). *La vida y las opiniones del caballero Tristram Shandy. Los sermones de Mr. Yorick*. J. Marías (trad.). Alfaguara.
- Sterne, L. (1980). *Tristram Shandy*. H. Anderson (ed.). Norton & Company.
- Stiening, G. (2021). »Der geheime Handgriff des Schöpfers«. Jacobis theonome Epistemologie. En C. Ortlieb y F. Vollhardt (eds.), *Friedrich Heinrich Jacobi (1743–1819). Romancier – Philosoph – Politiker* (pp. 171-190). De Gruyter. <https://doi.org/10.1515/9783110727340>
- Timm, H. (1971). Die Bedeutung der Spinozabriefe Jacobis für die Entwicklung der idealistischen Religionsphilosophie. En K.

- Hammacher (ed.), *Friedrich Heinrich Jacobi. Philosoph und Literat der Goethezeit* (pp. 35-84). Vittorio Klostermann.
- Timm, H. (1974). *Gott und die Freiheit. Studien zur Religionsphilosophie der Goethezeit. Band I: Der Spinozarenaissance*. Vittorio Klostermann.
- Verra, V. (1963). *F. H. Jacobi. Dall'Illuminismo all'Idealismo*. Edizioni di 'Filosofia'.
- Villacañas, J. L. (1989). *Nihilismo, especulación y cristianismo en F. H. Jacobi. Un ensayo sobre los orígenes del irracionalismo contemporáneo*. Anthropos-Universidad de Murcia.
- Wilde, N. (1894). *Friedrich Heinrich Jacobi. A Study in the Origin of German Realism*. [Tesis doctoral, Columbia College]. Library of Congress. <https://www.loc.gov/resource/gdcmassbookdig.friedrichheinric00wild/?st=gallery>

Filosofía en el espacio público

<https://doi.org/10.21555/top.v760.3338>

El liberalismo como religión y su repercusión global. Las interpretaciones críticas de John Gray y John Milbank en perspectiva comparada

Liberalism as Religion and Its Global Repercussion: A Comparative Perspective on the Critical Interpretations of John Gray and John Milbank

Luis Aránguiz Kahn

Universidad de Friburgo
Suiza

Luis.aranguiz.k@gmail.com

<https://orcid.org/0000-0001-7356-6679>

Recibido: 30 - 12 - 2024.

Aceptado: 27 - 03 - 2025.

Publicado en línea: 10 - 08 - 2026.

Cómo citar este artículo: Aránguiz Kahn, L. (2026). El liberalismo como religión y su repercusión global. Las interpretaciones críticas de John Gray y John Milbank en perspectiva comparada. *Tópicos, Revista de Filosofía*, 76, 551-580. <https://doi.org/10.21555/top.v760.3338>



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

Resumen

Este artículo analiza de manera comparada el pensamiento del filósofo John Gray y el teólogo socialista cristiano John Milbank —ambos ingleses contemporáneos— en torno a la relación entre liberalismo, religión y orden internacional. Aunque son de procedencias intelectuales distintas, ambos coinciden en sostener que dicha relación no consiste simplemente en la interacción de actores políticos y religiosos, sino sobre todo en una conjugación en la que ideología y religión se funden para dar origen a una religión y teología políticas precisas que delimitan las formas en que se conduce la vida humana a escala planetaria. Pese a esta semejanza, el artículo finaliza evidenciando las diferencias que existen entre Gray y Milbank sobre sus concepciones del mundo postliberal, desprendidas de sus valoraciones sobre liberalismo y religión.

Palabras clave: John Gray; John Milbank; postliberalismo; liberalismo; teología política; religión política; teología geopolítica; pluralismo de valores; política de la virtud; Eric Voegelin; Carl Schmitt.

Abstract

This paper makes a comparative analysis of the thought of the philosopher John Gray and the Christian socialist theologian John Milbank—both contemporary English authors—on the relationship between liberalism, religion, and international order. Although they come from different intellectual backgrounds, both agree that this relationship does not consist simply of the interaction between political and religious actors, but above all of a convergence in which ideology and religion merge to give rise to specific forms of political religion and theology that define the ways in which human life is conducted on a global scale. Despite this similarity, this paper ends by highlighting the differences that exist between Gray and Milbank regarding their conceptions of the post-liberal world derived from their assessments of liberalism and religion.

Keywords: John Gray; John Milbank; postliberalism; liberalism; political theology; political religions; geopolitical theology; value pluralism; politics of virtue; Eric Voegelin; Carl Schmitt.

Introducción

Entre las variadas críticas al liberalismo, hay dos que han destacado por referirse a su relación tensa con el ámbito de los valores de las sociedades. Una de ellas ha tratado sobre su postura frente a lo religioso, y la otra sobre el carácter expansivo que ha adoptado en el orden internacional. Aunque ambas cosas no son una ni la misma, este artículo mostrará cómo, en John Gray, filósofo político —ateo, además—, y John Milbank, teólogo socialista cristiano, estas se ven conjugadas y son para ellos, en última instancia, un problema teológico-político.

La preocupación liberal por la religión quedó bien reflejada en obras como *Liberalismo político*, de John Rawls. Ahí se indicó la necesidad de dejar fuera de la discusión pública las nociones comprensivas de la realidad, entre ellas las religiosas. A ello le ha seguido un dilatado debate que cruza nombres como Jürgen Habermas (2006) y Nicholas Wolterstorff (1997), entre tantos otros. Uno de los presupuestos discutidos refiere a la noción de que la religión es un fenómeno distinto del político, pues la primera corresponde a creencias basadas en una revelación sobrenatural, mientras que lo segundo debiese remitirse al campo de la razón. Naturalmente, esta es una forma de comprensión de lo político que descansa sobre los diversos desarrollos teóricos de la modernidad (Strauss, 2011). En vista a la diferencia fundamental que hay entre ambos campos, la tradición liberal ha apuntado a que los Estados no deben regirse por noción religiosa alguna. Mientras que en los años noventa la sociología giraba su atención a la “desprivatización” de la religión (Casanova, 1994) y el debate se ocupaba de analizar su injerencia en la vida pública, Gray y Milbank dieron un giro diferente y se preguntaron en qué sentido el liberalismo puede devenir en una forma de religión.

Las amplias expectativas ante la expansión de la democracia liberal con la égida del fin de la historia de Fukuyama (1989) fueron desde temprano matizadas por la idea de que el mundo se componía de espacios culturalmente diferentes, con tradiciones y religiones diversas, lo cual iba a poner límite a esta aspiración (Huntington, 1997). Una variada gama de autores de distintas posiciones, como el mismo Gray (2011a), posmarxistas y marxistas como Michael Hardt y Antonio Negri (2000) y Perry Anderson (2014), nacionalistas como Yoram Hazony (2018), entre otros, han indicado que las expectativas de expansión vienen a

ser un “imperialismo liberal” ya de Estados Unidos o de la Unión Europea, o de una alianza entre países liberales en general. En este arco de críticas, Gray y Milbank se distinguen porque ambos han buscado interpretar cierta forma de liberalismo y su expresión internacional como un fenómeno que responde a cuestiones religiosas. Así, sostener que el liberalismo puede llegar a ser una religión política, como hace Gray, resulta un problema para los propios liberales, porque significa develar el riesgo interno de convertirse en aquello a lo que buscan contener. Que no se trate de una religión oficial y reconocida no quiere decir que no pueda adoptar rasgos religiosos. De este modo, el debate entre liberalismo y religión se vuelca al interior mismo de la tradición liberal, lo que tiene como consecuencia una revisión arquitectónica de los elementos constituyentes de lo que se entiende por “liberalismo” y de sus transmutaciones religiosas. Milbank, por su parte, observa al liberalismo como una desviación herética del cristianismo, esto es, una forma de religión con ciertos distintivos teológicos.

Pese a sus diferencias de posición de origen, Gray y Milbank coinciden en diagnosticar los problemas del liberalismo a partir de un análisis que incorpora el elemento religioso. A su vez, ambos arriban a la propuesta de un pensamiento político postliberal. No obstante, aunque tienen estas coincidencias, sus conclusiones sobre el futuro son diferentes. Esto, sostenemos, se debe a su posición de origen, la cual determina las expectativas sobre el porvenir en cada caso. Al ponderar sus ideas prestando atención al escenario global actual, este artículo concluye que la posición de Gray se ajusta mejor a estas circunstancias que la de Milbank.

En lo sucesivo, se explora en primer lugar el pensamiento de Gray en cuanto al liberalismo, para luego revisar el análisis religioso que le aplica a su dimensión global. En segundo lugar, se explora el pensamiento de Milbank en el mismo orden: liberalismo y luego el análisis religioso que realiza sobre la dimensión global de aquel. En tercer lugar, y ya con una mirada panorámica de los principales aspectos de ambos pensadores en estas materias, se exploran de manera comparada las nociones de “postliberalismo” que ambos ofrecen ante la situación contemporánea. El artículo finaliza con algunas reflexiones sobre los alcances que tiene la postura de cada uno de manera comparada.

1. Gray

1.1. Crítica al liberalismo

El liberalismo ha ocupado un lugar central en la trayectoria de John Gray. Por ello, haremos un breve repaso de su toma de posición frente a la tradición liberal, la discusión filosófico-política de sus ideas y una perspectiva sistemática de su crítica al liberalismo. La pregunta que cabe responder es si acaso Gray es o no liberal y, si lo es en algún sentido, cuál es la cercanía que tiene con la tradición.

Por la distancia que toma Gray frente al liberalismo, hay quienes sostienen que se posicionaría como un antiliberal (Kelly, 2007; Crowder, 2007). Otra perspectiva sostiene que su posición sería más bien una forma de liberalismo conservador (Del Palacio, 2012). Aunque no hay acuerdo respecto a este punto, de todos modos, sí parece haberlo en cuanto al carácter pesimista y antiteórico de su pensamiento.

En lo político, Gray exhibe un realismo agonista y, por ello, también hay recepciones dispares de sus ideas. Puede apuntarse, por ejemplo, cómo su realismo rivaliza con las concepciones liberales de lo político, en la medida que se caracteriza por la asunción de la existencia de conflictos irreconciliables. No obstante, por el hecho de estar afirmado todavía en algunos presupuestos liberales, desde la perspectiva del realismo liberal de Matt Sleat (2013), Gray no alcanza a ser lo suficientemente realista debido justamente a su liberalismo.

El desarrollo del pensamiento de Gray en cuanto al liberalismo puede comprenderse en tres grandes etapas. Si bien en todas ellas se muestra como un crítico, hay matices claramente diferenciables. Se puede decir que entre mediados de los ochenta y noventa, presenta una posición “subjetivista” (Crowder, 2007, p. 64), caracterizada por el problema que supone elegir entre valores inconmensurables en una sociedad pluralista, y por el hecho de que la elección de uno u otro sería no racional — es decir, no se trata de elecciones racionales—. En los años noventa se observa un cambio a una visión “contextualista” (Crowder, 2007, p. 64) que acepta la elección racional, atendiendo a cómo los contextos de los seres humanos determinan sus elecciones, adquiriendo relevancia las tradiciones culturales como marcos de sentido en sociedades pluralistas. El tercer momento, visible en la primera década del siglo XXI, y que es

el que tiende a aparecer con mayor frecuencia, es el de la teoría del *modus vivendi* (Crowder, 2007, p. 65), esto es, un modo pragmático de resolver las colisiones que pueden darse entre las distintas tradiciones, caracterizado por la elección de ciertos mínimos comunes que permiten solucionar pacíficamente las diferencias.

Su postura sobre lo postliberal surge en el marco de la reflexión sobre comunidades políticas liberales, mientras que su crítica al imperialismo liberal se orienta hacia la comunidad internacional, pero en ambos casos el problema es la relación entre liberalismo y pluralidad de visiones de mundo. Esta continuidad se puede observar atendiendo a dos de sus obras, que trataremos: *Post-Liberalism*, de 1993, y *Black Mass*, de 2007. En ellas encontramos las críticas mencionadas, pero con distinto énfasis. Pese a la distancia de tiempo que las separa, se puede examinar cómo las nociones expuestas en el primer texto se proyectan a una escala mayor en el segundo. Al final, se verá que Gray es, pese a todo, un defensor de cierta forma de liberalismo. Lo que ocurre es que su oposición está primariamente contra una ideologización del liberalismo, pero no contra cierta forma de filosofía u orden liberales.

1.2. De filosofía política a religión política imperial

La crítica de Gray al liberalismo y su proyección a escala global se propone a partir de claves que podríamos denominar *grosso modo* “teológico-políticas”. Este último concepto lo entendemos como “una reflexión acerca del núcleo teológico de la política” (Scattola, 2008, p. 9). En este sentido, puede decirse que Gray desentraña una teología política del universalismo liberal, a la cual se opone. Respecto a concebir a Gray como teólogo político, Simon Critchley ha observado que, si su objetivo es derribar a los “ídolos” del “humanismo liberal” (Critchley, 2017), aquello lo hace adoptando una concepción del ser humano que es antagónica y, yendo entonces a la antropología de Gray, Critchley se encuentra con que “ofrece una redesccripción naturalizada, darwiniana, del pecado original. Dicho bruscamente, los seres humanos son simios asesinos. [...] Tristemente, también somos simios asesinos con anhelos metafísicos” (2017, p. 111). Somos lo que Gray (2011b) denomina *Homo sapiens*. Este pesimismo antropológico es el que llevaría a Gray al realismo y al conservadurismo, a desestimar toda noción de “progreso” como una ilusión y toda posibilidad de transformación del mundo. Son estas esperanzas, estos ídolos los que, proyectados a escala mundial, pueden llevar a un apocalipsis.

En su libro *Post-Liberalism*, Gray expone una crítica dirigida no a la filosofía política liberal en cuanto tal, sino en particular a lo que denomina “afirmaciones fundacionalistas del fundamentalismo liberal”¹ (1996, p. 284). Este giro de filosofía a fundamentalismo puede ilustrarse en cuatro tópicos que serán desarrollados posteriormente en el análisis de su posición postliberal: el individualismo moral o normativo, el universalismo, el meliorismo (*meliorism*) y el igualitarismo. Aunque Gray quisiera que Occidente hubiese entendido que el liberalismo no es universal, lo cierto es que las potencias, y en particular Estados Unidos, adoptaron una posición bastante distinta después de la caída del comunismo soviético. Si ya en *Post-Liberalism* había tenido la osadía de calificar de “fundamentalista” a la idea racional, ilustrada, dominante, como si fuera no solo una religión, sino una de sus peores patologías, casi quince años después, encontramos este argumento formulado de modo más sistemático en *Black Mass*. En sus términos:

[...] pasó a decirse que, en el futuro, solo habría un único tipo de régimen legítimo: el capitalismo democrático de estilo estadounidense [...]. Encabezados por Estados Unidos, los gobiernos occidentales se comprometieron a instaurar la democracia en todo el mundo: un sueño imposible que, en muchos países, solamente podía provocar el caos (2017, p. 50).

La aserción fundamental de la obra es que esta pretensión universalista del liberalismo es imposible de instaurar y que tiene un alto costo intentarlo. Además, agrega que esta pretensión es compartida por otras ideologías modernas diferentes solo en apariencia, puesto que su matriz común es el conjunto de ciertos elementos del cristianismo:

[...] la promesa cristiana de salvación universal fue heredada por sus sucesores laicos. Pero mientras que en el cristianismo la salvación se prometía únicamente para la vida del más allá, las religiones políticas modernas ofrecían la posibilidad de salvación en el futuro terrenal (2017, p. 49).

Gray utiliza el concepto de “religión política”, que, aunque no es presentado sistemáticamente, puede entenderse siguiendo a Voegelin

¹ Si la bibliografía final no indica lo contrario, las traducciones son de este autor.

—quien lo propuso originalmente— como la divinización del colectivo, cualquiera que fuere, que a su vez se vuelve la única o más importante fuente de sentido de la acción del individuo (Scattola, 2008, p. 27). El propio Voegelin (2014, p. 27) diría que de lo que se trata es de interpretar los movimientos de su tiempo no solo como políticos sino como religiosos. Si bien en *Black Mass* Gray cita a Voegelin solo una vez y para discutirle un punto, lo cierto es que la metodología de análisis que Gray aplica al liberalismo resulta prácticamente análoga a la que aplicó Voegelin a los casos de los nazis, los fascistas y los comunistas. Además, en otra ocasión, Gray (2011c) elaboró una crítica al comunismo soviético arrancando precisamente de esta idea de Voegelin. Por ello, la recurrencia al concepto de religión política, junto con otra serie de elementos como la reflexión en torno a lo apocalíptico, permite pensar que lo que efectúa Gray en esta obra es un ejercicio voegelineano de análisis del liberalismo. Aunque Gray hace explícita su recurrencia al análisis del milenarismo cristiano medieval de Norman Cohn (2004) para explorar sus alcances en la modernidad, esta digresión permite al menos interrogar si acaso la aparición —comparativamente desmedrada— de Voegelin hace honor a su grado de influencia en la obra.

La síntesis de la influencia Voegelin-Cohn se expresa en la siguiente fórmula: “los movimientos revolucionarios modernos son una continuación de la religión por otros medios” (Gray, 2017, p. 15).² La Revolución francesa, el comunismo, el fascismo-nazismo y, por supuesto, el liberalismo, reproducen, según Gray, algunos motivos cristianos fundamentales. Por ejemplo, “el cristianismo estableció la creencia de que la historia humana es un proceso teleológico” (2017, p. 18); ello puede encontrarse en el meliorismo liberal. Además, “el cristianismo inyectó la escatología en el corazón de la civilización occidental” (2017, p. 28), aquella noción de la búsqueda de un fin del tiempo —¿fin de la historia?—, la inauguración de un nuevo tiempo. Probablemente persiste en este aspecto el elemento universalista. Solo que, a diferencia del cristianismo, “mientras que los milenaristas creían que solo Dios podía rehacer el mundo, los revolucionarios modernos imaginaron que éste podía ser remodelado por la acción exclusiva de la

² Por el modo de formulación, esta oración no deja de recordar a Von Clausewitz y su idea de que la guerra es la continuación de la política por otros medios; y a MacIntyre, quien sostuvo que la política moderna es una guerra civil continuada por otros medios.

humanidad" (2017, p. 31). Estos elementos ya explican bastante bien esa enigmática línea con la que abre *Black Mass*, según la cual "la política de la Edad Contemporánea constituye otro capítulo más de la historia de la religión" (Gray, 2017, p. 13).

En el Occidente moderno, señala Gray, la violencia se justifica en función del progreso, que a su vez se afirma sobre la concepción escatológica de una utopía final. Esto es lo que permitiría entender por qué, por ejemplo, los Estados que más pretenden garantizar los derechos humanos no ven problema en ir a una guerra —asumiendo un importante costo humano— para defenderlos en países que no los respetan. El caso paradigmático que Gray comenta es el de la invasión a Irak en 2003 en el marco de lo acontecido tras el 9/11. El problema es el siguiente:

[...] las liberales son sociedades que vale la pena defender, ya que encarnan un tipo de vida civilizada en el que las creencias rivales pueden coexistir pacíficamente. Pero cuando se convierten en regímenes misioneros, ese logro corre el riesgo de dejar de ser tal. Las sociedades liberales realmente existentes se corrompen cuando van a la guerra para promover sus valores (2017, p. 279).

Aunque con otros términos, en *Black Mass* Gray afirma la diferenciación postliberal —como se verá luego— entre liberalismo como ideología y como práctica política enraizada en la filosofía liberal. La primera es la que conduce al régimen misionero, imperialista. La segunda es la que origina una vida civilizada, la de la sociedad civil. En su lenguaje, podría decirse que la religión política fundamentalista del imperialismo liberal concibe a "Estados Unidos como garante final de la libertad en el mundo" (2017, p. 241). La dificultad de esto consiste en que lleva al límite de lo posible la creencia en la universalidad del liberalismo democrático como modelo de orden y paz en un mundo irremediabilmente plural en valores intrínsecos y culturas. Tal es así que "a veces, no hay paz porque las personas no quieren paz. Tal vez deseen la victoria de la 'única fe verdadera', ya sea esta una religión tradicional o un sucesor laico de aquella, como el comunismo, la democracia o los derechos humanos universales" (2017, p. 272). De manera contraintuitiva, la filosofía que aspira a la paz mundial no puede llevar a otra cosa que al caos cuando se la ideologiza e impone por la fuerza. Aunque se pudiera

imponer de manera no agresiva, como un *soft power* (Nye, 2004), aun con ello sería conflictiva porque, de todos modos, colisionará con otras visiones irreductibles de la realidad.

Ahora bien, el elemento religioso del liberalismo no se ve únicamente en lo que toca a las creencias tras el imperialismo de Estados Unidos y los Estados aliados. Hay también, puede pensarse así, una activación de prácticas en torno a esta creencia. En efecto, “los Estados no actúan únicamente con el objeto de proteger sus intereses: también vehiculan mitos, fantasías y psicosis de masas” (Gray, 2017, p. 290). Existe una dimensión antropológica que va más allá de la pura ideología, a la cual el Estado puede apelar. Esta dimensión es activada para los fines de la creencia ya no de una religión establecida sino del Estado. Puesto que “los seres humanos tienen necesidades que no pueden satisfacerse por ningún medio racional” (2017, p. 291), el Estado apela a estas necesidades trascendentes como si él fuera una religión y como si el ciudadano fuera equivalente a un feligrés.

Llegado a este punto, la pregunta inevitable de Gray, para quien lo anterior en absoluto es aceptable, es cómo afrontarlo o solucionarlo. Afirma que “los mitos laicos reproducen la forma narrativa del género apocalíptico cristiano y, si existe algún modo de atemperar la violencia de la fe, este debe empezar por cuestionar dichos mitos” (Gray, 2017, p. 301). Aquí el lenguaje ya no es el del fundamentalismo o la religión política, sino el del liberalismo como un mito: no como falsedad, sino como una cierta creencia que le da sentido a una comunidad cualquiera. El liberalismo, cabe dejar nota, ha sido abordado de este modo también, por ejemplo, desde una perspectiva iliberal (Simpson, 2018) y una antropológica (Asad, 2003). Lo que resulta llamativo es que Gray aplica el pluralismo de valores, pero no para cuestionar las pretensiones liberales en una comunidad política, sino las pretensiones liberales a escala internacional. ¿Cómo el mito liberal podría darle sentido a la pluralidad de culturas y sus consecuentes formas políticas, que pueblan el planeta? No puede: “En el futuro, habrá Estados autoritarios y repúblicas liberales, democracias teocráticas y tiranías laicas, imperios, ciudades-Estado y numerosos regímenes mixtos, como los hubo en el pasado” (2017, p. 303). Gray adopta una clara posición realista, en términos de relaciones internacionales, fundada sobre el pluralismo de valores y entiende que, así como en la sociedad civil no se puede alcanzar un acuerdo ideal, en el ámbito internacional tampoco se puede.

Con frecuencia, ataca Gray incansablemente la “fe laica en el progreso” (2017, p. 44), la “fe en la violencia” (2017, p. 49), la “fe en la utopía” (2017, p. 269) y la “fe irracional en el futuro” (2017, p. 297). Al tratar al liberalismo como una religión más, un mito también, la propuesta de Gray finalmente no ha sido otra que una desmitologización del liberalismo a escala doméstica y global.

2. Milbank

2.1. Crítica al liberalismo

John Milbank es un teólogo perteneciente al ala anglocatólica de la Iglesia de Inglaterra. Esto significa que sus afinidades se encuentran más cerca de la tradición católica que del protestantismo en su variada gama de expresiones. Milbank (2009c), además, se considera parte de la tradición socialista cristiana, alejada del marxismo y el socialismo secular, antiestatista, de origen tanto católico como anglicano y particularmente inglesa. En el caso anglicano, esta tradición encuentra sus orígenes a mediados del siglo XIX, y propuso fundir elementos de filosofía socialista con teología cristiana (Forde, 2020). De ese modo, lo socialista y lo cristiano se combinan para buscar mayor justicia para trabajadores y pobres en las sociedades industriales. Manteniéndose en la izquierda (Milbank, 2009b), aunque distante de otras expresiones de izquierda cristiana, como la teología de la liberación (Milbank, 1997), Milbank llama a una reconsideración del elemento religioso en ese sector político.

Se lo ha considerado un antiliberal y *new traditionalist* (“nuevo tradicionalista”) por su enfoque político (Stout, 2004). Además, en su socialismo cristiano y en el de la corriente que ha encabezado, la ortodoxia radical, no se ha dejado de observar una tendencia a la utopía (Stout, 2004). En 1990 apareció su obra *Theology and Social Theory*, bajo el supuesto reconocido por el propio Milbank de que “solo una visión teológica podía desafiar la hegemonía emergente del neoliberalismo” en la era Thatcher (Milbank, 2006, p. xi). En dicha obra se aprecia la construcción de un argumento que coloca en oposición a las tradiciones cristiana y liberal. En efecto, la “razón secular”, señala en su prólogo a la segunda edición, “es vista no como lo que ella misma primariamente proclama ser, llámese lo secular, sino más bien como una heterodoxia disfrazada con varias franjas, como un paganismo revivido y un nihilismo

religioso" (2006, p. xiv). Su teología política consiste en recuperar el rol público de la religión frente al liberalismo y el secularismo, en especial frente a la "tiranía política" que representa para él el neoliberalismo.

Su comprensión política del liberalismo puede verse con claridad en el texto "Liberality versus Liberalism", de 2006. Siguiendo el planteamiento de *Theology and Social Theory*, apunta a la raíz teológica del liberalismo, señalando que "comienza con una naturalización encubierta del pecado original como egoísmo original: nuestro propio egotismo que nosotros pretendemos nutrir, y aún más, el egotismo del otro contra el que necesitamos protección" (Milbank, 2009b, p. 70). De aquí se desprende que una política "específicamente liberal" tenderá hacia:

[...] una supuesta protección contra los elementos extraños: el terrorista, el refugiado, la persona de otra raza, el extranjero, el criminal. El populismo parece ser de forma creciente una deriva inevitable de la democracia liberal sin calificativos. Se hace gala de una pretendida defensa de esta última para justificar la suspensión de libertades civiles y de las tomas de decisión democráticas. Por las razones que acabamos de ver, no se trata exactamente de una amenaza exterior y reaccionaria contra los valores liberales: por el contrario, es el mismo liberalismo el que tiende a cancelar aquellos valores de liberalidad (juicio justo, derecho a defensa, presunción de inocencia, *habeas corpus*, cierta libertad de discurso y de investigación, buen trato a los convictos) que había *asumido* (2009b, p. 70).

De ese modo, lo que pareciera ajeno al liberalismo, o una deformación, en realidad para Milbank es su consecuencia lógica, es el liberalismo puesto en acción, y "da ese giro siniestro cuando lo único que propugna es el mercado libre y el Estado-nación como unidad competitiva. El gobierno tenderá a convertirse enteramente en una función policial y militar" (2009b, p. 70). Por contraparte, la "liberalidad" es definida como un "credo de generosidad" (2009b, p. 77) que habría de imponerse a la lógica liberal del contrato. Así se pueden afrontar las consecuencias del "moderno" liberalismo como la pobreza, la desigualdad, la erosión de los cuerpos intermedios y la crisis ecológica. Para ello es fundamental, sostiene Milbank, que el cristianismo católico, en sus diversas formas,

retome su tradición crítica del liberalismo por radicalidad y no conservadurismo.

Ahora bien, su idea de que la izquierda debiese reconsiderar el elemento religioso no es solo para nutrir un relato político. Milbank arriba a proposiciones concretas como, por ejemplo, que en materia de comercio “necesitamos ver una emergencia de la banca cooperativa (supervisada quizás por organismos eclesiales, islámicos y judaicos) para regular y adjudicar las interacciones entre los muchos modos diferentes de esfuerzo cooperativo” (2009b, p. 80).

Posteriormente en 2015 desarrolló la distinción entre liberalismo y “política del alma” (*politics of the soul*). Apelando a la no existencia de un área neutral en lo que refiere a la fundamentación metafísica de principios, plantea que las asunciones liberales llevan inevitablemente al materialismo al dejar de lado la dimensión espiritual del ser humano, tal como la han entendido distintas culturas y filosofías. El liberalismo busca conseguir el mal menor, pero al costo de hacerlo sobre las tendencias humanas más negativas (Milbank, 2015). Además, aventura Milbank que, pese a que en Rousseau se encuentra una versión inocente del ser humano, en contraposición a Hobbes, aun así, la situación actual consiste en que se tiene una “combinación de liberalismo hobbesiano de derechas en economía y un liberalismo rousseauiano de izquierdas en cultura” (2015, p. 103). No obstante, se mantiene la falta de política del alma. La contraposición queda del siguiente modo:

Mientras que el cristianismo creía que la realidad era originalmente y en el fondo pacífica, y solo violenta debido a la irrupción del pecado, y con todo en la práctica usualmente había fomentado la guerra, el liberalismo revierte exactamente esto. En nombre de reducir el conflicto, el liberalismo sin embargo pensó que la realidad era inherentemente agonística y los humanos naturalmente egotistas e inclinados al conflicto (Milbank, 2015, p. 102).

Es por esto que Milbank no puede aceptar al liberalismo, pues implicaría reconocer que se puede fundar un orden sobre la “amoralidad” y el “vicio”. A su turno, esta perspectiva es determinante para comprender el modo en que aborda la cuestión del liberalismo a escala internacional.

2.1. De la apariencia secular a la teología geopolítica

Puede decirse que la crítica de Milbank al orden internacional liberal se sostiene sobre la cuestión de la soberanía. En 2002, refiriéndose a la situación global, señalaba:

Las nociones teocráticas de soberanía no son simplemente algo arcaico dentro del islam, que está contra nuestra modernidad occidental. En varios sentidos las nociones teocráticas son específicamente modernas en su positividad y formalidad (como indicó Carl Schmitt). Bush, en una crisis, ha apelado al supuesto destino divino de América, y es el judaísmo moderno el que ha desembocado en una forma estatista, sionista. Ahora está surgiendo una terrible simbiosis entre el sionismo y la lectura literalista americana protestante y no cristiana del Antiguo Testamento en la tradición puritana, que iguala al anglo-sajonismo [*Anglo-Saxondom*] con Israel (Milbank, 2002, p. 313).

Tanto para Gray como para Milbank los acontecimientos posteriores al 9/11 dejan al descubierto una forma particular de operar del liberalismo a escala global. Es de aquí que Milbank desarrolla su concepto de “teología geopolítica”, trabajado en un texto publicado en 2009 y titulado “Geopolitical Theology: Economy, Religion and Empire after 9/11”. Volveremos luego sobre la alusión a Carl Schmitt, pues nos parece capital para comprender su análisis del asunto.

En general, la relación entre geopolítica y religión puede encontrarse en a lo menos cuatro formas. La geopolítica puede entenderse como una “realidad”, según la cual hay religiones en conflicto en un marco territorial determinado. También puede ser una “disciplina”, esto es, un marco teórico específico para estudiar problemas de índole geopolítica en que en ciertos territorios hay rivalidades de poder entre grupos religiosos. Luego, puede ser un “pensamiento o visión”, es decir, el “deber ser” que se le asigna a un territorio o espacio geográfico desde una perspectiva religiosa. Finalmente, también puede entenderse como la formulación de estrategias por parte de actores para intervenir en cuestiones de orden geopolítico. Aquí la religión es utilizada en el marco de algún diseño para conseguir cierto fin en ese terreno (Portillo,

2013). Si hubiese que encuadrar las ideas de Milbank según estas cuatro distinciones, podría decirse que su pensamiento es una “disciplina” en el sentido que formula un marco teórico para entender las relaciones entre religión y geopolítica.

En el texto referido, Milbank (2009a) observa varios fenómenos en la escena internacional como la globalización, el capitalismo anárquico, los estados autoritarios, el neoimperialismo y el aparente retorno de la religión. En términos analíticos, toma distancia explícitamente de la lectura neomarxista del contexto internacional, pues considera que las relaciones económicas internacionales no se circunscriben solo a una dimensión material, sino a una religiosa igualmente. Aquí, aunque no menciona explícitamente la forma de su abordaje, es posible distinguir dos modos en que Milbank pone a interactuar la geopolítica con la religión: el primero de ellos consiste en examinar los fenómenos religiosos globales explícitos y su relación con los asuntos políticos. Así, menciona por ejemplo el apoyo judío al Estado de Israel en alianza con evangélicos; luego el evangelicalismo conservador y su vinculación con el capitalismo; les sigue el resurgimiento del islam y, finalmente, el resurgimiento del papado y el catolicismo.

El segundo modo consiste en perfilar lo que es, propiamente, una “teología geopolítica”. Su pregunta es: si capital, imperio y la forma de globalización actual son un “modo desviado de cristianismo” (Milbank, 2009a, p. 97), entonces es necesario examinar el componente religioso de aquello, entendiendo el significado de “religión” en un sentido amplio como lo que “une entre sí” (*binding together*), pues es el punto en que imaginación y práctica se enlazan (2009a, p. 98). De lo que se trata aquí es que el orden capitalista neoliberal global consta, para Milbank, de una matriz religiosa y, en cuanto tal, “si el capitalismo es una cuasi-religión, también permanece como una herejía cristiana” (2009a, p. 95).

Esta relación compleja entre religión y economía, en la cual economía en general es religión y viceversa (Milbank, 2009a), se define en el marco de la “teología geopolítica” en un sentido schmittiano. Si para Carl Schmitt (2009) los conceptos de la teoría moderna del Estado son conceptos teológicos secularizados, para Milbank puede decirse que los conceptos teológicos del orden económico global moderno son conceptos teológicos secularizados. En sus propios términos: “es el caso que la globalización, el gobierno universal o ‘imperio’, y una economía global compartida son, de hecho, los resultados del legado cristiano” (Milbank, 2009a, p. 99) y las relaciones internacionales del occidente cristiano tienen

orígenes teológicos (2009a, p. 106). Nociones como “gobierno universal” e “imperio” no son comprensibles sino atendiendo a la estrecha relación que hubo entre cristianismo —católico romano particularmente para el Occidente— y política. Es de ahí que se desprende, como una desviación teológica herética, el liberalismo y todo lo que Milbank asocia con él. La tendencia al universalismo terrestre, esto es, a querer expandirse por el orbe (es decir, implica él, globalización), es algo propiamente cristiano católico.

Estareflexiónsobreenelnúcleoteológico delageopolíticacontemporánea es la que, en efecto, nos lleva de regreso a la cuestión de la soberanía. Schmitt observó el componente teológico-político tras la soberanía estatal, y Milbank lo ve reflejado en un país como Estados Unidos. Si en 2002 Milbank afirmaba que las nociones teocráticas de soberanía son específicamente modernas, en 2009 sostiene que “el surgimiento de una soberanía política absoluta pura es una permutación” (2009a, p. 109) de modelos heréticos cristianos, refiriéndose a las herejías del arrianismo y nestorianismo. Es desde el núcleo teológico de la soberanía estatal que Milbank arriba a la idea de que el orden internacional actual responde a una herejía cristiana y que, por ende, solo puede ser reemplazada con cristianismo.

3. Dos postliberalismos

3.1. El postliberalismo de Gray: pluralismo de valores

El “postliberalismo” de Gray se desarrolla a partir de su concepto de “liberalismo”, expuesto a lo largo de varios trabajos, como *Liberalisms* (1989), *Two Faces of Liberalism* (2001) y *Post-Liberalism* (1996). Al discutir el problema de la identidad de una cierta tradición intelectual liberal, y realizando un balance de su propio pensamiento, señala que la interpretación que sostuvo antes, según la cual el liberalismo puede ser dividido en las versiones clásica y revisionista, es errónea (Gray, 1996, p. 285). En vista de esta autocrítica, resulta más apropiado seguir la noción de liberalismo que tiene Gray al momento de escribir *Post-Liberalism*. Señala el autor que:

[...] la posición defendida aquí es postliberal en cuanto que rechaza las afirmaciones fundacionalistas del fundamentalismo liberal. Es decir, niega que solo los regímenes liberales son legítimos para todos los

seres humanos. Los seres humanos han florecido en regímenes que no cobijan una sociedad civil liberal, y hay formas de florecimiento humano que son expulsadas en regímenes liberales. Los órdenes liberales no tienen, entonces, autoridad universal o apodíctica, contrariamente a la filosofía política liberal (Gray, 1996, p. 284).

Resulta sugestivo que el autor utilice un término convencionalmente asociado a las religiones para referirse a una forma no religiosa de pensar el orden político. Al escribir sobre fundamentalismo liberal como se podría escribir sobre fundamentalismo cristiano o musulmán, pareciera que Gray llama la atención, más que sobre un conjunto de ideas, sobre una actitud respecto a ellas. No son ideas a las que solo se considera verdaderas —lo cual de suyo ya es un problema intelectual para el liberalismo—: también son ideas que, en cuanto verdaderas, deben ser aceptadas por todos. De forma aún más clara, Gray dirá que “el peligro que enfrentamos en las sociedades civiles establecidas de Occidente es el de la corrosión de las practicas liberales por la ideología liberal” (1996, p. 327). El problema es la “ideología liberal”, no las “prácticas liberales” ni las instituciones liberales. Puede advertirse que estos elementos aparecen siete años después en *Two Faces of Liberalism*, en que Gray (2001, p. 31) retoma el empleo de la categoría “fundamentalismo” para cierta forma universalista de liberalismo ideológico y, de la misma manera, distingue la tolerancia liberal entre aquella que persigue alguna forma de vida ideal, y aquella que reconoce el pluralismo de valores (Gray, 2001, p. 12). A su juicio, “lo que es nuevo en el mundo moderno no es la aceptación de la diversidad en estilos de vida sino su hostilidad a las jerarquías” (2001, p. 4). Por lo tanto, lo que está siempre en juego es el deseo de jerarquizar.

Veamos, entonces, en qué se diferencian ideología liberal de práctica liberal. A juicio de Gray, hay cuatro grandes ideas que caracterizan al liberalismo y que pueden ser defendidas racionalmente cuando se contextualizan como elementos propios de las sociedades tardo o postmodernas: el individualismo moral o normativo, el universalismo, el meliorismo (*meliorism*) y el igualitarismo.

El individualismo moral o normativo es aquella idea que sostiene que no hay otro valor último excepto los estados mentales y sentimientos del individuo. El universalismo, por su parte, es la idea de que todo ser

humano tiene derechos y deberes que se le deben respetar, sin importar sus herencias culturales o circunstancias históricas. El meliorismo es la idea de que las instituciones humanas están abiertas al desarrollo indefinido mediante el ejercicio de la razón crítica. Hay aquí una cierta noción de “progreso”. Por último, el igualitarismo es la negación de cualquier jerarquía natural, moral o política entre los seres humanos, porque la especie humana es una comunidad moral de un solo estatus.

Gray, antes de su análisis, adelanta la conclusión:

[...] ninguna puede resistir la fuerza de una fuerte indeterminación e inconmensurabilidad radical de valores. Considerado como una posición en filosofía política, por lo tanto, el liberalismo es un proyecto fracasado [...] como perspectiva filosófica, está muerto. ¿Qué es, entonces, lo que queda vivo del liberalismo? El aspecto del liberalismo que aún queda vivo para nosotros, defenderé, es la concepción y la realidad histórica de la sociedad civil que nos ha sido legada (1996, p. 287).

Pues bien, estas concepciones resultan en una serie de prácticas contextuales e históricas —materializadas en instituciones— que son defendibles en un contexto de modernidad tardía, y el riesgo latente es que se las comprenda fuera de ese límite. Para el liberalismo que se concibe y busca la forma de vida ideal, las instituciones son aplicaciones de principios universales, mientras que para el liberalismo de pluralismo de valores son el medio para lograr la coexistencia pacífica (Gray, 2001, p. 12). ¿En qué fallan estas cuatro concepciones y qué es lo que se puede rescatar de ellas?

En lo que toca al meliorismo o “las ilusiones del progreso”, se rechaza la creencia en un único conjunto completo de principios que pueden configurar la realidad social. Hay conflictos entre libertades y conflictos entre valores inconmensurables. Se requiere, por tanto, de un objetivo más humilde para la filosofía política porque su fin es elucidar las prácticas y no aspirar a gobernarlas, afirma Gray (1996, p. 288), siguiendo en esto a Michael Oakeshott (1962). Al mismo tiempo, Gray (1996, p. 290) admitirá la teoría del pluralismo objetivo de Isaiah Berlin (1990) porque es un hecho que hay conflictos entre valores últimos que no se pueden resolver racionalmente. En otros términos, si el meliorismo sostiene que las instituciones humanas pueden progresar

indefinidamente mediante el uso crítico de la razón, desde Gray podría decirse que no hay un solo tipo de institución humana y, por tanto, que el progreso no significa lo mismo para todas, lo cual también implica que por tener cada una sus características propias, no puede haber tal cosa como unas mejor que otras.

Para explicar este problema, Gray utiliza como ejemplo el arte. En efecto, no puede decirse que haya una regla general para definir la calidad de una obra artística. En cada cultura estas reglas son distintas y, por lo tanto, el progreso o decadencia artística no se da dentro de la historia humana en general, sino dentro de tradiciones con sus propias normas internas. En sus palabras, “el tema de la inconmensurabilidad entre los bienes artísticos afirma [...] que no hay una forma de gran arte que sea la mejor, dado que no hay un estándar general por el que se pueda hacer tal juicio” (1996, p. 289). Esta noción de “pluralismo” acaba con la idea de la vida humana perfecta, porque muestra que hay variadas formas de “perfección” según el punto de vista de los valores que se adopten y, al situar todo aquello en los contextos particulares en que se da, el enfoque pluralista niega también la noción de una “historia humana” como un todo que puede tener significado omniabarcante. Tal significado solo puede existir en el marco de una u otra visión de mundo que la haga admisible. Así, no hay una mejora continua ni transversal de la humanidad como un todo, pues tal cosa solo ha de ser una presunción de una de las tantas visiones de mundo que se expresan en la pluralidad de la historia y, si hay tal progreso, solo puede ocurrir dentro de una civilización para la cual esa noción sea aceptable.

Los componentes universalista e igualitario del “síndrome liberal”, por su parte, son complementarios, señala Gray, debido a que el universalismo promueve que hay una serie de principios conferidos a todos los seres humanos y que esto a su vez comprende ciertas igualdades básicas, como enseña el igualitarismo. El pluralismo de valores da un golpe importante a ambas tendencias en la medida que hay una diversidad de bienes humanos y formas de florecimiento alcanzadas en las distintas sociedades. En sus términos:

[...] las políticas morales nunca se basan en derechos, debido a que los derechos nunca son ellos mismos fundacionales, sino más bien intermediarios entre reclamaciones sobre intereses humanos que son vitales para el bienestar y reclamaciones sobre obligaciones

que es razonable imponer sobre otros de acuerdo a estos intereses. En otras palabras, los derechos ganan su contenido desde los requerimientos del bienestar humano, y serán variables así como varían las demandas de bienestar humano (Gray, 1996, p. 303).

Finalmente, el individualismo normativo es entendido como cualquier teoría moral que no reconozca ningún valor intrínseco a cualquier bien colectivo y que, por tanto, considere que el valor de cualquier bien de este último tipo solo tiene un fin instrumental a aquello que sí tiene valor intrínseco: el individuo y su situación. Lo que ataca Gray es la idea de que solo esta forma de comprender la vida puede reportar un auténtico florecimiento. En efecto, “el individualismo nunca es lo fundamental en lo que tiene valor. Lo que tiene valor (en el mundo humano) es la forma de vida, incluso si, paradójicamente, esta es una forma de vida individualista” (1996, p. 311). Para Gray, nadie tiene la capacidad de ser el individuo radicalmente situado capaz de escoger libremente, puesto que incluso la condición posmoderna de profundo pluralismo no se elige, del mismo modo que no se elige nacer, por ejemplo, en un país de valores teocráticos. Así, no es el individualismo en cuanto tal lo que importa, sino la forma de vida, y aun cuando al individualismo se le reconoce su justo valor, aquello se hace entendiéndolo como un producto histórica y contextualmente situado.

Lo que queda del liberalismo, según Gray (1996, p. 314), es la sociedad civil, que tiene por características el ser tolerante a la diversidad de perspectivas religiosas y políticas y en la que el Estado no busca imponer una doctrina comprehensiva; esto es una forma de pervivencia del individualismo. Además, en ella tanto el gobierno como el ciudadano restringen su conducta al imperio de la ley (1996, p. 315), conservándose así el igualitarismo ante la ley, que es la única igualdad necesaria en una sociedad civil. En ella existe la institución de la propiedad privada (1996, p. 315) y se la puede identificar con el capitalismo de mercado (1996, p. 316). En lo que toca al meliorismo, una sociedad civil puede mejorar dentro del marco que sus propios términos le permiten. El universalismo, finalmente, sobrevive en la forma de la universalidad de la sociedad civil como condición para la prosperidad y paz de cualquier civilización moderna. Esta noción de “sociedad civil”, que es la posición postliberal de Gray, puede resumirse en una de sus famosas ideas: “no necesitamos valores comunes para vivir juntos en

paz. Necesitamos instituciones comunes en las que puedan coexistir muchas formas de vida” (2001, p. 40).

La perspectiva postliberal de Gray es un correctivo crítico que desmitologiza al liberalismo doméstico, y por ello permite luego desmitologizar el imperialismo liberal. Invita a renunciar a la idea de que la creencia en valores universales lleve lógicamente a la búsqueda de su imposición, y que la igualdad liberal no es el único valor que garantiza el florecimiento humano. Lleva a moderar las expectativas respecto al alcance del individualismo liberal porque hay contextos en que aquello no es reconocido como el valor último, y lleva a comprender que la idea de progreso concebida en el mundo occidental liberal contemporáneo pertenece a un contexto histórico que le permite existir. Lo que es progreso para el liberalismo democrático occidental, no lo es bajo otras formas de vida. De este modo, se produce un descentramiento, una puesta en perspectiva que disocia a la pretensión histórica de la realidad histórica del liberalismo. Se pueden conservar las prácticas liberales y sus beneficios a la vez que renunciar a su ideologización tanto a escala doméstica como global.

3.2. El postliberalismo de Milbank: política de la virtud

Las ideas de Gray han sido objeto de reflexión en la obra de John Milbank, lo cual queda reflejado particularmente en un libro escrito junto con el politólogo alemán Adrian Pabst. En *The Politics of Virtue*, Milbank y Pabst (2016) atienden las críticas y diagnósticos de Gray al liberalismo en su clave postliberal, pero no apuestan por un pesimismo y por la inconmensurabilidad de posturas, sino por reafirmar un comunitarismo de valores sustantivos. Aunque comparten la crítica de Gray, para ellos la situación postliberal no ha de conformarse con el pluralismo de valores, sino que ha de orientarse desde y más allá de una visión comunitarista del orden social en la que haya una “primacía de la libertad positiva y una visión sustantiva de un verdadero florecimiento humano” (Milbank y Pabst, 2016, p. 384). En efecto, lo que buscan es una:

[...] combinación de élites honorables y virtuosas con una mayor participación popular; un mayor sentido de deber cultural y jerarquía de valor y honor, junto con mucho más igualdad real y genuina libertad creativa en los ámbitos político y económico. Esto sería realizable mediante un enfoque mutualista renovado en ambos,

asuntos domésticos y exteriores, que sustituya a la dominancia del mercado, el Estado y la tecnocracia con la primacía de la sociedad, la cultura y las relaciones interpersonales (Milbank y Pabst, 2016, pp. 1-2).

Esto implica que el postliberalismo, teóricamente, no tiene solo un resultado. Más bien, se trataría de un “momento”, como lo llama Pabst (2020, p. 177) en otra ocasión. En lo que toca al debate británico contemporáneo en torno al Brexit, Pabst (2017) desarrolló su noción del postliberalismo como una alternativa que conjuga la preocupación por la sociedad sobre el mercado y el Estado y el énfasis en la importancia de las familias y grupos sociales como los primeros a los que se debe lealtad. Pero, además, dos años antes de *The Politics of Virtue*, Milbank y Pabst ya sostenían que la realidad actual de Inglaterra se podía observar en dos aspectos cruciales: primero, la expansión del Estado y del mercado a expensas de la sociedad civil, y segundo, la retirada de la Iglesia anglicana de sus labores caritativas, sociales, educacionales y culturales (Milbank y Pabst, 2014). Es a partir de esto que piensan en una política postliberal en la cual la Iglesia de Inglaterra puede jugar un rol determinante en fortalecer la sociedad civil frente al Estado y al mercado.

Al igual que con Gray, la noción de “postliberalismo” de Milbank y Pabst arranca de una valoración general del liberalismo. Puede decirse que la perspectiva que Milbank tiene sobre el liberalismo se proyecta plenamente en el trabajo escrito con Pabst. En efecto, la idea de que se está ante dos liberalismos relacionados, el hobbesiano de derechas en economía y el rousseauiano de izquierdas en cultura (Milbank, 2015), es el punto de partida de la obra con Pabst, pues ambos liberalismos siempre estuvieron en una “alianza secreta” (Milbank y Pabst, 2016, p. 1). Pero no solo eso, también señalan que los cuatro elementos que Gray engloba como aspectos liberales —junto con otros—, si bien puede que el liberalismo los presente de una manera distintiva, no pertenecen en absoluto al liberalismo. Aquí, entonces, recurren a la distinción entre liberalidad y liberalismo, aventurada ya por Milbank (2009b), explicando la liberalidad como el supuesto de que “las sociedades están más fundamentalmente unidas por la generosidad mutua que por el contrato” (Milbank y Pabst, 2016, p. 81). Enraizando esos valores en la cultura grecorromana, el Medievo y el Renacimiento, los autores sostienen que una:

[...] política de la virtud, necesariamente como política del alma, tiende a estar asegurada por existir al interior de un gobierno [*polity*] de comunidad “transpolítica”, orientado más allá del tiempo hacia la eternidad, y dedicado a ideales más rigurosos que la comunidad en general, los que sin embargo sirven para guiarla (Milbank y Pabst, 2016, p. 30).

Lo que el liberalismo hace es “minar el rol de una comunidad psíquica abarcadora” y dejar al individuo al arbitrio de una “colectividad anónima” (2016, p. 31). El liberalismo por definición es, entonces, independiente de otros elementos con los que se lo quiera caracterizar y de sus distintas vertientes, “la primacía del individuo” (2016, p. 33), con la cual se producen otros fenómenos:

Siempre aliado con esta primacía está el reemplazo de nociones de “bien” o “verdad” sustantivas con los derechos subjetivos como lo último, fundamentados subjetiva y voluntaristamente; la sustitución del contrato social formal en lugar de la unidad sustancial prescrita y el privilegio del progreso (hacia la libertad negativa) y las “leyes de la historia” (implicando la necesaria “racionalidad” y, por tanto, necesidad lógica de este progreso) por sobre la tradición y la contingencia (2016, p. 33).

Así, el liberalismo es “atomista” y oscila, señalan Milbank y Pabst, entre la subjetividad individual y la unidad colectiva. La alternativa postliberal de “política de la virtud” que ofrecen consiste en unir “más justicia económica con reciprocidad social” (2016, p. 3). Definido el postliberalismo como “la primacía del bien común” (2016, p. 72), los autores explican que esta perspectiva se compone de cinco elementos: cultura del honor, la comunidad como combinación de virtud y don, verdadero socialismo, pluralismo político y corporativismo genuino. Una cultura del honor, sobre todo en el ámbito económico y de negocios, no se alcanza solo con leyes: requiere un nuevo *ethos* que haga que dicha actividad trascienda el solo interés propio. Sobre la comunidad, por su parte, se propone una recuperación del compromiso ético con el grupo de pertenencia, de modo que la búsqueda de restablecimiento de comunidad de la clase trabajadora, con la búsqueda de más realización

holística de las clases medias, lleve a una combinación de trabajo con familia y comunidad. Con “verdadero socialismo” se refieren a retomar la reciprocidad que caracterizó al socialismo británico, es decir, prácticas sociales de asistencia mutua frente al Estado y al mercado. En cuanto al pluralismo político, refiere a la concepción de que los derechos no solo son individuales, sino que también debe reconocerse a los grupos sociales, es decir, a los cuerpos intermedios que se encuentran entre el Estado y el individuo. Por último, el corporativismo consiste en no separar lo político de lo social, ni lo económico de lo político, sino en la distribución de responsabilidad hacia las formaciones intermedias. En todo esto, no pierden de vista el componente jerárquico ni el religioso.

Liberalismo económico y liberalismo cultural potencian los derechos individuales, pero descuidan a los cuerpos intermedios, al tiempo que también están estrechamente relacionados con el Estado. El postliberalismo que proponen, entonces, busca ir más allá del individuo y del Estado, y apela a distintas modalidades de orden y organización social que fortalezcan a los cuerpos intermedios, los grupos sociales que componen el espacio que queda entre el individuo y el Estado.

Este postliberalismo tampoco descuida pronunciarse sobre el “imperialismo liberal” conducido por Estados Unidos a través de las Naciones Unidas. Respecto a la situación global, Milbank y Pabst plantean que, dado que las relaciones internacionales han estado marcadas por el secularismo en las últimas décadas, es conveniente “reinyectar” la tradición del realismo cristiano como marco de comprensión de dichas relaciones por poner énfasis en la primacía de las fuerzas culturales. Proponen la implementación de un modelo “asociativo” (2016, p. 361) basado en una comprensión metafísica que está constituida por el reconocimiento del origen trascendente de valores inmanentes. De ese modo, se antepone la cooperación al contrato. Así, además, se reconoce la pluralidad y la idea de bien común:

Tomados en conjunto, el compromiso con el bien común y el corporativismo constitucional por parte de potencias como Gran Bretaña y sus aliados, en teoría podría transformar el modelo dominante de globalización neoliberal. El foco en fines sustantivos compartidos puede corregir la fijación ya sea con relaciones instrumentales y transaccionales (intereses corporativos meramente nacionales o internacionales) y

con ataduras procedimentales (compromiso con reglas comunes y cuerpos reguladores) hacia la realidad de vínculos compartidos culturales y sociales que importan más en un mundo crecientemente globalizado (2016, p. 361)

Así las cosas, el postliberalismo de Milbank y Pabst a escala internacional propone, por un lado, que se reconozcan principios distintos de los liberales seculares, pero al mismo tiempo propone que el foco de la globalización neoliberal debe revertirse implementando un modelo alternativo basado en la aceptación de una comprensión metafísica de la realidad. En suma, tanto en lo doméstico como en lo global, Milbank invita a establecer valores sustantivos y jerarquías. Así, la política de la virtud se opone al pluralismo de valores de Gray, que no busca ni jerarquización de valores ni establecimiento de valores específicos, ni en lo global ni lo doméstico.

Consideraciones finales

Revisadas las posturas de Gray y Milbank frente al liberalismo a escala global, su relación con lo religioso y las propuestas postliberales de ambos, cabe una reflexión sobre la significancia de estas ideas a la luz de la escena contemporánea. Lo primero que resalta es el hecho de que Gray es, todavía, un tipo de liberal, mientras que Milbank es un decidido opositor. Pero de fondo, lo que yace es una oposición de posturas frente al ser humano: Gray es un pesimista antropológico, pero Milbank prefiere el optimismo. Sobre ello descansa la posibilidad de construir un mundo virtuoso. El primero busca liberar al liberalismo del riesgo de volverse religioso, pero el segundo busca combatirlo por ser una expresión religiosa herética. En un estilo como el de Voegelin, Gray analiza el problema del liberalismo como religión política, y en un estilo como el de Schmitt, Milbank analiza el problema de que los conceptos del mundo liberal sean de origen teológico.

Estas observaciones repercuten en la forma de pensar la realidad política doméstica y global. Gray prefiere el *modus vivendi* en ambos terrenos. Sabe que en las sociedades no se podrá estar siempre de acuerdo porque hay cuestiones fundamentales irreconciliables y, por ende, confía en instituciones que permitan el pluralismo de valores. Del mismo modo, sabe que la exportación imperial del liberalismo es una empresa utópica y acepta la realidad plural del escenario global.

Milbank, por su parte, busca abiertamente una mayor injerencia religiosa por la vía de una mayor aceptación de los elementos religiosos en las izquierdas y particularmente promoviendo su concepción de política de la virtud socialista cristiana en la política nacional. Del mismo modo, propone que el orden internacional esté orientado abiertamente por una concepción metafísica de la realidad. A escala doméstica y global, pluralismo de valores y política de la virtud se oponen, finalmente, porque el primero no busca establecer una jerarquía valórica, mientras que la segunda se plantea ella misma como una escala de valores. Una admite que el florecimiento humano puede alcanzarse de maneras diferentes, mientras la otra cree ser ella la mejor forma de alcanzar dicho florecimiento. Si el liberalismo se ha vuelto una religión política como piensa Gray, o si es una heterodoxia cristiana como piensa Milbank, las opciones que se ofrecen son dos: desmitologizarlo para mantener un cierto *modus vivendi*, o reemplazarlo por otra religión, un tipo de interpretación política no herética del cristianismo.

Es aquí que cabe el ejercicio evaluativo pues, aunque el problema teológico-político tratado se remite al campo académico, no deja de tener significancia en el escenario contemporáneo del mundo occidental —dentro del cual también consideramos a Latinoamérica— y global. Dentro del mundo occidental, las instituciones liberales han logrado sostener con diverso éxito el pluralismo dentro de las distintas sociedades. No obstante, fenómenos como autoritarismos diversos, el populismo y la injerencia pública de ciertas expresiones religiosas insatisfechas con el pluralismo, parecieran poner en entredicho dichas instituciones. A nivel global, no se observa solo el ascenso de modelos alternativos al liberalismo democrático, sino que en realidad históricamente siempre ha habido culturas y tradiciones políticas distintas que son reivindicadas por Estados, y esto se hace evidente en la situación de multipolaridad actual. Así las cosas, tanto a nivel internacional como dentro de los países occidentales el liberalismo democrático se ve bajo asedio.

En este contexto, para ponderar cuál de los dos pensadores ofrece un marco de pensamiento para abordar de manera más apropiada la realidad de la pluralidad, resulta útil leerlos a ambos desde la perspectiva del otro. Si bien, de acuerdo con el fundamentalismo liberal, las ideas de Milbank deberían ser excluidas del debate, lo cierto es que desde la mirada de Gray son admisibles dentro del *modus vivendi* de pluralismo de valores, tanto en lo doméstico como en lo internacional. El socialismo cristiano a la inglesa y la mirada metafísica de participación

internacional pueden convivir con otras nociones. Evidentemente, en este modelo las ideas de Milbank no podrían posicionarse como valores sustantivos para todos, pero tampoco habría impedimento para que las practicara quienes las aceptan. Ahora bien, si las ideas de Gray se miran desde el marco de Milbank, el resultado es diferente. Las ideas del primero no son admisibles porque, aunque Gray desmitologiza el liberalismo y lo conserva como una práctica, para Milbank el liberalismo en sí mismo sigue siendo un problema fundamental. En este sentido, la política de la virtud tiene un componente exclusivista que, hipotéticamente implementado, tendría dificultades con el pluralismo de valores porque se concibe a sí mismo como un esquema de valores que debe ser predominante. Esto no quiere decir que la política de la virtud no pueda contribuir a la convivencia, pero el escenario actual no es propicio para que ideas como esas moldeen el conjunto de una sociedad.

A nivel internacional, la idea de Milbank de recuperar el realismo cristiano y la dimensión metafísica de la realidad frente a la globalización neoliberal puede encontrar resonancia en aquellas regiones en que el liberalismo y la secularización no son bien recibidos. Su limitación está, más bien, dentro del mundo occidental secularizado, en el que una perspectiva así difícilmente sea vista como otra cosa que una interpretación occidental del cristianismo, la que además no es extensiva a todos los que se llaman cristianos. A nivel doméstico, la política cristiana no concita las adhesiones democráticas de décadas pasadas —salvo las expresiones conservadoras de los últimos años—. Si, como él piensa, la unidad sustancial prescrita fue sustituida por el contrato social, entonces uno de los mayores problemas de su posición es precisamente que necesita que la idea de la unidad sustancial vuelva a ser plausible. En este contexto, la viabilidad de un proyecto socialista cristiano requeriría previamente de un cambio profundo de estado de cosas en cuanto al valor asignado a las religiones y su relación con la vida pública que no parece estar dentro de las expectativas de las sociedades modernas.

Por contraparte, las ideas realistas de Gray en materia internacional parecen más viables porque, si bien no se identifica con una tradición valórico-religiosa, sin embargo, muestra la suficiente amplitud como para dirigirse a ellas y a una amplia audiencia gracias al marco plural de su pensamiento. A nivel doméstico, aunque no propone un cambio radical de estado de cosas, reconoce la importancia de una corrección

y mejoramiento de lo que ya se tiene. No puede desconocerse que las instituciones liberales han ayudado a una relativa paz social en las sociedades occidentales y lejos de buscar un reemplazo, lo que propone es mantenerlas desde una mirada realista. Así, no pide, con ánimo utópico, llegar a una unidad sustancial en lo social ni tampoco buscar una alternativa a la globalización neoliberal. Su propuesta, más modesta, es buscar el mayor bien posible dentro de los márgenes del legado liberal que ha conformado al mundo occidental, y ello se logra mediante la renuncia al fundamentalismo liberal.

Al buscar la matriz religiosa del liberalismo, Milbank inevitablemente acaba por tratarlo como una religión. Si esto se lee desde Gray, podría entenderse como un ataque a la religión política liberal. De ser así, las perspectivas de ambos autores podrían complementarse. Pero si la discusión, como la piensa Milbank, es entre su tipo de cristianismo y el liberalismo sin distinciones como las que hace Gray, entonces no hay solución posible. Uno tendrá que prevalecer sobre el otro. En cambio, si la discusión es entre ideología liberal y práctica liberal, tal como hizo Gray, aquello deja suficiente espacio para el desarrollo y convivencia de ideas distintas, pues una aproximación realista como la suya entiende la pluralidad de valores sin dejar de defender por ello la existencia de la sociedad civil moderna con sus propios y legítimos criterios como una forma específica de sociedad. Puede haber coexistencia pacífica respetando las instituciones liberales y evitando la imposición de los valores de unos sobre otros. Pero la condición de posibilidad de este tipo de sociedad ya no obedece solo a buscar que se sujeten a ella quienes profesan valores distintos de los liberales sino, paradójicamente, a pedirlo de quienes los profesan religiosamente.

Bibliografía

- Anderson, P. (2014). *Imperium et consilium*. Akal.
- Asad, T. (2003). *Formations of the Secular*. Stanford University Press.
- Berlin, I. (1990). *The Crooked Timber of Humanity*. John Murray.
- Casanova, J. (1994). *Public Religions in the Modern World*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226190204.001.0001>
- Cohn, N. (2004). *The Pursuit of the Millenium*. Pimlico.
- Critchley, S. (2017). *La fe de los que no tienen fe*. B. Thoilliez (trad.). Trotta.
- Crowder, G. (2007). Gray and the Politics of Pluralism. En J. Horton y G. Newey (eds.), *The Political Theory of John Gray* (pp. 59-76). Routledge.

- Del Palacio, J. (2012). Conservadurismo británico contemporáneo: John Gray y la teoría política del *modus vivendi*. *Isegoría*, 47, 601-614. <https://doi.org/10.3989/isegoria.2012.047.12>
- Forde, J. (2020). *Anglican Socialism and Welfare: John Milbank's "Blue Socialist" Thinking and the Church of England's Approach to Welfare since 2008*. [Tesis de doctorado, Universidad de Manchester]. Research Explorer. <https://research.manchester.ac.uk/en/studentTheses/anglican-socialism-and-welfare-john-milbanks-blue-socialist-think/>
- Fukuyama, F. (1989). The End of History? *The National Interest*, 16, 3-18. <https://doi.org/10.1080/07366988909450540>
- Gray, J. (1989). *Liberalisms*. Routledge.
- Gray, J. (1996). What Is Dead and What Is Living in Liberalism? En J. Gray, *Post-Liberalism* (pp. 283-328). Routledge.
- Gray, J. (2001). *Las dos caras del liberalismo*. M. Salomon (trad.). Paidós.
- Gray, J. (2011a). El fin de la historia, ¿otra vez? En J. Gray, *Anatomía de Gray* (pp. 271-278). A. Santos Mosquera (trad.). Paidós. 271-278.
- Gray, J. (2011b). El *homo sapiens* y la extinción en masa. En J. Gray, *Anatomía de Gray* (pp. 487-494). A. Santos Mosquera (trad.). Paidós.
- Gray, J. (2011c). Orígenes culturales del comunismo soviético. En J. Gray, *Anatomía de Gray* (pp. 257-260). A. Santos Mosquera (trad.). Paidós.
- Gray, J. (2017). *Misa negra. La religión apocalíptica y la muerte de la utopía*. A. Santos Mosquera (trad.). Sexto Piso.
- Habermas, J. (2006). *Entre naturalismo y religión*. P. Fabra, D. Gamper, F. J. Gil Martín, J. L. López de Lizaga, P. Madrigal y J. C. Velasco (trads.). Paidós.
- Hardt, M. y Negri, A. (2000). *Imperio*. A. Bixio (trad.). Paidós.
- Hazon, Y. (2018). *The Virtue of Nationalism*. Basic Books.
- Huntington, S. (1997). *El choque de civilizaciones*. J. P. Tosa Abadía (trad.). Paidós.
- Kelly, P. (2007). The Social Theory of Anti-Liberalism. En J. Horton y G. Newey (eds.), *The Political Theory of John Gray* (pp. 25-42). Routledge.
- Milbank, J. y Pabst, A. (2014). The Anglican Polity and the Politics of the Common Good. *Crucible: The Christian Journal of Social Ethics*, 1, 7-15.
- Milbank, J. y Pabst, A. (2016). *The Politics of Virtue*. Rowman & Littlefield. <https://doi.org/10.5040/9798881817411>
- Milbank, J. (1997). *The Word Made Strange*. Blackwell.
- Milbank, J. (2002). Sovereignty, Empire, Capital and Terror. *South Atlantic Quarterly*, 101(2), 304-324. <https://doi.org/10.1215/00382876-101-2-305>

- Milbank, J. (2006). *Theology and Social Theory: Beyond Secular Reason*. Blackwell. <https://doi.org/10.1002/9780470694121>
- Milbank, J. (2009a). Geopolitical Theology: Economy, Religion and Empire after 9/11. En M. Morgan (ed.), *The Impact of 9/11 on Religion and Philosophy* (pp. 85-112). Palgrave Macmillan. https://doi.org/10.1057/9780230101609_7
- Milbank, J. (2009b). Liberalidad contra liberalismo. En J. Milbank y A. Pabst, *El pensamiento de John Milbank* (pp. 67-94). S. Montiel Gómez (trad.). Editorial Nuevo Inicio.
- Milbank, J. (2009c). *The Future of Love*. Cascade Books.
- Milbank, J. (2015). The Politics of the Soul. *Open Insight*, 6(9), 91-108. <https://doi.org/10.23924/oi.v6n9a2015.pp91-108.127>
- Nye, J. (2004). *Soft Power*. Public Affairs.
- Oakeshott, M. (1962). *Rationalism in Politics*. Basic Books.
- Pabst, A. (2017). Postliberalism: The New Centre Ground of British Politics. *The Political Quarterly*, 88(3), 500-509.
- Pabst, A. (2020). The Post-liberal Moment. *Telos*, 190, 177-180. <https://doi.org/10.1111/1467-923X.12363>
- Portillo, A. (2013). Una aproximación a la comprensión de la geopolítica de las religiones. *Fermentum. Revista Venezolana de Sociología y Antropología*, 23(67), 135-148.
- Rawls, J. (1995). *Liberalismo político*. S. R. Madero Báez (trad.). FCE.
- Scattola, M. (2008). *Teología política*. H. Cardoso (trad.). Ediciones Nueva Visión.
- Schmitt, C. (2009). *Teología Política*. F. J. Conde y J. Navarro Pérez (trads.). Trotta.
- Simpson, P. (2018). *Political Illiberalism*. Routledge.
- Sleat, M. (2013). *Liberal Realism*. Manchester University Press. <https://doi.org/10.7228/manchester/9780719088902.001.0001>
- Stout, J. (2004). *Democracy and Tradition*. Princeton University Press. <https://doi.org/10.1515/9781400825868>
- Strauss, L. (2011). *Leo Strauss. El filósofo en la ciudad*. C. Hilb (ed.). Prometeo Libros.
- Voegelin, E. (2014). *Religiones políticas*. M. Abella Martínez y P. García Guirao (trads.). Trotta.
- Wolterstorff, N. (1997). The Role of Religion in Decision and Discussion of Political Issues. En R. Audi y N. Wolterstorff (eds.), *Religion in the Public Square: The Place of Religious Convictions in Political Debate* (pp. 67-120). Rowman and Littlefield.

Reseñas

Koch, K. (2023). *Denken in Zwecken. Bedeutung und Status der Teleologie in der theoretischen Philosophie Kants und Hegels*. Felix Meiner Verlag. 310 pp.

La monografía de Karen Koch sobre el polivalente concepto de teleología en Kant y en Hegel constituye una aportación muy valiosa tanto para la *Kant-Forschung* como para la *Hegel-Forschung*. Es un texto valioso por varias razones: su actualidad, su claridad y su originalidad. Su actualidad reside en que a lo largo de trescientas páginas mantiene una discusión sostenida con autoras y autores que en los últimos años se han ocupado directa o indirectamente del concepto, tales como Ginsborg y Massimi o Ng y Kreines. Constituye así una fuente de consulta invaluable que rastrea y expone con probidad las discusiones a las que ha dado pie el multifacético concepto. Ofrece además una lectura propia, original y propositiva del concepto tanto al interior de los sistemas filosóficos de cada uno de los autores como respecto del vínculo entre ambos, reavivando la larga discusión que surge cuando se aborda la filosofía hegeliana en relación con la kantiana y que sugiere la pregunta por las continuidades y/o las discontinuidades.

El libro se divide del siguiente modo: 1) “Introducción”, 2) “La posibilidad del caos empírico”, 3) “La presuposición de la cognoscibilidad de la naturaleza”, 4) “El agnosticismo de Kant. Hacia la concepción del concepto de ‘fin natural’”, 5) “Los límites del mecanismo”, 6) “La realidad efectiva de la finalidad interna”, 7) “Conocimiento finito y absoluto en la *Ciencia de la lógica*”, 8) “Las concepciones de la finalidad en Kant y Hegel en comparación” y 9) “Conclusión”. Del capítulo 2 al 4, el texto se enfoca en la filosofía teórica kantiana; del 5 al 7, Koch explora los problemas inherentes a la filosofía hegeliana. En el último capítulo, la autora revela los puntos en los que ella marca las intersecciones del concepto en los dos autores.

De acuerdo con mi lectura, tres son las preguntas más importantes a las que el libro ofrece una respuesta innovadora:

1) ¿Cuál es la relación entre el principio teleológico y el principio de la unidad sistemática en Kant?

2) ¿Cuál es la relación entre finalidad externa y finalidad interna en la *Ciencia de la lógica*?

3) Si se admite que Hegel no es un filósofo trascendental (cfr. Koch, 2023, p. 269), ¿puede decirse que la filosofía hegeliana de la teleología *funda* la teleología en Kant?

La respuesta afirmativa a esta pregunta es, a mi entender, una de las más arriesgadas pero originales del texto.

La palabra que la autora usa —“fundación” (*Fundierung*)—, no es, considero, la más adecuada para describir el papel de la filosofía hegeliana en relación con la kantiana, pues, como es sabido, Hegel ofrece objeciones a la categoría de “fundación” o de “fundamento” (*Grund*) en la *Ciencia de la lógica*, en donde esta categoría, que aparece en la doctrina de la esencia, marca la oposición irresuelta entre un fundamento y lo fundado, manteniendo a ambos polos en su unilateralidad (Hegel, *WdL*, GW12, p. 65).

A pesar de que la palabra pueda sugerir una relación entre lo fundado y el fundamento, el significado que la autora le da a la palabra “fundación” no es ese, sino que se aclara mediante una analogía con la relación entre teleología externa e interna que Hegel describe en su *Ciencia de la lógica*. En este sentido, así como en la *Ciencia de la lógica* la teleología subjetiva es la presuposición necesaria para comprender cabalmente la teleología externa, su importancia y su *verdad* (aunque sea un capítulo posterior), así, el concepto de “finalidad subjetiva” en Hegel debe ser la presuposición necesaria para la comprensión del concepto de “teleología” en Kant. Por eso la autora afirma: “Por ello, Hegel puede afirmar, con Kant y contra Kant, que la finalidad externa presupone la interna” (Koch, 2023, p. 296)

Esta original lectura intenta ir más allá de las lecturas tradicionales —por ejemplo, la de Horstmann (1991, p. 19)—, que ven en las posturas de ambos autores una mera oposición, como la que sugiere que mientras que para Kant la finalidad solo estaría vinculada al entendimiento y, en este sentido, sería meramente subjetiva, Hegel estaría en contra de esta interpretación “subjetivista” de la teleología y propondría su efectividad (cfr. p. 16).

1. La tesis sobre la relación entre el principio teleológico y el principio de la unidad sistemática de la razón en Kant

Debe recordarse que el objetivo principal que persiguen los cuidadosos análisis que realiza la autora al concentrarse en la versión kantiana del problema de la teleología es elucidar el lugar sistemático

que ocupan los juicios teleológicos para la producción del conocimiento. Su enfoque es teórico, no práctico.

Koch defiende la idea de que, a pesar de que el propio concepto de “finalidad” es *incomprensible* (de ahí la defensa del agnosticismo kantiano), ello no obsta para que podamos afirmar que el principio de causalidad final es trascendental ni para afirmar que constituye, por ello, una condición imprescindible para la producción de conocimiento. Qué tipo de conocimiento contribuye a producir dicho principio es lo importante y lo que justifica que la autora lo vincule estrechamente con el principio de la unidad sistemática de la razón. Ciertamente, no se trata de la determinación de objetos particulares, sino de su comprensión a partir del orden general de la naturaleza. Contrariamente a como lo percibe Ginsborg, para Koch (pp. 121-129) el principio de sistematicidad y el principio de finalidad no son el mismo principio. Son, más bien, dos principios distintos que se vinculan mediante su función opuesta pero complementaria: mientras que el principio de sistematicidad de la razón, que es exclusivamente regulativo, realiza funciones inferenciales que nos permiten unificar los conocimientos provistos por el entendimiento, pero carece de anclaje en la experiencia y por ello permanece indeterminado (lo cual nos permite, *per analogiam*, referirnos a la naturaleza como dicho orden), el principio de finalidad permitiría una inducción que haría posible el anclaje con el objeto de la experiencia (cfr. p. 60). En este contexto, Koch (p. 61) nos recuerda que, para Kant, el hecho de que podamos aprehender sistemáticamente el todo de la naturaleza no significa por ello que podamos conocer a la naturaleza como una totalidad, pero afirma que es gracias “[...] al uso de ambos principios que la naturaleza constituye para nosotros una unidad sistemática susceptible de ser conocida” (p. 63). Y agrega: “Un sistema, de acuerdo con Kant, se fundamenta en una idea. Esta prescribe el ‘para qué’ del orden sistemático. Y esta finalidad, en el contexto teórico, es la aspiración hacia la unidad de nuestros conocimientos [*Erkenntnisse*], y con ello del saber [*Wissen*]. Aspiramos a vincular nuestros conocimientos para lograr un saber articulado sobre la naturaleza” (p. 63). El sistema cumple una función: la de ordenar sus elementos constitutivos. Dicho orden es una relación que la autora denomina “inferencial” (p. 69) o relación de “arriba abajo” (*top-down*) (p. 69).

Ahora, un aspecto importante para la preparación del suelo en donde se instalará la discusión con Hegel es el de la diferencia que Kant establece entre finalidad externa e interna. Koch (p. 291) considera que si

bien Hegel recurre al mismo vocabulario que Kant, no está movilizando los mismos contenidos semánticos.

Para Kant, hablamos de finalidad externa cuando juzgamos una relación “causa-efecto” como una relación “fin-medio”, pero para Hegel esta relación solo es establecida mediante la intencionalidad y Hegel la entiende como finalidad subjetiva. Koch (p. 139, n. 55) precisa que la finalidad externa en Kant se refiere a procesos más que a objetos. En mi opinión, el uso de la palabra “proceso”, que aparece en contextos similares, quizás no sea la mejor elección terminológica, ya que un proceso implica modificaciones en todos los elementos de la relación procesual desde una perspectiva que también conlleva una temporalidad distinta. Considero que para entender esta relación entre fin y medio habría que hablar de dinamicidad (más que de procesualidad), pues en los sistemas dinámicos el tiempo es discreto y se establece claramente la diferencia sustancial de los elementos en juego. (cfr. Poli, 2015, p. 32). En este sentido, sería más preciso hablar de la relación dinámica (secuencial) entre fin y medio, ya que se trata de acciones instrumentales y en ellas hay una clara precedencia de la intención. El juicio de finalidad externa describiría entonces aquel conjunto de acciones secuenciales que se realizan en vistas a obtener una utilidad de algo *mediante* otra cosa. (cfr. Kant, KdU, AA V: 367). No obstante, estos juicios no tienen que ver con la constitución de los objetos (cfr. Koch, 2023, p. 139) ni con los agentes. Los juicios teleológicos externos son solo hipotéticos.

El uso más significativo del juicio teleológico—lo que Kant llama “juicio teleológico absoluto” (cfr. Koch, 2023, p. 141, n. 60)—es el interno, y consiste en enjuiciar un objeto como un fin (esto es lo que Koch supone que es la finalidad interior). Esta es una determinación constitutiva del objeto natural. La estructura de la idea de “fin”, que implica que el todo tiene prioridad sobre las partes, es la que nos permite juzgar a un objeto natural como un organismo —y no a la inversa, como sugieren algunas lecturas quizá inadvertidamente *naturalistas* (Koch, 2023, pp. 167-168)—. El concepto de “organismo” tiene aquí un carácter trascendental. Koch recuerda que Kant, en *Sobre el uso de los principios teleológicos en la filosofía* (ÜGTP, AA VIII: 179), lo define como la autoproducción de una totalidad cuyas partes se encuentran interactuando recíprocamente como fines y medios. Koch (2023, p. 170) concluye que es posible pensar objetos autoorganizados cuando pensamos de manera causal-final porque tenemos una imagen de la naturaleza como un todo organizado,

mostrando con ello que la posibilidad de concebir fines reside en la idea de la unidad sistemática.

2. ¿Cuál es la relación entre lo que Hegel llama, en la *Ciencia de la lógica*, finalidad externa y finalidad interna?

La pregunta que guía el apartado dedicado a las consideraciones hegelianas sobre la teleología en la *Ciencia de la lógica* es: ¿cuál es el significado de la enigmática expresión “la teleología es la verdad del mecanismo” (Hegel, WdL, GW 12: 154)? La respuesta que la autora ofrecerá a esta pregunta también ofrecerá indicios para comprender el significado de la tesis que sostiene Koch sobre la relación “fundante” entre la tesis hegeliana y la tesis kantiana sobre la teleología. Como Koch (2023, p. 180) reconoce, esta pregunta hegeliana y los conceptos que implica —a saber, la diferencia entre mecanismo y finalidad— remiten a la diferencia entre finalidad externa y finalidad interna que Hegel retoma, explícitamente (aunque no al pie de la letra) del propio Kant. Me concentraré aquí en las diferencias que la autora resalta.

Una diferencia importante que la autora reconoce entre los dos paradigmas reside en que mientras que Kant se ocupa del juicio teleológico y el mecanismo solo en relación con lo que Kant llama “Naturaleza”, Hegel, por su parte, extenderá la presencia de estas categorías a la realidad efectiva (*Wirklichkeit*) (cfr. Koch, 2023, p. 180, n. 13), lo cual, sin que la autora lo indique explícitamente, nos sitúa en una esfera de discusión distinta a la mera discusión sobre la función epistémica de los juicios teleológicos. Con la *Ciencia de la lógica* nos encontramos en una discusión que, si bien implica el carácter “pensado” de la realidad efectiva, nos remite al problema ontológico de la constitución (pensante) de los objetos.

En el capítulo sobre el mecanismo, recuerda Koch (p. 192), Hegel considera que el objeto no se constituye por las relaciones mecánicas en las que está implicado, sino que es un simple portador de ellas. Es la determinación ontológica del objeto mecánico la que define nuestro conocimiento de él, lo cual también se opone a la configuración del proceso de conocimiento en Kant, aunque la postura de Hegel no implica ningún realismo naturalista. Koch advierte que el conocimiento del objeto mecánico no es “incorrecto”, pero sí insuficiente. La insuficiencia de este conocimiento no reside en la adecuación incompleta de un juicio a un objeto, sino en la naturaleza misma del objeto. Un objeto mecánico

carece, podríamos decir, de consistencia ontológica, pues no tiene en sí mismo las relaciones causal-efectuales que lo hacen ser objeto mecánico. Ello debe servir para señalar, al menos brevemente, que hay para Hegel un conocimiento suficiente de la realidad, incluso la mecánica, que no por ello “falsea” al conocimiento anterior (cfr. Koch, 2023, p. 193).

De esta caracterización del objeto mecánico como él mismo “insuficiente” se sigue la respuesta que ofrece Koch a la pregunta por la expresión hegeliana —“la teleología es la verdad del mecanismo”—. Como sostiene la autora, la teleología ofrece estructuras más favorables a la individuación que el mecanismo. Con ellas, los propios objetos mecánicos pueden alcanzar también la consistencia o individuación de la que carecían al ser solo portadores de vínculos contingentes entre causa y efecto.

El paradigma por excelencia de las relaciones teleológicas es el de la intencionalidad subjetiva. Los objetos mecánicos solo son tales cuando se individualizan a partir de la acción teleológica subjetiva (cfr. Koch, 2023, p. 203). Es a partir de aquí que la autora defiende la idea, propia de la *lógica retrospectiva* de la *Ciencia de la lógica*, de que es el movimiento posterior el que fundamenta, en el sentido de que cohesiona y da consistencia, al anterior.

La explicación concreta de esa fundamentación *à rebours* la encontramos en el organismo. Un organismo ve finalidad en los objetos en función de sus necesidades. No existe un mundo de medios y fines *per se*. Al mundo orgánico no le precede un mundo de relaciones mecánicas al que él deba adaptarse. El organismo, de acuerdo con la tesis de Koch (p. 233), es el que, en función de sus procesos adaptativos y asimilatorios, se relaciona con su medio de manera tal que produce relaciones de instrumentalización. Con su actividad intencional crea, o *funda*, al mundo mecánico de relaciones entre medios y fines. Pero al mismo tiempo es este mundo de la acción intencional el que le da a la subjetividad orgánica su propio afuera, un afuera que la individualiza (cfr. p. 234). El papel preponderante de lo vivo en la tesis central de Koch comienza a delinearse en este momento. Y ya a partir de aquí puede comprenderse que la Vida, como momento posterior a la teleología en la *Ciencia de la lógica*, constituya el pivote que hará bascular la respuesta a la pregunta por el sentido de la expresión “la teleología es la verdad del mecanismo” hacia una dimensión netamente biologizante y hacia un tipo particular de epistemología.

Aquí hubiera sido pertinente que la autora aludiera a uno de los aspectos característicos de la filosofía postkantiana de la primera hora: la voluntad de estrechar los lazos entre la razón teórica y la razón práctica. Hegel, de acuerdo con la lectura de Koch, estaría colocando la idea en el corazón mismo de la realidad efectiva en tanto que es la idea la que vive y conoce a través de relaciones orgánicas. Aquí me parece que se trata de relaciones, al menos, proto-prácticas, con el medio. Solo como idea de conocimiento *en el mundo de las relaciones con el medio* tenemos conocimiento de los objetos, pues solo cuando los organismos están adentrados en sus procesos autoproductivos (en sus relaciones teleológicas con el medio) producen la realidad. El conocimiento es verdadero aquí porque lo que se conoce es proceso (cfr. p. 263).

Ahora podemos acercarnos a la respuesta a la tercera pregunta que considero que este libro responde.

3. Si se admite que Hegel no es un filósofo trascendental, ¿cómo debe leerse la filosofía hegeliana de la teleología en relación con la propuesta kantiana

La respuesta que ofrece Koch ya la hemos anunciado. Se trata, de acuerdo con ella (p. 268), de una fundación. El carácter fundacional de la reflexión hegeliana residirá en el papel preponderante que Koch le atribuye al concepto de “Vida”, mismo que le hace sugerir algo que denomino una especie de ecología de las relaciones teleológicas motivadas por la agencia inherente a los organismos en sus medios.

Koch sostiene que la importancia del conocimiento “performativo” que el organismo tiene de sí en su proceso de adaptación y asimilación al medio a través de sus relaciones instrumentales reside en que contribuye a “situar” dicho conocimiento. Para Koch, Hegel ofrece una fundamentación al proponer que debemos asumir un vínculo entre el lugar en el todo y el conjunto de estructuras finalísticas internas que constituyen la vida. Esta suposición es necesaria para comprender los objetos mecánicos (cfr. p. 274). Frente a la preminencia de la legalidad causal-mecánica de la naturaleza tal y como se describe en las analogías de la experiencia, misma que, como se mostró en el análisis que la autora realizó de la función epistémica de los juicios teleológicos, funda el uso regulativo del principio teleológico (ya que se vincula con la idea de unidad sistemática de la razón), la autora propone “encarnar” la

finalidad y hacer de esta finalidad subjetiva “vivida” y “conocida” el punto de partida de nuestra empresa de conocimiento.

Bibliografía

- Hegel, G. W. F. (1999). [WdL]. *Wissenschaft der Logik. Erster Band. Die objektive Logik. Zweites Buch. Die Lehre vom Wesen* (1813). GM 12. Felix Meiner Verlag.
- Horstmann, R.-P. (1991). *Die Grenzen der Vernunft. Eine Untersuchung zu Zielen und Motiven des Deutschen Idealismus*.
- Kant, I. (1913a). [KdU]. *Kritik der Urteilskraft*. En *Gesammelte Schriften. Band V* (pp. 167-515). De Gruyter.
- Kant, I. (1913b). [ÜGTP]. *Über den Gebrauch teleologischer Prinzipien in der Philosophie*. En *Gesammelte Schriften. Band VIII* (157-183). De Gruyter.
- Poli, R. (2015). *The Difference between Dynamical Systems and Process Theories*. En V. Petrov y A. C. Scarfe (eds.), *Dynamic Being: Essays in Process-Relational Ontology* (pp. 28-43). Cambridge Scholars Publishings.

Zaida Olvera
Universidad Veracruzana
zolvera@uv.mx

Aouad, M., Di Vincenzo, S. y Fadlallah, H. (2023). *Averroes' Middle Commentary on Aristotle's Metaphysics: Critical Edition of the Arabic Version, French Translation and English Introduction*. Brill. 632 pp.

Este volumen forma parte del proyecto auspiciado por la Unión Académica Internacional para la edición y traducción de las obras de Averroes, de manera que la contraportada dice: *Averrois Opera. Series A: Averrois Arabicus. XXXVI. Commentarium Medium in Aristotelis Metaphysicam*. Averroes (m. 595/1198) interesó a los cristianos de la Edad Media pero también a las comunidades judías de la época en su calidad de comentador de Aristóteles. Los latinos se interesaron por el llamado comentario mayor (*tafsīr*) de una obra de Aristóteles, siempre que lo hubiera, y si no, por el comentario medio (*talḥīṣ*), mientras que los filósofos judíos optaron por los comentarios medios. Los llamados comentarios menores tuvieron una suerte dispar.

Maroun Aouad ya se había ocupado de Averroes/Ibn Rushd, cuando editó, tradujo y estudió el *Commentaire moyen de la Rhétorique* (de Aristóteles), publicado por Vrin en 2003, e indirectamente con el *Livre de la Rhétorique du philosophe et médecin Ibn Ṭumlūs (Alphagiag bin Thalmus)*, m. 620/1223, también publicado por Vrin, en 2007. En el curso del proyecto europeo PhiC y de su continuador, el francés PhASIF, Aouad descubrió un manuscrito de Averroes, mal catalogado, organizado en tres carpetas y repartido en dos cajas, en la biblioteca Qarawīyīn de Fez, que identificó con el comentario medio (*talḥīṣ*) a la *Metafísica*, *MCMet*. Hasta aquel momento, el original árabe del comentario medio se consideraba perdido, aunque se conservaban dos traducciones hebreas, una primera de Zerahya ben Shealtiel (sobre 1284) y una segunda de Qalonymos ben Qalonymos (de 1317), que desplazó a la primera por razones de inteligibilidad. Si Julio Manardo atribuyó correctamente a Elia del Medigo la autoría de los textos que las ediciones de Venecia presentan como comentario medio —sin terminar, pero que puede ser también del comentario mayor (Aristóteles y Averroes, 1562)—, obtendríamos una versión latina hecha a partir del hebreo de Qalonymos.¹

Maroun Aouad comprobó que en el *Budd al- 'Arif*, de Muḥammad ibn 'Abd al-Ḥaqq Ibn Sab'īn (m. 1269 o 1271) plagiaba pasajes completos del comentario medio, de manera que, con ayuda de estos textos plagiados,

¹ Di Vincenzo (*MCMet*, pp. 53-54) indica que solamente hay dos fragmentos que con seguridad son de Elia.

de la comparación con la traducción hebrea, de una prueba “fuerte” del mismo Averroes² y de alguna anotación marginal, demuestra la autenticidad del texto averroico. La tarea de edición era difícil porque Maroun Aouad y Hamidé Falallah tenían delante 1260 fragmentos de papel en mal estado que pertenecen a ocho de las catorce secciones comentando los libros Delta (4, parte), Epsilon (5, completo), Zêta (6, completo), Êta (7, completo), Thêta (8, parte), Mü (12, parte) y Nü (13, parte).³ El capítulo primero nos ofrece una descripción paleográfica detallada de estos fragmentos que contienen el único testimonio del *MCMet* en árabe, una descripción que muestra la enorme capacidad de los editores para descifrarlos.

El equipo contaba con la colaboración de Silvia Di Vincenzo, que en el capítulo segundo reconstruye “la tradición textual del *MCMet* en árabe, hebreo y latín” (*MCMet*, pp. 38-114). Desde el principio, Di Vincenzo habla de dos ramas; llama *alpha* a la tradición de las dos traducciones hebreas y *bêta* a la tradición del único manuscrito árabe, que abrevia F, por Fez: los plagios de Ibn Sab‘în siguen la segunda. Compara las dos tradiciones (*MCMet*, pp. 85-112) y demuestra cómo Averroes revisó su propio texto, de manera que la versión *bêta* es más completa y refleja mayor proximidad al comentario mayor. No entro en más detalles porque este capítulo merece una reseña aparte en un foro especializado, y un recuerdo a los estudios de Mauro Zonta.

En el capítulo tercero, los tres autores reconocen el valor de las traducciones de Zerahya y Qalonymos, que utilizaron como testimonios en el aparato crítico, obviamente, solo cuando la revisión de Averroes en la versión *bêta* no les afectaba. Además, pudieron consultar algunas citas del comentario medio que contiene la obra de Ibn Sab‘în (1978), y los pasajes correspondientes del comentario mayor, editado por Maurice Bouyges, excepto los libros XIII y XIV porque faltan en este (*MCMet*, pp. 115-116).⁴ Este aparato crítico se complementa con a) referencias al comentario mayor y b) notas aclaratorias.

² Se trata de una referencia de Averroes a su comentario medio, que hace en el comentario mayor (p. 9).

³ Los números árabes entre paréntesis corresponden a la numeración del comentario medio, en el que el libro α , *alpha elatton*, es desconocido.

⁴ Kappa falta tanto en la edición del gran comentario o mayor, de Bouyges, como en la del medio, de Aouad.

El resultado es una edición impecable de estas partes del comentario medio conservadas en árabe. La acompaña la traducción, anotada, a cargo de Maroun Aouad, y unos índices de nombres, obras, comunidades, lugares, y términos técnicos citados en el cuerpo árabe y francés, así como de nombres citados en los estudios, extremadamente detallados.

La pregunta que a menudo se formula en trabajos de esta clase, que exigen un esfuerzo tan extraordinario de dedicación y tiempo, está relacionada con el valor que puede tener el resultado en la investigación del pensamiento, en este caso, de Averroes. Quisiera contestarla comparando dos fragmentos, uno del comentario mayor y otro del medio, partiendo del supuesto de que el mayor es posterior al medio. Estos dos fragmentos explican unas palabras de Aristóteles sobre el concepto de “ente”, uno de los conceptos fundamentales en su metafísica, y se encuentran al comienzo del libro Zêta (VII), donde Aristóteles habla de los sentidos del término “ente” y que Averroes leía en su versión árabe que en sentido literal dice:⁵

“Existencia” se dice de muchas maneras que antes hemos especificado en la discusión donde hemos mencionado de cuántas maneras se dice la cosa. Entre ellas, está aquello que indica qué es la cosa y que ello es esta cosa, y entre ellas, está aquello que indica que esta cosa es una cualidad, una cantidad u otra de las cosas que en la discusión se predicán de manera parecida (Averroes, 1938-1952, 745, 12-15).

La distinción principal es entre esencia (τί ἐστι) e individuo (τόδε τι), como ya señalaba Ross (1924, II, 2, pp. 159-160), y la distinción entre sustancia y accidentes es secundaria. Mi traducción literal quiere coincidir con la que Michael Scot hizo al latín, y que fue presentada como *translatio nova* en las ediciones de Venecia:

Ens dicitur multis modis, ut prius diximus in sermone, in quo diximus quot modis dicitur aliquid. quoddam enim significat quid est aliquid, et quod est hoc, et quoddam significat quod est quale, aut quantum, aut

⁵ Existen dos traducciones francesas de esta parte: la de Laurence Bauloye (1997a), con una extensa anotación, y la de Nabil Elsakhawi (2000).

aliud eorum, quae praedicantur in sermone secundum hunc modum (Aristóteles y Averroes, 1562, 153r^o).

Averroes interpreta este pasaje tanto en el comentario mayor como en el medio. En el mayor, Averroes lo hace dando por supuesto que el término “existencia” (*huwīya*) y “existente” (*mawjūd*) son sinónimos del *to on* griego: uno, *huwīya*, es un abstracto derivado de *hūwa*, “ello es”, y el otro, *mawjūd*, es un participio/adjetivo sustantivado, “lo encontrado o que se puede encontrar”,⁶ y probablemente la doble traducción refleja la distinción apuntada por Ross. A continuación, traduzco el pasaje en el cual Averroes reproduce palabras de Aristóteles y las explica:

Dice [Aristóteles]: “Puesto que se ha demostrado que el término ‘existencia’ (*huwīya*) y ‘existente’ (*mawjūd*) se predicán de muchas maneras según hemos mencionado en el libro [Delta: Dāl] donde hemos explicado de cuántas maneras se dicen los términos utilizados en esta ciencia”. Lo hemos especificado antes y hemos dicho que el nombre “existente” comprende aquello que indica la quiddidad (*māhīya*) de la sustancia determinada y la misma sustancia determinada, que comprende [también] aquello que indica un accidente en este individuo determinado que subsiste por sí mismo y que comprende aquello del nombre de existente que indica un accidente en estas sustancias, bien sea una cualidad, una cantidad u otra de las cosas que se predicán de las sustancias sin que la predicación exprese sus esencias (*dhawāt*) ni dé respuesta a qué es la sustancia determinada. Esto es lo que [Aristóteles] expresaba con sus palabras “y entre ellas, está aquello que indica que esta cosa es una cualidad, una cantidad u otra de las cosas que en la discusión se predicán de manera parecida”. [Aristóteles] quería decir que el nombre “existente” (en las sustancias) comprende aquello que indica una cualidad, una cantidad u otro de los accidentes que no se predicán de ello porque den a conocer su sustancia, sino que se predicán de ello solo

⁶ Ver los comentarios de Bauloye (1997b, pp. 51-53) sobre la distinción.

en el aspecto de cómo es ello, cuánto es ello, dónde está ello, o cuándo está ello, o por las clases de predicación de las demás categorías (Averroes, 1938-1952, 746).

Averroes comenta, pues, con detenimiento el pasaje *Met.* 1028a10-13 bastantes años después de haber redactado el medio. La edición de M. Aouad ha reconstruido el texto árabe del segundo, y este pasaje es el comienzo del libro 6, puesto que, como hemos dicho, el comentario medio ignoraba la existencia del *Alpha meizon*, empezando con la frase “El estudio de la verdad es difícil bajo un aspecto y fácil bajo otro” (993a30), circunstancia que tuvo eco en los comentadores hebreos.⁷ M. Aouad traduce:

6.3. [Aristote] a dit: *L'existant et l'être sont dits de nombreuses espèces, comme nous l'avons expliqué en détail dans le propos antérieur.* En effet, certaines indiquent ce qu'est l'objet de désignation qui n'est pas dans un substrat, certaines indiquent une qualité dans l'objet de désignation ou une quantité ou une autre des choses qui ne font pas connaître, de l'objet de désignation qui n'est pas dans un substrat, ce qu'il est – il s'agit du reste des catégories. Puisqu'il en est ainsi, il est évident que l'existant premier et l'être premier parmi ces [espèces], je veux dire celle qui est antérieure, est celui qui indique, de l'objet de désignation qui n'est pas dans un substrat, ce qu'il est car cette [espèce] est celle qui indique sa substance alors que les autres [espèces] indiquent des choses extérieures à son essence (*MCMet*, pp. 208-211).

Averroes se interesó entonces por el papel del sujeto o sustrato que Aristóteles menciona a continuación contraponiendo acciones a sujetos agentes, como caminar al hombre que camina, y se pregunta cuál es más existente. Responde que los actores poseen más ser porque lo que está debajo de estos es algo determinado (1028a25-26). Tanto la traducción árabe (Averroes, 1938-1952, 749, 15-750, 1) como la de Michael Scot —“quoniam subjectum eorum est aliquod terminatum” (Aristóteles y Averroes, 1562, 154v^o)— coinciden en dar este sentido. El comentario

⁷ Ver el artículo de Halper (2020) sobre esta frase, y Halper (2020, pp. 139-140) para su traducción al hebreo.

medio, en cambio, no contempla esta contraposición, sino que enlaza con el libro anterior —o sea, Epsilon— para razonar de otra manera.

Averroes se centra en dos clases de seres: la de los seres determinados carentes de sustrato —es decir, las esencias— y la de los predicados de los seres, es decir, las diez categorías tradicionales, para concluir que la primera clase tiene primacía sobre la segunda. Acto seguido introduce la clase del existente en un sustrato como “realización” de la primera. Encuentro la transición falta de una explicación, que no es fácil, puesto que aquí se plantea el problema de las formas.⁸ El comentario mayor es más prudente, en el sentido de que se centra en la oposición de la sustancia tanto *huwīya* como *mawjūd* frente a las categorías predicamentales.

Esto es solo una pequeña muestra, pero esperamos que sirva para defender la necesidad de trabajos como el de Maroun Aouad y sus dos colaboradoras para comprender adecuadamente a pensadores como Averroes que, a diferencia de Avicena, nunca consideró haberlo entendido todo bien y para siempre, sino que necesitaba repensar los problemas de la filosofía y lo reflejaba en sus escritos.

Referencias

- Aouad, M., Di Vincenzo, S. y Fadlallah, H. (2023). [MCMet]. *Averroes' Middle Commentary on Aristotle's Metaphysics: Critical Edition of the Arabic Version, French Translation and English Introduction*. Brill. <https://doi.org/10.1163/9789004515765>
- Aristóteles y Averroes. (1562). *Aristotelis Metaphysicorum libri XIII cum Averrois Cordubensis in eosdem commentariis et epitome. Volumen Octavum*. Apud Iunctas.
- Averroes. (1938-1952). *Tafsīr mā ba'd al-ṭabī'at*. M. Bouyges (ed.). Cuatro volúmenes. Imprimerie Catholique.
- Bauloye, L. (1997a). Averroès: Grand commentaire de la *Métaphysique* d'Aristote, Z1 et Z2. Traduction et notes. *Bulletin d'Études Orientales*, 49, 52-74.
- Bauloye, L. (1997b). *La question de l'essence. Averroès et Thomas d'Aquin, commentateurs d'Aristote : Métaphysique Z1*. Peeters.

⁸ Recordemos que es una cuestión relacionada con la de la definición, reenvío al artículo de M. di Giovanni (2003).

- Di Giovanni, M. (2003). La definizione delle sostanze sensibili nel *Commento grande* di Averroè a *Metafisica* Z 10. *Documenti e Studi Sulla Tradizione Filosofica Medievale*, 14, 27-63.
- Elsakhawi, N. (2000). *Etude du livre Zay (Dzêta) de la Métaphysique d'Aristote dans sa version arabe et son commentaire par Averroès*. Presses Universitaires du Septentrion.
- Halper, Y. (2020). In One Sense Easy, in Another Difficult: Reverberations of the Opening of Aristotle's *Metaphysics* α ἑλαττον in Medieval and Renaissance Hebrew Literature. *Revue des Études Juives*, 179, 133-160.
- Ibn Sab'īn, M. i. 'A. a.-Ḥ. (1978). *Budd al-'Arīf*. G. Kattoura (ed.). Dār al-Andalus.
- Ross, W. D. (1924). *Aristotle's Metaphysics: A Revised Text with Introduction and Commentary*. Dos volúmenes. Clarendon Press.

Josep Puig
 Universidad Complutense de Madrid
 puigmont@ucm.es

Gerlach, S. (2023). *Schellings Philosophie der Potenz 1798-1854*. Mohr Siebeck. 300 pp.

La investigación sobre Schelling ha mostrado que el concepto de “potencia” (*Potenz*) constituye una noción recurrente que aparece a lo largo de su obra (cfr. Hemmerle, 1966; Schrödter, 1986; Grün, 1993). No obstante, un aspecto clave de dicha noción es su carácter dinámico, el cual se despliega dentro del marco de la naturaleza (*Natur*) y la mente (*Geist*). En este contexto, la potencia constituye una fuerza inherente que se estructura a través de un movimiento gradual, manifestándose en diversas formas que corresponden a distintos grados de desarrollo. Schelling describió lo anterior como la “potenciación constante del yo” (AA I, 9, 1, p. 146).¹ Es precisamente este aspecto el que ha llamado la atención del trabajo de Stefan Gerlach, que se enfoca en una interpretación sistemática del concepto de potencia en Schelling.

El libro forma parte de la colección *Collegium Metaphysicum* de la reconocida editorial alemana Mohr Siebeck. La primera parte se centra en las “Potencias de la naturaleza y del espíritu” (1789-1800). La segunda parte examina las “Potencias del absoluto – La filosofía de la identidad” (1801-1806). La tercera parte se dedica al “Desarrollo de una auténtica doctrina de los poderes” (1809-1821), mientras que la cuarta aborda las “Potencias del ser en la filosofía tardía” (1827-1854). De este modo, el libro ofrece un estudio ambicioso que abarca los diferentes momentos del pensamiento de Schelling. El autor del libro señala cómo estos diversos usos no constituyen una noción ambigua y vacía, sino que representan una riqueza semántica que adquiere sentido dentro del propio *corpus* de Schelling y, en especial, en su diálogo con otros sistemas filosóficos. Además, Gerlach ofrece una visión exhaustiva y detallada de uno de los conceptos más recurrentes y complejos en el pensamiento de Schelling. Su trabajo analiza la semántica, el uso, los campos de aplicación y la posición sistemática del campo conceptual de “potencia, potencias y potenciación”.

La reconstrucción del concepto de “potencia” por parte de Gerlach convierte dicha noción en el hilo conductor de una reflexión que atraviesa la “filosofía de la naturaleza” (*Naturphilosophie*), la “filosofía

¹ La cita se refiere a la edición de las obras completas de Schelling, conocida como la *Akademie-Ausgabe* (AA), *Schelling: Historisch-kritische Ausgabe*, editada por Frommann-Holzboog.

de la Identidad” y el “pensamiento tardío”. No se trata de un término estable, en términos sistemáticos, sino que las “potencias” (*Potenzen*), en plural, cambian de registro e intensidad según el contexto —vida, lógica, historia, libertad, teología—. Esta movilidad no es un defecto, sino la fuente misma de su potencia explicativa (cfr. Gerlach, 2023, pp. 2-3). Gerlach lo expresa de manera clara: antes de comprender a Schelling, es necesario ordenar y diferenciar los usos del término, ya que su multiplicidad semántica es tanto el problema como la riqueza de la doctrina. En este aspecto puede notarse una notable virtud del texto: intentar sistematizar el concepto mediante sus diversos usos. En lo que sigue nos detendremos en los aspectos más relevantes del texto de Gerlach.

En el capítulo primero —“Potencias de la naturaleza y del espíritu” (1798-1800)— se registra el origen profundo del aparato conceptual de las potencias. El autor registra que Schelling recoge la “teoría de la excitación” de John Brown, pero no la replica en un sentido crítico, sino todo lo contrario, es decir, la transforma. Junto al de este autor también se encuentran nombres como Carl August Eschenmayer, así como la influencia de la matemática al respecto (cfr. pp. 33-37). En este sentido, es posible identificar, por así llamarlo, dos respuestas insuficientes: una que reduce la vida a lo puramente químico, y otra que la explica exclusivamente a través de estímulos externos. Schelling postula que el principio vital reside en la tensión entre dos principios opuestos: uno positivo, supraindividual, y otro negativo, interior al individuo (cfr. AA I, p. 6). Esta dialéctica inicial actúa ya como un esqueleto teórico: la vida se concibe como el equilibrio dinámico de contrarios, lo que abre la posibilidad de pensar la totalidad como un juego de potencias (cfr. p. 17). Es en este germen biológico donde la idea de “potencia” adquiere su primera estructura operativa: no es simplemente energía, sino una forma de relación y diferenciación. Esta reflexión ya deja ver el trabajo de Gerlach, pues abre el espectro de la riqueza conceptual de la noción de “potencia”, a la vez que muestra sus usos en plural (cfr. pp. 51-53).

Más adelante, el autor ofrece una relevante interpretación del concepto en el capítulo cuarto —“*Die Darstellung meines Systems* (1801)”—. En dicho capítulo, se señala que Schelling logra integrar los niveles lógico, ontológico y cosmológico, preparando el terreno para su doctrina de las potencias. En este capítulo, Gerlach plantea una tesis aclaradora: toma la fórmula lógica $A = A$ y la interpreta como experiencia ontológica

(cfr. pp. 86-88). El primer A a la izquierda del signo equivale al sujeto, mientras que el segundo A a la derecha representa al objeto; la igualdad deja de ser una tautología y se convierte en el modelo del conocimiento que se reconoce a sí mismo. Sin embargo, la originalidad de Schelling, según Gerlach, radica en la torsión que imprime a la fórmula: “solo se da en la forma de reconocer su identidad consigo misma” (p. 85). Desde esta perspectiva, las potencias se configuran como las modalidades en que el “Uno” se despliega sin perder su unidad. Dicho de otro modo, la potencia pasa de ser un instrumento explicativo de la naturaleza a una forma de manifestación del “Absoluto”; cada potencia es, por así decirlo, una manera en que la unidad se hace pluralidad. Por ello, se debe resaltar que el capítulo cuarto constituye el punto de inflexión donde lo lógico se convierte en ontológico y cosmológico.

Es pertinente al respecto señalar el interesante desglose del concepto en el *Tratado de la libertad*, en el capítulo sexto del libro. En 1809, con el *Freiheitsschrift*, la doctrina de las potencias adquiere una dimensión ética y existencial. La potencia deja de ser solo una estructura del mundo y se convierte en condición de la libertad humana. La posibilidad del mal se revela como un correlato necesario de la libertad: si la voluntad es potencia, entonces puede orientarse tanto hacia la afirmación creativa como hacia la negación destructiva. Gerlach muestra que, en este contexto, la “doctrina de las potencias” (*Potenzenlehre*) (cfr. pp. 158-160) adquiere un tono trágico; la filosofía se transforma también en una reflexión sobre el precio de la libertad. No se trata de una teorización abstracta, según el autor, sino de una anatomía del conflicto moral en la que la potencia explica por qué la responsabilidad humana es irreductible.

El capítulo octavo, “Las elaboraciones de la madurez” —*Weltalter*, 1811-1815 y los desarrollos posteriores— constituye el intento de convertir el desarrollo de las potencias en un sistema histórico y mítico. En los bosquejos de “Las edades del mundo” (*Weltalter*), las potencias se presentan como épocas, mitos y fuerzas que se encarnan en procesos culturales: la historia deja de ser una mera cronología y se interpreta como un despliegue de potencias. Es en este contexto donde Schelling, según Gerlach, pone a prueba la eficacia interpretativa de su aparato conceptual: las potencias permiten leer la cultura, el mito y la religión como fases de un movimiento unitario (cfr. pp. 218-220).

Más adelante, Gerlach se dedica a reflexionar sobre el concepto en textos posteriores. En la fase tardía (1827-1854) de Schelling, Gerlach

manifiesta de que es posible hallar la ambición final de la “doctrina de las potencias” (*Potenzenlehre*), que reclama el rango de ontología primera. Las potencias se convierten en las “potencias originarias” (*Urmächte*) (p. 226) del ser: no son simplemente categorías, sino principios fundamentales que ordenan la realidad, desde lo natural hasta lo divino. El autor acentúa que Schelling intenta formalizar dicha doctrina mediante notaciones, series, operadores y, al mismo tiempo, la deja abierta a la experiencia histórica y religiosa. La filosofía, por ende, se entiende como una interpretación de la revelación y como participación en la automanifestación del “Absoluto”. En otras palabras, la “doctrina de las potencias” es la propuesta de pensar el ser como un proceso, y el libro de Gerlach muestra cómo esa propuesta se impone a lo largo de más de medio siglo de textos y variantes terminológicas.

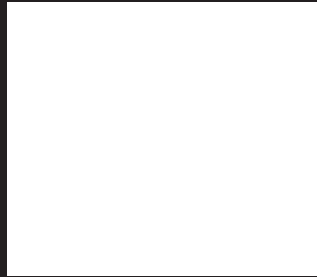
Tales intentos sistemáticos, brevemente esbozados, constituyen una prueba clara de la meta ambiciosa del libro y de la manera en que este logra cumplir con tal propósito. Y en ello radica la riqueza de la obra. Gerlach no se limita a acumular citas o referencias cercanas: desenmaraña la historia conceptual, diferencia los sentidos y muestra tanto las discontinuidades como las continuidades. El valor sistemático de su reconstrucción radica en que convierte un laberinto terminológico en un mapa útil: permite comprender cómo una palabra —*Potenz*— llega a sustentar toda una reflexión sobre la metafísica. Para el lector interesado en ahondar en el pensamiento de Schelling, la obra ofrece algo más que referencias bibliográficas cruzadas. Sin duda, se puede afirmar que el trabajo de Gerlach propone que dicha noción puede aclararse bajo una mirada en la que se combine ya sean la unidad y multiplicidad, ya sean la lógica, cosmología, libertad y teología.

A través de la noción de “potencia”, Schelling nos enseña a pensar la realidad como un movimiento “autopotenciador”. El trabajo de Gerlach pone de manifiesto una lectura atenta y, sobre todo, fructífera de la obra de Schelling. Además, el libro también tiene la virtud de contar con un estilo claro y accesible para el lector. Quien se acerque a este estudio se familiarizará con la investigación sobre Schelling y disfrutará de una explicación general, pero a la vez minuciosa, de su sistema filosófico. Reflexionar sobre la potencia en el contexto de nociones que siguen siendo relevantes en la discusión filosófica contemporánea —como naturaleza, espíritu y divinidad— constituye otro de los impulsos clave de esta obra. Por todo ello, este texto debe ser recomendado a cualquier estudioso de Schelling.

Bibliografía

- Grün, K. J. (1993). Ursprung und methodische Bedeutung der Potenzenlehre Schellings. *Archiv für Begriffsgeschichte*, 36, 174-195.
- Hemmerle, K. (1966). Zum Verständnis der Potenzenlehre in Schellings Spätphilosophie. *Philosophisches Jahrbuch*, 74(1), 99-125.
- Schelling, F. W. J. (2005). *Historisch-kritische Ausgabe. Reihe I: Werke. Band 1,9,1: System des transscendentalen Idealismus (1800)*. Frommann-Holzboog.
- Schrödter, H. (1986). Die Grundlagen der Lehre Schellings von den Potenzen in seiner ‚Reinrationalen Philosophie‘. *Zeitschrift für philosophische Forschung*, 40(4), 562-586.

Osman Choque-Aliaga
Universidad San Francisco Xavier de Chuquisaca
Albert-Ludwigs-Universität Freiburg
choque.osman@usfx.bo



TÓPICOS

REVISTA DE FILOSOFÍA

