

Juan Villacrés Bolaños  
EDITOR



# Introducción a la **FILOSOFÍA ANALÍTICA**

**VOLUMEN 2**

COLECCIÓN FILOSÓFICA ACTUAL

Editorial Universitaria  
Universidad Central del Ecuador

Filosófica Editorial

Facultad de Ciencias Sociales y Humanas  
Universidad Central del Ecuador





**INTRODUCCIÓN A LA FILOSOFÍA ANALÍTICA**  
Volumen II



La filosofía ocupa un lugar insustituible a la hora de cuestionar los prejuicios y certezas impuestas por el sentido común, las creencias culturales y las costumbres. En lugar de ofrecer respuestas definitivas, expande horizontes y cultiva la duda crítica. Nos libera del dogmatismo y mantiene viva nuestra capacidad de asombro, al estar vinculada al mundo, a la vida, provocando sorpresa y enriqueciendo nuestras percepciones. Además, revela aspectos inadvertidos de lo cotidiano, profundizando nuestra comprensión más allá de lo evidente o asumido como incuestionable.

En un contexto generalizado de fragmentación, desinformación y agotamiento social, el pensamiento filosófico se convierte en un acto de resistencia creativa, permitiendo concebir posibilidades más allá de los límites definidos por las estructuras dominantes de la época.

La filosofía, lejos de ser un artefacto puramente intelectual, es una práctica indispensable y transformadora. La COLECCIÓN FILOSÓFICA ACTUAL surge como una apuesta crucial para impulsar la reflexión crítica desde Ecuador hacia América Latina y otras regiones del mundo. Esta colección busca consolidar un espacio de diálogo interdisciplinario, convirtiendo la filosofía en un recurso activo ante las crisis contemporáneas. Así, la filosofía asume su tarea esencial: liberarnos de certezas que limitan nuestra comprensión, permitiéndonos no solo habitar el presente con mayor lucidez, sino también imaginar y construir futuros posibles de forma reflexiva.

COLECCIÓN FILOSÓFICA ACTUAL  
Facultad de Ciencias Sociales y Humanas  
Universidad Central del Ecuador

Prof. Santiago M. Zarria

Prof. Martín Aulestia Calero



Juan Villacrés Bolaños  
(Editor)

**INTRODUCCIÓN A LA FILOSOFÍA ANALÍTICA**  
**Volúmen II**

María Pía Méndez Mateluna  
Joaquim Giannotti  
Hernán Accorinti  
Olimpia Lombardi  
Angélica María Pena  
Elberto A. Plazas  
Michael Schon  
Miguel Ángel Sebastián  
Ányerson Stiths Gómez-Tabares  
Santiago Arango-Muñoz

COLECCIÓN FILOSÓFICA ACTUAL

FILOSÓFICA EDITORIAL

UNIVERSIDAD CENTRAL DEL ECUADOR

FACULTAD DE CIENCIAS SOCIALES Y HUMANAS

EDITORIAL UNIVERSITARIA

UNIVERSIDAD CENTRAL DEL ECUADOR

COLECCIÓN FILOSÓFICA ACTUAL

La Colección Filosófica Actual constituye un proyecto editorial y de investigación desarrollado entre Filosófica, Fundación de Estudios Filosóficos, Políticos y Culturales, y la Facultad de Ciencias Sociales y Humanas de la Universidad Central del Ecuador.

INTRODUCCIÓN A LA FILOSOFÍA ANALÍTICA  
Volumen II

Juan Villacrés Bolaños  
Editor

1.ª edición | octubre de 2026  
ISBN | 978-9907-840-05-6

EDITORIAL UNIVERSITARIA  
Universidad Central del Ecuador  
Ciudadela Universitaria, av. América, s. n.  
Quito, Ecuador  
editorial@uce.edu.ec

FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
Universidad Central del Ecuador  
COLECCIÓN FILOSÓFICA ACTUAL  
CONSEJO EDITORIAL  
Santiago M. Zarria  
Goethe-Universität, Frankfurt am Main,  
Alemania | Universidad Central del Ecuador  
Martín Aulestia Calero  
Universidad Central del Ecuador  
FILOSÓFICA EDITORIAL

Filosófica, Fundación de Estudios Filosóficos,  
Políticos y Culturales  
Quito, Ecuador  
filosofica@fundacionfilosofica.com  
www.fundacionfilosofica.com

FILOSÓFICA  
Fundación de Estudios Filosóficos, Políticos y Culturales  
CONSEJO ACADÉMICO  
Fernando Barredo  
Pontificia Universidad Católica del Ecuador  
David Cortez  
RWU Hochschule Ravensburg-Weingarten, Alemania  
Fernando Wirtz  
Universität Tübingen, Alemania | Universidad de Kioto  
Micaela A. Díaz,  
Lee University, Cleveland, TN, EE.UU.

Publicación apoyada por:

- ICALA, Stipendienwerk Lateinamerika-Deutschland e.V.
- Facultad de Ciencias Sociales y Humanas de la Universidad Central del Ecuador



**DERECHOS Y ACCESO**  
Esta publicación se encuentra disponible para consulta y descarga gratuita. Los contenidos pueden utilizarse para fines no comerciales, siempre que se reconozca adecuadamente la autoría y se cite la fuente original. En caso de realizar modificaciones, adaptaciones o transformaciones, deberá preservarse el principio de libre acceso al contenido.  
La autorización de acceso abierto no implica renuncia a los derechos morales de autor ni permite atribuir a los autores, editores o instituciones participantes modificaciones realizadas por terceros.



# Índice

|                   |    |
|-------------------|----|
| INTRODUCCIÓN..... | 11 |
|-------------------|----|

## PRIMERA SECCIÓN METAFÍSICA ANALÍTICA

|                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------|-----|
| CAPÍTULO I. UNA INTRODUCCIÓN NO IMPARCIAL A LA FUNDAMENTACIÓN<br>METAFÍSICA: DETERMINACIÓN Y FUNDAMENTALIDAD..... | 31  |
| <i>María Pía Méndez Mateluna &amp; Joaquim Giannotti</i>                                                          |     |
| CAPÍTULO II. LOS MÚLTIPLES ROSTROS DEL DETERMINISMO.....                                                          | 73  |
| <i>Hernán Accorinti &amp; Olimpia Lombardi</i>                                                                    |     |
| CAPÍTULO III. PERSISTENCIA E IDENTIDAD PERSONAL.....                                                              | 117 |
| <i>Angélica María Pena-Martínez</i>                                                                               |     |

## SEGUNDA SECCIÓN FILOSOFÍA DE LA MENTE

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| CAPÍTULO IV. LOS ESTADOS MENTALES COMO ESTADOS DE ACTITUD<br>PROPOSICIONAL.....                                | 153 |
| <i>Elberto A. Plazas</i>                                                                                       |     |
| CAPÍTULO V. THE MIND-BODY PROBLEM IN THE ANALYTIC TRADITION.....                                               | 197 |
| <i>Michael Schon</i>                                                                                           |     |
| CAPÍTULO VI. EL CARÁCTER SUBJETIVO DE LA EXPERIENCIA:<br>AUTOCONSCIENCIA E INFORMACIÓN DE PRIMERA PERSONA..... | 239 |
| <i>Miguel Ángel Sebastián</i>                                                                                  |     |
| CAPÍTULO VII. EL DEBATE ACTUAL SOBRE LA LECTURA DE MENTES EN BEBÉS.....                                        | 275 |
| <i>Anyerson Stiths Gómez-Tabares &amp; Santiago Arango-Muñoz</i>                                               |     |



## INTRODUCCIÓN

Hacia finales del siglo XIX, el idealismo era la corriente dominante entre los filósofos de habla inglesa. Como bien señaló Quinton (1971), su ingreso formal en la tradición británica puede fecharse al año 1876 y no fue hasta aproximadamente 1945 que los departamentos de filosofía dejaron de tener un profesorado que, mayoritariamente, se identificaba con el idealismo. En el ambiente intelectual de esa época se daba por hecho que la realidad era fundamentalmente inmaterial. Sin embargo, fue en 1903, año en el que Russell publicó *The Principles of Mathematics* y Moore su *Refutation of Idealism* cuando los inmateralistas británicos sintieron la primera embestida. Según Candlish (2009), la batalla decisiva se daría entre Russell y Bradley en torno a la naturaleza de las relaciones, lo que a su vez enfrentaría el pluralismo de Russell (asumido en la obra mencionada) con el monismo de Bradley.

*The Principles of Mathematics* fue una obra de profunda influencia donde se integra a un argumento metafísico técnicas formales. Para darnos una idea de su relevancia consideremos la opinión que Popper compartió a Magee acerca de esta obra:

*The Principles of Mathematics*, a mi parecer, es uno de los libros más grandiosos y asombrosos jamás escritos... Su logro no tiene parangón. Russell redescubrió la lógica y la teoría de números de Frege, y sentó las bases de la aritmética y el análisis. Ofreció la primera definición clara y sencilla de los números reales sobre esta base (una mejora con respecto a las de Cantor, Dedekind y Peano). Y no solo aportó una teoría de la geometría, sino también un nuevo enfoque a la mecánica... Fue el primero en ofrecer una teoría satisfactoria de los números irracionales, un problema que había intrigado a la humanidad durante 2400 años. Y dio una solución verdaderamente brillante al antiguo problema del movimiento<sup>1</sup>. (Magee, 1971, pp. 143-145)

En esta obra, Russell asume que cada proposición está conformada por las entidades invocadas por sus palabras (como si usara un lenguaje Lagadoniano). Donde, la proposición es en sí misma una unidad en la que las entidades que la conforman estarían como fun-

didas en ésta. Así, la proposición: <Pablo Palacio escribió Débora> es una entidad distinta al agregado de entidades: Pablo Palacio, escribir, y Débora. Esto se debe a que en su posición pluralista, los bloques fundamentales de la realidad son una diversidad de entidades que al relacionarse entre sí producen nuevas entidades; donde cada una de éstas en vez de ser un agregado, es en sí misma una unidad. Según Russell estas unidades pueden ser invocadas a través de proposiciones relacionales. Así, por ejemplo, en una proposición, el verbo es el que encarnaría la mencionada unificación. Sin embargo, Russell reconoce que la característica unificadora del verbo no abarca su naturaleza.

El verbo, cuando se usa como verbo, encarna la unidad de la proposición y, por lo tanto, se distingue del verbo considerado como término, aunque no sé cómo explicar con claridad la naturaleza precisa de esa distinción. (Russell, 1903/1937, p. 50)

Precisamente, Bradley muestra la incompatibilidad entre la característica unificadora del verbo, y su rol como *término*. Esto ya que, dicho rol, sería el que tienen las relaciones externas. Una relación externa es aquella cuyo ser y existencia es independiente del relata (i.e. aquello que está siendo relacionado por la relación). Por ejemplo, si Juan ama a Joselyn, la relación *amor* sería externa si ésta puede existir y ser lo que es independientemente de si Juan y Joselyn existen. En *Appearance and Reality*, Bradley (1893) proponía que para que una relación externa  $R$  (una relación no instanciada) vincule a dos entidades  $A$  y  $B$ , ésta necesitaría de otras dos relaciones  $R_1$  y  $R_2$  que vinculen  $R$  con  $A$ , y  $R$  con  $B$ ; respectivamente. Es decir, puesto que las naturalezas tanto de  $A$ ,  $B$  como  $R$  carecen de algo que vincule estas entidades entre sí, entonces  $R$  necesitaría un par de relaciones extra para que  $A$ ,  $B$  y  $R$  se vinculen. Sin embargo, si las relaciones son externas, cada relación extra necesitaría de un par adicional de relaciones, y así sucesivamente ad infinitum. Siendo así, una relación externa  $R$  nunca podría terminar vinculando  $A$  y  $B$ .

En otras palabras, Bradley argumenta que no es posible construir a través de relaciones externas una entidad donde sus com-

ponentes estén fundidas formando una unidad. Así, Bradley manifiesta:

La posición fundamental del señor Russell se me ha mantenido incomprensible. Por una parte, me veo llevado a pensar que defiende un pluralismo estricto, que no admite nada por fuera de términos simples y relaciones externas. Por otra parte, el señor Russell parece afirmar de modo enfático... ideas que un pluralismo semejante seguramente ha de repudiar. Él se apoya de manera constante en unidades que son complejas y que no se pueden analizar en términos y relaciones. A mi entender, estas dos posiciones son irreconciliables, pues la segunda, según mi comprensión, contradice de lleno a la primera. Si existen tales unidades, y, más aún, si tales unidades son fundamentales, entonces, sin duda, el pluralismo queda en principio abandonado por falso. (Bradley, 1910/1914, p. 281)

Russell por su parte contraatacó apuntando a las relaciones concebidas como internas—i.e. la postura en que la relación depende del relata--, puesto que consideró que el monismo de Bradley estaba soportado por la existencia de éstas. Según Russell, la posición del monista que supone que las relaciones son internas es la siguiente. Si una relación interna  $R$  está instanciada sobre un relata  $a, b$ , estos tres elementos forman un todo unificado, donde la proposición que apunta a esta totalidad no manifiesta algo acerca de  $a$  o  $b$ , sino que exhibe dicha totalidad. En sus propias palabras:

La teoría monista sostiene que toda proposición relacional ‘ $aRb$ ’ debe resolverse en una proposición referente al todo que  $a$  y  $b$  componen... Se nos dice, por parte de quienes defienden esta opinión [la referencia incluye a Bradley], que el todo contiene diversidad en su interior, que sintetiza diferencias... Por mi parte, me es imposible atribuir a estas frases ningún significado preciso. (1903/1937, pp. 219-220)

Luego de esta afirmación, Russell considera que la relación asimétrica ‘mayor que’ en la proposición  $\langle a$  es mayor que  $b \rangle$ , y expresa que, para los monistas mencionados, dicha proposición en realidad se entendería como  $\langle$ La totalidad formada por  $a$  y  $b$  contiene di-

versidad de magnitud>. Esta segunda proposición, sin embargo, es incapaz de reflejar la asimetría que, por otro lado, la primera sí exhibe. Notemos que, en una relación asimétrica no se puede cambiar el orden del relata sin modificar su valor de verdad. De manera que: si  $a > b$  es verdadero, entonces  $b > a$  será falso. En este sentido las relaciones internas no podrían garantizar la existencia de relaciones asimétricas. Por lo que, sería erróneo considerar que las relaciones internas son esenciales para la naturaleza de la realidad.

Sin embargo, concuerdo con Candlish (2009) en que Bradley no defendía la existencia de relaciones internas. El, más bien, consideraba que las relaciones son irreales. Es decir que, en principio sería inadecuado partir al Absoluto (la unidad que es lo único que realmente existe en su monismo) en distintas partes y erróneamente categorizarlas como fundamentales, de manera que estas al relacionarse produzcan situaciones o hechos. En palabras de Bradley:

En resumidas cuentas, lejos de admitir que el Monismo requiere que todas las verdades puedan interpretarse como la predicación de cualidades del Todo, el Monismo, tal como yo lo concibo, sostiene que toda predicación, sin importar cuál sea, es, en última instancia, falsa e irreal. (Bradley, 1935, p. 672)

Aquí, vale notar que, Russell (1903) consideraba que toda proposición expresa la existencia de una relación. Por lo que, en la cita anterior, Bradley al responder a Russell y manifestar la irrealidad de la predicación (a través de proposiciones) está expresando la irrealidad de las relaciones. En este sentido, la postura de Bradley no estaría siendo atacada por los argumentos de Russell en contra de las relaciones internas. Tomando en cuenta esto, concuerdo con Candlish (2009) en que para explicar el salto del idealismo a la corriente analítica en la filosofía anglo-parlante no es suficiente el relato este-reotípico (como él le llama) en el que, Russell ganó la disputa acerca de la naturaleza de las relaciones, y al hacerlo supera al idealismo consolidando la corriente analítica. En este sentido prefiero tomar una posición cauta en la línea de Lewis cuando afirma:

Las teorías filosóficas nunca son refutadas de manera concluyente. La teoría sobrevive a su refutación a un precio... Pero, al final del día, una vez descubiertos todos los argumentos astutos, las distinciones y los contraejemplos, es de prever que aún nos enfrentaremos a la pregunta de qué precios vale la pena pagar. (Lewis, 1983, p. x)

Según Candlish (2009) los factores que van más allá de los argumentos esgrimidos en los debates mencionados incluyen, por ejemplo, la posición de Russell al abandonar el cristianismo y la adopción momentánea del idealismo. Los aspectos místicos de dicha corriente pudieron tentar a Russell en su dimensión espiritual. En el ensayo 'Seems, ¿Madam? Nay, it is' presentado a la sociedad Cambridge Apostles en 1899, Russell sugiere que no pudo encontrar en la metafísica idealista el consuelo que de otra forma hubiese buscado en la religión, y aquello fue parte del desencanto con dicha corriente.

Se supone que McTaggart aún atribuye a la filosofía el poder de brindar consuelo y alivio —el último poder de los impotentes. Es esta última posesión de la que, esta noche, deseo despojar a la decrepita progenitora de nuestros dioses modernos (Russell, 1899/2004, p. 49)

En particular, la armonía de la única entidad realmente existente según el idealismo monista, con el que Russell coqueteó brevemente, no se alcanzaría ni después de morir, ni en el presente. Esto ya que, en tal postura idealista, nuestra cognición solo nos da acceso a la apariencia de lo plural, mientras lo real estaría fuera de nuestro alcance en la vida mundana. Así, Russell menciona:

Incluso el puro interés intelectual, del cual surge la metafísica, es un interés por explicar el mundo de la apariencia. Pero, en lugar de explicar de verdad este mundo sensible, palpable y actual, la metafísica construye otro mundo fundamentalmente diferente; tan distinto, tan desconectado de la experiencia real, que el mundo de la vida cotidiana permanece completamente inafectado por él y sigue su curso como si no existiera mundo de la Realidad en absoluto. (Russell, 1899/2004, p. 53)

Adicionalmente, Candlish (2009) identifica algunos factores de orden socio-político. Señala, por ejemplo, que el impacto de la primera guerra mundial desmotivó a la gente a tener una aproximación inmaterial o espiritual a la realidad. Así como también apunta al declive del imperio británico cuya dirección moral y espiritual estaba soportada por algunos idealistas hegelianos, y señala también el recelo, postguerra, a la influencia de la filosofía idealista prusiana en el pensamiento anglófono. Estos factores se añaden a la fuerte influencia que causó Russell con su aprecio a las matemáticas, y el ya mencionado distanciamiento del idealismo.

Sin embargo, más allá de todos los factores mencionados hasta aquí, fue la aproximación Russelliana a los problemas filosóficos donde se valora y adopta el lenguaje, la metodología, y la claridad de la ciencia lo que marcó la dirección de la corriente analítica. Russell visionó una nueva forma de hacer filosofía con una precisión y claridad de pensamiento y expresión sin paralelo. Tomemos, por ejemplo, el clásico *On Denoting*, donde Russell arroja luz sobre aspectos ante los que el lenguaje hegeliano no ofrecía respuesta. El siguiente extracto ilustra el tipo de claridad que busca la corriente analítica, donde la teoría de la denotación resuelve el problema sugerido a continuación:

En virtud del principio del tercero excluido, o bien “A es B” o bien “A no es B” debe ser verdadero. Por lo tanto, o bien “el actual rey de Francia es calvo” o bien “el actual rey de Francia no es calvo” debe ser verdadero. Sin embargo, si enumeráramos las cosas que son calvas, y luego las cosas que no son calvas, no encontraríamos al actual rey de Francia en ninguna de las dos listas. Los hegelianos, que aman las síntesis, probablemente concluirán que lleva peluca. (Russell, 1905, p. 485)

El desarrollo de la lógica proposicional por parte de Frege, junto con los aportes que, sobre ésta, realizaron tanto Russell como Whitehead dotaron de un poder expresivo y deductivo a la corriente analítica, que no estaba disponible previamente. Estos desarrollos han sido seminales para campos como la metalógica—la cual por



e.g. permitió el desarrollo de los teoremas de incompletitud de Gödel--, las ciencias cognitivas, o la inteligencia artificial.

En la actualidad, la corriente analítica es fundamental en las facultades de filosofía en el mundo anglófono. Por ejemplo, según *Philosophical Gourmet Report* (2025) editado por Brian Leiter, todas las facultades de filosofía de las universidades Ivy League se autodenominan como analíticas. Los intereses de la corriente se han ampliado al punto de que ahora ya se han trabajado sobre temas que en los primeros años de la corriente eran abordados casi exclusivamente por otras corrientes filosóficas. Por ejemplo, ciertos aspectos de la filosofía budista han sido tratados formalmente por dialeteistas como Graham Priest (2014), lo inefable ha sido trabajado con gran claridad por Silvia Jonas (2016), o el monismo ha sido defendido por Schaffer (2010).

La constitución de estos dos volúmenes es parte de un esfuerzo por despertar interés en la filosofía analítica en las personas de habla hispana que no están aún familiarizadas con esta corriente. Sobre todo, se busca dotar de algunas herramientas agudas de pensamiento a sus lectoras y lectores; las cuales, sin duda, sirven como bases para profundizar en la forma analítica de tratar los problemas filosóficos. Si bien es cierto, la constante amenaza de los conflictos sociales, ambientales y políticos que se cierne sobre Latinoamérica sugiere a la corriente continental como el campo teórico sobre el que se piensen estos conflictos, sin embargo, la filosofía analítica provee de una precisión que beneficiaría la argumentación y análisis de éstos. De hecho, cuando propuse la conformación de esta obra, tenía en mente más que nada el entregarse con la mente clara a los distintos problemas de la realidad y a la amplitud de la experiencia humana, la cual no se limita a lo social o político.

En este volumen la primera sección trata temas metafísicos imprescindibles incluyendo relaciones de determinación metafísica, determinismo, metafísica de la persistencia, e identidad personal. La segunda sección aborda temas fundamentales de la filosofía de la mente los cuales comprenden estados mentales como estados de

actitud proposicional, el problema cuerpo-mente, subjetividad, la consciencia, y lectura de mentes.

En el primer capítulo María Pía Méndez Mateluna y Joaquim Giannotti analizan con detalle un tema ineludible en la metafísica analítica, a saber, la fundamentación. El capítulo se enfoca en relaciones en las que el ser y/o existencia de una entidad se determina por el ser y/o existencia de otra. Es claro, por ejemplo, que mi identidad biológica y existencia se debe a la de mis padres. En este ejemplo la dependencia se da entre particulares concretos. Sin embargo, los hechos (o las proposiciones que los manifiestan), también pueden estar conectados a través de relaciones de determinación. Así, el hecho de que el conjunto  $\{2\}$  existe debe su existencia al hecho de que el número 2 existe. Esta es una relación de determinación explicativa—puesto que el primer hecho se explicaría con el segundo—, que además no es causalidad; esto ya que, la existencia del segundo hecho no sucede temporalmente al primero. Son este tipo de relaciones, principalmente, entre hechos que son consideradas fundamentación. El punto de vista ortodoxo considera que sus propiedades estructurales son asimetría e irreflexividad. Son asimétricas porque la existencia del número 2 fundamenta la existencia del singleton  $\{2\}$ , pero no es que la existencia del segundo fundamenta la del primero; e irreflexivas porque una entidad no se fundamentaría a sí misma. Méndez y Giannotti distinguen con claridad porqué el análisis en términos de fundamentación no puede ser reemplazado por análisis realizados en términos de e.g. necesidad, superveniencia, o dependencia ontológica. Por ejemplo, se podría pensar que un hecho X depende de otro Y si y sólo si es *necesario* que X exista si Y existe. Desde el punto de vista modal de la necesidad, el condicional: X existe si Y existe sería necesario siempre que el condicional sea verdadero en todos los mundos posibles. Brevemente, un mundo posible puede considerarse como una situación contrafactual acerca de la totalidad del universo. Siendo así, tanto el número 2 como su singleton al coexistir en todos los mundos posibles cumplirían con ser mutuamente dependientes, puesto que el condicional mencio-

nado es verdadero en todos los mundos posibles para cuando  $X$  e  $Y$  se concretan ya sea en 2 o en  $\{2\}$ . Es decir que, este análisis en términos de necesidad no da cuenta de la direccionalidad de la dependencia que, por otra parte, la asimetría de la fundamentación sí manifiesta. Adicionalmente, Méndez y Giannotti muestran como dentro del fundacionismo metafísico—postura que asume que una entidad depende de otra y esta de otra en una cadena de dependencias que termina en una, o un grupo de entidades no fundamentadas—, la fundamentación puede jugar un rol esencial tanto para dar cuenta cuándo se llega a la base (o último eslabón) de la cadena de dependencia, como para identificar qué eslabón va antes que otro en la cadena jerárquica de dependencias.

En el segundo capítulo Hernán Accorinti y Olimpia Lombardi nos llevan por un recorrido histórico analizando la noción de determinismo. A través de este viaje argumentan que podemos distinguir un fenómeno determinista cuando en una sucesión de acontecimientos, el evento posterior  $Ep$  depende del anterior  $Ea$  de manera que éste no podría haber sido sucedido por un evento diferente a  $Ep$ . Así, por ejemplo, la conexión nomológica (a manera de ley) entre eventos, nos permiten reconocer cuándo estamos tratando con un fenómeno determinista, en contraste a estar ante uno fatalista. Esto ya que el segundo correspondería a una sucesión de eventos que no necesitan depender uno de otro a través de una ley; sino que más bien están destinados a sucederse por una fuerza trascendente que ineludiblemente los ordena (casi caprichosamente) de manera sucesiva. Sin embargo, tal como los autores señalan, esto no significa que todo fenómeno nomológico sea determinista. Esto se debe a que, por ejemplo, existen sistemas físicos que evolucionan de manera tal que no todas sus propiedades para un tiempo posterior están unívocamente determinadas por el estado del sistema en un instante previo. Esto se puede observar en la mecánica cuántica, donde (en general, cuando no se realiza una medición) tan solo es probable que propiedades como e.g. velocidad o posición tengan cierto valor para un instante dado, a pesar de que el estado del sistema, expresado en la llamada función de onda, evo-

luciona temporalmente siguiendo una ley (ésta se manifiesta a través de una ecuación diferencial como la ecuación de Schrödinger). Adicionalmente, tal como los autores mencionan, si bien es cierto hay expresiones matemáticas que nos ayudan a predecir, esto no implica que el determinismo pueda ser subsumido en la predictibilidad. Una de las razones es que el predecir implica superar el desconocimiento a partir de la información con la que se cuenta, lo cual no necesariamente significa que se va a inferir el futuro a partir del pasado. Puesto que el determinismo en un sentido epistémico implicaría tener la capacidad de conocer la evolución temporal de los eventos, se tiene que predictibilidad no es idéntica con determinismo epistémico. Finalmente, los autores argumentan a partir de la filosofía de la Física que la evolución de la totalidad del universo no necesariamente sería determinista incluso si los sucesos individuales que conforman lo que es el caso en el universo sí lo fueran

Angélica María Pena-Martínez en el tercer capítulo analiza qué es lo que constituye la identidad de una persona en el transcurso del tiempo. A lo largo de nuestra vida los rasgos que en general consideraríamos constitutivos de nuestra identidad cambian. Así, por ejemplo, la persona que está leyendo esto en este instante tiene una constitución material y psicológica muy diferente a la que tuvo cuando fue bebé. Siendo así, ¿Qué es lo que nos permite identificarnos tanto con la persona que somos actualmente, como con la persona que fuimos al nacer? En la primera parte del capítulo la autora argumenta que este tipo de identificación la tenemos cuando hay un cambio gradual—al pasar de un instante al inmediatamente posterior—en el contenido mental (aquello de lo que se tratan nuestras experiencias mentales) y en los roles que nuestros estados mentales tienen en la red de experiencias que tenemos (por ejemplo, una imagen mental tiene el rol de evocar cierto estado anímico). Mientras que, en la segunda parte, Angélica María sostiene que una persona persiste siendo lo que es, a través del tiempo, debido a que el individuo tridimensional X que la persona es en algún instante—un estadio temporal de la persona—tiene

una contraparte temporal Y—otro estadio temporal--en el instante posterior. Donde X e Y deben ser similares, continuos y estar relacionados causalmente. Estas tres características se esperarían de los estados mentales tanto de X como de Y, puesto que para la autora éstos son los aspectos distintivos de la identidad personal. Con esta perspectiva, Angélica María responde a los problemas de la fusión y fisión en el contexto de la identidad personal. El primero se da cuando imaginamos que se fusiona el cerebro de un individuo A, con el cuerpo de otro B, y nos preguntamos si el sujeto resultante preserva la identidad personal de alguno de los dos anteriores. El segundo problema consiste en imaginar que la psicología de un individuo es traspasada al cerebro de otras dos personas, y nos preguntamos si la identidad del primero es la misma que la de los otros dos.

La intersección entre la filosofía de la mente y la filosofía del lenguaje es abordada agudamente por Elberto Plazas en el cuarto capítulo. Consideremos, por ejemplo, que: Ibeth cree que Sócrates no escribió nada. Siendo así, será que ¿Ibeth cree que el maestro de Platón no escribió nada? Por un lado, parece ser que, si en una ocasión Ibeth cree lo primero, y en otra cree lo mismo manifestado por: el maestro de Platón no escribió nada, entonces en ambas ocasiones tendría el mismo estado mental. Sin embargo, también podríamos pensar que el estado mental de Ibeth al creer lo primero fuera el mismo que tuviera si creyese lo segundo, solo en caso conozca que Sócrates fue el maestro de Platón. El autor muestra los problemas que surgen cuando se aborda este tipo de casos considerando que actitudes mentales como: creer, temer, pensar, etc., pueden tenerse hacia entidades no subjetivas y de acceso común como las proposiciones (concebidas como la historia común que cuentan distintas oraciones, como cuando se afirma “estoy sobre el piso” o “el piso está debajo mío”). Dichos problemas se analizan a través de la relación entre intencionalidad e intensionalidad. La primera (con c) es la característica de los fenómenos mentales de tratarse acerca de algo, la segunda (con s) se refiere a los contextos del lenguaje en los que dejan de cumplirse aspectos caracterís-

ticos de proposiciones como: Sócrates no escribió nada. Así, por ejemplo, si en esta proposición cambiamos el término ‘Sócrates’ por ‘el maestro de Platón’ ambas proposiciones serían verdaderas ya que se refieren a la misma persona. Sin embargo, al poner dicha proposición luego de ‘Ibeth cree que’, notamos que la proposición resultante puede ser verdadera para cuando forma parte de ésta el término ‘Sócrates’ y falsa cuando se lo cambia por ‘el maestro de Platón’. El autor muestra cómo estos problemas se atenúan en la teoría de Fodor, quien parte de la suposición de que lo mental se caracteriza por su función, y que la acción de pensar concebida como computar se actualiza en procesos del sistema nervioso. Así, los representantes de la realidad—concebidos en otras teorías como las proposiciones—en este caso son códigos encarnados en el sistema nervioso. De manera que, las creencias se tienen acerca de estos representantes, y no acerca de la realidad en sí misma. En general, el recorrido por el que nos lleva Elbertio ilumina cómo la filosofía analítica trabaja sobre el lenguaje para entender lo mental.

En el capítulo quinto Michael Schon discute críticamente el desarrollo del problema mente-cuerpo, aquel en el que buscamos entender ¿Cómo se relaciona lo mental (incluyendo la consciencia) con lo material? Aquí, Schon contrasta dos maneras de afrontar esta pregunta, por un lado, la postura dualista en la que la mente puede existir sin la materia; y, por otro lado, la perspectiva materialista donde necesariamente lo mental existe si lo material existe. En la comparación, el autor resalta las similitudes entre algunos debates contemporáneos en filosofía de la mente, y los que tuvo Descartes con Arnaud y la princesa Elisabeth de Bohemia. Fue Descartes el que—contrariamente con lo que era usual en el siglo XVII—defendió que, aunque no se nos presenten separados el cuerpo de la mente, estos no solo son claramente distinguibles, sino que lo mental podría haber existido sin lo material. Para sostener esto Descartes asumió que aquello que se concibe claramente como distinto—en este caso la mente y el cuerpo—, posiblemente es distinto; y que esto, implicaría que dichas entidades (mente y cuerpo) en efecto son distintas. Por lo que, la una

podría haber existido sin la otra. El autor muestra, por ejemplo, ciertas similitudes en esta argumentación y la de Kripke al rechazar la teoría materialista de la identidad mente-cerebro. Dicha teoría identifica estados mentales con procesos físico-químicos así, por ejemplo, considera que no hay nada por encima de cierta activación de las fibras C que sea dolor. U.T. Place y J.J.C. Smart defendieron esta teoría argumentando que la siguiente afirmación: i) dolor = activación de fibras C, está motivada científicamente al igual que: ii) rayo = descarga electrostática. Place y Smart consideraban que verdades como ii) que se descubren empíricamente (a través de la información sensorial) son contingentes. Esto quiere decir que, las entidades identificadas tanto en i) como en ii) podrían haber sido no idénticas en una situación hipotética. Por otro lado, Kripke argumenta que: si dos entidades son realmente idénticas, la identidad debe mantenerse en todas las situaciones hipotéticas en las que éstas participen. Por lo que ii) al ser verdadera, es verdadera necesariamente (i.e. en todas las posibles situaciones) y no contingentemente como Place y Smart defendían. El autor muestra como parte del argumento de Kripke se puede leer como una reformulación del argumento de Descartes. Además, explica que dicha argumentación aplica también a otros tipos de materialismo (fiscalismo) donde no se asevera que lo mental y lo material son idénticos, sino más bien que el ser y existencia de lo primero depende de lo segundo. Así, el autor revisa distintas posturas que han iluminado el problema, viajando temporalmente desde su nacimiento hasta discusiones contemporáneas del debate sobre el dualismo versus fiscalismo.

Las características de la consciencia como la experiencia cualitativa, y la información en primera persona son analizadas en el capítulo sexto por Miguel Ángel Sebastián. En particular, el autor discute si la primera puede existir sin la segunda; y defiende que, más bien, la consciencia de uno mismo (información en primera persona) es un rasgo inseparable de la experiencia consciente al que no se lo ha analizado en la misma medida que al qualia. Estos rasgos son aspectos fundamentales de notorias discusiones acerca

de la conciencia por parte de filósofos como Jackson y Nagel. Así, el qualia es la experiencia cualitativa que, por ejemplo, gana María al salir de su cuarto en el famoso experimento mental de Frank Jackson. María es el ser humano más inteligente, y conoce el estado del arte de las ciencias cuando éstas han llegado a explicar todo lo científicamente explicable; por esto, ella conoce todo lo que la ciencia tiene que decir acerca del color rojo. María nunca ha visto dicho color, ya que ha vivido en un cuarto donde absolutamente todo es blanco o negro. Un día sale del cuarto y mira el rojo de una rosa, la información que gana al hacerlo, y que no le proveyó la ciencia, corresponde al qualia. Por otro lado, Nagel enfatiza en un rasgo distintivo de la conciencia—apuntando a lo que se siente ser una criatura consciente—con su célebre pregunta ¿Qué se siente ser un murciélago? Miguel Ángel defiende que tanto el qualia de lo rojo y lo que se siente ser un murciélago aportan información que puede ser accedida por distintos sujetos, pero que quién la recibe obtiene información perspectiva; es decir, se informa de cómo es el mundo para dicho sujeto, no de cómo es el mundo independientemente de la existencia de éste. El autor defiende que el aspecto perspectivo de la conciencia es ineludible incluso en experiencias de despersonalización, como aquellas debidas a: estados meditativos, condiciones mentales como el síndrome de Cotard (donde uno está convencido que no existe, o que se está muerto), o la ingesta de psicodélicos. Miguel Ángel afirma que éstas no son auténticas experiencias de disolución de la información en primera persona; más bien, argumenta que, cuando son reportadas estas tratan de acomodarse para encajar en el relato de la disolución del yo. La posición de Miguel contrasta con la de Hume; a quien, por otro lado, en introspección se le presenta una experiencia--como, por ejemplo, la que se tiene al oler café—que no viene acompañada de experimentar la existencia de un sujeto (el mismo) que se encuentra en dicha experiencia olfativa. De hecho, el autor argumenta que ante cualquier experiencia consciente X, el sujeto también es consciente de que alguien está participando de dicha experiencia X, por lo que X le permite ser consciente de sí mismo.



En el capítulo séptimo Ányerson Gómez-Tabares y Santiago Arango-Muñoz discuten cuándo y cómo se adquiere la capacidad de comprender que otros pueden albergar, por ejemplo, conocimientos y creencias que difieren de las de uno. Tradicionalmente, se consideraba que esta capacidad conocida como lectura de mentes emerge alrededor de los cuatro años, sin embargo, estudios recientes apuntan a que es alrededor de los quince meses que ya se la adquiere. Los experimentos para detectar si en efecto se tiene dicha capacidad se han realizado en niños de alrededor cuatro años atestiguando situaciones (e.g. recreadas con muñecos) donde, por ejemplo, un individuo X ubica un objeto en una locación L1 (un cajón, o una canasta) dentro de una habitación antes de abandonarla, luego otra persona Y cambia la ubicación del objeto a otra locación L2 (dentro de la habitación) sin que X se dé cuenta, finalmente X regresa a la habitación. Aquí, preguntamos al niño ¿Dónde esperarás que X busque al objeto? Si el niño espera que X busque el objeto en L1, esto indicaría que el niño es capaz de comprender que sus estados mentales—como el conocer que el objeto está en L2—pueden diferir de los estados mentales de X (quién en efecto ignora dicho hecho, mientras que el niño se percató de tal desconocimiento). En este caso, la interacción (preguntas y respuestas) de los investigadores con el niño es verbal. Por lo que, los recursos mentales que usa el niño para generar sus expectativas son lingüísticos o proposicionales. Es decir, en esta edad los niños ya comprenden y manejan conceptos. Por otro lado, una infanta de alrededor de quince meses luego de atestiguar las situaciones manifestaría si el accionar de X es contrario a sus expectativas sosteniendo prolongadamente su mirada sobre cierta locación. Así, si la expectativa no conceptual, de la niña es que X al regresar al cuarto busque el objeto en L1—lo cual indicaría que la niña en efecto comprende que otros pueden tener estados mentales distintos a los suyos—, pero éste más bien acude a L2, entonces la niña mantendrá la mirada sobre L2 durante más tiempo del habitual. Para dar cuenta de este comportamiento, los autores examinan dos propuestas cognitivas que permitirían gestionar la información a

un infante de esa edad sin usar de lleno conceptos. Santiago y An-yenson proponen una forma diagramática--con mapas visuales--en la que la información es gestionada no conceptualmente, tan solo recurriendo a patrones enriqueciendo así las dos propuestas antes mencionadas. Es decir, detallan cómo puede ser representada por medio de diagramas visuales la información que nos transmite nuestros estados mentales. En este sentido, los autores muestran cómo interactúan las propuestas filosóficas acerca del contenido y representaciones mentales con la consolidación de avances en las ciencias cognitivas.

Juan Villacrés Bolaños  
EDITOR

## Referencias

- Bradley, F.H. (1983). *Appearance and Reality: A Metaphysical Essay*, Oxford University Press.
- Bradley, F. H. (1914). *Essays on truth and reality*. Cambridge University Press. (Trabajo originalmente publicado en 1910).
- Bradley, F.H. (1935). Relations. En *Collected Essays*. Clarendon Press (pp. 629-676). (Trabajo originalmente escrito en 1924).
- Candlish, S. (2009). *The Russell/Bradley dispute and its significance for twentieth century philosophy*. Springer.
- Jonas, S. (2016). *Ineffability and its metaphysics: The unspeakable in art, religion, and philosophy*. Palgrave Macmillan.
- Leiter, B. (s. f.). Analytic and Continental Philosophy. *The Philosophical Gourmet Report*. Recuperado 12 de noviembre de 2025 de <https://philosophicalgourmet.com/analytic-and-continental-philosophy/>
- Lewis, D. (1983). *Philosophical Papers Volume I*. Oxford University Press.
- Magee, B. (1971). *Modern British Philosophy*. London: Secker & Warburg.
- Priest, G. (2014). *One: Being an Investigation into the Unity of Reality and of its Parts, including the Singular Object which is Nothingness*. Oxford University Press (UK).

- Quinton, Anthony (1971). Absolute Idealism. *Proceedings of the British Academy* 57, 303–29
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479-493.
- Russell, B. (1937). *The Principles of Mathematics* (2nd ed.). W. W. Norton & Company. (Trabajo original publicado en 1903).
- Russell, B. (2004). Seems, ¿Madam? Nay, it is. En *Why I am not a Christian and other essays on religion and related subjects* (pp. 48-57). Routledge. (Ensayo presentado en 1899).
- Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical review*, 119(1), 31-76.



**PRIMERA SECCIÓN**

**METAFÍSICA ANALÍTICA**



## CAPITULO I

# UNA INTRODUCCIÓN NO IMPARCIAL A LA FUNDAMENTACIÓN METAFÍSICA: DETERMINACIÓN Y FUNDAMENTALIDAD

*María Pía Méndez Mateluna<sup>1</sup> & Joaquim Giannotti<sup>2</sup>*

## RESUMEN

Este capítulo es una guía para conocer los elementos básicos de la fundamentación metafísica. Ofrece una revisión accesible de sus

- 
- 1 María Pía Méndez Mateluna es Profesora Asistente en la Escuela de Gobierno y Administración Pública de la Universidad Mayor (Chile). Es investigadora responsable del Proyecto FONDECYT de Postdoctorado 3240284 de la Agencia Nacional de Investigación y Desarrollo (ANID) "Machismo en América Latina: Un Marco Analítico para Explorar la Opresión de Género". Su investigación se enfoca sobre temas de filosofía política feminista, epistemología social y política, machismo, autonomía, epistocracia, género.
  - 2 Joaquim Giannotti es Profesor Asistente del Núcleo de Ciencias Sociales y Artes de la Universidad Mayor (Chile). Es Investigador Asociado del African Centre for Epistemology and Philosophy of Science (ACEPS) de la Universidad de Johannesburg. Sus áreas de especialización son en la metafísica de la ciencia y las áreas metafísicas de la filosofía de la ciencia, fundamentalidad, fundación, y la ontología de los poderes.

características y usos, comparando este concepto con otras formas de dependencia que se pueden encontrar en la literatura. Enfatiza dos roles teóricos centrales que la fundamentación supuestamente juega en la teorización filosófica: (i) dar cuenta de una forma distintiva de determinación no causal y (ii) entender mejor la estructura jerárquica de la realidad. El capítulo apunta a persuadir a las lectoras de la utilidad de la fundamentación discutiendo cómo pueden mitigarse algunas objeciones dirigidas a esos roles.

## ABSTRACT

This chapter is a guide to the basics of metaphysical grounding. It offers an accessible overview of its features and uses, comparing this concept with other forms of dependency one can find in the literature. It emphasizes two major theoretical roles grounding is claimed to play in philosophical theorizing: (i) accounting for a distinctive form of non-causal determination and (ii) illuminating the hierarchical structure of reality. The chapter aims to persuade the reader of the usefulness of grounding by discussing how some objections targeting these roles can be mitigated.

### 1. Reseña del capítulo

Nuestro mundo despliega estructura. Muchas cosas, el sentido amplio de la palabra, parecen exhibir varias dependencias de otras. Que un material sea frágil depende de su organización molecular. Que una reunión de personas sea un asado depende de las actitudes de sus participantes, de la presencia de una parrilla, y así. Que una persona esté consciente depende, al menos parcialmente, del buen funcionamiento de su cerebro. Se arguye que el concepto de ‘fundamentación metafísica’ (*‘metaphysical grounding’*) es un instrumento poderoso para teorizar acerca del tipo de dependencia que no puede entenderse adecuadamente en términos de causas y efectos, o en términos de conexiones necesarias. El objetivo de este capítulo es guiar a través de los elementos centrales de la fundamentación a la lectora que desconoce la noción. Ilustraremos los puntos claves



para articular una teoría de la fundamentación, enfocándonos en los dos principales roles que se sostiene que esta juega para nuestra teorización filosófica: (i) dar cuenta de una forma distintiva de determinación no causal, y (ii) contribuir a la comprensión de las nociones de fundamentalidad y prioridad.

La estructura es la siguiente. En la sección 2, presentamos la fundamentación utilizando ejemplos. En la sección 3, ofrecemos una breve historia del concepto. La sección 4 entrega una revisión de las características de la fundamentación y elecciones teóricas clave que se pueden hacer al articular una teoría. En la sección 5, nos enfocamos en el carácter determinativo de la fundamentación al discutir sus conexiones con relaciones de necesidad, superveniencia, dependencia ontológica, y esencia. En la sección 6, discutimos cómo se puede analizar la fundamentalidad y la prioridad en términos de fundamentación. La sección 7 presenta algunas objeciones contra los frutos que puede rendir y la unidad de la fundamentación. Concluimos, en la sección 8, indicando dos áreas poco exploradas a las que la fundamentación podría contribuir.

## **2. Fundamentación por medio de ejemplos**

En una primera aproximación, la fundamentación es una forma de determinación entre elementos donde el elemento determinado se obtiene o se genera en virtud de y se explica en términos de algunos otros elementos.

La noción de fundamentación se suele entender como irreducible a conceptos más básicos (e.g., Schaffer, 2009; Rosen, 2010; Audi, 2012). Una manera popular de presentarla es por medio de ejemplos que intentan entregar una idea intuitiva de que una cosa ocurra en virtud de otra. Seguiremos esta práctica, ya que transmite eficazmente la idea de que este fenómeno de dependencia ‘en virtud’ se puede encontrar asociado a una diversidad de temas y no está vinculado a unos específicos. Aquí hay una lista corta y no exhaustiva de afirmaciones ‘en virtud de’ que se pueden a menudo encontrar asociadas a la fundamentación:

- a. El hecho de que la pelota es roja y redonda ocurre en virtud del hecho de que es roja y del hecho que es redonda.
- b. Que la pelota sea roja es en virtud de que es escarlata.
- c. Lo incorrecto de una acción ocurre en virtud del que es hecha con la sola intención de causar daño.
- d. El hecho que el singleton {Sócrates} exista ocurre en virtud del hecho que Sócrates existe.
- e. El hecho de que algo exista ocurre en virtud del hecho que María existe.
- f. Un objeto compuesto existe en virtud del hecho que sus partes existen.
- g. Hechos acerca de los estados mentales de una persona ocurren en virtud de hechos acerca de los estados neuronales de esa persona.
- h. Las leyes de la naturaleza ocurren en virtud de distribuciones de cualidades.
- i. Una persona es una mujer ocurre en virtud del hecho que ella ocupa cierto rol social.
- j. El hecho de que una obra de arte sea bella ocurre en virtud de hechos sobre las actitudes de quienes la ven.

No nos preocupemos respecto de si estas afirmaciones son correctas. En cambio, deberíamos enfocarnos en lo que estos ejemplos buscan expresar. Quienes proponen a la fundamentación sostienen que, a pesar de las diferencias superficiales, estos ejemplos, si fueran correctos, muestran las características distintivas de la fundamentación. En cada caso, las defensoras de la fundamentación argumentarían que la conexión entre los ítems tiene un carácter de necesidad subyacente, un elemento explicativo, y un carácter no causal pero determinativo. Como explicaremos, estos tres ingredientes son constitutivos de lo que es la fundamentación.

Considera la lista (a) – (j). Las conexiones en estos ejemplos son modales en el sentido de que, si fueran verdaderas, los hechos que se obtendrían en virtud de los otros hechos lo serían necesariamente. En algunos casos, la necesidad es metafísica o lógica: la

conexión se mantiene en todos los mundos posibles. Por ejemplo, consideremos el ejemplo (e). Incluso si no se dice explícitamente, es lógicamente necesario que, si María existe, entonces algo existe. Y que la pelota sea roja es en virtud de que es escarlata, entonces es metafísicamente necesario que, si una pelota es escarlata, también es roja (b). La modalidad de las conexiones en (g) y (j) es algo más débil que una de necesidad metafísica o lógica, ya que son más plausibles en todos y sólo en aquellos mundos posibles que tengan las mismas leyes de la naturaleza o prácticas sociales. Adicionalmente, las conexiones son explicativas. Podríamos decir, muy plausiblemente, que si que la pelota sea roja es en virtud de que es escarlata, entonces la pelota es roja porque es escarlata. Y podríamos decir que una acción es incorrecta porque es hecha con la sola intención de causar daño (c). La característica más distintiva de estas conexiones, sin embargo, es que despliegan una dependencia determinativa que es difícil de explicar de manera precisa pero que va más allá de alguna forma de necesidad. Podríamos decir, por ejemplo, el hecho de que algo exista ocurre en virtud del hecho de que María (un ser humano común y corriente) exista, en el sentido de que el primer hecho depende del segundo, y que es necesario que, si María existe, algo existe. Pero el hecho de que María existe no es, en el sentido ordinario de la palabra, una causa del hecho de que algo existe. No se trata de un proceso físico que se desarrolla a través del tiempo, como la rotura de una ventana causada por el anterior lanzamiento de una piedra por parte de Suzie. Al golpear la ventana con una cierta aceleración, la piedra transfiere su energía al vidrio, lo que provoca su rotura<sup>3</sup>. No es obvio si el hecho de que María exista transfiere alguna energía al hecho de que algo existe. Más aún, la dependencia es co-temporal a diferencia del caso de la rotura de la ventana. Pero, el hecho de que María existe parece determinar el hecho de que algo existe. El carácter deter-

---

3 Algunos filósofos han identificado los procesos de transferencia de energía con la causalidad (p. ej., Salmon 1984, 1997; Dowe (1992)). La identificación es, por supuesto, controvertida. Aquí el ejemplo pretende destacar una diferencia intuitiva entre los procesos físicos causales y los fundamentales.

minante y no causal de la conexión entre unos elementos y otros es un indicador inicial de fundamentación.

### 3. Una ampliamente simplificada y breve historia de la fundamentación

Una vez que entendemos intuitivamente este concepto, podemos proceder a la tarea de refinar la noción de fundamentación. Sin embargo, para apreciar el atractivo de esta noción y sus supuestos benéficos, es útil comenzar con una breve contextualización histórica.

Preguntas concernientes a los fundamentos de ítems de interés puede encontrarse desde la antigüedad<sup>4</sup>. Sócrates preguntando qué hace que un acto sea piadoso, que puede encontrarse en el *Eutyphro* de Platón, suele presentarse como una pregunta paradigmática sobre qué fundamenta qué. Algo como la fundamentación aparece en el trabajo de Aristóteles, particularmente en su *Posterior Analytics*. Preguntas similares acerca de los fundamentos de ciertas cosas sobrevivieron la filosofía medieval occidental y la filosofía moderna temprana, particularmente en conexión con la discusión sobre el Principio de Razón Suficiente iniciado en el trabajo de Leibniz (cf. Amijee, 2023). Unos pocos siglos más tarde, Bolzano ofrece una lectura de la fundamentación en su *Teoría de la Ciencia* (1985/1837) que anticipa muchos aspectos de la discusión actual sobre la naturaleza de la fundamentación. Junto con otros factores, el relato prosigue diciendo, que los sentimientos anti metafísicos que tuvieron su cúspide con Carnap y Quine, llevaron a la desaparición de la fundamentación hasta su resurgimiento reciente en el siglo XXI. ¿Cómo llegamos de un panorama Quineano desértico al prolífico ambiente metafísico actual?

La resurrección de la fundamentación se le acredita principalmente al trabajo influyente de Kit Fine (2001, 2012), Jonathan

---

<sup>4</sup> Las dependencias explicativas no causales tienen orígenes antiguos también en tradiciones budistas (Priest, 2018) y en la metafísica china (Li & Perkins, 2015; Perkins, 2023). Según nuestro conocimiento, la historia de la fundamentación en la filosofía latinoamericana es un tema no explorado.

Schaffer (2009), y Gideon Rosen (2010). Pero su consolidación como un tema legítimo de discusión ha sido promovido a regañadientes por objeciones influyentes en contra de su fecundidad teórica, particularmente en el artículo seminal ‘No work for a theory of grounding’ de Jessica M. Wilson (discutiremos algunas objeciones en mayor detalle en la sección 7). Para comprender mejor el auge de la fundamentación en la filosofía analítica contemporánea, sin embargo, tenemos que considerar su lugar en el giro post modal en la filosofía analítica.

En la filosofía analítica occidental anglófona, durante los años 50 y 60, las herramientas preferidas para teorizar eran el significado y el análisis. Se pretendía que la teorización filosófica adecuada se hiciera a través del análisis extensional: las nociones de interés debían sustituirse por predicados con la misma extensión o nombres co-rreferenciales. Las figuras principales de este enfoque fueron Tarski, Quine y Davidson. Desde finales de los años 60, hemos observado un cambio de enfoque en la modalidad: el uso de teorías y marcos de mundos posibles para el análisis metafísico se ha vuelto popular. Desde finales de los años 60 hasta los años 90, el atractivo del análisis modal, particularmente a través del aparato de mundos posibles, de nociones clave como esencia, posibilidad y necesidad, semántica, conocimiento, causalidad y contenido mental (solo para dar algunos ejemplos) han florecido. Algunos de los principales autores que llevaron a cabo este enfoque de la filosofía fueron Montague, Stalnaker, Lewis, Hintikka y Kripke. Dos ejemplos que encarnan el espíritu modalizador de este enfoque son *Naming and Necessity* (Kripke 1980) de Kripke y *On the Plurality of Worlds* (Lewis 1986) de David Lewis.

Desde el principio de los 90s, hemos estado presenciando una revolución post modal. El leitmotiv de este movimiento es que los conceptos modales (necesidad y posibilidad) son demasiado toscos para investigar ciertos fenómenos de interés. El famoso ejemplo de Kit Fine (1994) de Sócrates y singleton {Sócrates} es representativo del espíritu post modal, y puede ayudar a la lectora a comprender mejor la manera en que la fundamentación no es puramente

una noción modal. Supongamos que analizamos la dependencia en términos de necesidad tal que  $x$  depende de  $y$  si y sólo si es necesario que  $x$  exista, si  $y$  existe (cf. Fine 1995, p. 2). Según Fine, podemos estar de acuerdo en que hay una conexión necesaria entre la existencia de Sócrates y la existencia del singleton. En cada mundo posible donde Sócrates existe, su singleton lo hace ya que siempre que  $x$  existe, también existe  $\{x\}$ . Sin embargo, lo opuesto también es cierto: en cada mundo posible donde  $\{\text{Sócrates}\}$  existe, los filósofos griegos también lo hacen. El problema es que el análisis de la dependencia en términos de necesidad no logra recuperar la asimetría intuitiva de la dependencia de  $\{\text{Sócrates}\}$  respecto de la existencia de Sócrates, y no viceversa. Por lo tanto, según Fine, deberíamos reconocer una noción de dependencia que escape a una definición puramente en termino de necesidad o propiedades necesarias si deseamos capturar la dependencia asimétrica relevante, una que no puede entenderse exhaustivamente en términos puramente modales.

Teniendo en cuenta esta breve historia del concepto, podemos pasar a la descripción de las características clave y las principales opciones teóricas que se podrían hacer al diseñar una teoría de la fundamentación.

#### 4. Fundamentación: Las bases

Si queremos emplear la fundamentación de manera eficaz y responsable en nuestra teorización filosófica, no podemos basarnos en ejemplos, como los presentados en la sección 2. Debemos suplementar la fundamentación con un conjunto de características y especificaciones. Por motivos de espacio, no podemos cubrir exhaustivamente cada opción teórica posible de tomar. Para una discusión más profunda de algunas características técnicas, redirigimos la lectora a Fine (2012).

## 4.1 ¿Operador o Relación?

¿Cómo deberíamos representar afirmaciones que involucran fundamentación? Hay dos posturas principales en la literatura. Una es el *operacionalismo*, y la otra es el *relacionalismo*.

El acercamiento operacionalista sostiene que la manera más precisa de representar la fundamentación es por medio de un conector sentencial o un operador flanqueado por sentencias. Es común encontrar ‘<’ como el operador de la fundamentación (e.g., Fine 2012). Típicamente, la sentencia o sentencias en el lado izquierdo de ‘<’ son las que están haciendo el trabajo de fundamentar. La sentencia o sentencias del lado derecho de ‘<’ son las que son fundamentadas por la sentencia o sentencias del lado izquierdo. Por ejemplo, (j) y (k) serían afirmaciones de fundamentación representadas de la siguiente forma, de acuerdo con el operacionalismo:

(j) Sócrates existe < la existencia de {Sócrates}

(k) La pelota es roja, la pelota es redonda < la pelota es roja y redonda

De acuerdo con la postura relacionalista, las afirmaciones de fundamentación son representadas por medio de un predicado relacional, tal como ‘...está fundamentado en...’ que conecta a algunas entidades. Según una postura popular pero no obligatoria, los elementos de la relación de fundamentación son hechos. Los ejemplos (j) y (k) serían representados, según el relacionalismo, como (l) y (m), respectivamente:

(l) el hecho de que {Sócrates} existe está fundamentado en el hecho de que Sócrates existe

(m) el hecho de que la pelota es roja y redonda está fundamentado en el hecho de que la pelota es roja y en el hecho de que la pelota es redonda.

La elección entre operacionalismo y relacionalismo no es meramente representacional. En cambio, codifica diferentes compromisos, particularmente de tipo ontológico, asociados con la fundamenta-

ción. El acercamiento relacionalista es, al menos superficialmente, ontológicamente más demandante que el operacionalismo, porque requiere un compromiso con los elementos de la fundamentación sumado a la relación de fundamentación en sí, que normalmente se considera irreducible a nociones más básicas. En cambio, la concepción operacionalista no necesita incurrir en un compromiso ontológico de ese tipo. La operacionalista puede argumentar que el operador ‘ $\prec$ ’ es sólo una forma filosóficamente útil de representar una operación entre oraciones de una manera análoga a la que se pueden introducir otros operadores en una formalización de un lenguaje dado para una variedad de propósitos teóricos. Así, la elección entre operacionalismo y relacionismo está inevitablemente influenciada por la ontología favorita de cada persona. Sostenemos, sin embargo, que, si se pretende que la fundamentación capture un fenómeno mundano genuino en lugar de meramente representacional, es probable que la operacionalista necesite un compromiso con una relación más allá del operador.

#### 4.2 ¿Hechos o Verdades u Otras Entidades?

La elección entre operacionalismo y relacionalismo limita la selección de entidades, entendidas en sentido amplio, que pueden entrar en relaciones de fundamentación. Sin embargo, todavía se necesitan algunas especificaciones adicionales. Quienes prefieren el operacionalismo, probablemente preferirán entidades representacionales, capaces de ser portadores de verdad, tales como sentencias o proposiciones (e.g., Fine, 2012). Quienes prefieren el relacionalismo preferirán pensar que la fundamentación enlaza pedazos de la realidad en lugar de representaciones de la realidad, tales como hechos u objetos y propiedades.

Desafortunadamente, no siempre es claro lo que tienen en mente quienes teorizan sobre fundamentación cuando hablan de proposiciones y hechos. DeRosset (2023, pp. 40–46) distingue entre concepciones *finas* (*thin*) y *gruesas* (*chunky*) de hechos. En la concepción fina, la existencia del hecho de que  $p$  se implica por el hecho de que  $p$  es verdadero. En consecuencia, uno podría verse



tentado a identificar hechos sutiles con proposiciones verdaderas. En la concepción gruesa, los hechos son entidades estructuradas que están de algún modo constituidas por objetos, propiedades, conectivos, etcétera (e.g., Rosen 2010, p. 114). No representan el contenido de oraciones verdaderas, ya que son elementos independientes de la mente que podrían no poder expresarse en oraciones. Además de las concepciones fina y densa basada en hechos, hay una postura más liberal según la cual las entidades de cualquier categoría pueden sostener relaciones de fundamentación (e.g., Schaffer 2009). Para esta postura, es la misma María quien fundamenta la existencia de algo en lugar de que el hecho de que María existe fundamente el hecho de que algo existe.

Nuevamente, la elección entre estas opciones está influenciada por varios compromisos de fondo y consideraciones independientes. Por ejemplo, alguien que piensa que las concepciones estructuradas de hechos o estado de cosas son problemáticas (cf. Macbride 2022) ciertamente evitaría elegir estos ítems. Si bien no negamos las dificultades para completar una buena explicación metafísica de los hechos, creemos que un liberalismo sin restricciones acerca de los elementos de la fundamentación causa una pérdida de precisión respecto de lo que se supone que la fundamentación sistematice.

### 4.3 ¿'Worldly' o Representacional?

Una elección relacionada concierne a la pregunta sobre si la fundamentación es no representacional (*wordly* en inglés) o representacional. Usando la distinción previa, podríamos decir que la fundamentación es *worldly* o no representacional si involucra hechos gruesos (*chunky* en inglés), como el hecho de que Milán está en Italia. Dos ejemplos de esta postura son Correia (2010) y Audi (2012). La fundamentación es representacional si no es *worldly*—a saber, si no involucra pedazos objetivos de la realidad. Algunos ejemplos de esta postura son Schnieder (2010), Correia (2017), y Litland (2017).

Las concepciones no representacionales y representacionales están en desacuerdo respecto de si hay casos en que dos sentencias corresponden a hechos equivalentes en términos de la funda-

mentación. Dos hechos son equivalentes en términos de la fundamentación si y sólo si fundamentan los mismos hechos y están fundamentados por los mismos hechos. Mantengamos las cosas simples ilustrando la diferencia con un ejemplo. En la concepción no representacional, ‘A’ y ‘A o A’ describen el mismo hecho entendido como un pedazo objetivo de la realidad: a saber, A. En la concepción representacional, en cambio, las sentencias describen diferentes hechos. Esta concepción es sensible a la forma en que se expresa el contenido de las oraciones. Puesto que ‘A’ y ‘A o A’ representan su contenido de manera diferente, existen diferencias correspondientes en los hechos que denotan o a los que corresponden. Una forma de ilustrar la diferencia es considerar la irreflexividad de la fundamentación. Típicamente, el hecho de que  $p$  es un fundamento del hecho de que  $p$  o  $q$ . En la concepción worldly, hemos dicho que ‘A’ y ‘A o A’ son el mismo hecho. Por lo tanto, el hecho de que A no puede ser un fundamento para el hecho de que A o A si la fundamentación es irreflexiva. Por el contrario, en la perspectiva representacional, dado que los dos hechos son diferentes, es posible mantener, como es canónico hacerlo (por ejemplo, Fine 2012), que el hecho de que A es un fundamento del hecho de que A o A.

Un factor motivante para la concepción no representacional, por ejemplo, podría ser la creencia de que calza mejor con una investigación sobre las ‘junturas’ de la realidad porque concierne a hechos gruesos (e.g., Correia, 2010, pp. 258–259). Nótese, sin embargo, que se pueden aceptar ambas concepciones (e.g., Correia, 2020) o cuestionar la seriedad de la distinción (deRosset, 2023). Advertimos a la lectora que quiera investigar más este tema que la discusión puede convertirse en una muy técnica puesto que involucra metafísica de los hechos y su individuación. Este es un ejemplo de cómo una teoría de la fundamentación trae consigo un ‘pack’ de compromisos.

#### 4.4 ¿Factiva o no factiva?

Otra distinción concierne a si la fundamentación es factiva o no factiva. En la concepción factiva, si un ítem A fundamenta un ítem B, entonces ambos A y B existen. Por ejemplo, A o B pueden ser afirmaciones verdaderas o hechos existentes (*obtaining*). Suponiendo que la existencia de las patas de la silla es un fundamento de la existencia de la silla, podríamos decir que es cierto, según la concepción factiva, que tanto las patas como la silla existen. Queda por determinar qué otras condiciones deben satisfacer algunos hechos para que exista una relación de fundamentación. El enfoque típico es introducir la relación de fundamentación como una afirmación justificada a través de sus beneficios teóricos. En las siguientes secciones, hablaremos más sobre estos temas.

En la postura no factiva, A puede fundamentar B incluso si A o B son afirmaciones falsas o hechos inexistentes. Por ejemplo, en la concepción no fáctica, el hecho de que el actual rey de Francia sea calvo podría ser motivo de que el actual rey de Francia sea calvo o de que Montevideo esté en Chile. La distinción puede parecer irrelevante. Pero su importancia puede resaltarse al considerar el elemento explicativo de la fundamentación (sección 2). Una lectura plausible podría decir que las conexiones explicativas deben ser factivas. Si esta afirmación es correcta, la postura sobre la factividad o no factividad de la explicación influencia la elección entre estos dos ‘sabores’ de fundamentación.

#### 4.5 ¿Orden Parcial Estricto?

La fundamentación está motivada por el deseo de sistematizar la estructura de la realidad. Recuerden el ejemplo de Sócrates y {Sócrates}. Es muy intuitivo pensar que es la existencia de Sócrates la que fundamenta la existencia de {Sócrates}, y no viceversa. Esta direccionalidad, como explicaremos en la sección 6, hace de la fundamentación una herramienta útil para analizar los conceptos de fundamentalidad y prioridad. Pero lo que permite a la fundamentación jugar este rol son algunas características estructurales.

En la lectura ortodoxa, la fundamentación induce un orden parcial estricto entre los ítems que gobierna (defensores de esta lectura son, por ejemplo, Schaffer, 2009; Rosen, 2010; Audi 2012; Fine, 2012; Raven, 2012). Decir que la fundamentación es un orden parcial estricto significa que impone relaciones que son irreflexivas, asimétricas, y transitivas.

Un argumento popular para avalar la lectura ortodoxa es que la fundamentación es explicativa. Las conexiones de la fundamentación son irreflexivas, asimétricas, y transitivas porque las conexiones explicativas tienen estas características. Hay dos posturas opuestas acerca del carácter explicativo de la fundamentación. El *Unionismo* sostiene que la fundamentación es sólo una forma de explicación (e.g., Raven, 2012, 2015). El *Separatismo* sostiene que la fundamentación es explicativa sin ser idéntica a la explicación (e.g., Schaffer, 2009). Una motivación diferente y menos popular para creer en las características ortodoxas concierne a la similitud con la causación: puesto que la causación es irreflexiva, asimétrica, y transitiva, también lo es la fundamentación (e.g., Schaffer, 2016; Wilson, 2018).

Las características estructurales de la lectura ortodoxa pueden cuestionarse de varias maneras. Alguien podría debilitar la idea de apelar a la explicación para justificarlas, al apuntar al difícilmente negable carácter subjetivo de la explicación, siendo que se supone que la fundamentación sea objetiva (e.g., Maurin, 2019). Por ejemplo, si Alberto no sabe qué son las bárbulas, una explicación de la diferencia entre plumas plumáceas y pennáceas en términos de bárbulas no es particularmente reveladora. Por el contrario, para María, que es una ornitóloga competente, la explicación es clara e informativa. Alguien más podría argüir que las explicaciones holísticas, como la explicación de la interdependencia de los valores de masa, volumen y densidad de un fluido homogéneo, requieren de una revisión del compromiso ortodoxo con la asimetría de la fundamentación (e.g., Thompson, 2018). Finalmente, se podría cuestionar la fiabilidad de usar a la causación como una guía para la fundamentación. Por ejemplo, se puede argumentar que la

supuesta diacronicidad de la causalidad, es decir, la idea de que los efectos son temporalmente posteriores a sus causas, la hace suficientemente diferente de la fundamentación, que normalmente se considera una relación co-temporal, como para poner en duda la utilidad de tomar la primera como guía para la segunda (e.g., Bernstein, 2016).

Cada característica estructural enfrenta, adicionalmente, más objeciones específicas. Aquí no podemos reconstruir los detalles. Pero podemos agregar que los argumentos comparten una estructura similar. Típicamente invocan una posibilidad metafísica que representa un contra ejemplo ya sea para la irreflexividad (e.g., Jenkins, 2011; Bliss, 2020), la asimetría (e.g., Wilson, 2014; Thompson, 2018), o la transitividad (e.g., Schaffer, 2012). Los contraejemplos involucran casos de dioses autodependientes, quarks mutuamente dependientes y fallas de transitividad que involucran propiedades distribucionales, como la de estar punteado. El debate, que permanece abierto (e.g., Wilson, próxima publicación), es si estos contraejemplos expresan posibilidades metafísicas.

¿Debiésemos adoptar la ortodoxia de la fundamentación? Como en otras áreas de indagación, es mejor, pensamos, mantener la mente abierta. Los casos metafísicos tenderán a seguir siendo controversiales. No obstante, algunos fenómenos naturales, que pueden plausiblemente entenderse en términos de fundamentación, representan un cuestionamiento más convincente contra la postura ortodoxa. Por ejemplo, en la teoría cuántica, la mutua dependencia no causal entre componentes entrelazados puede plausiblemente entenderse como un caso de fundamentación simétrica (Giannotti, 2021; cf. Ismael & Schaffer, 2020). En este caso, si es correcto, la fundamentación demuestra ser útil para mejorar nuestra comprensión de la naturaleza. La aceptación de instancias de fundamentación heterodoxa no nos obliga a abandonar la capacidad explicativa de esta noción. Sin embargo, sí hace más difícil identificar su fuente, ya que no puede derivarse de las características formales de la explicación. La tarea de explicar qué da cuenta

de la capacidad explicativa de la fundamentación sigue siendo un asunto no resuelto (e.g., Cameron, 2022 Rubenstein 2024).

#### 4.6 Total y parcial

Una distinción diferente pero ampliamente aceptada concierne a dos tipos de fundamentación: total y parcial. Intuitivamente, una fundamentación total es algo como una explicación completa de un cierto fenómeno o una causa completa de un cierto efecto. Una fundamentación parcial es algo como una explicación parcial o una causa contribuyente. Según una lectura estándar, podemos definir una fundamentación parcial en términos de su pertenencia a una fundamentación total (e.g., Fine 2012, p. 50). Este principio es llamado a veces ‘completabilidad’: A es un fundamento parcial de B si y sólo si hay una posiblemente vacía pluralidad de ítems  $\Gamma$  tal que  $\Gamma$  con A son un fundamento total de B.

Al adoptar el principio de completabilidad, podemos asumir sólo una noción primitiva de fundamentación total y analizar la variedad parcial en términos de ella. Sin embargo, el principio ha sido recientemente cuestionado por consideraciones metafísicas (Dixon, 2016; Leuenberger, 2020; Trogon & Witmer, 2021, Werner, 2021) y científicas (Giannotti, 2022).

Los cuestionamientos a la completabilidad de la fundamentación comparten una estructura común. Identifican un escenario metafísica o naturalmente posible en el cual algunos hechos están parcialmente fundamentados, pero les faltan fundamentos totales. Por ejemplo, Leuenberger (2020) analiza un supuesto escenario metafísicamente posible en el que algunas entidades pueden instanciar una propiedad “schmarge” positiva o negativa dependiendo de su constitución. El caso está construido de tal manera que es apropiado afirmar que el “schmarge” positivo o negativo de un objeto se fundamenta en la constitución del objeto. Sin embargo, no hay nada en este mundo posible que pueda explicar la polaridad específica. Leuenberger sostiene que la polaridad específica del “schmarge” está meramente parcialmente fundamentada y, dado que este caso es metafísicamente posible, deberíamos aceptar la noción

de meros fundamentos parciales. La idea de ítems que están meramente parcialmente fundamentados tiene profundas implicaciones concernientes a análisis de fundamentalidad basados en la fundamentación. Regresaremos a este tema en la sección 6.

## 5. Determinación

La lectora ha sido expuesta a las principales distinciones y sabores de la fundamentación. Resumiendo lo que se ha dicho hasta ahora, la fundamentación es una relación o conexión que posee varias características, la cual tiene un carácter determinativo y explicativo<sup>5</sup>.

Es el carácter determinado de la fundamentación, es decir, el hecho de que algunos hechos existan en virtud de otros, lo que se afirma que distingue esta relación de otras nociones metafísicas de dependencia. Esta característica supuestamente supera las limitaciones de otras dependencias no determinativas, y es usualmente promocionada como uno de los principales beneficios de agregar a la fundamentación a nuestro maletín de herramientas filosóficas. No es obvio, sin embargo, que otras relaciones no puedan poseer especificaciones análogas a las analizadas en la sección anterior. En consecuencia, el objetivo de esta sección es comparar la fundamentación con otros conceptos de dependencia con el objetivo de enfatizar su distinción.

### 5.1 Necesidad

Quienes proponen la fundamentación sostienen que las dependencias modales, es decir, en términos de una entidad necesariamente dependiente de otras, dejan sin responder muchas preguntas filosóficas importantes. Tomemos a la necesidad (*necessitation*), por

---

5 No discutiremos la relación entre fundamentación y *truthmaking*. Esta última es la relación entre un ítem capaz de ser verdadero, tal como una proposición, y algo más en virtud de lo cual el primer ítem es verdadero, así como un hecho que existe. La lectora interesada en esto puede revisar Correia & Schnieder (2012, pp. 25). Tampoco, exploramos la conexión entre fundamentación y causalidad. La lectora puede estudiar Schaffer (2016) y Wilson (2018).

ejemplo. Es necesario que  $\{\text{Sócrates}\}$  exista cuando Sócrates existe. Pero establecer una conexión modal no responde la pregunta sobre qué depende de qué. Deja fuera, por ejemplo, la dependencia intuitiva de  $\{\text{Sócrates}\}$  en Sócrates.

¿Cómo conecta la fundamentación con la necesidad? Deberíamos comenzar por mencionar que la fundamentación es más que necesidad. No hay una relación de fundamentación correspondiente a cada conexión de necesidad. Es necesario, por ejemplo, que, si Sócrates existe, entonces  $2+2=4$ . Pero la existencia de Sócrates no determina que  $2 + 2 = 4$ . Sin embargo, una lectura estándar dice que la fundamentación induce conexiones necesarias. Skiles (2020) llama a esta concepción de fundamentación ‘*necesitismo*’ (‘*necessitarianism*’). Para la lectura necesitista, toda vez que A fundamenta totalmente a B, es necesario que, si A existe (*obtains*), B también (un ejemplo de esta postura es Rosen, 2010). Una lectura que contrasta con esta perspectiva es, como lo llama Skiles (2020), el *contingentismo* (*contingentism*). Según esta postura, toda vez que A fundamenta totalmente a B, la existencia de A no necesita la existencia de B. Los argumentos a favor del contingentismo típicamente invocan casos de bloqueadores: ítems cuya presencia merma la necesaria conexión entre A y B a pesar de su conexión de fundamentación (e.g., Leuenberger, 2014a; Skiles, 2015). Por ejemplo, Leuenberger (2014) invoca el ejemplo de un “cromaplasma”, una propiedad que, si se ejemplifica, corta el vínculo entre un estado neuronal específico y la experiencia fenomenológica asociado con la percepción de un objeto rojo. En mundos posibles donde se instancia el “cromaplasma”, un sujeto que observe un objeto rojo estaría en el estado cerebral apropiado, pero la experiencia fenoménica estaría bloqueada.

¿La fundamentación es necesitista o contingentista? Shlomit Wygoda Cohen (2020) argumenta que la respuesta depende de si incluimos bloqueadores en la base de conexión a tierra completa. Para adaptar el ejemplo de Leuenberger, podríamos decir que la experiencia fenomenológica de un objeto rojo está totalmente fundada en el estado cerebral y la ausencia de cromaplasma. Sugerimos que



el debate entre necesidad y contingentismo es un debate sobre lo que debemos incluir en la base completa de hechos bajo estudio.

## 5.2 Superveniencia

Una dependencia famosa y familiar es la superveniencia. Una formulación inicial de superveniencia es esta: los ítems-A supervienen sobre los ítems-B sólo en caso de que los ítems-A no puedan variar sin que haya algunas variaciones en los ítems-B. Decir, por ejemplo, que las propiedades mentales supervienen sobre las propiedades físicas significa que no puede haber un cambio en las propiedades mentales sin un cambio en las propiedades físicas. Como con la necesidad, quienes proponen la fundamentación arguyen que la superveniencia no logra abordar importantes preguntas puesto que traza sólo una forma de covariación necesaria. Podemos aceptar que hay superveniencia entre la fragilidad de un jarrón y su estructura molecular. Pero la superveniencia no logra abordar el sentido en que la estructura molecular del jarrón determina su fragilidad.

La ortodoxia dice que la fundamentación es más que superveniencia. Pero estas nociones no son mutuamente excluyentes. Por ejemplo, alguien podría verse tentada a analizar la fundamentación en termino de superveniencia. Por ejemplo, McDaniel (2013) ha propuesto un análisis de la fundamentación en términos de superveniencia más la noción Lewisiana de naturalidad (*naturalness*) (Lewis, 1983). Según esta propuesta, reformulada ligeramente para que se ajuste a la discusión, un hecho  $f$  fundamenta un hecho  $g$  si y sólo si las propiedades involucradas en  $g$  supervienen sobre las propiedades involucradas en  $f$ , y las propiedades involucradas en  $f$  son más naturales que las involucradas en  $g$ <sup>6</sup>. Bricker (2016) propuso un análisis en términos de superveniencia más fundamentalidad, entendida en términos de perfecta naturalidad. Según este enfoque, si algunos hechos  $\Gamma$  son fundamentales (o perfectamente

---

6 Lewis (1983) sostiene que las propiedades naturales son simples, intrínsecas, y no manipuladas. Son las propiedades que ‘caracterizan a la naturaleza en sus juntas’ (propia traducción) y aparecen en una descripción completa y mínima de la realidad.

naturales) y  $g$  sobreviene sobre  $\Gamma$  entonces  $\Gamma$  fundamenta  $g$ . Como sostiene Kovacs (2020, pp. 349 – 350), ambas propuestas son problemáticas. Consideremos algunos hechos matemáticos necesarios, como el hecho de que 3 es un número impar. Tal hecho sobrevendría trivialmente sobre hechos que involucran propiedades naturales. Sin embargo, parece improbable sostener, por ejemplo, que el hecho de que un electrón tenga carga negativa sea una base del hecho de que tres es un número impar.

La fuente del problema es que la superveniencia es una noción más débil que la fundamentación: es una mera covariación necesaria entre dos elementos o grupos de elementos. Por el contrario, el carácter determinante de la fundamentación impone una restricción de pertinencia o relevancia entre los hechos que la fundamentan y aquellos que son fundamentados. Tal restricción, como establece la objeción, no se interpreta adecuadamente en términos de superveniencia. Nuestra sugerencia es considerar la fundamentación y la superveniencia como nociones complementarias en lugar de intentar reducir la primera a la segunda. La presencia de una relación de superveniencia es un indicador potencial, aunque falible, de la presencia de una conexión de fundamentación, lo que merece una investigación más profunda.

### 5.3. Dependencia Ontológica

Bajo el nombre de ‘dependencia ontológica’, hay una diversa familia de relaciones que abordan la idea de una cosa debiendo su existencia a la existencia de otras cosas. Por ejemplo, podemos decir que Batilde existe en virtud de que sus partes existen. Las dependencias ontológicas específicas varían según sus elementos o necesidad (Lowe, 1994; Koslicki, 2012). Aquí utilizaremos ‘dependencia ontológica’ como una etiqueta general y abordemos la pregunta sobre cómo se relacionan el fundamento y la dependencia ontológica. Una diferencia inicial entre estas nociones es que típicamente relacionan distintos tipos de ítems. La fundamentación típicamente ocurre entre hechos o verdades o proposiciones. En cambio, la dependencia ontológica generalmente involucra objetos y propieda-

des. Sin embargo, dado que es posible liberalizar la fundamentación y concebirla como una relación que involucra objetos y propiedades (por ejemplo, Schaffer 2009), la diferencia de relata es superficial.

Una manera distinta de delinear una distinción entre fundamentación y dependencia ontológica es destacando que esta última no necesita ser determinativa. En cambio, sugerimos que el aspecto determinativo de la fundamentación es plausiblemente constitutivo de la noción. Para explicar esta diferencia, podemos considerar una formulación clásica de dependencia ontológica en términos modales, es decir, en términos de coexistencia necesaria. Una cosa  $x$  depende ontológicamente de  $y$  solo en caso de que, si  $y$  no existiera,  $x$  tampoco hubiese existido (e.g., Lowe 1994). Consideremos ahora, por ejemplo, una postura que diga que una sustancia no puede existir sin ejemplificar algunas propiedades. Según la formulación anterior, la sustancia en cuestión sería ontológicamente dependiente de algunas propiedades. Sin embargo, tal dependencia no es suficiente para establecer que una sustancia está fundamentada en el sentido de estar determinada o hecha existir por las propiedades. Es coherente sostener, para la postura de este ejemplo, que nada determina o fundamenta las sustancias.

Podemos aún preguntarnos si podríamos definir la dependencia ontológica en términos de fundamentación. De hecho, las aproximaciones definicionales no son nuevas (e.g., Schnieder, 2004, 2006, 2020a, 2020b; Correia, 2005). Por ejemplo, podemos definir la dependencia ontológica en términos del fundamento de esta forma: Necesariamente, si una entidad  $x$  depende ontológicamente de una entidad  $y$ , entonces el hecho de que  $x$  exista se funda en algún hecho acerca de  $y$ . Como Schnieder (2020a, 2020b) explica, esta formulación inicial puede hacerse más rigurosa en términos formales. En lugar de sumergirnos en una discusión más técnica a cerca de como formalizar la definición anterior, consideremos una posible complicación con las posturas definicionales. Por ejemplo, Koslicki (2013, p. 47) y Orilia (2009, p. 334) sostienen que las cosas no están situadas en relaciones de dependencia ontológica porque hay una relación explicativa, como la fundamentación, entre

ellas. Sino que es al revés. Hay una relación explicativa entre ellas porque se sitúan en alguna forma de dependencia ontológica. Si la objeción es correcta, las aproximaciones definicionales necesitan algo de ajuste (ver, por ejemplo, Schnieder 2020a, pp. 115–119 para algunas formas de resistir la objeción).

Hay, además, un problema más profundo con el problema de definición que señalamos. Ya que la fundamentación es determinativa, lo adecuado de la definición resultante está limitado a formas determinativas o productivas de dependencia ontológica: tipos de dependencias donde algunas entidades son creadas o generadas por otras. Por ejemplo, Lowe (1994) analiza una forma de dependencia de la identidad en la que la identidad de una cosa depende de la identidad de otra. Tal vez la identidad del fresco de la Capilla Sixtina dependa de la identidad de Michelangelo. Sin embargo, en este caso parece incorrecto sostener que la identidad de Michelangelo produjo la identidad del fresco. Aunque es esperable que haya desacuerdos acerca de este ejemplo, el punto que queremos hacer debiese ser menos controversial: tener que ser cautas cuando nos encontramos con el término ‘dependencia ontológica’, ya que puede referirse a conexiones determinativas y no determinativas.

## 5.4 Esencia

El concepto de esencia es un actor central en la revolución post modal en filosofía analítica. Su resurgimiento moderno es ampliamente acreditado a Fine (1994, 1995), quien caracteriza la esencia de un objeto como aquello en virtud de lo cual podemos decir qué es ese objeto. La esencia es post modal en el sentido de que no puede entenderse adecuadamente en términos de necesidad—sostiene Fine. La propiedad de *ser humana* es parte de la esencia de Amalia. Pero afirmar que es necesario que Amalia tenga esta propiedad no lo captura, ya que hay muchas propiedades necesarias que no trazan características esenciales. Por ejemplo, es necesario que Amalia pertenezca a {Amalia}. Pero muchas personas compartirían la intuición de que pertenecer a su conjunto unitario no es parte de la esencia de Amalia.

Entender la relación entre fundamentación y esencia es quizás una de las tareas más importantes para avanzar en nuestra comprensión de ambas nociones. En la literatura, podemos identificar tres grupos heterogéneos de acercamientos: *esencia primero*, *fundamentación primero*, y *posturas unificadoras*. Posturas pertenecientes a cada familia difieren de manera significativa respecto de los detalles más finos. Aquí podemos sólo delinear a grandes rasgos esas diferencias.

La postura de la esencia primero (e.g., Rosen, 2010, 2017; Fine, 2012; Dasgupta, 2014) busca explicar existencia de conexiones de fundamentación en términos de la esencia de los ítems enlazados. Esquemáticamente, podríamos decir que toda vez que  $x$  fundamenta a  $y$ , hay algo sobre la esencia de  $x$  o  $y$  o ambas que explica el hecho de que  $x$  fundamenta a  $y$ . Estas aproximaciones tienen dos importantes ventajas. Primero, ofrecen una explicación sobre por qué algunas conexiones de fundamentación ocurren. Según este enfoque, si el hecho de que la pelota sea roja se fundamenta en el hecho de que es escarlata, esta conexión de fundamentación se explica por la esencia de uno de estos hechos (o sus constituyentes). Segundo, porque las esencias son plausiblemente capaces de dar la ‘última explicación’ en el sentido de que cuando apelamos a la esencia, hemos llegado al final de la ruta explicativa (e.g., Glazier, 2017). Pero, a pesar de estas ventajas, hay tres preocupaciones que surgen en torno a la postura de las esencias primero. Para comenzar, debiese aceptarse que las esencias son tanto metafísica como epistémicamente misteriosas. Es notoriamente difícil decir qué es una esencia de algún modo revelador. Y esta puede ser una razón para evitarla. Una segunda preocupación aborda a las posturas de esencia primero que no buscan reducir la fundamentación. Una segunda preocupación, en cambio, abarca a las posturas reductivas de esencia primero (e.g., Zylstra, 2019). Estas aproximaciones explican la fundamentación en términos de esencia. Algunas de las personas que proponen la fundamentación podrían argüir que reducir la fundamentación a una esencia no logra abordar el rol especial que la primera supuestamente debiese cumplir en nuestra teoriza-

ción. Por ejemplo, como explicamos en la próxima sección, la fundamentación tiene una conexión especial con la fundamentalidad y la prioridad. En cambio, no es obvio e incluso un error, según Kit Fine (2020), pensar que la esencia induce relaciones de prioridad. Por ejemplo, el hecho de que ser humana sea esencial para Amalia no la hace inmediatamente menos fundamental que la propiedad de ser humana.

Las lecturas de fundamentación primero buscan explicar la esencia de una cosa o verdades esenciales acerca de ella en términos de fundamentación (e.g., Correia, 2013; De Rizzo, 2022). Estas aproximaciones tienen el atractivo de expandir la fecundidad de la fundamentación. Sumado a comprender las dependencias determinativas, la fundamentación puede emplearse para analizar un concepto central para la filosofía occidental. Si se encontrara un análisis exitoso, éste reivindicaría la aplicabilidad de la fundamentación para dilucidar otros conceptos filosóficos. Lamentablemente, junto con las objeciones a propuestas específicas (e.g., Carnino, 2014), estas posturas atraen problemas similares a los que se presentan contra las posturas de esencia primero. Alguien que piensa que la esencia es metafísica y epistémicamente misteriosa probablemente sostendrá lo mismo respecto de la fundamentación. Y las lecturas reductivas de fundamentación primero pueden objetarse en virtud de que no ponen atención al especial rol explicativo de la esencia para capturar lo que hace que algo tenga la naturaleza que tiene.

Quizás, las posturas unificadoras son un camino intermedio más prometedor. Estas posturas apuntan a combinar la fundamentación y la esencia de maneras útiles sin establecer una prioridad explicativa entre ellas. Un ejemplo de esta forma es la lectura de Kit Fine (2015), según la cual la esencia y la fundamentación son condiciones metafísicas complementarias necesarias y suficientes. Según Fine (2015, p. 296), por ejemplo, por el hecho de que singularmente Sócrates es esencialmente un conjunto cuyo único miembro es Sócrates puede analizarse que es constitutivamente necesario que el conjunto sea uno cuyo único miembro es Sócrates. Y el hecho de que la existencia de Sócrates fundamenta la existencia

del singleton Sócrates puede interpretarse en el sentido de que es constitutivamente suficiente que si Sócrates existe, su singleton existe. En este enfoque, investigar la esencia y los fundamentos de ciertos elementos son formas complementarias de dar cuenta de la constitución o naturaleza de un elemento (Fine 2015, p. 311) que no se excluyen entre sí.

## 6. Fundamentalidad

Hasta ahora hemos distinguido las características de la fundamentación y contrastado su carácter determinista con otras nociones metafísicas. Captar fenómenos de determinación no causal es uno de los dos papeles teóricos principales que se afirma que desempeña la fundamentación en nuestra teorización. El segundo, que trataremos en esta sección, es el análisis de los conceptos de fundamentalidad y prioridad. En consecuencia, se afirma que la fundamentación es una herramienta útil para investigar una imagen popular de la realidad como si contuviera una jerarquía de hechos, donde algunos son más fundamentales que otros y algunos son absolutamente fundamentales.

En una primera aproximación, decir que algo es fundamental expresa la idea de que tal cosa es básica: no necesita que se dé cuenta de cualquier otra cosa para explicarla. Si la física va por buen camino, algunos hechos sobre campos o cuerdas son plausiblemente fundamentales. La prioridad es la relación que expresa que una cosa es más fundamental que otra: lo primero es más básico que lo segundo. Para utilizar otro ejemplo de la física, parece intuitivo que el hecho de que un electrón tenga carga negativa es más fundamental que el hecho de que un electrón tenga carga negativa o positiva. Aunque estos ejemplos pueden parecer intuitivos, el desafío consiste en especificar los conceptos de fundamentalidad y prioridad en términos informativos. Las defensoras de la fundamentación argumentan que esta última noción es útil para completar esta tarea (p. ej., Schaffer 2009). Debería quedar claro de inmediato que las definiciones de fundamentalidad y prioridad en términos de fundamentación están restringidas a la categoría de elementos a los que se aplica este último concepto. Así, por ejemplo, si uno piensa que la fundamentación

relaciona hechos, las definiciones resultantes serán para la fundamentalidad de hechos y la fundamentalidad de prioridades. Quienes defienden la fundamentación no necesitan negar que hay muchos otros tipos de fundamentalidad, como la fundamentalidad de objetos o la fundamentalidad de conceptos, que escapan a un análisis en términos de fundamentación (por ejemplo, Wilson 2014). Dado que hemos sugerido que la fundamentación se entiende mejor como la relación de hechos, también proponemos que las definiciones de fundamentalidad y prioridad en termino de fundamentación son útiles para una concepción fáctica de estas nociones. La presentación, sin embargo, puede ampliarse adecuadamente a otros aspectos de fundamentación.

La idea de que la fundamentación es un concepto útil para definir la fundamentalidad y la prioridad está asociada con la idea de que la realidad está compuesta por una jerarquía de hechos que tiene en su base una “capa” de hechos básicos que no se explican en términos de nada más. En un modelo típico, que refleja la idea de que los dominios científicos están igualmente organizados en niveles de dependencias explicativas (por ejemplo, Nagel 1961), algunos hechos microfísicos son fundamentales. Tales hechos fundamentan hechos físicos menos fundamentales, que a su vez fundamentan hechos químicos, y así para otros hechos menos fundamentales. Como era de esperar, tanto la tesis de que la realidad está organizada en cadenas prioritarias de hechos como la de que tiene algunos hechos fundamentales son controvertidas y requieren una defensa adicional (véase Bliss 2024 para una discusión accesible de lo que podría motivar estas tesis). Aquí centrémonos en las definiciones de fundamentalidad y prioridad en termino de fundamentación.

Una definición inicial de un hecho fundamental es que  $x$  es fundamental si y solo si no hay nada que fundamente  $x$  (p. ej., Schaffer 2009, p. 373). Esta definición captura la idea preteórica de que los hechos fundamentales no deben explicarse en términos de ningún otro. Complementariamente, una caracterización inicial de la prioridad fáctica es la siguiente: si  $x$  fundamenta  $y$ , entonces  $x$  es más fundamental que (o previa a)  $y$  y  $y$  es menos fundamental que  $x$  (cf.



Schaffer 2009, p. 373). Hay una conexión natural entre estas nociones. Los ítems fundamentados deben su existencia a sus fundamentos. Los ítems fundamentados son plausiblemente derivativos porque son constituidos, en algún sentido metafísico, en términos de sus fundamentos. En este sentido, los fundamentos son previos a lo que fundamentan. Una formulación inicial que conecta la fundamentación con la prioridad es esta: si  $x$  fundamenta  $y$ , entonces  $x$  es más fundamental que (o previa a)  $y$  e  $y$  es menos fundamental que  $x$  (cf. Schaffer 2009, p. 373). Esta formulación sólo enuncia una condición suficiente para la prioridad. Sin embargo, captura la idea de que lo que está fundamentado se genera  $y$ , por lo tanto, se explica en términos de sus fundamentos. Ahora vamos a discutir que ambas formulaciones sufren limitaciones importantes.

Empecemos con la idea de que ser fundamental es ser “ungrounded”: una cosa fundamental es algo que no es determinado por nada. Esta caracterización de la fundamentalidad como falta de fundamentación (o *ungroundedness* en inglés) intuitivamente expresa la idea de que los ítems fundamentales son los más básicos porque no puede darse cuenta de ellos en términos de nada más.

Lamentablemente, esta lectura simple es demasiado simple. No puede permitirnos utilizar la fundamentación para interpretar casos de entidades fundamentales pero mutuamente dependientes o de entidades fundamentales autodependientes. (e.g., Wilson 2014; Tahko 2023). Tal vez Dios es fundamental pero dependiente en sí mismo. A lo mejor algunos eventos fundamentales cuánticos son todos interdependientes entre sí. No se puede dar cuenta de esas posibilidades en términos de fundamentación. Una opción para defender la lectura simple es rechazar la posibilidad metafísica de entidades auto fundadas o mutuamente fundadas invocando a las características estructurales ortodoxas de la fundamentación, a saber, irreflexividad y asimetría. Pero esta respuesta puede traer una limitación importante al caracterizar lo fundamental en términos de no fundamentación, ya que un defensor de la fundamentación no necesita descartar desde el comienzo que los casos de entidades fundamentales auto dependientes o mutuamente dependientes

sean imposibles. Para considerar los casos problemáticos, podríamos adoptar una postura más sofisticada (o menos simple) y definir lo fundamental como sigue:  $x$  es fundamental si y sólo si para cada  $y$ , si  $y$  fundamenta  $x$ , entonces  $x$  fundamenta  $y$  (e.g., Bennett 2017, p. 136; Giannotti, 2020, p. 7; la definición debería permitir que  $x$  sea idéntica a  $y$ ). La lectura sofisticada puede acoger a entidades auto fundamentadas y mutuamente fundamentadas.

Alguien podría resistirse a la necesidad de complicar la definición de fundamentalidad en términos de fundamento, rechazando la posibilidad de entidades mutuamente dependientes. Sin embargo, esta estrategia, a falta de consideraciones auxiliares, cierra una posibilidad metafísica viva que no nos parece del todo inverosímil. Por tanto, la explicación sofisticada es una definición más prometedora.

La posibilidad de ítems meramente parcialmente fundamentados (sección 4.8) es una fuente distinta de problemas de hechos fundamentales como aquellos que carecen de fundamentos. Tales entidades están no fundamentadas, pero no completamente ¿Son fundamentales? Una definición de lo fundamental como no fundamentado es inadecuada para abordar el modo en que estas entidades son fundamentales en un sentido más débil que aquellas que están completamente no fundamentadas. Nuevamente, quien defienda la lectura simple podría rechazar la posibilidad de entidades meramente parcialmente fundamentadas. Sin embargo, casos metafísicos (e.g., Dixon, 2016; Leuenberger, 2020; Trogdon and Witmer 2021) y científicos (McKenzie, 2017; Giannotti 2022) nos dan razones para pensar que la idea de ítems que están sólo parcialmente fundamentados es tanto metafísica como nomológicamente (o naturalmente) posible. Estas consideraciones pretenden mostrar la falsedad de un principio de completabilidad, que permite definir la fundamentación parcial en término de fundamentación completa (4.8). Estamos de acuerdo en que es mejor renunciar a una definición de fundamentación parcial en términos de fundamentación plena que rechazar la posibilidad de ítems meramente parcialmente fundamentados. Las objetoras probablemente argumentarán que renunciar al principio de completitud es inaceptable. Sin embargo,

al hacerlo, no erosionamos nuestra comprensión de la fundamentación como una conexión determinante con fuerza explicativa y de necesidad. Es importante enfatizar, no obstante, que la teoría de la fundamentación por sí sola no establece que haya entidades ni fuerte ni débilmente fundamentales en nuestra realidad.

Ahora pasemos a analizar la limitación de caracterizar la prioridad en términos de fundamentación. Como dijimos, la formulación de que si  $x$  fundamenta  $y$ , entonces  $x$  es más fundamental que  $y$  ofrece sólo una condición suficiente para que  $x$  sea más fundamental que  $y$ . Encontrar una condición necesaria es mucho más complicado. Para ver por qué, encontraras inmediatamente una dificultad si definieras la prioridad como sigue:  $x$  es más fundamental que  $y$  si y sólo si  $x$  fundamenta  $y$ . Un electrón en Santiago plausiblemente es más fundamental que un pudú en Valdivia. Pero intuitivamente el electrón en Santiago no fundamenta el pudú en Valdivia. Además, en este caso, podríamos imponer varias cláusulas para mejorar la definición. Por ejemplo, Karen Bennett (2017, p. 161) ofrece esta propuesta refinada que aquí formulamos en términos de fundamentación<sup>7</sup>:

$x$  es más fundamental que  $y$  si y sólo si:

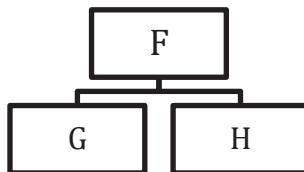
- 1)  $x$  está a menos pasos de distancia de la entidad(es) fundamental(es) que termina su cadena única que lo que  $y$  está de la entidad(es) fundamental(es) que termina su cadena única,
- 2) o  $x$  fundamenta al menos parcialmente a  $y$ ,
- 3) o  $x$  se sitúa en lo ancestral de una relación de fundamentación con  $y$ ,

---

<sup>7</sup> Bennett (2017) sostiene que esta definición cubre más relaciones ‘constructivas’, no sólo fundamentación. Relaciones constructivas incluyen composición, constitución, y realización—sólo por nombrar algunas. Esta formulación supone que existe una variedad de cadenas de dependencia. Sin embargo, como sostiene Correia (2021), la formulación de Bennett enfrenta algunas dificultades para dar cuenta de los casos en los que algunos elementos de las cadenas, aunque no todos, dependen de elementos de diferentes cadenas.

- 4) o  $x$  es absolutamente fundamental e  $y$  no lo es, o  $x$  pertenece a algún tipo  $K$  e  $y$  pertenece a un tipo  $K^*$  tal que
  - i. ni  $K$  ni  $K^*$  incluye tanto a miembros fundamentados como a miembros no fundamentados,
  - ii.  $y$  no pertenece a  $K$  y  $x$  no pertenece a  $K^*$ ,  $y$
  - iii.  $K^*$ s son típica y normalmente fundamentados en  $K$ s.

La complejidad de una definición puede ser un costo aceptable para alguien que quiere preservar un análisis de prioridad en términos de fundamentación. Pero como está, la formulación de Bennett enfrenta tres problemas (para una discusión más detallada, ver Correia, 2021b). Por lo tanto, debe ser mejorada. Primero, la definición de Bennett trata con cadenas. Pero hemos visto que un ítem puede estar fundamentado en muchas otras. Estas estructuras se entienden mejor como ‘árboles’ más que cadenas. Un árbol de fundamentación se ve como aparece en el diagrama de abajo, donde  $F$  está fundamentado en cada  $G$  y  $H$ :



Un ejemplo de un árbol de fundamentación puede ser el hecho de que Batilde es una filósofa esté fundamentado en varios hechos, tal como el hecho que Batilde trabaja como profesora en un departamento de filosofía, el hecho de que Batilde desarrolla investigaciones filosóficas, y muchos otros. Un segundo problema es que la formulación de Bennett presupone que la cadena de fundamentación termina. Pero puede ser deseable tener una formulación de fundamentalidad relativa que deje abierta la posibilidad de árboles de fundamentación interminables. Tercero, pero menos obvio, la idea de pasos de fundamentación se entiende plausiblemente en términos discretos en el sentido de que cada elemento de una ca-

dena de dependencia puede ordenarse numéricamente. Pero no es obvio que la fundamentación es discreta. ¿Deberíamos renunciar al prospecto de una definición de prioridad basada en la fundamentación? Sería extraño que la fundamentación nos sirva para un análisis de la fundamentalidad absoluta pero no para la prioridad. Deberíamos entonces explorar una versión mejorada de la definición de Bennett. Lamentablemente, no está claro cómo se puede mejorar la explicación de Bennett. Correia (2021) ha ofrecido una propuesta que define la fundamentalidad relativa de algunos hechos analizando dónde aparece este hecho en una jerarquía de tipos de hechos inducida por la fundamentación. Se parte de las relaciones de fundamentación entre tipos de hechos para recuperar el estatus de fundamentalidad de un hecho objetivo (ya sea absolutamente fundamental, menos o más fundamental o tan fundamental como otro hecho). Sin embargo, como podemos observar en la definición de Bennett, existe otra noción de fundamentalidad relativa. Se trata de la idea de que el estatus de fundamentalidad de un hecho está determinado, en términos de teoría de la fundamentación, por la distancia que lo separa de los hechos no fundamentados o fundamentales. No es obvio si estos dos tipos de fundamentalidad relativa pueden unificarse. Resolver este problema sigue siendo una tarea para una explicación de la fundamentalidad basada en la fundamentación.

## 7. Escepticismo y desunión

Hemos sugerido que abordar la determinación y ayudar a entender mejor la fundamentalidad absoluta/relativa son algunas de las mayores ventajas teóricas de la fundamentación. En esta sección delineamos dos grupos de objeciones dirigidas a la utilidad de esta noción. El primer grupo de objeciones sostiene que la fundamentación es muy poco específica para ser útil. El segundo grupo apunta al supuesto poder unificador de la fundamentación, es decir, la idea de que una variedad de fenómenos de dependencia determinativa no causal puede entenderse en términos de fundamentación. Ambas objeciones de ninguna manera alcanzan a agotar los argumentos existentes en contra de la fundamentación. Por razones de espacio

no vamos a decir mucho, por ejemplo, sobre el escepticismo de la ‘vieja escuela’ respecto de la fundamentación. Según las escépticas de la vieja escuela, la fundamentación es ‘oscura’ o ininteligible (e.g., Howfber, 2009; Daly, 2012; ver Kovacs, 2023 para más información sobre escepticismo en torno a la fundamentación). Puesto que ya hemos discutido algunas objeciones contra un análisis de fundamentalidad y prioridad basada en la fundamentación, no vamos a ahondar en temas de esa área. Invitamos a las lectoras interesadas, no obstante, a revisar las críticas seminales e influyentes de Jessica Wilson (2014) y Kathrine Koslicki (2015) sobre fundamentación que contienen detalles adicionales y más objeciones.

Las objetoras del primer grupo arguyen que la fundamentación y afirmaciones respecto de qué fundamenta que son demasiado genéricas para ser informativas o explicativas. Decir que lo físico fundamenta lo mental, diría la objetora de este primer grupo, no revela algo esclarecedor acerca de la dependencia de lo mental sobre lo físico. Wilson (2014) sostiene que lo que esclarece la dependencia en cuestión son una plétora de relaciones de fundamentación ‘pequeñas-f’ más específicas (‘small-g’ en inglés), que incluyen, ‘tipo de identidad, *token* pero no *type* identidad, realización funcional, la clásica relación mereológica parte-todo, la relación de composición causal, la relación de membresía de conjunto, la relación de subconjunto, y la relación determinable-determinado, entre otras’ (Wilson, 2014, p. 539). Este tipo de objeción es apremiante: si la fundamentación no es útil para esclarecer dependencia de interés, pero otras relaciones ‘pequeñas-f’ hacen un mejor trabajo, no tenemos motivos para promoverla siquiera.

Una respuesta prometedora es rechazar la rivalidad entre fundamentación y relaciones ‘pequeñas-f’. Una podría argüir que la fundamentación y las dependencias más específicas juegan diferentes roles, los que no obstante son complementarios. Ya hemos presentado los roles de abordar determinaciones y elucidar prioridad. No es obvio que las dependencias ‘pequeñas -f’ puedan hacer lo mismo. Alternativamente, una podría sostener que es útil tener una noción más ‘gruesa’ (no específica) de fundamentación como

punto de partida para investigar fenómenos que muestran una conexión determinativa pero no puramente modal. La especificación hacia una dependencia más ‘depurada’ de dependencia es un asunto subsecuente (e.g., Nolan, *forthcoming*). Para usar el ejemplo de Nolan (*forthcoming*), es útil afirmar que fumar causa cáncer incluso en ausencia de la identificación del mecanismo específico que conecta a la nicotina en el torrente sanguíneo con el cáncer, y el descubrimiento de esto último no merma la utilidad de lo primero en promover preguntarse por lo último. Aunque los detalles de esta respuesta aún deben desplegarse en detalle, la respuesta es inicialmente atrayente y ello nos permite ser optimistas respecto de su éxito. Una respuesta diferente a las objeciones del primer grupo concierne al poder unificador de la fundamentación, el cual abordamos a continuación.

Promover la fundamentación se justifica a veces a través de su poder para unificar superficialmente dependencias heterogéneas no-causales bajo una sola noción. Unificacionistas sostienen que lo que explica las similitudes estructurales compartidas entre muchas relaciones pequeñas-*f* es que son idénticas a la fundamentación o que son especificaciones de ella (Schaffer, 2009; Rosen, 2010; Audi, 2012 pueden considerarse unificacionistas).

Las objetoras niegan que podamos obtener claridad al juntar dependencias específicas y explicativamente autónomas. Contra la postura unificacionista, hay varios ‘sabores’ de pluralismo acerca de la fundamentación. Algunas como Wilson (2014), sostienen que hay muchas dependencias que están, finalmente, desunificadas. Otras, como Bennett (2017), arguyen que las muchas dependencias comparten algunos parecidos de familia pero que están en última instancia desconectadas. Algo interesante es que incluso algunas defensoras de la fundamentación adoptan una versión de pluralismo. Por ejemplo, Fine (2012) cree que la fundamentación está en juego en varios campos investigativos que involucran explicaciones determinativas no-causales, incluyendo ciencia y ética. Pero Fine no cree que el tipo de fundamentación amada y querida por (algunas) metafísicas es el mismo tipo de fundamentación empleada por

científicas y éticas. La razón, según Fine, es que diferentes tipos de fundamentación se asocian con diferentes tipos de modalidades últimamente irreducible la una a otra: metafísica, normativa, natural. De este modo, deberíamos distinguir entre fundamentación metafísica, fundamentación normativa, y fundamentación natural.

Una estrategia de responder a la objeción contra los poderes unificadores de la fundamentación es sostener que muchas dependencias ‘pequeñas-f’ comparten las mismas características estructurales de fundamentación, y que esta es una razón para pensar que están unificadas (e.g., Berker 2018). Una respuesta diferente e intrigante es argüir que el reconocimiento de una fundamentación unificadora no merma el rol explicativo especial de las dependencias más específicas. Una manera sugerente de pensar acerca de la relación entre fundamentación y las otras dependencias ‘pequeñas-f’ en analogía con la relación género-especie (*genus-species*). La fundamentación es como un género; las dependencias ‘pequeñas-f’ son como especies (propuestas en esa línea pueden encontrarse en Schaffer, 2016; Bennett, 2017; Kortabarria, 2023). Aunque evocativa, los detalles de la postura género-especie siguen estando poco explorados. Por ahora, hay trabajo pendiente para defender los méritos de la fundamentación como un unificador metafísico. En particular, debería demostrarse cómo conectar a la fundamentación con dependencias más específicas de maneras interesantes, y debería darse una explicación precisa sobre qué ventajas nos trae la unificación. Estas tareas son desafiantes, pero no imposible.

## 8. El futuro de la fundamentación

Llegando al final de este capítulo, decidimos reservar un espacio para sugerir dos principales áreas de investigación que nos parece ofrecen un terreno prometedor, sin embargo, subestimado para la aplicabilidad de la fundamentación, particularmente para elucidar posiciones, conceptos, e hipótesis. Un área es la metafísica social y la filosofía de las ciencias sociales. Y la otra es la metafísica de la ciencia y la filosofía de la ciencia.



La idea de que la fundamentación puede elucidar ciertos aspectos de la ciencia tiene pedigrí antiguo en la filosofía de occidente. Podemos encontrarla en el trabajo de Aristóteles y culminando en la teoría de la ciencia de Bolzano (ver Roski y Schnieder 2022). Por ejemplo, McKenzie (2020) sostiene que la fundamentación más que la dependencia ontológica no determinativa es una herramienta prometedora para elucidar la relación entre estructuras de simetría y campos fermiónicos y bosónicos. Alastair Wilson (2022) plantea que la fundamentación puede ayudarnos a dar sentido a la fundamentalidad y los niveles en mecánica cuántica Everetiana. Más recientemente, Kortabarría y Giannotti (2024) propusieron que la fundamentación puede utilizarse para elucidar explicaciones determinativas, pero no causales, de relevancia científica tales como la explicación topológica, explicaciones de entrelazamiento cuántico, y explicaciones de equilibrio. Sin embargo persisten las dudas respecto de si la metafísica de la fundamentación es suficientemente maleable como para ser de ayuda para mejorar nuestra comprensión de teorías científicas, particularmente de física cuántica (e.g., McKenzie 2023). De manera más general, se requiere más trabajo para establecer que la fundamentación mejora nuestra comprensión de las dependencias no causales de interés científico.

En áreas conectadas con la metafísica social, podemos encontrar propuestas que analizan conceptos socialmente relevantes, tales como la raza y el género, en términos de fundamentación, particularmente en el contexto del construccionismo social (Schaffer 2017; Griffith 2018). Según esta lectura, las propiedades como *ser mujer* o *ser latino/a* dependen de varios factores sociales, prácticas, y normas, en lugar de, digamos, ocurrencias únicamente biológicas. No obstante, hay desafíos importantes respecto de la productividad de la fundamentación para esta área que se mantienen. Aquí un ejemplo. Elizabeth Barnes (2014) ha sostenido que discusiones en torno a la naturaleza del género no concierne a sus fundamentos. En cambio, concierne a lo que el género *es* en lugar de *lo que hace que sea el caso de que* alguien sea una instancia de cierto género. Por ejemplo, tanto Haslanger (2000) como Jenkins (2016) están

de acuerdo en que el género de alguien está fundamentado en jerarquías sociales e interpersonales. Pero Jenkins arguye que deberíamos permitir una dimensión de identidad conectada al género, y tal dimensión no se entiende, captura o explica adecuadamente en términos de sus fundamentos, ya que se refiere a cómo un individuo se relaciona con un grupo de género (cf. Passinsky, 2021). Creemos que la aplicabilidad de la fundamentación se beneficiaría de una exposición más clara sobre cómo podría ayudar a las metas meliorativas que muchas indagaciones en estas áreas socialmente relevantes comparten. Tanto Haslanger como Jenkins, para continuar con el ejemplo, están embarcadas en proyectos que apuntan a desacreditar conceptos que son utilizados para oprimir y discriminar a grupos sociales específicos ¿Cómo puede ayudar la fundamentación, si es que es de algún modo capaz, a alcanzar las metas de estos proyectos meliorativos? Dejamos esta pregunta como un punto de discusión para la lectora.

Cerramos en una nota optimista. La fundamentación es una herramienta bastante flexible. Podemos modificarla y ‘afinarla’ para que sirva a nuestras metas explicativas. Mientras más la usamos, más podemos aprender cómo nos beneficia en nuestras prácticas explicativas. No hay una razón lo suficientemente fuerte para pensar que futuras investigaciones no vindiquen la utilidad de la fundamentación para explorar dominios de investigación científicos y sociales.

## Referencias

- Amijee, F. (2023). Inquiry and metaphysical rationalism. *Australasian Journal of Philosophy*, 101 (4), 809 – 823.
- Audi, P. (2012). Grounding: Toward a Theory of the In-Virtue-Of Relation. *Journal of Philosophy*, 109 (12), 685-711.
- Barnes, E. (2014). Going Beyond the Fundamental: Feminism in Contemporary Metaphysics. *Proceedings of the Aristotelian Society*, 114 (3pt3), 335-351.
- Bennett, K. (2017). *Making Things Up*. Oxford University Press.
- Bliss, R. (2020). Fundamentality. In M. J. Raven (ed) *The Routledge Handbook of Metaphysical Grounding* (pp. 336 – 347). Routledge.

- Bernstein, S. (2016). Grounding is not Causation. *Philosophical Perspectives*, 30 (1), 21–38.
- Berto, F. & Nolan, D. (2023). Hyperintensionality. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/hyperintensionality/>
- Berker, S. (2018). The unity of grounding. *Mind*, 127, 729–77
- Bricker, P. (2006). The Relation between General and Particular: Entailment vs. Supervenience. *Oxford Studies in Metaphysics*, 2, 251–287.
- Bolzano, B. (1985ff/1837). Wissenschaftslehre. In J. Berg & E. Winter (Eds.), *Bernard Bolzano Gesamtausgabe*, Reihe 1, Bd. 11–14. Stuttgart-Bad Cannstatt: Frommann-Holzboog.
- Cameron, R. P. (2022). *Chains of Being*. Oxford: Oxford University Press.
- Carnino, P. (2014). On the Reduction of Grounding to Essence. *Studia Philosophica Estonica*, 7 (2), 56–71.
- Cohen, Shlomit Wygoda (2018). Not All Partial Grounds Partly Ground: Some Useful Distinctions in the Theory of Grounding. *Philosophy and Phenomenological Research*, 100 (1), 75–92.
- Correia, F. (2010). Grounding and Truth-Functions. *Logique et Analyse*, 53(211), 251–279.
- Correia, F. (2017). Real Definitions. *Philosophical Issues*, 27(1), 52–73.
- Correia, F. (2020). Granularity. In M. J. Raven (ed) *The Routledge Handbook of Metaphysical Grounding* (pp. 228 – 243). Routledge.
- Correia, F. (2021). Fundamentality from grounding trees. *Synthese* 199 (3-4), 5965–5994.
- Correia, F., & Schnieder, B. (2012). Grounding: An opinionated introduction. In B. Schnieder & F. Correia (Eds.), *Metaphysical grounding* (pp. 1–36). Cambridge University Press.
- Daly, C. (2012). Scepticism about grounding. In B. Schnieder & F. Correia (Eds.), *Metaphysical grounding* (pp. 81 – 100). Cambridge University Press.
- Dasgupta, S. (2014). On the Plurality of Grounds. *Philosopher's Imprint*, 14(20), 1–28.
- DeRosset, L. (2023). *Fundamental Things*. Oxford: Oxford University Press.
- De Rizzo, J. (2022). A ground-theoretical modal definition of essence. *Analysis* 82 (1), 32–41.

- Dixon, T. S. (2016). Grounding and Supplementation. *Erkenntnis* 81 (2), 375-389.
- Fine, K. (1994). Essence and Modality. *Philosophical Perspectives*, 8, 1-16.
- Fine, K. (1995). Ontological Dependence. *Proceedings of the Aristotelian Society*, 95, 269-290.
- Fine K. (2001). The question of realism. *Philosophers' Imprint*, 1, 1-30.
- Fine, K. (2012). Guide to Ground. In B. Schnieder & F. Correia (Eds.), *Metaphysical grounding* (pp. 37 – 80). Cambridge University Press.
- Fine, K. (2015). Unified Foundations of Essence and Ground. *Journal of the American Philosophical Association*, 1, 296-311.
- Fine, K. (2020). Comments on Jessica Wilson's "Essence and Dependence". In M. Dumitru (ed.) *Metaphysics, Meaning, and Modality: Themes from Kit Fine* (pp. 471 – 475). New York: Oxford University Press.
- Giannotti, J. (2021). Fundamental Yet Grounded. *Theoria*, 87 (3), 578-599.
- Giannotti, J. (2022). Brutalist fundamentalism: radical and moderate. *Synthese*, 200 (3), 1-19.
- Glazier, M. (2017) Essentialist Explanation. *Philosophical Studies*, 174, 2871-89.
- Griffith, A. (2018) "Social Construction and Grounding," *Philosophy and Phenomenological Research* 97, 393-409
- Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Noûs* 34, 31-55.
- Hofweber, T. (2009). Ambitious, Yet Modest, Metaphysics. In D. Chalmers, D. Manley, & R. Wasserman (eds.) *Metametaphysics: New Essays on the Foundations of Ontology* (pp. 260 – 289). Oxford University Press.
- Ismael, J., and Schaffer, J. (2020). Quantum Holism: Nonseparability as Common Ground *Synthese*, 197 (10), 4131-4160
- Jenkins, C. (2011). Is Metaphysical Dependence Irreflexive? *The Monist*, 94 (2), 267-276.
- Jenkins, K. (2016). Amelioration and Inclusion: Gender Identity and the Concept of Woman. *Ethics*, 126 (2), 394-421.
- Kortabarra, M. (2023). A Defense on the Usefulness of 'Big-G' Grounding. *Metaphysica: International Journal for Ontology and Metaphysics*, 24 (1), 147-174.

- Kortabarria, M. & Giannotti, J. (2024). Scientific Explanation as a Guide to Ground. *Synthese*, 203 (3), 1-27.
- Koslicki, K. (2012). Varieties of Ontological Dependence. In B. Schnieder & F. Correia (Eds.), *Metaphysical grounding* (pp. 186 – 213). Cambridge University Press.
- Koslicki, K. (2013). Ontological Dependence: An Opinionated Survey. In B. Schnieder, M. Hoeltje & A. Steinberg (eds.), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence* (pp. 31-64.). Munich: Philosophia Verlag.
- Koslicki, K. (2015). The Coarse-Grainedness of Grounding. *Oxford Studies in Metaphysics*, 9, 306-344.
- Koslicki, K. (2020). Skeptical doubts. In M. J. Raven (ed) *The Routledge Handbook of Metaphysical Grounding* (pp. 164 – 179). Routledge.
- Kovacs, D. M. (2023). Varieties of Grounding Skepticism. *The Monist*, 106 (3), 301-316.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Li, Chenyang & Perkins, F. (eds.) (2015). *Chinese Metaphysics and its Problems*. Cambridge University Press.
- Litland, J. E. (2017). Grounding grounding. *Oxford Studies in Metaphysics*, 10, 279 – 316.
- Lewis, D. K. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61, 343–77
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell.
- Leuenberger, S. (2014). “Grounding and Necessity”, *Inquiry: An Interdisciplinary Journal of Philosophy*, 57, 151–74
- Leuenberger, S. (2020). The fundamental: Ungrounded or all-grounding? *Philosophical Studies*, 177, (9):2647-2669.
- Lowe, E. J. (1994). Ontological Dependency. *Philosophical Papers*, 23 (1), 31-48.
- MacBride, F. & Janssen-Lauret, F. (2022). Why Lewis Would Have Rejected Grounding. In H. Beebe & A. R. J. Fisher (eds.), *Perspectives on the Philosophy of David K Lewis* (pp. 66 – 91). Oxford University Press.
- McDaniel, K. (2013). Degrees of Being. *Philosophers’ Imprint*, 13 (19): 1–18.

- McKenzie, K. (2017). Against Brute Fundamentalism. *Dialectica*, 71 (2), 231-261.
- McKenzie, K. (2020). Structuralism in the Idiom of Determination. *British Journal for the Philosophy of Science*, 71 (2), 497-522.
- McKenzie, K. (2023). *Fundamentality and Grounding*. Cambridge University Press.
- Maurin, A-S. (2019). Grounding and metaphysical explanation: it's complicated. *Philosophical Studies*, 176 (6), 1573-1594.
- Nagel, E. (1961). *The Structure of Science. Problems in the Logic of Explanation*. New York: Harcourt, Brace & World, Inc.
- Nolan, Daniel (forthcoming). Grounding, Explanation, and the Tasks of Metaphysics. In Aaron Segal & Nick Stang (eds.), *Systematic Metaphysics: Historical and Contemporary Perspectives*. Oxford University Press.
- Orilia, F. (2009). Bradley's Regress and Ungrounded Dependence Chains. *Dialectica* 63, 333–341.
- Passinsky, A. (2021). Finean Feminist Metaphysics. *Inquiry*, 64 (9), 937-954.
- Perkins, F. (2023) Metaphysics in Chinese Philosophy. The Stanford Encyclopedia of Philosophy (<https://plato.stanford.edu/archives/win2023/entries/chinese-metaphysics>)
- Poggiolesi, F. (2020). Logics. In M. J. Raven (ed.) *The Routledge Handbook of Metaphysical Grounding* (pp. 213 – 227).
- Priest, G. (2018). Buddhist Dependence. In R. Bliss & his (eds.) *Reality and its structure: essays in fundamentality* (pp. 126 – 140). Oxford University Press.
- Raven, M. J. (2012). In Defence of Ground. *Australasian Journal of Philosophy*, 90, 687–701.
- Raven, M. J. (2015). Ground. *Philosophy Compass* 10 (5), 322-333.
- Raven, M. J. (2020) (ed.) *The Routledge Handbook of Metaphysical Grounding*. Routledge.
- Reutlinger, A. & Saatsi, J. (eds.) (2018). *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford University Press.
- Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In B. Hale & A. Hoffman (eds.), *Modality: Metaphysics, Logic, and Epistemology* (pp. 109–135). Oxford University Press.
- Rosen, G. (2017). Ground by Law. *Philosophical Issues*, 27 (1), 279-301.

- Roski, St. & Schnieder, B. (eds.) (2022). *Bolzano's philosophy of grounding: translations and studies*. Oxford University Press.
- Rubenstein, Ezra (2024). Two approaches to metaphysical explanation. *Noûs*, 58 (4), 1107-1136.
- Schaffer, J. (2009). On What Grounds What. In D. Chalmers, D. Manley & R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology* (pp. 347–383). Oxford University Press.
- Schaffer, J. (2012). Grounding, transitivity, and contrastivity. In F Correia & B Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality* (pp. 1122 – 138). Cambridge University Press.
- Schaffer, J (2016). Grounding in the image of causation. *Philosophical Studies*, 173 (1), 49-100.
- Schaffer, Jonathan (2017). Social construction as grounding; or: fundamentality for feminists, a reply to Barnes and Mikkola. *Philosophical Studies*, 174 (10), 2449-2465.
- Schnieder, B. (2004). A Note on Particularized Qualities and Bearer-Uniqueness. *Ratio*, 17, 218–228.
- Schnieder, B. (2006). Truth-making without truth-makers. *Synthese*, 152, 21–46.
- Schnieder, B. (2010). A Logic for 'Because'. *Review of Symbolic Logic*, 4(3), 445–465.
- Schnieder, B. (2020a). Grounding and dependence. *Synthese* 197 (1),95-124.
- Schnieder, B, (2020b) Dependence. In M. J. Raven (ed.) *The Routledge Handbook of Metaphysical Grounding* (pp. 107 – 120). Routledge.
- Skiles, A. (2015). Against Grounding Necessitarianism. *Erkenntnis*, 80, 717–51
- Skiles, A. (2020). Necessity. In M. J. Raven (ed.) *The Routledge Handbook of Metaphysical Grounding* (pp. 147 – 163). Routledge.
- Tahko, T. (2023). Fundamentality. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/fundamentality/>
- Thompson, N. (2018). Metaphysical Interdependence, Epistemic Coherentism, and Holistic Explanation. In R. Bliss & G. Priest (eds.) *Reality and its structure: essays in fundamentality* (pp. 107 – 126). Oxford University Press.

- Trogon, Kelly & Witmer, D. Gene (2021). Full and partial grounding. *Journal of the American Philosophical Association*, 7 (2):252-271.
- Werner, J. (2021). Arbitrary grounding. *Philosophical Studies*, 179 (3), 911-931.
- Wilson, A. (2018). *Metaphysical Causation*. *Noûs*, 52 (4), 723-751.
- Wilson, A. (2022). Fundamentality and Levels in Everettian Quantum Mechanics. In Valia Allori (ed.), *Quantum Mechanics and Fundamentality* (pp. 397-415). Springer.
- Wilson, J. M. (2014). No Work for a Theory of Grounding. *Inquiry: An Interdisciplinary Journal of Philosophy* 57 (5-6):535-579.
- Zylstra, J. (2019). The essence of grounding. *Synthese*, 196 (12), 5137-5152.



## CAPITULO II

### LOS MÚLTIPLES ROSTROS DEL DETERMINISMO

*Hernán Accorinti<sup>1</sup> y Olimpia Lombardi<sup>2</sup>*

#### RESUMEN

El problema del determinismo es una de las cuestiones tradicionales de la metafísica occidental, que hace su aparición ya en los orígenes de la reflexión filosófica de Occidente, y recorre la historia de la filosofía hasta la Modernidad, cuando comienza a considerarse en términos científicos. Es así que con la física newtoniana se instala la idea de una realidad determinista y mecanicista, la visión del

---

1 Hernán Accorinti es doctor en Filosofía por la Universidad de Buenos Aires, e investigador del CONICET.

2 Olimpia Lombardi es doctora en Filosofía por la Universidad de Buenos Aires (UBA), investigadora superior del CONICET y directora del Grupo de Filosofía de las Ciencias en el Instituto de Filosofía de la UBA. Especialista en filosofía de la física y de la química, es autora de más de 150 publicaciones internacionales y forma parte del cuerpo editorial de revistas como *Foundations of Physics* y *Philosophy of Physics*.

universo como un enorme mecanismo de relojería cuyos estados se siguen inexorablemente a partir del estado inicial, en el instante de la creación. Esta visión se mantiene hasta la irrupción de la mecánica cuántica a principios del s. XX.

Sin embargo, a través de tan largo recorrido, la noción de determinismo no siempre fue concebida del mismo modo, y a veces fue asociada con nociones emparentadas pero diferentes. Esta situación ha conducido a desacuerdos como consecuencia de discrepancias meramente terminológicas en lugar de ser resultado de divergencias verdaderamente filosóficas. Por este motivo, en el presente trabajo abordaremos una elucidación detallada del concepto de determinismo, distinguiendo en primer lugar sus sentidos matemático, epistémico y ontológico, para luego diferenciarlo de nociones asociadas como legalismo y fatalismo. A continuación, se señalará la estrecha relación entre el modo de concebir el determinismo y la interpretación de la posibilidad. Esto permitirá discutir los nexos que usualmente se establecen entre determinismo y predictibilidad, así como los presupuestos reductivos que subyacen al debate sobre el determinismo. Finalmente, se presentará una distinción muy pocas veces señalada en la bibliografía sobre el tema, entre determinismo individual y determinismo universal.

**Palabras clave:** determinismo, metafísica occidental, legalismo, fatalismo.

## ABSTRACT

The problem of determinism is one of the traditional questions of Western metaphysics, which makes its appearance at the very origins of Western philosophical reflection, and runs through the history of philosophy until the Modern era, when it begins to be considered in scientific terms. Thus, with Newtonian physics, the idea of a deterministic and mechanistic reality is established, the vision of the universe as an enormous clockwork mechanism whose states inexorably follow from the initial state, at the instant of creation. This vision is maintained until the emergence of quantum mechanics at the beginning of the 20th century.

However, throughout such a long journey, the notion of determinism was not always conceived in the same way, and was sometimes associated with related but different notions. This situation has led to disagreements as a consequence of merely terminological discrepancies rather than as a result of truly philosophical divergences. For this reason, in the present work we will address a detailed elucidation of the concept of determinism, distinguishing first its mathematical, epistemic, and ontological senses, and then differentiating it from associated notions such as legalism and fatalism. Next, the close relationship between the way of conceiving determinism and the interpretation of possibility will be pointed out. This will allow for a discussion of the links that are usually established between determinism and predictability, as well as the reductive presuppositions that underlie the debate on determinism. Finally, a distinction rarely noted in the literature on the subject will be presented, between individual determinism and universal determinism.

**Keywords:** determinism, Western metaphysics, legalism, fatalism.

## 1. Introducción

El problema del determinismo es una de las cuestiones tradicionales de la metafísica occidental, que hace su aparición ya en los orígenes de la reflexión filosófica de Occidente, y recorre la historia de la filosofía. Este problema se expresa en múltiples planos, desde un plano moral cuando la pregunta por la determinación de las acciones humanas se supedita a la pregunta por el libre albedrío y la responsabilidad ética y jurídica que tiene un individuo toda vez que este se encuentra constreñido por un concatenamiento de acontecimientos que sobrepasa su estricta individualidad, hasta un plano religioso donde se busca una suerte de síntesis entre un Dios todopoderoso que todo lo sabe y todo lo puede y el creyente que, desde la misma doctrina, necesita de una autonomía en sus acciones que ameriten la razón misma del juicio final. Como se señalará en la Sección 3, estrictamente es en la Modernidad que este problema

empieza a ser abordado y respondido en términos estrictamente científicos en autores como Leibniz y Laplace. Es así que con la física newtoniana se instala la idea de una realidad determinista y mecanicista, la visión del universo como un enorme mecanismo de relojería cuyos estados se siguen inexorablemente a partir del estado inicial, en el instante de la creación. Esta visión se mantiene hasta la irrupción de la mecánica cuántica a principios del s. XX.

Sin embargo, a través de tan largo recorrido, puede observarse que quizás la razón por la cual esta discusión continúa generando fuertes controversias no se debe tanto a que el problema puede abordarse desde los diferentes planos mencionados, sino a que la noción misma de determinismo no siempre fue concebida del mismo modo. De hecho, como veremos, a veces fue asociada con nociones emparentadas pero diferentes. Esta situación ha conducido a desacuerdos como consecuencia de discrepancias meramente terminológicas, en lugar de ser resultado de divergencias verdaderamente filosóficas. Es decir, si bien sería impropio adscribir a la tesis de tinte wittgensteiniana según la cual “la mayor parte de los interrogantes y proposiciones de los filósofos estriba en nuestra falta de comprensión de nuestra lógica lingüística” (Wittgenstein, 1921, p. 66), haciendo de este modo que las discusiones filosóficas resulten superfluas producto de una incomprensión del lenguaje, confiamos que en este caso una elucidación del concepto que permita diferenciar sus sentidos y tejer el tipo de vínculo existente entre el determinismo y otras nociones usualmente emparentadas pero no idénticas, permitirá echar luz a la discusión acotando el alcance y el sentido estricto que entraña el determinismo. Por este motivo, en el presente trabajo en primer lugar haremos un breve recorrido histórico que permita identificar los orígenes de esta longeva discusión. En segundo lugar, identificaremos los problemas que ha tenido la tradición para caracterizar de un modo claro y no controversial la tesis principal del determinismo. Luego, como un intento de superación de este problema comenzaremos plenamente con el trabajo de corte netamente analítico distinguiendo, en una primera instancia, sus sentidos matemático, epis-

témico y ontológico; para luego, en una segunda instancia, diferenciar la noción de determinismo de otras nociones asociadas como lo son el legalismo, el fatalismo, la causalidad y la predictibilidad. Finalmente, se presentará una distinción, muy pocas veces señalada en la bibliografía sobre el tema, entre determinismo individual y determinismo universal, que permitirá poner de manifiesto la posibilidad de determinismo individual en todos los sistemas del universo, pero sin determinismo universal.

## 2. El determinismo en la historia del pensamiento

En tanto cuestión metafísica por excelencia, el problema del determinismo hace su aparición ya desde los orígenes de la reflexión filosófica en la Grecia clásica. Una de las primeras formulaciones de una tesis de tipo determinista se encuentra en los atomistas griegos: según Leucipo, “ninguna cosa sucede sin razón, sino que todas acaecen por una razón y por necesidad” (citado en Mondolfo 1942 [1983], p. 115). En este sentido, el atomismo clásico encuentra cierta resonancia en la tradición mítica y literaria griega: la trágica historia de Edipo es uno de los mejores ejemplos de la idea de un destino necesario e inexorable del cual el hombre no puede escapar.

Tal vez haya sido esta consecuencia fatalista el aspecto que motivó el profundo rechazo de Aristóteles hacia la doctrina del atomismo. Si bien admitía la legalidad de la naturaleza en su devenir de la potencia al acto, Aristóteles no aceptaba que los hechos se encontraran predeterminados. Es precisamente esta tesis la que funciona como premisa en su famoso argumento acerca de los futuros contingentes: si en todo par de enunciados acerca del futuro, uno es la negación del otro, uno de ellos es verdadero y el otro es falso, entonces todo lo que sucede, sucede por necesidad; pero no todo lo que sucede, sucede por necesidad, pues hay hechos contingentes; por lo tanto, no puede afirmarse la validez de la Ley del Tercero Excluido para los enunciados acerca del futuro:

Quiero decir, por ejemplo, que es necesario (ἀνάγκη) que vaya a haber (μὲν ἐνέσθαι) una batalla naval mañana o no la vaya a haber (μὴ ἐνέσθαι),

pero ni tendrá lugar mañana una batalla naval necesariamente (*ἀνάγκαιον*) ni no tendrá lugar necesariamente, aunque necesariamente tendrá lugar o no tendrá lugar. (Aristóteles 1984, p. 18 b15-25).

Si bien el epicureísmo adoptó la doctrina atomista de Demócrito y Leucipo, rechazó su estricto determinismo: en su caída vertical, los átomos experimentan una desviación espontánea e intrínseca que Epicuro (1995) denominaba *parenklisis*. En su poema *De Rerum Natura*, Lucrecio se refirió a tal desviación como *clinamen*, adjudicándole dos objetivos centrales: permitir el choque entre los átomos y, con ello, la formación de los objetos físicos, y destruir la oprimente visión de una fatalidad ineluctable, haciendo posible la libertad humana (García Gual, 1981, p. 110-113). Casi contemporáneo al epicureísmo, el estoicismo se le enfrentó en múltiples aspectos, y la cuestión del modo de determinación del futuro por el pasado no fue la excepción. Los estoicos recuperaron, así, la tradición griega del fatalismo, postulando la concatenación causal de todos los eventos del universo:

Los sucesos anteriores son causa de aquéllos que les siguen, y en esta manera todas las cosas van ligadas unas con otras, y así no sucede cosa alguna en el mundo que no sea enteramente consecuencia de aquélla y ligada a la misma como a su causa [...]. Si pudiese haber un hombre que percibiera el encadenamiento de todas las causas, nada podría engañarle. Porque quien abarca las causas de los hechos futuros, es necesario que abarque todo cuanto habrá de ser. (palabras de un antiguo estoico anónimo, citadas en Long 1975 [1984], p. 163).

Durante la Edad Media, el problema del determinismo se insertó en la discusión teológica, encarnándose en la doctrina de la *predestinatio*: la predestinación se sigue de un modo natural de la omnisciencia divina. Desde San Agustín, el problema que desveló a los teólogos medievales fue el de compatibilizar la predestinación con la libertad humana, lo cual resultaba imprescindible para justificar la responsabilidad del hombre y, con ello, la dimensión ética del Juicio Final.

Con la Modernidad, el problema de reconciliar una metafísica determinista —consecuencia de los poderes absolutos de Dios— con la posibilidad del libre albedrío no desapareció, si bien se trasladó del plano teológico al filosófico. Es con este propósito que Spinoza, al tiempo que sostuvo un fuerte determinismo, reformuló el concepto de libertad: ser libre consiste fundamentalmente en seguir la propia naturaleza, en tanto que esta naturaleza se encuentra en estrecha relación con el resto de lo real. En otras palabras, para Spinoza (1980) necesidad y libertad no son excluyentes sino que se complementan, dado que la libertad no se entiende como libre albedrío sino como una autodeterminación entendida en términos de aquello que se realiza en función de su propia naturaleza. Por su parte, Leibniz intentó compatibilizar la libertad humana con el determinismo moral resultado de la absoluta bondad de Dios distinguiendo entre necesidad absoluta y necesidad hipotética: la necesidad absoluta es de orden lógico y, por tanto, su negación es imposible; pero algo puede ser hipotéticamente necesario en relación con la suprema bondad divina y su elección del mejor de los mundos posibles. Según Leibniz, las acciones de los hombres son necesarias sólo en el sentido de que existe una razón suficiente para ellas; pero esto es consistente con la libertad, puesto que no son necesarias en sentido absoluto (Leibniz, 1914).

Pero fue con Galileo y con la mecánica moderna que la cuestión del determinismo se instaló específicamente en la filosofía de la naturaleza, inaugurando la tendencia —que perdura hasta nuestros días a través de lo que se denomina epistemología naturalizada (Giere, 1999)— según la cual los filósofos adoptan la teoría científica fundamental de su época como guía para su creencia en la verdad o falsedad del determinismo. La nueva concepción determinista, basada en la física, no hizo más que traducir el viejo problema de la compatibilidad entre predestinación y libre albedrío a la oposición entre determinismo físico y libertad. Por ejemplo, Hobbes (1998) adhería a un determinismo mecánico que incluía al ser humano: el comportamiento del hombre es comportamiento de materia y, por tanto, debe ser explicado mediante los mismos

principios generales que se aplican a la materia. No obstante, consideraba el determinismo consistente con la libertad humana en la medida en que definía libertad como ausencia de coerciones o restricciones externas para la acción, algo que incluso los seres inanimados pueden poseer (Hobbes, 1998).

Finalmente, cuando logró imponerse la física newtoniana, se instala de un modo definitivo la idea de una realidad determinista y mecanicista, la visión del universo como un enorme mecanismo de relojería cuyos estados se siguen inexorablemente a partir del estado inicial, en el instante de la creación (Leibniz, 1990; Laplace, 1814). Kant, firme defensor de la mecánica de Newton (Kant, 2004), admite no obstante que la tesis determinista sólo es aplicable al mundo de los fenómenos, pero no al mundo nouménico de lo moral, al cual pertenece la libertad humana. El tradicional conflicto entre libertad y determinismo, formulado en la Tercera Antinomia, es un conflicto aparente: el hombre es libre, no porque pueda evadirse de los nexos causales que lo involucran, sino porque no es enteramente una realidad natural y, en consecuencia, puede introducir en el mundo el inicio de nuevas cadenas causales (Kant, 2003).

Diversos autores coinciden en considerar que el siglo XIX marca el apogeo de la concepción determinista (ver, por ejemplo, Février, 1955; Jauch, 1973,; Butterfield, 1998): este constituye el marco metafísico dentro del cual todo científico intenta organizar sus conceptos teóricos. En esta época, la mayor parte del público cultivado considera la mecánica clásica como la mejor teoría física, e incluso como un fundamento teórico no revisable: la idea central es que, con su teoría, Newton ha formulado el esquema definitivo para la explicación del mundo físico. Tal esquema se resume en la Segunda Ley de Newton, según la cual la fuerza ejercida sobre un cuerpo es igual al producto entre su masa y su aceleración: así, conocida la masa del cuerpo y la fuerza que actúa sobre él, puede calcularse el movimiento del cuerpo a partir de su estado inicial. En resumen, formular la explicación física de un fenómeno consiste en ajustarlo a este esquema, fijando las fuerzas que intervienen; el



caso paradigmático es el de la gravitación, donde el propio Newton propuso la forma de la fuerza responsable de los fenómenos gravitatorios y calculó, con asombrosa precisión, las trayectorias de planetas y otros cuerpos celestes.

Pero la misma época que presencia el apogeo de la concepción determinista produce los gérmenes de su declinación. Según el historiador de la ciencia Stephen Brush (1976), la confianza en el determinismo comienza a mermar ya en la segunda mitad del siglo XIX, con la irrupción de los métodos estadísticos en la física a través de la teoría cinética de los gases. No obstante, en general se sigue considerando que la probabilidad cumple exclusivamente la función de permitir el tratamiento de sistemas muy complejos y con un enorme número de grados de libertad: la estadística se concibe como expresión de la ignorancia acerca del proceso perfectamente determinista que sigue un sinnúmero de elementos inobservables. Esta interpretación epistémica de la probabilidad es la que se encuentra implícita en el párrafo que abre el Capítulo V de *El Origen de las Especies*:

Hasta aquí he hablado a veces como si las variaciones [...] fuesen debidas a la casualidad. Esto, por supuesto, es una expresión completamente incorrecta, pero sirve para reconocer llanamente nuestra ignorancia de la causa de cada variación particular. (Darwin, 1859, p. 189).

El golpe de gracia a la visión determinista del universo sobreviene con la irrupción de la mecánica cuántica y las dificultades para ser interpretada en términos de los esquemas clásicos. A partir de ese momento se produce un cambio radical en la actitud de científicos y filósofos hacia el determinismo: de ser considerado paradigma de cientificidad, el determinismo pasa a convertirse en el ingrediente central de una cosmovisión ampliamente superada. René Thom resume claramente este nuevo clima intelectual:

Una escuela de epistemología contemporánea pretende que el mundo de las antiguas verdades de la ciencia clásica ha muerto, que el determinismo «laplaciano» *ha vivido* ya demasiado, y que se abre la era de

una nueva ciencia en la cual podrá florecer la innovación, en la cual el libre albedrío del hombre podrá expansionarse a su gusto lejos de las restricciones obsoletas del determinismo mecanicista. (Wagensberg, 1992, p. 63).

Sin embargo, la decadencia circunstancial de un concepto no implica su intrascendencia y aun cuando el determinismo no monopoliza el espíritu de la época como otrora supo hacer, la pregunta respecto del modo en que están concatenados los acontecimientos perdura. Es más, dado que, como en cualquier antagonismo, es en la contraposición donde los opuestos se identifican como tales, para comprender los alcances del indeterminismo imperante desde el apogeo de la mecánica cuántica, resulta imprescindible realizar una elucidación de los límites conceptuales que subyacen al determinismo. Precisamente eso es lo que se hará en los siguientes apartados.

### **3. El núcleo de la idea de determinismo**

A través de su larga historia, la doctrina determinista fue concebida desde muy diferentes perspectivas y defendida sobre la base de los argumentos más diversos. Algunos han adherido al determinismo como consecuencia de sus posturas acerca de la libertad humana o de los poderes divinos. Otros hallaron nexos significativos entre el determinismo y cuestiones lógicas inherentes al tiempo. Pero, desde el siglo XVII, el determinismo encontró su fundamento teórico en el desarrollo de la mecánica clásica, cuyo éxito llevó al apogeo la cosmovisión determinista de corte mecanicista.

No obstante este supuesto compartido acerca del carácter determinista de lo real, la idea no logró plasmarse en una definición única y precisa; por el contrario, la tesis determinista intentó expresarse, con mayor o menor éxito, a través de múltiples definiciones en las cuales el determinismo solía asociarse a otras nociones como la de causalidad universal —según la cual todo evento tiene una causa— o la de predictibilidad en principio —que sostiene que, dado el conocimiento del presente y de las leyes, es posible predecir todo el futuro—. De esta multiplicidad de caracteriza-

ciones, tal vez la que mejor captura el núcleo central de la idea de determinismo sea la formulada por William James:

¿Qué sostiene el determinismo? Sostiene que aquellas partes del universo ya establecidas, fijan y sentencian absolutamente lo que serán las demás partes. El futuro no tiene posibilidades ocultas en su seno: la parte que llamamos presente es compatible sólo con una totalidad. Cualquier otro complemento futuro que el fijado desde la eternidad es imposible. (James, 1897, p. 150).

El tono poético de esta caracterización no minimiza sus ventajas: su amplitud le permite adecuarse a las más diversas ontologías, sin presuponer el carácter material o mecanicista de lo real; además, prescinde totalmente de nociones emparentadas con el determinismo, cuya elucidación introduciría dificultades adicionales. Con esas pocas palabras, James logra recoger la idea básica que subyace a toda tesis determinista: el futuro no se encuentra abierto a la posibilidad; el presente fija unívocamente el devenir futuro.

No es este, sin embargo, el fragmento que acude a la mente cuando se habla de determinismo. Por el contrario, la definición incansablemente repetida en toda discusión en torno al problema del determinismo es la formulada por Laplace en la introducción de su famoso *Essai Philosophique sur les Probabilités*:

Así pues, hemos de considerar el estado actual del universo como el efecto de su estado previo y como la causa del que ha de seguirle. Una inteligencia que en un momento determinado conociera todas las fuerzas que animan a la naturaleza, así como la situación respectiva de los seres que la componen, si además fuera lo suficientemente amplia como para someter a análisis tales datos, podría abarcar en una sola fórmula los movimientos de los cuerpos más grandes del universo y los del átomo más ligero; nada le resultaría incierto y tanto el futuro como el pasado estarían presentes ante sus ojos. (Laplace, 1814, p. 25).

Curiosamente, mucho menos conocido es un pasaje donde Leibniz, mucho antes que Laplace, expresa su convicción determinista a través de un recurso muy similar:

Todo acaece matemáticamente, esto es, infaliblemente, en todo el ancho mundo, de suerte que, si alguien pudiese tener una percepción suficiente de las partes inferiores de las cosas y tuviese bastante memoria y entendimiento para captar todas las circunstancias y tenerlas en cuenta, sería un profeta y vería lo futuro en lo presente, como en un espejo. (Leibniz, 1990, p. 14).

Si bien sobre la base de una concepción metafísica matematicista, Leibniz también expresa la idea de un ser capaz de poseer un conocimiento perfecto de la evolución determinista del universo.

Desde su formulación, el célebre pasaje de Laplace ha recibido tantas interpretaciones como autores lo han citado, desde las más burdas, que acusan a Laplace de haberse jactado de su propia habilidad para reconstruir la historia universal (Crutchfield et al., 1987, p. 16), hasta las más sofisticadas, como la de Ernst Cassirer, quien reflexiona acerca de las capacidades que debería poseer la inteligencia laplaciana para que nada le resultara incierto. Según Cassirer, la idea misma del ser mítico que imagina Laplace encierra dos nociones de conocimiento mutuamente incompatibles: la noción de conocimiento intuitivo, necesario para determinar el estado presente del universo con precisión absoluta, y la noción de conocimiento discursivo que, en su grado más alto de perfección, permite calcular todos los estados futuros de la historia universal (Cassirer, 1936, p. 9). De aquí Cassirer concluye que la fórmula laplaciana no constituye una adecuada expresión de la tesis determinista, puesto que conduce a un inexorable *ignorabimus* al cual debe resignarse el conocimiento científico.

Pero, ¿es realmente el conocimiento de lo real aquello a lo que Laplace se refiere? Si bien el texto asocia determinismo y predictibilidad, dependiendo de qué capacidades posea la superinteligencia laplaciana se obtienen diferentes sentidos de la noción de predictibilidad y, con ello, de la noción de determinismo. Incluso, algunas interpretaciones pueden desviar de la idea central de determinismo: si el ser laplaciano poseyera poderes divinos, podría conocer el futuro de un modo intuitivo mediante su precognición, aun en un universo carente de toda regularidad; por otra parte, si

su capacidad de cálculo se identificara con la noción matemática de computabilidad, entonces podría suceder que los estados futuros resultaran no computables —no calculables de un modo efectivo, es decir, mediante un algoritmo que en un número finito de pasos efectúa el cálculo (Barry Cooper, 2004)— en un universo que, no obstante, cumpliera con el núcleo central de la idea de determinismo. La tensión señalada por Cassirer entre conocimiento intuitivo y discursivo es una manifestación más de las dificultades derivadas del intento de precisar la capacidad predictiva de la superinteligencia laplaciana.

Todos estos inconvenientes tienden a desaparecer si se abandona el plano epistémico y se interpreta el texto de Laplace en términos ontológicos. Un primer motivo para ello son las palabras del propio autor inmediatamente a continuación del célebre pasaje:

El espíritu humano ofrece, en la perfección que ha sabido dar a la astronomía, un débil esbozo de esta inteligencia [...]. Todos sus esfuerzos por buscar la verdad tienden a aproximarlos continuamente a la inteligencia que acabamos de imaginar, pero de la que siempre permanecerá infinitamente alejado. (Laplace, 1814, pp. 25-26).

Con su ser dotado de poderes esencialmente suprahumanos, Laplace no alude al conocimiento alcanzable por el hombre sino a una predictibilidad “en principio”, basada en la conexión unívoca entre los sucesivos estados del universo. Pero cuando se traslada al plano epistémico para referirse al máximo conocimiento que el ser humano puede alcanzar, de inmediato introduce el cálculo de probabilidades: allí la probabilidad es interpretada como medida de la ignorancia acerca de la evolución precisa que describen los sistemas reales, regidos por leyes deterministas. En otras palabras, la superinteligencia laplaciana cumple el papel de recurso retórico o, quizás, de *Gedankenexperiment* diseñado deliberadamente para esclarecer el contenido de una noción perteneciente al plano ontológico. Esta interpretación, que combina determinismo —en un sentido ontológico— e impredecibilidad, recibe nuevo apoyo en otro pasaje de la misma obra acerca de las probabilidades:

La curva descrita por una simple molécula de aire o de vapor está determinada de una forma tan exacta como las órbitas de los planetas. Entre ellos no hay más diferencia que la derivada de nuestra ignorancia. La probabilidad es relativa en parte a esta ignorancia y en parte a nuestros conocimientos.” (Laplace, 1814, p. 27).

Aquí Laplace abandona toda consideración acerca de la predictibilidad, sea humana o divina, para referirse al carácter determinista de toda evolución física en tanto regida por las leyes newtonianas del movimiento. En este devenir de lo real, donde cada estado se encuentra rígidamente determinado por el estado previo, la probabilidad no mide un grado de indeterminación sino que expresa la diferencia entre el conocimiento humano y la realidad determinista.

Los inconvenientes para alcanzar una definición aceptable de determinismo persistieron aún mucho tiempo después de los primeros intentos de Leibniz y Laplace. Incluso en el siglo XX muchas elucidaciones continuaron apoyándose en el concepto de predictibilidad. Así, según Louis de Broglie, la tesis determinista es

la creencia de que el mundo físico es una inmensa máquina cuya evolución está inexorablemente determinada, de tal modo que un conocimiento exacto de su estado actual deberá permitir la previsión de todos sus estados futuros. (de Broglie, 1939, p. 262).

Tal vez consciente de la posibilidad de predicciones estadísticas, Max Planck (1948) matiza la caracterización anterior exigiendo que los eventos puedan ser predichos con certeza. Desde una perspectiva diferente, quienes intentan definir el concepto de determinismo evitando aludir a la predictibilidad a menudo terminan recurriendo a la no menos conflictiva noción de determinación; este es el caso de Percy Bridgman, quien afirma:

[...] por determinismo entendemos la creencia de que el futuro del universo completo, o de una parte aislada de él, está determinada en términos de una descripción completa de su condición presente. (Bridgman, 1927, p. 209).

Bertrand Russell es aún más cuidadoso en evitar toda circularidad sin caer en cuestiones epistémicas relativas a la predictibilidad:

Una «ley causal», según uso el término, puede definirse como un principio general en virtud del cual, con suficientes datos acerca de ciertas regiones del espacio-tiempo, es posible inferir algo sobre ciertas otras regiones del espacio-tiempo. (Russell, 1948, p. 317).

Aquí Russell funda su caracterización de ley causal (que aquí denominamos ley determinista) en la noción lógica de inferencia. Sin embargo, si se incluye la posibilidad de efectuar inferencias estadísticas a partir de leyes causales, en la caracterización de Russell el concepto de causalidad/determinismo colapsa con el de legalidad natural: toda ley natural sería determinista.

Una cuestión que se agrega al difícil problema de hallar una caracterización adecuada del concepto de determinismo es la que se refiere al modo en que se lo debe interpretar, esto es, como una tesis metafísica o como una hipótesis científica. Tradicionalmente el determinismo fue considerado una tesis metafísica: científicos y filósofos no habrían tenido inconveniente alguno en admitir ingredientes puramente ontológicos en su pensamiento. Pero la situación se modificó durante el siglo XX con la gran influencia del positivismo lógico y su declarado rechazo hacia la metafísica: el determinismo se convirtió, así, en una hipótesis científica que debía someterse al estricto tribunal de los criterios positivos de la ciencia. En su conocida obra acerca del determinismo, Paulette Février es un claro exponente de esta nueva perspectiva:

[...] es importante por consiguiente llegar a establecer criterios positivos que permitan aislar una concepción del determinismo por la cual se haga este principio verificable en el nivel de la experiencia. (Février, 1955, p. 19).

Incluso el propio Karl Popper, a pesar de su explícito enfrentamiento con las tesis del Círculo de Viena, no logra escapar a su influencia cuando intenta sustituir la doctrina metafísica del determinismo

por el concepto preciso y testeable de determinismo científico basado, una vez más, en la noción de predictibilidad (Popper, 1982). Pero en la actualidad ya no somos prisioneros de las desmesuradamente rígidas pretensiones positivistas: durante las últimas décadas, muchos autores (por ejemplo, Earman, 1986; 2004; Belot, 1995; Wüthrich, 2011) han vuelto a investigar el problema del determinismo en el plano ontológico y, de este modo, la discusión ha recordado su cauce filosófico original.

#### 4. Los múltiples sentidos de determinismo

En filosofía, es frecuente que las discusiones se enturbien debido a los diferentes sentidos que adoptan los términos involucrados en la argumentación. Por este motivo, resulta conveniente comenzar por algunas consideraciones terminológicas que allanarán el camino de la ulterior investigación.

En su uso técnico, el término ‘determinación’ adopta al menos dos sentidos claramente diferentes:

- *Determinación como posesión de propiedades definidas.* Este sentido es de uso corriente en filosofía: esto es lo que “*determinatio*” significa en latín y así se emplea el término en varios idiomas europeos, especialmente el alemán (*Determination*). En esta acepción, determinado es aquello que tiene propiedades definidas y, por tanto, puede ser caracterizado de un modo inequívoco.
- *Determinación como conexión constante y unívoca.* Este es un sentido de uso corriente en ciencia: conexión constante y unívoca entre eventos o entre objetos reales o ideales.

Respecto del problema del determinismo, no es el primer sentido de determinación el que entra en juego. El concepto de determinismo encierra la idea de conexión constante y unívoca, pero integra un elemento adicional: el tiempo. La idea básica de determinismo, aquella que James intentaba recoger con su caracterización, apunta a un aspecto dinámico de la realidad, esto es, la sucesión temporal unívoca de los eventos que conforman la historia del universo.



Si bien estas consideraciones permiten ir acotando la noción de determinismo, aún no conducen a un sentido único y preciso. En efecto, puede comprobarse que el término ‘determinista’ se aplica a entidades muy diversas, adoptando así diferentes acepciones:

- En un primer sentido, que denominaremos *matemático*, el predicado ‘determinista’ se aplica a ecuaciones dinámicas: se dice que una ecuación dinámica es determinista cuando, dado el valor de las variables dependientes en un cierto instante, fija unívocamente el valor de dichas variables en todo instante posterior.
- En un segundo sentido, que denominaremos *epistémico*, el predicado ‘determinista’ se aplica al conocimiento acerca de un sistema real: se dice que poseemos un conocimiento determinista acerca de un sistema cuando el conocimiento del estado de dicho sistema en un instante dado permite conocer unívocamente su estado en todo instante posterior. Sin embargo, expresado de este modo, el sentido epistémico del predicado ‘determinista’ resulta demasiado exigente: la práctica científica señala que la determinación empírica del estado de un sistema en un cierto instante brinda, para cada variable de estado, no sólo un dado valor sino un inevitable error que depende de la precisión del instrumento de medición utilizado; además, la Teoría de Propagación de Errores permite calcular la evolución de las imprecisiones iniciales con el transcurso del tiempo. En consecuencia, a fin de evitar su vacuidad parece conveniente debilitar el sentido epistémico según la siguiente versión: se dice que poseemos un conocimiento determinista acerca de un sistema cuando el conocimiento de su estado en un dado instante permite conocer unívocamente su estado en todo instante posterior dentro de un margen de error acotado.
- En un tercer sentido, que denominaremos *ontológico*, el predicado ‘determinista’ se aplica a los sistemas reales, sean físicos, biológicos, sociales, etc. Se dice que un sis-

tema es determinista cuando, dado su estado en un cierto instante, su evolución para todo instante posterior resulta físicamente —no lógicamente— necesaria. En otras palabras, si el sistema se encuentra en el estado  $e_1$  en el instante  $t_1$ , las leyes —interpretadas como regularidades ontológicas— hacen imposible que se encuentre en un estado diferente de  $e_2$  en  $t_2$ ; tales evoluciones también se consideran ontológicamente deterministas.

Según cuál sea la acepción del término ‘determinista’ que se adopte, surgen los diferentes sentidos con los que suele interpretarse la noción de determinismo en la discusión filosófica:

- *Determinismo metodológico*: posición metodológica que considera que la tarea de la ciencia consiste en alcanzar descripciones matemáticamente deterministas de los fenómenos reales, esto es, descripciones basadas en ecuaciones deterministas, pues tal tipo de ecuaciones suministra la máxima información acerca de los estados futuros de los sistemas reales.
- *Determinismo epistémico*: posición gnoseológica que considera que es posible alcanzar un conocimiento determinista de todos los sistemas reales —sea en la versión fuerte o débil del predicado ‘determinista’ en su sentido epistémico—.
- *Determinismo ontológico*: posición metafísica según la cual todos los sistemas reales son ontológicamente deterministas. Desde una perspectiva totalizadora, el universo entero es concebido como un sistema determinista. Quienes adoptan esta posición tienden a interpretar aquellas leyes científicas que se expresan bajo la forma de ecuaciones deterministas como regularidades que, inscriptas en el plano ontológico, rigen el comportamiento de los sistemas reales.

Si se reflexiona acerca de estas definiciones provisionales a la luz de la historia de la filosofía, puede comprobarse que, salvo en épocas recientes, el problema del determinismo se formuló en términos metafísicos: cuando James se refería a un futuro sin posibilidades ocultas en su seno no aludía a cuestiones metodológicas o epistémicas, sino a las características de la realidad misma (véanse también las posiciones de Leibniz y Laplace comentadas en la sección anterior). En otras palabras, históricamente el núcleo central de la noción de determinismo apunta a una idea referida al plano ontológico (por ejemplo, Tomás de Aquino 1952-1965, Spinoza 1980). Los sentidos metodológico y epistémico son subsidiarios respecto del núcleo metafísico del concepto: sólo ante las dificultades para lograr una caracterización precisa de la noción de determinismo, ingresan a la discusión consideraciones respecto del conocimiento y de la metodología de la ciencia.

Un análisis superficial del problema permitiría suponer que, sobre la base de las distinciones formuladas, su solución no presenta dificultades: un sistema real es ontológicamente determinista si las ecuaciones dinámicas que describen su evolución temporal son matemáticamente deterministas, y esto último se comprueba examinando las características formales de tales ecuaciones. Sin embargo, la solución dista de ser tan sencilla: estas primeras definiciones del concepto de determinismo en sus diferentes sentidos brindan un punto de partida para la investigación pero, a su vez, abren un nuevo abanico de cuestiones.

Con el giro lingüístico que experimentó la filosofía durante el siglo XX, los conceptos que tradicionalmente pertenecían al ámbito de la metafísica fueron abordados desde la perspectiva del análisis del lenguaje: en este sentido, la noción de determinismo no fue la excepción. Para muchos autores contemporáneos, la clave para formular una adecuada definición del concepto reside en admitir que el determinismo es una propiedad de teorías (ver, por ejemplo, Dear 1961; Butterfield 1998). Sobre la base de la tradicional identificación entre teoría y conjunto deductivamente cerrado de enunciados de un lenguaje formal, ya Ernest Nagel adoptaba una

perspectiva similar cuando intentaba una caracterización puramente lógico-formal de determinismo:

Supongamos que se ha establecido un conjunto de enunciados generales  $L$  tal que, dado el estado de  $S$  en algún instante inicial, con ayuda de  $L$  puede deducirse un estado único de  $S$  en algún otro instante. Por consiguiente, en principio es posible calcular el estado de  $S$  para cualquier instante, dados  $L$  y el estado de  $S$  para algún instante inicial. (Nagel, 1961, p. 261).

En otras palabras, una teoría  $T$  es determinista si, dada la descripción de estado inicial  $s(t_1)$  en el instante  $t_1$ , la descripción de estado  $s(t_2)$  en cualquier instante posterior  $t_2$  es lógicamente deducible de  $T$ . Si bien aparentemente inobjetable, esta perspectiva lógico-formal es rechazada por John Earman (1971) y Richard Montague (1974): una descripción de estado es un enunciado del lenguaje de  $T$ ; pero mientras los estados de un sistema pueden formar un conjunto no-numerable, los lenguajes estándar contienen sólo un infinito numerable de enunciados. Por ejemplo, en mecánica clásica el estado de un sistema es definido por el par ordenado  $(q, p)$ , donde  $q$  es la posición y  $p$  es el momento cinético, y ambas magnitudes se representan matemáticamente mediante números reales, que un lenguaje lógico no puede representar mediante un infinito numerable de enunciados. Esto muestra que el enfoque puramente lógico-formal no resulta suficiente siquiera para obtener una caracterización adecuada de la noción de determinismo en su acepción matemática.

Pero la cuestión se oscurece aún más cuando se pretende extraer conclusiones ontológicas acerca del carácter determinista de un sistema a partir de un análisis exclusivamente lógico-formal como el que propone Nagel. En física, la evolución temporal de un sistema suele representarse mediante ecuaciones diferenciales que gobiernan el modo en que ciertas magnitudes varían con el tiempo; por lo tanto, en estos casos el problema del determinismo metodológico se reduce a analizar si tales ecuaciones diferenciales poseen solución única para todo instante futuro dadas las condiciones iniciales, y esta es una propiedad puramente matemática de

las ecuaciones en cuestión. Supóngase ahora un sistema cuya evolución temporal se expresa mediante la ecuación diferencial  $dp/dt = kp$ , donde  $k$  es una constante; su solución única,  $p(t) = p(0) e^{kt}$ , determina unívocamente el valor de  $p(t)$  en cualquier instante futuro  $t$  dada la condición inicial  $p(0)$  en el instante  $t=0$ . Sin embargo, nada puede afirmarse aún en el plano ontológico: la variable  $p$  podría representar la probabilidad de ocurrencia de cierto tipo de eventos y, por tanto, si bien la ecuación dinámica determina unívocamente los valores de  $p$  para todo instante futuro, expresa la evolución de una probabilidad a través del tiempo y no la sucesión temporal unívoca de las propiedades del sistema en el plano ontológico. Por ejemplo, la ecuación de Schrödinger es “determinista” en el sentido de fijar unívocamente la sucesión de estados cuánticos de un sistema, pero cada estado cuántico no determina las propiedades del sistema en el instante correspondiente sino solo su probabilidad: la ecuación de Schrödinger describe la evolución de probabilidades.

Hasta aquí, en estos dos últimos apartados, se ha intentado acotar la noción de determinismo identificando los alcances de las definiciones tradicionales que se han brindado, y como consecuencia de dicho análisis se impuso la necesidad de diferenciar los múltiples sentidos del determinismo con el afán de clarificar la discusión y evitar que una discusión conceptual se empantane como producto de una confusión terminológica. En las siguientes secciones se analizarán las relaciones existentes entre el determinismo y nociones usualmente emparentadas.

## **5. El determinismo y sus coterráneos: fatalismo, legalismo, causalismo y predictibilidad**

### ***5.1. Determinismo, fatalismo y legalismo***

A través de su larga historia, el concepto de determinismo no sólo sufrió múltiples metamorfosis, sino que fue tradicionalmente asociado con otras nociones ligeramente emparentadas pero en modo alguno equivalentes; entre ellas, las nociones de fatalismo y lega-

lismo son algunas de las nociones que con mayor frecuencia se han identificado con la idea de determinismo.

El *fatalismo* es una antigua doctrina según la cual existe un hado o destino, a veces incluso por encima de los propios dioses, cuyo poder supranatural e ineluctable dirige el curso de los acontecimientos. Esta es la idea que recorre la tragedia griega y que Sófocles, en su *Antígona*, expresa del siguiente modo:

[...] la fuerza del destino es algo tremendo: ni la dicha, ni Ares, ni una fortaleza, ni las negras naves zarandeadas por la mar conseguirán esquivarla. (Sófocles, *Antígona*: Acto IV, Estrofa I).

La misma idea se trasluce de las palabras de Omar Kayyam:

La vida es tan solo un tablero, cuyos cuadros blancos son los días y los negros son las noches, con el que el hado se divierte con los humanos. Como si fueran piezas de ajedrez nos mueve a su antojo.” (citado en Wagensberg, 1992, p. 15).

Si bien el hado puede afectar toda la realidad, en general el fatalismo concierne principalmente a las acciones humanas: Edipo estaba predestinado a matar a su padre Laertes y a casarse con su madre Yocasta; a pesar de sus desesperados intentos por escapar a su destino, todos sus actos se conjuraron, bajo la guía del *fatum*, para conducirlo al trágico desenlace de su vida.

La identificación entre determinismo y fatalismo ha sido muy frecuente en la historia del pensamiento. Esta errónea asimilación, común entre los románticos, es la que inspiró a Ralph Emerson para escribir: “*El libro de la Naturaleza es el libro del Destino*” (Emerson, 1903, p. 20). Para el fatalismo, los acontecimientos no se encuentran vinculados entre sí directamente sino a través de un poder exterior a ellos: el futuro está escrito por el libre albedrío de un *fatum* trascendente y sobrenatural. La necesidad que el fatalismo afirma es alej, no sujeto a las leyes naturales, y por ello sus pautas, si las tiene, se consideran inescrutables. Para el fatalista los acontecimientos son incondicionales, esto es, ocurren con independencia de las circuns-

tancias: como lo preestablecido debe suceder, nada podrá evitar que suceda, nada podrá oponerse a los designios insondables del *fatum*. El determinismo, por el contrario, es legal: no existe entidad trascendente alguna que vincule los acontecimientos entre sí; éstos se suceden unívocamente a través del tiempo según la ley natural que los rige. Pero la mayor diferencia entre las doctrinas determinista y fatalista reside en que, para el determinismo, la necesidad no es absoluta sino condicional: no existe evento alguno que ocurra independientemente de la historia previa del sistema o del universo; que un evento suceda en el futuro depende de que ocurra en el pasado otro evento con el cual se encuentra nomológica y unívocamente conectado. En otras palabras, aun en el caso de un sistema ontológicamente determinista, no existe principio alguno que predetermine sus condiciones iniciales.

No obstante, el carácter legal del determinismo no debe conducir a la identificación entre determinismo y *legalismo*, según el cual todos los eventos de la naturaleza se encuentran sujetos a leyes naturales. Quienes asimilan ambos conceptos creen necesario defender el determinismo como ingrediente esencial del método científico, sin el cual la ciencia misma sería imposible. Henri Poincaré parece adoptar esta perspectiva cuando afirma: “la ciencia es determinista; es a priori; postula el determinismo porque sin él no podría existir” (Poincaré, 1913, p. 244). Esta tesis, que podría haber sido defendida con cierto fundamento en la época del apogeo de la mecánica clásica, continúa no obstante siendo sostenida mucho más recientemente por científicos como Jorge Wagensberg:

El segundo principio que el científico asume tácitamente para elaborar ciencia podría llamarse la hipótesis determinista [...]: los sucesos del mundo no son independientes entre sí, exhiben cierta regularidad [...]. Dicho de otro modo, en virtud del principio determinista adquiere sentido nada menos que el concepto de ley de la naturaleza. Porque en la naturaleza no todo es posible; de todos los sucesos virtuales que podrían ser —sea el caos— no todos son. (Wagensberg, 1992, p. 12).

Si bien oscilando entre los planos metodológico y ontológico, aquí Wagensberg supone que el determinismo afirma exclusivamente la conexión legal entre acontecimientos: los hechos naturales se encuentran vinculados entre sí a través de leyes naturales. Sin embargo, el determinismo impone una exigencia más, esto es, su conexión unívoca. Las leyes probabilísticas, como el llamado “colapso de la función de onda” en mecánica cuántica, aun cuando adjudican probabilidad nula a ciertos sucesos “virtuales”, no fijan unívocamente el evento futuro dado el evento presente, sino que atribuyen una distribución de probabilidades sobre todos los eventos futuros posibles. Esta identificación entre determinismo y legalismo, más extendida que cuanto podría suponerse en principio, también aparece en el ámbito de las ciencias humanas; si bien, en general, el determinismo es considerado como un obstáculo para el libre albedrío y para la responsabilidad, hay autores que asumen una perspectiva opuesta cuando suponen que abandonar el determinismo equivale a renunciar a la posibilidad de una ciencia del comportamiento humano. En este sentido, Burrhus Frederic Skinner afirma:

Si vamos a utilizar los métodos científicos en el campo de los asuntos humanos, hemos de suponer que la conducta está determinada y regida por leyes. Hemos de esperar descubrir que lo que el hombre hace es el resultado de unas condiciones específicas, y que una vez descubiertas estas podemos anticipar y, hasta cierto punto, determinar sus acciones. (Skinner, 1953, p. 6; citado de la versión en castellano).

Una vez más, Skinner presenta la falsa dicotomía determinismo versus inexistencia de leyes naturales, lo cual vacía de contenido el concepto mismo de ley natural indeterminista. En consecuencia, debido a que existen leyes naturales indeterministas, como las de la física cuántica, no resulta apropiado asociar el determinismo con la legalidad.

## **5.2. Determinismo y causalismo**

Más allá de las ideas de fatalismo y legalismo, la noción más veces citada en toda discusión acerca del determinismo es, sin duda, la



noción de causalidad. En términos más precisos, la tesis determinista suele ser identificada con el *causalismo*, doctrina que afirma la validez del Principio de Causalidad, comúnmente expresado como: “todo evento tiene una causa” o, de un modo no equivalente, “las mismas causas producen los mismos efectos”. El problema de la causación es una de las cuestiones filosóficas de más larga tradición. Si bien con una interpretación del concepto de causa más amplia que la actual, los nexos causales constituían ya la amalgama ontológica del universo aristotélico. Según la interpretación tradicional de la filosofía aristotélica, para Aristóteles la causa se dice de cuatro maneras. Esta doctrina fue reformulada durante la Baja Edad Media por la escolástica en términos de cuatro causas: sobre la noción de causalidad construyó Santo Tomás una de sus famosas pruebas de la existencia de Dios. Con el nacimiento de la ciencia moderna, las causas final, formal y material desaparecieron de la discusión filosófica, de modo tal que el concepto de causa pasó a identificarse con el de causa *efficientes*. En este sentido, en *Il Saggiatore* Galileo definía causa como la condición necesaria y suficiente para la aparición de algo: “aquello, y no otra cosa debe llamarse causa, a cuya presencia sigue el efecto y a cuya eliminación el efecto desaparece” (Galileo, 1890-1909, Vol. 6, p. 265). Con el empirismo, la causación abandona el plano ontológico para convertirse en una categoría epistémica. Este cambio de perspectiva ya se manifiesta en Locke quien, en su *Ensayo sobre el Entendimiento Humano*, afirma:

Aquello que produce cualquier idea simple o compleja, lo denotamos por el nombre general de causa, y aquello que es producido, por el nombre de efecto. (Locke, 1690, Libro II, Cap.XXVI, Secc.I).

Con Hume, el viraje se completa y la relación causal se asimila a la mera conjunción constante entre la causa y el efecto (Hume, 1739, Libro I, Parte III, Secc.II-III; en tiempos recientes se ha brindado una nueva interpretación según la cual Hume no sería un antirrealista nomológico sino un realista escéptico: ver, por ejemplo, Wri-gh, 1983 y Strawson, 1989).

Hasta este momento de la historia de la filosofía, el problema del determinismo no era concebido como un problema diferente al de la causación, y así se lo siguió considerando por mucho tiempo más. El propio Laplace se expresaba en términos de causas y efectos en el famoso pasaje de su *Theorie Analytique des Probabilités*. Fue recién durante el siglo XX que determinismo y causalismo comienzan a ser vistos como doctrinas que, si bien relacionadas, se basan en nociones conceptualmente diferentes. Gaston Bachelard es particularmente explícito en este punto cuando afirma que “la causalidad y el determinismo no son absolutamente sinónimos” (Bachelard, 1935, p. 110). Frente a esta dualidad de conceptos, los diferentes autores han adoptado las posiciones más diversas. Por ejemplo, Paulette Février considera la causalidad como una categoría del *a posteriori*, relacionada con los encadenamientos pasados entre acontecimientos, mientras que concibe el determinismo como una categoría del *a priori*, relativa al porvenir y, por tanto, a la idea de previsión (Février, 1955, p. 251). Algunos autores ven el nexo causal como más general que la relación de determinación; ésta es la perspectiva que adopta Rudolph Carnap cuando sostiene que

[...] el determinismo es una tesis especial acerca de la estructura causal del mundo. Es una tesis según la cual esta estructura causal es tan fuerte que, dada una descripción total del mundo en un instante dado, entonces, con ayuda de las leyes, puede calcularse todo suceso pasado o futuro.” (Carnap, 1966, p. 288).

La posibilidad de nexos causales no deterministas es también admitida por Louis de Broglie, quien expone claramente su postura:

Consideremos un fenómeno *A* al cual le sigue siempre cualesquiera de los fenómenos  $B_1, B_2, B_3, \dots$ . Si, además, ninguno de los fenómenos  $B_1, B_2, B_3, \dots$  se produce cuando *A* no se produce, se podrá decir, adoptando una definición amplia de causalidad, que *A* es la causa de los fenómenos  $B_1, B_2, B_3, \dots$ , y esta definición estará de acuerdo con el viejo adagio: *sublata causa, tollitur effectus*. Con esta definición existirá, entonces, un vínculo de causalidad entre el fenómeno *A* y los fenómenos  $B_1, B_2, B_3, \dots$ , pero no habrá ya determinismo [...]

cuando no podemos prever de ningún modo cuál de los fenómenos  $B_1, B_2, B_3, \dots$  se va a producir cuando  $A$  se produce. El determinismo sólo reaparece en el caso límite, cuando existe un solo fenómeno  $B$ . (de Broglie, 1941, p. 64).

Una posición totalmente opuesta es la que adopta Mario Bunge al afirmar:

El principio causal es un caso particular del principio de determinación: se aplica, esencialmente, cuando la determinación se realiza en forma unívoca o inequívoca por condiciones externas.” (Bunge, 1959, p. 38).

Esta perspectiva se funda en una peculiar concepción del concepto de determinación, que incluye un elemento de productividad: lo determinante produce, genera lo determinado (Bunge, 1959, p. 356). Pero, en la realidad natural, los estados de un sistema no poseen capacidad productiva (Bunge, 1959, p. 83). Sobre esta base, el autor postula un concepto de determinismo diferente del basado únicamente en la sucesión temporal unívoca entre los estados de un sistema.

Por último, es necesario alertar contra la inadecuada identificación entre determinismo y *mecanicismo*. Tal asimilación, que se remonta al atomismo griego de Demócrito y Leucipo, podía ser sostenida razonablemente en la época del apogeo de la mecánica clásica, cuando se creía que la mecánica agotaba toda posibilidad de descripción del universo físico. Pero en la actualidad, con el reconocimiento de múltiples fenómenos físicos deterministas, pero no mecánicos, la identificación entre determinismo y mecanicismo resulta ya inaceptable. Mientras el mecanicismo exige la sucesión temporal unívoca entre los estados mecánicos de cualquier sistema real, el determinismo no impone restricción alguna acerca de qué tipo de estado deba constituir la evolución dinámica del sistema. El electromagnetismo de Maxwell brinda un claro ejemplo de teoría que, sin aludir a propiedades mecánicas, describe evoluciones dinámicas de sistemas deterministas.

### 5.3. Determinismo y predictibilidad

La identificación entre los conceptos de determinismo y predictibilidad tiene una larga historia: las cuestiones asociadas con los poderes de un ser humano o sobrehumano aparecen recurrentemente toda vez que se intenta precisar el concepto de determinismo más allá de la formulación de su idea central. El propio Laplace postula una superinteligencia capaz de reconstruir el pasado y el futuro del universo a partir de su estado presente. Pero en la famosa definición laplaciana, el recurso a un “genio” superpoderoso cumple la función de esclarecer un concepto que, no obstante, refiere al plano ontológico. Esta visión de una realidad intrínsecamente determinista con independencia de todo observador humano o sobrehumano fue la que prevaleció hasta el siglo XIX, aun cuando los intentos de elucidación del concepto de determinismo continuaran aludiendo con frecuencia a las posibilidades de predicción.

Esta perspectiva sufrió un fuerte viraje desde comienzos del siglo XX, con la enorme influencia del positivismo lógico y su insistencia en la búsqueda de criterios verificacionistas para admitir la legitimidad de cualquier concepto. Sobre la base del principio verificacionista del significado, según el cual un enunciado tiene sentido si es analítico o si puede ser verificado empíricamente, la noción de determinismo carece de sentido si no se la refiere a un criterio empírico para establecer inequívocamente el carácter determinista o indeterminista de un sistema. Así, de un modo consistente con su rechazo por las cuestiones metafísicas, el positivismo lógico borra el contenido ontológico de la noción de determinismo que pasa a ser identificada, ahora sí conceptualmente, con la noción de predictibilidad. Esta perspectiva trasciende los límites originales del Círculo de Viena, ejerciendo una enorme influencia sobre el pensamiento de los científicos naturales hasta muy avanzado el siglo XX.

La interpretación positivista-lógica del concepto de determinismo hace entrar en escena una nueva cuestión: el problema de la determinación empírica de las condiciones iniciales de un sistema. El determinismo entendido como predictibilidad requiere

que el estado inicial del sistema pueda fijarse por vía empírica; los casos donde esto no sea posible en la práctica —paradigmáticamente, el lanzamiento de dados— resultarían, por tanto, indeterministas. Tal exigencia es lo que permite a Virgil Hinshaw (1959) sostener que ni siquiera el propio Laplace fue un determinista estricto, pues no suponía que fuéramos capaces de predecir el resultado de un lanzamiento de dados individual: en tal situación existen magnitudes físicas cuyo valor inicial no puede obtenerse por medición sobre el sistema.

A pesar de su declarada oposición al positivismo lógico, el propio Popper cae bajo su influencia cuando se trata de la cuestión del determinismo. Popper admite la legitimidad del concepto metafísico de determinismo pues, a diferencia de los positivistas, rechaza el supuesto según el cual no-testeable implica no-significativo. No obstante, propone un concepto científico de determinismo que se identifica con la noción de predictibilidad y que, a diferencia de su contrapartida metafísica, posee contenido empírico: si bien ambos sentidos de determinismo no se asimilan, el determinismo científico implica el metafísico (Popper, 1950). Desde esta perspectiva, Popper reinterpreta la definición de Laplace: el genio laplaciano no debe concebirse como un Dios omnisciente e inmaterial, sino como un súper-científico corpóreo que interactúa con el sistema cuyo comportamiento intenta predecir. De este modo ingresa a la discusión el problema de las condiciones iniciales: la superinteligencia, al igual que el científico humano, no puede fijar el estado inicial del sistema con una precisión absoluta; sólo le es posible trabajar con un grado finito de precisión (Popper, 1982). Esta noción de determinismo permite a Popper afirmar que existen sistemas clásicos indeterministas, en contra del tradicional supuesto acerca del carácter determinista de la mecánica clásica:

Mi tesis es que muchos sistemas físicos, incluyendo la física clásica y la física cuántica, son indeterministas en un sentido tal vez más fundamental que el que se adscribe usualmente al indeterminismo de la física cuántica. (Popper, 1950, p. 117).

En principio, la propuesta popperiana de introducir una nueva noción de determinismo, más fuerte que la noción ontológica original, nada tiene de objetable: se trata de dos conceptos diferentes vinculados a través de una relación de implicación. En el marco de la epistemología de Popper, esta distinción podría incluso defenderse argumentando que responde al criterio de demarcación entre el ámbito científico y el ámbito metafísico. No obstante, la argumentación se oscurece cuando Popper revela su propósito último, escapar de:

[...] la pesadilla del determinista físico [donde] todo lo que ocurre en este mundo está predeterminado físicamente, incluyendo nuestros movimientos y, por tanto, nuestros actos. Así pues, todos nuestros pensamientos, sentimientos y esfuerzos [...] son, si no meras ilusiones, a lo sumo subproductos superfluos (“epifenómenos”) de acontecimientos físicos. (Popper, 1972, p. 203).

Popper, justamente, pretende evitar la concepción según la cual:

[...] el mundo en su conjunto, con todo lo que hay en él, es un inmenso autómatas y, por tanto, nosotros no somos más que diminutos engranajes o, a lo sumo, sub-autómatas suyos.” (Popper, 1972, p. 207).

Pero para eludir esta lúgubre visión es necesario adherir a un indeterminismo ontológico y no meramente científico: no es suficiente la imposibilidad de predecir nuestras acciones, puesto que es en el plano ontológico que la pesadilla se genera; no es suficiente creer que somos libres, debemos ser libres. Si el determinismo científico implica el ontológico, sólo una falacia formal permite concluir que el indeterminismo científico implica el indeterminismo ontológico. Puesto que no es legítimo inferir conclusiones ontológicas a partir de consideraciones epistémicas, John Earman (1986, pp. 7-8) considera que la formulación de un concepto epistémico de determinismo constituye un abuso del lenguaje, dado que existe un término, ‘predictibilidad’ más adecuado para recoger las cuestiones epistémicas. Sobre esta base, Earman recomienda depurar la discu-

sión acerca del determinismo de toda connotación epistémica que pueda conducir a confusiones y falacias argumentativas.

Pero la inadecuada transposición entre los planos ontológico y gnoseológico no es el único inconveniente que genera la asimilación entre determinismo y predictibilidad. Una dificultad adicional es la que surge cuando se examina en detalle la noción misma de predicción. El caso más sencillo de predicción científica es aquél donde las leyes que intervienen no incluyen variables estadísticas. En este caso, la predicción es la contrapartida de la explicación según el modelo nomológico-deductivo: ambas operaciones comparten la misma estructura lógica; predecir es deducir enunciados relativos a hechos desconocidos, a partir de leyes y de enunciados singulares referidos a la particular situación bajo estudio (Hempel, 1965, pp. 247-251). Por ejemplo, conociendo el volumen  $V_G$  y la temperatura  $T_G$  de un gas que se aproxima a un gas perfecto, a partir de la conocida relación entre presión, volumen y temperatura,  $PV/T=k$ , donde  $k$  es una constante, el físico puede predecir la presión  $P_G$  del gas. Sin embargo, esta posibilidad de predicción nada permite afirmar acerca del carácter determinista o indeterminista del sistema; la ley  $PV/T=k$  afirma la relación funcional entre tres magnitudes del gas en estado de equilibrio, pero no describe la evolución temporal del sistema. La asimilación entre determinismo y predictibilidad pasa por alto que la noción más general de predicción no alude exclusivamente a la inferencia del futuro a partir del pasado, sino a toda *inferencia de lo desconocido a partir de lo conocido*. Las únicas predicciones que guardan alguna relación con el problema del determinismo son las que se basan en leyes dinámicas, esto es, leyes que permiten calcular estados futuros a partir del estado en el instante inicial —en algunos casos, más las condiciones de contorno—; pero este tipo de predicción, si bien suele ser la única forma de predicción considerada por algunos filósofos, es sólo un caso particular de una noción de predicción mucho más amplia.

De estas consideraciones puede concluirse que, si la identificación entre predictibilidad y determinismo ontológico involucra una incorrecta asimilación entre diferentes planos de análisis,

tampoco es adecuada la identificación entre predictibilidad y determinismo epistémico, puesto que la noción de predicción no se limita al conocimiento de los estados futuros del sistema. No obstante, podría pensarse que la noción de determinismo epistémico se asimila a un concepto restringido de predictibilidad que podría denominarse “futurístico”. Pero aun en este sentido estrecho, el concepto de predicción al cual se alude resulta más amplio que el requerido para caracterizar el determinismo epistémico: dado el estado actual de la ciencia, no puede ignorarse la posibilidad de efectuar predicciones estadísticas acerca del futuro aún en sistemas respecto de los cuales poseemos un conocimiento indeterminista —adoptando aquí el sentido epistémico del término ‘determinista’—. En tales casos, no podrá determinarse unívocamente el estado en el cual se encontrará el sistema luego de un cierto intervalo, pero sí se podrá predecir la distribución de probabilidades correspondientes a los distintos estados posibles. En efecto, aún quienes admiten que sólo pueden efectuarse predicciones estadísticas de los fenómenos subatómicos, no dudan en reconocer los éxitos predictivos de la mecánica cuántica. En definitiva, si se pretende asociar determinismo epistémico y predictibilidad, es necesario recurrir a un sentido aún más restringido de predicción, que podría denominarse “predicción futurística unívoca”. Esto muestra que, contrariamente a lo que parece suponer Earman, no es tan sencillo prescindir del concepto de determinismo epistémico reemplazándolo por el de predictibilidad: resulta interesante retener el sentido epistémico del término ‘determinista’ en la medida en que alude a un tipo muy peculiar de predicción.

Sobre la base de estas consideraciones, ¿cómo se relacionan determinismo ontológico y predictibilidad? En primer lugar, debe recordarse nuevamente que la pregunta sólo tiene sentido respecto de la predictibilidad de tipo futurístico. Pero aún en este caso, no puede ignorarse que incluso en sistemas ontológicamente indeterministas pueden efectuarse predicciones estadísticas, siempre que se cuente con las leyes estadísticas que rigen su comportamiento dinámico. Por lo tanto, el indeterminismo ontológico no implica



total impredecibilidad sino únicamente impredecibilidad futurística unívoca. Esto pone de manifiesto, una vez más, que la noción de determinismo epistémico no resulta superflua, puesto que permite expresar la misma idea de un modo conceptualmente más sencillo —indeterminismo ontológico implica indeterminismo gnoseológico—, poniendo claramente de manifiesto la relación de implicación unidireccional entre conceptos análogos pero pertenecientes a diferentes planos de análisis.

## 6. Dos formas de determinismo

Recordemos por un momento algunas de las caracterizaciones de determinismo suministradas en las secciones anteriores:

- Laplace nos habla de la conexión entre el estado actual del *universo* como el efecto de su estado previo y la causa del estado posterior.
- Leibniz se refiere al ancho *mundo* y a cómo lo que hay en él acaece matemáticamente, esto es, infaliblemente.
- De Broglie alude al *mundo* físico como una inmensa máquina cuya evolución está inexorablemente determinada.
- James describe un *universo* donde las partes ya establecidas, fijan absolutamente lo que serán las demás partes.

Todas las caracterizaciones anteriores se aplican a la realidad como un todo: al “universo” en Laplace, al “ancho mundo” en Leibniz, al “mundo físico” en de Broglie, nuevamente al “universo” en James. Se trata de observar la historia de lo real en su conjunto, para afirmar que todos sus estados quedan unívocamente determinados por sus estados anteriores.

Sin embargo, otro grupo de caracterizaciones de la noción de determinismo también presentadas en las secciones anteriores divergen del anterior como se resalta a continuación :

- Leucipo afirma que ninguna *cosa* sucede sin razón, sino que todas acaecen por necesidad.
- En su famoso argumento, Aristóteles se refiere a la *batalla naval* de mañana.

- El antiguo estoico habla de *sucesos* anteriores que son causa de aquellos que les siguen.
- Laplace se refiere al movimiento totalmente determinado de una única *molécula* de aire o de vapor.

En este segundo grupo de caracterizaciones de la noción de determinismo los autores ya no se refieren al mundo o al universo como un todo, sino al modo en que los objetos o sucesos particulares transitan una historia completamente determinada: son las “cosas” para Leucipo y el ignoto estoico, la “batalla naval” para Aristóteles o una “simple molécula” para Laplace en este caso.

La diferencia entre los dos grupos de caracterizaciones muestra que parece haber dos enfoques generales que conducen a dos conceptos diferentes de determinismo. Según el primer concepto, que podría denominarse ‘*determinismo universal*’ y simbolizarse como  $\mathbf{Det}_U$ , el predicado determinista se aplica al universo  $U$  en su conjunto,  $\mathbf{Det}_U(U)$ , y establece la concatenación unívoca entre sus estados. Según el segundo concepto, que podría denominarse ‘*determinismo individual*’ y simbolizarse como  $\mathbf{Det}_I$ , el predicado determinista se aplica a cada uno de los sistemas u objetos  $s$  del universo,  $\mathbf{Det}_I(s)$ , y refiere a la historia de cada uno de esos objetos, unívocamente determinada por su estado inicial. En su aplicación a la relatividad general, Earman (2008) introduce una distinción análoga entre “ramificación de ensamble” y “ramificación individual”. Con la terminología ‘universal-individual’ se simbolizarán los correspondientes conceptos de indeterminismo como  $\mathbf{Indet}_U(U) = \neg \mathbf{Det}_U(U)$  e  $\mathbf{Indet}_I(s) = \neg \mathbf{Det}_I(s)$ .

Dada la diferencia entre las dos nociones, se impone la pregunta acerca de la relación entre ambas. La respuesta que parece natural es la que afirma que, si todos los sistemas del universo transitan historias deterministas, entonces el universo completo es determinista:

$$\forall s \mathbf{Det}_I(s) \Rightarrow \mathbf{Det}_U(U)$$

Si este fuese el caso, en realidad no se trataría estrictamente de dos conceptos diferentes, sino de una forma general del concepto que surge derivativamente cuando se lo puede aplicar a todos

los casos particulares. Sin embargo, existen situaciones —modelos del universo— en los cuales todos los sistemas individuales describen historias deterministas y, no obstante, los estados del universo completo no se encuentran unívocamente determinados por sus estados previos, es decir,

$$\forall s \text{ } Det_I(s) \wedge \neg Det_U(U)$$

Como veremos a continuación estas situaciones se manifiestan tanto en el ámbito clásico como en el relativista.

## 6.1. Mecánica clásica

Como es bien sabido, en el marco de la mecánica clásica no existe un límite máximo de velocidad que las entidades físicas no puedan superar, por lo tanto, una partícula puede adquirir una velocidad indefinidamente grande. Tal característica de la mecánica clásica hace posible una situación en la cual una partícula aumenta su velocidad indefinidamente, de modo tal que su coordenada espacial tiende a infinito a medida que el tiempo tiende a un instante finito determinado  $t^*$ . Esta situación, que suele denominarse ‘*solución de escape*’, es la de un universo que contiene una única partícula, la cual “desaparece” en el infinito espacial en el instante  $t^*$  (ver Earman, 1986). Dado que las ecuaciones de la mecánica clásica son  $t$ -invariantes, la imagen temporalmente especular de la gráfica anterior también representa una evolución mecánicamente posible: un universo que permanece vacío hasta  $t = t^*$ , y en el cual luego aparece repentinamente una partícula desde el infinito espacial, lo cual constituye una clara violación del determinismo. En efecto, llamando  $e_0$  al estado del universo completo en un instante  $t_0 < t^*$ , para  $t_1 > t^*$  existen dos estados condicionalmente posibles respecto de  $e_0$ :  $e_1$ , correspondiente al universo que permanece vacío para todo instante posterior a  $t^*$ , y  $e_1$ , correspondiente al universo que contiene la partícula proveniente del infinito espacial (véase Figura III.4 en Earman, 1986).

La pregunta acerca del determinismo de la mecánica clásica ha dado origen a un interesante ámbito de discusión respecto de las situaciones en las cuales el determinismo clásico parece encon-

trar sus límites (ver, por ejemplo, Wilson, 1989; Hutchison, 1993; Lombardi, 2002; Norton 2003, 2008). No entraremos aquí en estos debates. El objetivo de la presente subsección ha sido mostrar, en el ámbito de la mecánica clásica, la posibilidad de que un estado del universo completo no determine unívocamente su estado en todo tiempo posterior, lo cual construye un caso de *Indet<sub>U</sub>*. Sin embargo, en este universo indeterminista en sentido universal, las trayectorias de los objetos no se “bifurcan” en el tiempo: la partícula que proviene del infinito espacial en el tiempo  $t^*$  —cuyo comportamiento es precisamente el responsable del indeterminismo universal— describe una trayectoria única y sin diferentes posibilidades futuras, esto es, una trayectoria determinista en sentido individual: *Det<sub>I</sub>*. Por lo tanto, éste es un caso en el cual todos los sistemas individuales —en este caso, la partícula única— describen historias deterministas y, no obstante, no todos los estados del universo completo se encuentran unívocamente determinados por sus estados previos:  $\forall s \text{ } Det_I(s) \wedge \neg Det_U(U)$ . En definitiva, con independencia de los debates acerca de las condiciones físicas necesarias para que situaciones de este tipo se hagan efectivas, este resultado pone de manifiesto claramente que los dos sentidos de determinismo son conceptualmente diferentes.

## 6.2. Mecánica relativista

Una característica central de la relatividad general es su covariancia general, es decir, la invariancia de las ecuaciones de campo frente a transformaciones arbitrarias de coordenadas. En 1913 Einstein notó que, si las ecuaciones que gobiernan la gravedad cumplen con la covariancia general, entonces la métrica no queda determinada de manera unívoca por las fuentes gravitacionales, esto es, los cuerpos con masa. En particular, consideró un espacio-tiempo lleno de materia excepto en una región, el “agujero”, que quedaba libre de materia, y luego se preguntó si la especificación completa de la métrica y la distribución de materia fuera del agujero fijaría la métrica dentro: su respuesta fue negativa. En términos formales: dado un espacio-tiempo  $(M, g)$  y un difeomorfismo  $h$  de  $M$  sobre sí mismo

que (i) conduce a una nueva métrica  $g'$  tal que  $g = g'$  fuera del agujero y  $g \neq g'$  dentro del agujero, y (ii) conecta las dos métricas de una manera “suave”, entonces si  $(M, g)$  es un modelo de las ecuaciones de campo,  $(M, g')$  también lo es.

Sobre la base de este resultado, John Earman y John Norton (1987) formularon el ya célebre “argumento del agujero”. La pregunta que está en la base del argumento es si ambos modelos,  $(M, g)$  y  $(M, g')$ , representan diferentes universos o no. Aquí se abre lo que los autores denominan “el dilema del indeterminismo”:

- De acuerdo con el substancialismo, según el cual el espacio-tiempo puede existir independientemente de los objetos que contiene, los modelos  $(M, g)$  y  $(M, g')$  representan diferentes universos. Esto conduce a un fuerte indeterminismo. En efecto, supongamos que  $(M, g')$  es un espacio-tiempo con tiempo global, tal que el agujero se encuentra en el futuro de cierto tiempo global  $t_0$ ; entonces, se tienen dos universos que coinciden hasta el tiempo  $t_0$  y difieren en el futuro de  $t_0$ .
- De acuerdo con el relacionalismo, según el cual el espacio-tiempo está definido por las relaciones espacio-temporales entre los objetos y eventos que contiene,  $(M, g)$  y  $(M, g')$  representan el mismo universo. En otras palabras, el relacionalista admite la equivalencia de Leibniz, según la cual dos modelos relacionados por un difeomorfismo representan el mismo universo. Por lo tanto, no existe indeterminismo en los universos relativistas.

Earman y Norton se inclinan decididamente por la alternativa relacionalista, ya que los dos modelos son observacionalmente indistinguibles y considerar que representan diferentes universos conduce a un indeterminismo radical.

El “argumento del agujero” ha sido extensamente comentado y muchas veces objetado desde distintas perspectivas (para una revisión de algunas de las críticas, ver Norton, 2019), pero aquí no nos detendremos en ello. Lo que resulta relevante a la luz del presente

trabajo es que la noción de determinismo involucrada en este caso es el indeterminismo universal: el resultado de Einstein, para el substancialista, conduce al **Indet<sub>U</sub>**, mientras que para el relacionalista, no resulta un desafío al **Det<sub>U</sub>**. Esto significa que el concepto de determinismo universal tiene sentido en el ámbito de la relatividad general en la medida en que permite distinguir las consecuencias que se obtienen desde distintas posiciones metafísicas respecto del espacio-tiempo. Pero, por otra parte, la existencia o no de “agujeros” en el espacio-tiempo no afecta el carácter determinista individual de los cuerpos en sus trayectorias particulares. Por ejemplo, tanto en el espacio-tiempo original  $(M, g)$  como en el espacio-tiempo con agujero  $(M, g')$ , un objeto describe una historia única sin diferentes alternativas posibles. Por lo tanto, para el substancialista puede darse significativamente una situación en la cual todos los objetos individuales del universo describen historias deterministas y, no obstante, no se cumple el determinismo universal:  $\forall s \text{ Det}_I(s) \wedge \neg \text{Det}_U(U)$ . Como en el caso clásico, independientemente de los argumentos físicos o metafísicos que se esgriman en favor o en contra de estas formas de determinismo, la posibilidad de determinismo individual e indeterminismo universal pone de manifiesto la diferencia conceptual entre ambas.

## 7. Consideraciones finales

Si los eventos futuros están unívocamente determinados por las condiciones que lo anteceden; si la acción humana, en tanto evento del mundo, está trágicamente determinada o si resulta posible compatibilizar la libertad necesaria para cualquier tipo de responsabilidad ética con los requisitos propios de la ciencia moderna, es una discusión donde confluyen argumentos de distinto tipo y que involucra insumos de diferentes ámbitos: científicos, éticos, metafísicos y epistémicos. El espíritu de este trabajo no fue dirimir en favor o en contra de dichos argumentos sino, por el contrario, sentar las bases para que los mismos se encaucen sobre los carriles adecuados. Para ello, en primer lugar, se analizaron diferentes definiciones que se han esgrimido en la historia del pensamiento con

el objeto de identificar las dificultades para precisar de un modo claro y no circular la noción de determinismo. Se mencionó que algunas definiciones descansan en conceptos emparentados pero no idénticos, como el concepto de predictibilidad o causalismo. Asimismo se estableció la dificultad conceptual existente en torno al modo en que Laplace define el término, dado que el mismo puede interpretarse en términos epistémicos o metafísicos. En función de ello se consideró que quizás la definición más prometedora es la brindada por James, dado que no sólo no recurre a conceptos tan problemáticos como el de causalismo o predictibilidad sino que, además, tiene la ventaja de no restringirse, en la definición misma, a una ontología particular. En segundo lugar, la identificación de las diferentes entidades a las cuales puede aplicarse el término ‘determinista’ permitió reconocer los múltiples sentidos del concepto en cuestión. Es así que el aspecto matemático, epistémico u ontológico que el concepto codifica contribuiría a que se impongan, respectivamente, diferentes sentidos: un determinismo de corte metodológico, un determinismo de corte gnoseológico y, por último, un determinismo de corte ontológico. El análisis de las definiciones a la luz de estas clasificaciones nos permitió concluir, por un lado, que en la literatura la discusión ha sido eminentemente establecida en el plano ontológico y que sólo ante las dificultades que esta perspectiva de análisis acarrea se instituyeron como posibles vías de escape las perspectivas metodológica o epistémica. Pero, a su vez, que los planos de análisis resultan independientes dado que ni se cumple que el determinismo ontológico imponga un determinismo metodológico o epistémico, ni se cumple su inversa.

Asimismo se intentó disociar nociones que usualmente se presentan como complementarias. Para ello se mostraron las diferencias conceptuales entre fatalismo y determinismo aduciendo que, en tanto que el fatalismo supone una necesidad sin legalidad y sin condicionamiento alguno, el determinismo supone cierta legalidad. En efecto, el concatenamiento de los eventos no se asume como un fenómeno inescrutable, ya que el devenir de los acontecimientos está, a diferencia de lo que sucede con el fatalismo,

íntimamente ligado al comportamiento de los eventos que lo anteceden. Por otra parte, se adujo que esto no implica asimilar las nociones de legalidad y determinismo en tanto que la existencia de leyes indeterministas socava cualquier tipo de identidad.

A su vez, se argumentó, desde diferentes perspectivas y sustentándose en diferentes fuentes, y en contra de varias definiciones establecidas, que el causalismo y el determinismo son nociones conceptualmente diferentes. Y por último se señaló que resulta impropio asociar predictibilidad con determinismo sin más, dado que no sólo eso supondría reducir la noción de determinismo a la noción de determinismo epistémico y, consecuentemente, confundir el plano ontológico con el plano gnoseológico, sino también porque ello no contemplaría la existencia de predicciones cuyo carácter o bien no es en absoluto de índole determinista ni indeterminista, o bien que pueden suponer ontologías indeterministas.

Aun sabiendo que queda mucho camino por recorrer, confiamos en que estas clarificaciones conceptuales, junto a las mencionadas al final del último apartado entre determinismo universal e individual, resulten prometedoras en tanto puedan coadyuvar a un análisis refinado de la discusión que le impida empantanarse en un farragoso y estéril terreno muchas veces generado más por discrepancias del ámbito terminológico que del conceptual.

## Referencias

- Aristóteles. (1984). *The Complete Works of Aristotle. Revised Oxford*. (J. Barnes (Ed. y Trad.). Princeton: Princeton University Press.
- Bachelard, G. (1935). *Le Nouvel Esprit Scientifique*. Alcan.
- Barry Cooper, S. (2004). *Computability Theory*. Chapman and Hall/CRC.
- Belot, G. (1995). Determinism and ontology. *International Studies in the Philosophy of Science*, 9, 85-101.
- Bridgman, P. (1927) [1949]. *The Logic of Modern Physics*. Macmillan.
- Brush, S. G. (1976). Irreversibility and indeterminism: Fourier to Heisenberg. *Journal of the History of Ideas*, 37, 603-630.
- Bunge, M. (1959) [1978]. *Causalidad*. Eudeba.



- Butterfield, J. (1998). Determinism and indeterminism. En Edward Craig (Ed.), *Routledge Encyclopedia of Philosophy* (Vol. 3 pp. 33-39). Routledge.
- Carnap, R. (1966) [1969]. *Fundamentación Lógica de la Física*. Editorial Sudamericana.
- Cassirer, E. (1936) [1956]. *Determinism and Indeterminism in Modern Physics*. Yale University Press.
- Crutchfield, J. P., Farmer, J. D., Packard, N. H. y Shaw, R. S. (1987). "Caos", *Investigación y Ciencia*, 125, 16-29.
- Darwin, C. (1859) [1983]. *El Origen de las Especies*. Sarpe.
- Dear, G. F. (1961). Determinism in classical physics. *The British Journal for the Philosophy of Science*, 11, 289-304.
- de Broglie, L. (1939) [1944]. *Materia y Luz*. Espasa-Calpe.
- de Broglie, L. (1941). *Continu et Discontinu en Physique Moderne*. Albin.
- Earman, J. (1971). Laplacian determinism, or is this any way to run a universe?. *The Journal of Philosophy*, 68, 729-744.
- Earman, J. (1986). *A Primer on Determinism*. Dordrecht: Reidel Publishing Company.
- Earman, J. (2004). Determinism: what we have learned and what we still don't know. En J. K. Campbell, M. O'Rourke y D. Shier (Eds.), *Freedom and Determinism* (pp. 21-46). Bradford Book/MIT Press.
- Earman, J. (2008). Pruning some branches from «branching spacetimes». En D. Dieks (Ed.), *The Ontology of Spacetime II* (pp. 186-205). Elsevier.
- Earman, J. and Norton, J. D. (1987). What price spacetime substantivalism? The hole story. *British Journal for the Philosophy of Science*, 38, 515-525.
- Emerson, R. W. (1903). *Works*, Vol. VI. Routledge.
- Février, P. (1955) [1957]. *Determinismo e Indeterminismo*. Universidad Nacional Autónoma de México, Dirección General de Publicaciones.
- Epicuro (1995). *Obras Completas*. Ediciones Cátedra.
- Galileo Galilei (1890-1909). *Opere*. Edizione Nazionale.
- García Gual, C. (1981) [1985]. *Epicuro*. Alianza Editorial.
- Giere, R. (1999). *Science without Laws*. University of Chicago Press.
- Hempel, C. (1965) [1988]. *La Explicación Científica*. Paidós.
- Hinshaw, V. (1959). Determinism versus continuity. *Philosophy of Science*, 26, 310-324.
- Hobbes, T. (1998). *Leviatán*. Fondo de Cultura Económica.

- Hume, D. (1739) [1981]. *Tratado sobre la Naturaleza Humana*. Orbis.
- Hutchison, K. (1993). Is classical mechanics really time-reversible and deterministic?. *The British Journal for the Philosophy of Science*, 44, 307-323.
- James, W. (1897) [1956]. "The dilemma of determinism", pp. 145-183 en *The Will to Believe*. New York: Dover Publications.
- Jauch, J. M. (1973). Determinism in classical and quantal physics. *Dialectica*, 27, 13-26.
- Laplace, P. S. (1814) [1985]. *Ensayo Filosófico sobre las Probabilidades*. Alianza Editorial.
- Leibniz, G. W. (1990). *Escritos en torno a la Libertad, el Azar y el Destino*. (C.Roldán, trad.). Tecnos.
- Leibniz, G. W. (2014). *Teodicea*. Biblioteca Nueva.
- Locke, J. (1690) [1986]. *Ensayo sobre el Entendimiento Humano*. Fondo de Cultura Económica.
- Lombardi, O. (2002). ¿Es la mecánica clásica una teoría determinista?. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, 17, 5-34.
- Long, A. A. (1975) [1984]. *La Filosofía Helenística*. Alianza Editorial.
- Mondolfo, R. (1942) [1983]. *El Pensamiento Antiguo. Historia de la Filosofía Greco-Romana*. Vol. 1. Losada.
- Montague, R. (1974). Deterministic theories (pp. 303-359). En *Formal Philosophy* (pp. 303-359). Yale University Press.
- Nagel, E. (1961) [1989]. *La Estructura de la Ciencia*. Paidós.
- Norton, J. D. (2019). The hole argument En E. N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy* (Summer 2019 edition), <https://plato.stanford.edu/entries/spacetime-holearg/>.
- Planck, M. (1948) [1949]. *Scientific Autobiography and Other Papers*. Philosophical Library.
- Poincaré, H. (1913) [1947]. *Ultimos Pensamientos*. Espasa-Calpe.
- Popper, K. (1950). Indeterminism in quantum physics and in classical physics. *The British Journal for the Philosophy of Science*, 1, 117-133, 173-195.
- Popper, K. (1972) [1992]. *Conocimiento Objetivo*. Tecnos.
- Popper, K. (1982) [1986]. *El Universo Abierto*. Tecnos.
- Russell, B. (1948) [1983]. *El Conocimiento Humano*. Orbis.
- Skinner, B. F. (1953). *Science and Human Behavior*. Free Press. Versión en castellano.

- Skinner, B. F. (1971). *Ciencia y Conducta Humana*. Galton, Centro de Investigaciones Psicológicas.
- Sófocles (1998). *Tragedias Completas*. Cátedra.
- Spinoza, B. (1980). *Ética Demostrada Según el Orden Geométrico*. Ed. Orbis.
- Strawson, G. (1989). *The Secret Connexion: Causation, Realism, and David Hume*. Clarendon Press.
- Tomás de Aquino (1952-1965). *Summa Theologiae, Summa contra gentes, Compendium theologiae*. Edición Marietti, Taurini-Romae.
- Wagensberg, J. (Ed.) (1992). *Proceso al Azar*. Tusquets.
- Wilson, M. (1989). John Earman's *A Primer on Determinism*. *Philosophy of Science*, 56, 502-532.
- Wittgenstein, L. (1921) [1981]. *Tractatus Logico-Philosophicus*. Alianza.
- Wright, J. P. (1983). *The Sceptical Realism of David Hume*. Manchester University Press.
- Wüthrich (2011). Can the world be shown to be indeterministic after all?, En C. Beisbart and S. Hartmann (Eds.). *Probabilities in Physics* (pp. 365-389). Oxford University Press.



## CAPITULO III

### PERSISTENCIA E IDENTIDAD PERSONAL

*Angélica María Pena-Martínez*<sup>1</sup>

#### RESUMEN

El objetivo de este artículo es presentar la solución exdurantista para el problema de la persistencia de las personas. Para lograr este objetivo, primero expondré el problema con el uso de dos experimentos mentales clásicos: los de fisión y fusión. Y en segundo lugar, sostendré que si bien una la noción de continuidad psicológica que asuma el funcionalismo materialista es necesaria, no es suficiente para dar cuenta de cómo persisten las personas. Finalmente mostraré que la teoría tetradimensional para la persistencia exdurantista, también conocida como la “concepción de estadios”, puede com-

---

1 Angélica María Pena-Martínez es candidata al doctorado en Filosofía de la Ciencia por la Universidad Nacional Autónoma de México (UNAM). Es licenciada en Filosofía por la Facultad de Filosofía y Letras y maestra en Filosofía de la Ciencia por la misma universidad. Se especializa en metafísica, filosofía del tiempo y ontología temporal. Su investigación se centra en los problemas metafísicos asociados con los viajes en el tiempo, así como en cuestiones de persistencia e identidad personal. Actualmente desarrolla su investigación doctoral en filosofía del tiempo y se desempeña como docente de asignatura en la Universidad Nacional Rosario Castellanos.

plementar la noción de continuidad psicológica que respete el funcionalismo materialista con relaciones de contrapartes temporales para dar cuenta de los casos de fusión y fisión.

## ABSTRACT

The aim of this paper is to present an exdurantist solution to the problem of people persistence. To achieve this aim, first I will illustrate the problem with the use of two mental experiments: fission and fusion. Secondly, I will claim that although a notion of psychological continuity that assumes materialist functionalism is necessary, it is not sufficient to explain how people persist. Finally, I will show that the tetradimensional view for exdurantist persistence, also known as “the stage view”, can complement the notion of psychological continuity in a way that respect a materialist functionalist view with temporal counterpart relations that explain cases of fusion and fission.

## 1. Introducción

La identidad personal (en adelante, “IP”) es un tema central en filosofía debido a sus múltiples implicaciones. Nos invita a reflexionar sobre cuestiones sumamente importantes, por ejemplo: ¿qué significa ser una persona?, ¿cómo puede alguien seguir siendo la misma persona a pesar de los múltiples cambios que experimenta?, etc. Además, la IP está profundamente ligada a debates sobre la responsabilidad moral, especialmente cuando atribuimos acciones a individuos en distintos momentos de sus vidas. Su relevancia trasciende el ámbito filosófico, influyendo en nuestra comprensión de las decisiones humanas y su impacto en el mundo que nos rodea. No es de extrañar, entonces, que la IP haya dejado una huella significativa en la cultura popular. Consideremos un ejemplo que ilustra esta conexión.

En “*El extraño caso del Dr. Jekyll y el Sr. Hyde*” de Robert Louis Stevenson (1886), el Dr. Jekyll crea un brebaje que, al ser bebido por él mismo, lo transforma en “alguien distinto”: el Sr. Hyde. El Sr. Hyde

comete una serie de crímenes que levantan la sospecha del abogado Utterson, quien eventualmente descubre la verdad acerca de su “identidad”. La apariencia física del Dr. Jekyll y el Sr. Hyde es muy diferente. El Dr. Jekyll es irreconocible ante la mirada de alguien familiarizado con él si está bajo los efectos del brebaje. Al final de la historia, el Dr. Jekyll reconoce haber cometido los crímenes y se atribuye la responsabilidad de lo que hizo en la forma del Sr. Hyde. La historia es intencionadamente ambigua y evoca preguntas interesantes acerca de la dualidad de los individuos, y acerca de qué es lo que constituye a las personas. Por ejemplo, no es evidentemente claro que el Dr. Jekyll deba ser juzgado por los crímenes del Sr. Hyde. El Dr. Jekyll y el Sr. Hyde podrían ser considerados diferentes personas. El brebaje claramente cambió al Dr. Jekyll, pero él parece recordar los crímenes de Sr. Hyde, e incluso siente culpa por haberlos cometido. No obstante, sin los efectos del brebaje, muy dudosamente hubiese desatado esos instintos criminales.

La pregunta que levanta la historia anterior es: ¿qué razones filosóficas tenemos disponibles para argumentar que el Dr. Jekyll es (o no es) la misma persona que Mr. Hyde? En este texto esbozaré algunas teorías acerca de IP que nos darán pistas para contener esta interrogante, que se relaciona con el problema (más específico) de cómo es que las personas persisten en el tiempo a pesar del cambio.

La estructura de este texto es la siguiente. En la segunda sección desarrollaré algunas de las nociones preteóricas y teóricas que son necesarias para abordar el problema de la persistencia de las personas. En la tercera sección, presentaré dos de los experimentos mentales clásicos en la literatura de IP: fusión y fisión. En la cuarta sección, explicaré a grandes rasgos en qué consisten las teorías de continuidad física y psicológica para ver cómo es que éstas solucionan los experimentos mentales expuestos. En la quinta sección, introduciré el modelo exdurantista de la persistencia para ver cómo es que éste puede ayudar a complementar las teorías de continuidad para dar cuenta de IP y de nuevo, contrastaré las soluciones a los experimentos mentales.

## 2. Identidad Personal y lo que Importa

Es preciso distinguir dos cuestiones: la pregunta acerca de la naturaleza de IP y la pregunta acerca de lo que importa (Lewis 1983; Parfit 1984). La primera es una pregunta metafísica, mientras que la segunda tiene un corte normativo. La pregunta sobre la naturaleza de IP puede formularse de las siguientes maneras: ¿qué hace que  $x$  sea una, y la misma persona a través del tiempo? Es decir, ¿cuáles son las condiciones de posibilidad que permiten decir que  $x$  en  $t_0$  es la misma persona que  $x$  en  $t_1$ ? Estas preguntas son acerca de la continuidad de una persona en el tiempo, y se relacionan con la pregunta: ¿Qué es lo que hace a una entidad una persona? La pregunta acerca de lo que importa tiene relación con lo que nos interesa mantener o rescatar para sostener que la persona de mañana tiene lo que importa para que sea la misma. Intuitivamente, lo que nos importa mantener tiene que ver con el tipo de características que consideramos valiosas, por ejemplo, la personalidad, los valores, creencias, deseos, etc.

Una noción preteórica acerca de IP es que no hay diferencia entre la pregunta sobre IP y la pregunta acerca de lo que importa. Esto quiere decir que intuitivamente las condiciones de posibilidad para IP rescatan lo que importa. Esto es cierto, pero sólo en casos normales: en casos cotidianos, ambas cuestiones van de la mano. Podemos pensar que ser continuos psicológica o físicamente a través del tiempo es suficiente para IP y que esto es justamente (para la mayoría de personas) lo que importa. Para poner a prueba esta intuición, consideremos el siguiente caso.

Supongamos que la condición suficiente y necesaria involucrada en IP de  $x$  tiene que ver con la continuidad de su cuerpo. El cuerpo de  $x$  sería la clave para que  $x$  sea la misma persona a través del tiempo. De acuerdo con la idea de que el cuerpo es el criterio suficiente y necesario para la IP, diríamos que si  $x$  tiene el mismo cuerpo que  $y$ , entonces,  $x$  es la misma persona que  $y$ . Pero ahora consideremos la siguiente situación: supongamos que el cuerpo de  $x$  en  $t_0$  es el mismo cuerpo que el de  $y$  en  $t_1$ . También supongamos que,  $y$  no tiene ninguno de los estados mentales, ni la personalidad,



etc., que  $x$  tenía en  $t_0$ . Este caso sería nomológicamente posible; es análogo a un caso de amnesia total. Las personas con amnesia total no recuerdan nada de lo que eran, ni de lo que creían, deseaban, nada... Sin duda estos casos son muy tristes, pero si suponemos que la concepción del cuerpo es correcta, entonces, parece que no deberíamos preocuparnos: ¡dado que  $y$  en  $t_1$  tiene el mismo cuerpo que  $x$  tenía en  $t_0$ , sería la misma persona que  $x$ ! Algunos podrían sentirse tentados a decir que la solución anterior no es del todo convincente, pues incluso si la continuidad del cuerpo es suficiente para IP, esto podría no ser suficiente para ofrecernos lo que nos importa acerca de qué es ser una persona. Alguien podría decir que lo que importa es mantener la psicología: si  $x$  es continua psicológicamente, entonces, tendremos lo que importa y sólo así estaremos satisfechos.

El ejemplo anterior podría sugerirnos que hay una diferencia entre la pregunta metafísica acerca de IP y la pregunta normativa. Por lo tanto, “¿seré la misma persona a través del tiempo?” es distinta de la pregunta “¿qué es lo que me importa mantener para sobrevivir?” (Parfit 1984; Kagan 2012). De este modo, puede haber casos en los que conservar el cuerpo sea un criterio suficiente para IP, pero que no sea suficiente para mantener lo que importa.

A continuación, expondré dos casos exóticos que reflejan con más claridad por qué en algunos escenarios lo que importa para IP no va de la mano con las condiciones de posibilidad para IP. Estos escenarios muestran, por un lado, que no es obvio que adoptar un criterio de posibilidad para IP garantice mantener lo que importa y, por otro lado, que la identidad numérica no es necesariamente lo que importa.

### 3. Dos Experimentos Mentales: Fusión y Fisión

Existen varios escenarios contrafácticos en la literatura sobre IP, pero hay dos particularmente relevantes que ilustran cómo surgen algunos problemas relacionados al considerar ciertos criterios, como el de continuidad física o continuidad psicológica, como condiciones de posibilidad para IP. Estos son los escenarios de fusión y fisión.

Antes de presentar ambos escenarios, es pertinente comentar algo sobre la importancia de los experimentos mentales para esta literatura. Alguien puede pensar que no es importante considerar estos escenarios ya que parecen más que nada de ciencia ficción y que no se relacionan con cómo es la realidad. Pero hay dos cuestiones a considerar.

La primera de ellas es una cuestión epistemológica. Según Parfit (1984, p. 200), estos experimentos mentales muestran que tenemos distintas reacciones que generan creencias acerca de nosotros mismos; acerca de nuestra continuidad; o acerca de lo que nos parece importante mantener para nuestra IP. Al plantearnos estos escenarios exóticos ponemos a prueba nuestras intuiciones o creencias acerca de los casos más sencillos.

La segunda cuestión es de corte metafísico y metodológico. Los experimentos mentales son de importancia en filosofía porque nos ayudan a ver cuál es la naturaleza de las nociones conceptuales que utilizamos para responder algunas cuestiones como la continuidad de las personas en el tiempo. Los experimentos mentales son planteados para saber cuánto es posible “estirar” nuestros conceptos y averiguar qué tan fuerte son para explicar distintos fenómenos. Adicionalmente, los experimentos mentales pueden ayudarnos a ver qué tan atinadas son nuestras nociones preteóricas. Algo que hay que tener en mente es que, a final de cuentas, los experimentos mentales se desarrollan dentro del espacio de posibilidad lógico, por lo que no podemos descartarlos tan fácilmente.

### **3.1. El caso de la Fusión**

Si dos objetos que son numéricamente distintos se fusionan, en un tiempo posterior a la fusión, tendremos un único objeto. Dicho objeto tendrá como partes a los dos objetos anteriores a la fusión. Un caso de fusión tiene la siguiente forma: supongamos que en  $t_0$  tenemos dos objetos. En algún punto posterior a  $t_0$  se los logra fusionar. El resultado, en un tiempo posterior  $t_1$  es un tercer objeto, numéricamente diferente a cualquier de los dos objetos anteriores.

Consideremos el siguiente experimento mental basado en alguno de los casos discutidos por Shoemaker (1963). Supongamos que A y B son dos hermanos gemelos con personalidades muy distintas. El gemelo A es un justiciero y el gemelo B es un villano. Un día los gemelos tienen un accidente en el que el cerebro del gemelo B y el cuerpo del gemelo A quedan destrozados. Un tal Filósofo Loco somete los restos funcionales de cada uno de los gemelos a una cirugía. Como resultado de la cirugía, surge una nueva persona en condiciones perfectas: esta persona, C, está compuesta por el cuerpo (sin cerebro) de B y el cerebro del gemelo A. Consideremos las siguientes preguntas: ¿Quién es esta nueva persona C?, ¿es C alguno de los gemelos que sobrevivió a la operación?, ¿cuál gemelo sobrevivió?

Tenemos cuatro posibles respuestas para estas preguntas:

(R1) C es A

(R2) C es B

(R3) C es A y B

(R4) C no es ni A ni B

### 3.2. El Caso de la Fisión

El caso de la fisión se desarrolla en la dirección opuesta al caso anterior. La idea es que un objeto puede dividirse o fisionarse, similarmente a como lo hacen las amibas. De esta manera, el objeto inicial, posterior a la división, es parte de los dos objetos en los que se divide. El caso de la fisión ha sido discutido en la literatura por algunos filósofos como Parfit (1984). Parfit discute la posibilidad de fisión con el uso de algunos escenarios posibles que involucran la posibilidad de la teletransportación. Hay uno en el que la teletransportación funciona satisfactoriamente y otro en el que no lo hace. El escenario más problemático, y que involucra fisión, es este último.

Tomemos el siguiente caso de teletransportación simple primero. Supongamos que el Dr. Jekyll es forzado a entrar a una máquina teletransportadora bajo la influencia del brebaje que lo transforma en el Sr. Hyde (supongamos que el Sr. Hyde tiene la intención macabra de deshacerse del Dr. Jekyll). El Sr. Hyde pide

la ayuda de un asistente para presionar el botón. Una vez que el asistente presiona el botón de la máquina, ésta creará un duplicado del Dr. Jekyll en otra región espacio temporal. El duplicado del Dr. Jekyll, llamémoslo, Dr. Jekyll\*, tiene toda la información relevante que posee el original, pero la materia que lo compone no es la misma que la que componía al Dr. Jekyll. La máquina teletransportadora destruye al Dr. Jekyll original durante el proceso de duplicación. La cuestión aquí es si podemos decir que Dr. Jekyll sobrevivió a la teletransportación. Hay por lo menos dos respuestas posibles de considerar (Parfit, 1984). La primera de ellas es que la teletransportación implica que el Dr. Jekyll no sobrevive y, por lo tanto, el Dr. Jekyll\* no es la misma persona que el Dr. Jekyll, sino simplemente una réplica de él. La segunda respuesta consiste en decir que, en virtud de que el Dr. Jekyll\* preserva lo que importa, a saber, su personalidad, apariencia, etc., aunque no es la misma persona numéricamente hablando, no hay problema en decir que son la misma persona desde un sentido cualitativo.

Ahora, consideremos un caso un poco más complicado que el anterior: la máquina teletransportadora funciona mal y en lugar de destruir al Dr. Jekyll, la máquina genera a otra persona: Dr. Jekyll\*. Tanto el Dr. Jekyll como el Dr. Jekyll\* viven después del experimento sus vidas normales. La cuestión aquí es la siguiente: ¿el Dr. Jekyll sobrevivió?; ¿qué relación hay entre el Dr. Jekyll y su duplicado?; ¿cuál de las dos personas es mejor candidata para ser la misma persona que se introdujo en la máquina?

La historia puede empeorar aún más. Imaginemos un caso de teletransportación que resulta en una fisión. El escenario es similar a los anteriores, pero cuando el asistente aprieta el botón de la máquina, ésta genera dos duplicados cualitativamente indiscernibles, pero compuestos de materia distinta. Llamemos al primer duplicado “Dr. Jek” y al segundo “Dr. Kyll”. En este caso, la teletransportación destruye al Dr. Jekyll. De esta manera, la pregunta que aquí nos interesa responder es: ¿hay algún candidato mejor que otro para ser la persona que continúa siendo el Dr. Jekyll? De nuevo, tenemos cuatro respuestas posibles:

- (R1) Dr. Jekyll es el Dr. Jek.
- (R2) Dr. Jekyll es el Dr. Kyll.
- (R3) Dr. Jekyll es el Dr. Jek y Dr. Kyll.
- (R4) Dr. Jekyll no es ni el Dr. Jek ni el Dr. Kyll.

## 4. Teorías de Continuidad para la Identidad Personal

Con las herramientas que tenemos hasta el momento no es posible dar una respuesta teóricamente justificada para cada uno de los casos presentados. El objetivo de esta sección es presentar dos de las teorías de continuidad más establecidas en la literatura clásica de IP. Si bien el objetivo de esta sección es meramente expositivo, más adelante veremos qué respuestas ofrece cada una de estas posiciones a los casos de fusión y fisión.

### 4.1. Continuidad Psicológica

Hay teorías filosóficas que sostienen que, de todos los aspectos relevantes para IP, hay algunos más sustantivos para dar cuenta de nuestra persistencia. Estas teorías son reductivas en el sentido de que reducen lo que importa mantener para IP a elementos o características muy específicas (véase Parfit, 1984). La idea general de las teorías reductivas es que para que tengamos IP, sólo es necesario mantener algunos rasgos importantes. Uno de los rasgos que muy plausiblemente sean relevantes para IP son los psicológicos. De acuerdo con esta concepción, que las personas persistan en el tiempo sólo consiste en que las personas mantengan algún tipo de continuidad psicológica a lo largo de su existencia.

Un antecedente de este tipo de teorías es la teoría lockeana acerca de IP que se concentra en un rasgo psicológico muy específico: la memoria. De acuerdo con Locke (1690), si yo recuerdo haber escrito este texto, yo soy la misma persona que escribió este texto. Si el Dr. Jekyll recuerda haber cometido todos los crímenes imputados, es la misma persona y merece ser castigada. Para Locke, la memoria es de suma importancia pues gracias a ella podemos distinguarnos de otras personas. Locke creía que  $x$  y  $y$  son dos personas distintas si y sólo si las memorias de  $x$  son distintas a las de  $y$ .

Claramente, el criterio lockeano de la memoria, al ser tan general, se presta a objeciones. Uno de los críticos más famosos a la visión Lockean de IP es Reid (1764; 1785), para quien la memoria no puede ser un criterio—metafísicamente hablando—ni suficiente ni necesario para IP. Para Reid, es imposible dar cuenta de cómo es que las personas persisten considerando algo tan simple como la memoria. Si bien Reid cree que la memoria es un rasgo psicológico importante, él no la considera esencial para dar cuenta de nuestra persistencia (Copenhaver, 2014).

Hay tres razones que pueden rastrearse en la crítica de Reid contra la idea de que la memoria es lo fundamental para IP. La primera consiste en que no siempre recordamos todo lo que hemos hecho o experimentado. La segunda razón es que la memoria, como rasgo aislado, no podría ser el único rasgo relevante para IP. Finalmente, la tercera razón de Reid tiene relación con la naturaleza de las memorias. De acuerdo con él, para que un estado mental sea una memoria genuina, la persona que recuerda algo debe ser idéntica a la que experimentó ese algo. Pero lo anterior es circular: para definir lo que es una memoria presuponemos IP y para explicar IP presuponemos un concepto de memoria genuina.

Como consecuencia de las críticas anteriores, y como lo sostienen otros autores, si bien no es posible dar cuenta de IP con un solo rasgo psicológico como la memoria, se pueden encontrar otras alternativas. Una buena estrategia para definir una teoría psicológica para IP es considerar una noción de continuidad psicológica de manera más amplia, i.e., sin un rasgo psicológico específico como criterio de IP. Este tipo de visiones pueden incluir una variedad de elementos psicológicos relevantes, como, por ejemplo: los deseos, la memoria, las creencias, la personalidad (asumiendo que ésta depende de los estados psicológicos), etc.

Considerando esta estrategia, podemos definir una teoría de continuidad psicológica de la siguiente manera:

(Teoría de Continuidad Psicológica): A y B son *la misma persona* sii A y B son *continuos psicológicamente*, i.e., sii los estados psicológicos de B son continuos con los estados psicológicos de A.

La definición anterior es algo vaga, pues no es del todo claro qué quiere decir que los estados psicológicos de una persona sean continuos. Pero que la noción de “continuidad psicológica” no sea del todo precisa tiene que ver con que las nociones de “propiedad psicológica” o “estado psicológico” no son precisas. Haciendo a un lado este pequeño problema, podemos tomar la siguiente caracterización propuesta por Lewis (1976): si existen ciertos patrones psicológicos que se mantienen de un tiempo a otro, es decir, si hay un vínculo psicológico fuerte entre mis estados psicológicos actuales y los pasados (o futuros), entonces, habrá continuidad psicológica. Por ejemplo, un estado psicológico que tengo en  $t_1$  de una creencia de que hay un monstruo bajo la cama está conectado con otro estado de sensación de miedo en  $t_2$ . Ambos estados psicológicos me impulsan a levantarme de la cama, prender la luz y cerciorarme de que no hay tal monstruo bajo la cama en  $t_3$ . Dado que revisé que no hay nada bajo de la cama, me formo otro estado psicológico de creencia de que no hay un monstruo bajo de la cama en  $t_4$ . La creencia de que hay un monstruo continúa de  $t_1$  a  $t_2$ , y cambia en  $t_4$ . En  $t_4$  podría tener un estado psicológico de recordar que en  $t_1$  tenía miedo de un monstruo. Todos estos estados psicológicos manifiestan cierta continuidad que explica mi comportamiento. De esta manera, la definición mencionada es reforzada con una caracterización más precisa de qué tipo de conexión hay entre un estado psicológico y otro.

Hay otras dos condiciones que, como veremos más adelante, son necesarias para IP a la luz de las teorías tetradimensionales de la persistencia, estas son la dependencia causal y la similitud. Lewis (1976) llama a la suma de estas dos condiciones “conexión psicológica”. La condición de causalidad, así como la de similitud serán fundamentales porque podemos tener casos en los que aparentemente habría IP pero que, si se evaluamos el caso a la luz del criterio de causalidad o similitud, tal apariencia desvanece. Por cuestiones de claridad en la presentación de la teoría de continuidad psicológica de forma independiente de la visión tetradimensionalista, me quedaré por el momento con la noción de continuidad

como fue descrita y posteriormente haré un intento por refinar la noción en cuestión.

## 4.2. Continuidad Física

Hay otros autores que piensan que los rasgos relevantes para dar cuenta de IP son físicos. Algunos de estos autores se inclinarían a decir, por ejemplo, que yo soy la misma que ayer porque tengo el mismo cuerpo o, por lo menos, que éste luce exactamente igual que ayer. Pero, consideremos el caso del Dr. Jekyll y Mr. Hyde mencionado en la Introducción: los cuerpos de ambos lucían totalmente distintos antes y después de tomar el brebaje que convierte al Dr. Jekyll en el Sr. Hyde, ¿diríamos que son personas distintas? O sin ir tan lejos, consideremos la posibilidad de que en unos años me extraigan algún órgano de mi cuerpo: esta extracción haría que mi cuerpo sea distinto en sentido estricto al de hoy y, en este caso, parece difícil aceptar que eso es suficiente para que deje de haber IP. Estas teorías corporales son conocidas en la literatura como teorías de continuidad corporal y, como podemos ver a simple vista, son problemáticas. Básicamente, es necesario argumentar por qué hay partes del cuerpo de una persona que no son lo suficientemente relevantes o esenciales para su IP.

Como opción alternativa a la teoría anterior, podemos considerar una parte física mínima que puede ser privilegiada para dar cuenta de IP. Esta opción, quizá la más aceptada entre los teóricos de las ciencias cognitivas, psicólogos y algunos filósofos de la mente, podría ser el cerebro. Parece intuitivo pensar que el cerebro es el órgano encargado de la mayoría de funciones vitales y de los rasgos psicológicos. Esto podría llevarnos a pensar que el cerebro puede cumplir el rol de la estructura física mínima funcional que podríamos rastrear para dar cuenta de IP.

De esta manera, mi cerebro, y no cualquier otro órgano, por ejemplo, sería aquello que tenemos que rastrear para el criterio de continuidad física. Una teoría que defiende algo así sería una teoría de continuidad cerebral.



Pero de nuevo, ¿será relevante que la totalidad de mi cerebro se mantenga para respetar el criterio de continuidad? Recordemos que la estrategia que seguí en el caso de la continuidad psicológica fue optar por una noción de continuidad general y no muy específica. Las nociones de continuidad concretas o demasiado específicas pueden ser problemáticas porque no ayudan en algunos casos relevantes tales como los de fusión y fisión anteriormente mencionados. El caso de la continuidad cerebral sería un caso de continuidad física muy concreto: establece que la continuidad física sólo depende de la continuidad del (mismo) cerebro físico. Sin embargo, hay casos en los cuales a pesar de no haber continuidad cerebral en términos de mismidad de materia (como el caso del Dr. Jekyll y sus réplicas el Dr. Jek y el Dr. Kyll, mencionados en el ejemplo de la fisión con teletransportación, o en el caso simple de teletransportación) parece que hay persistencia a pesar de no haber continuidad física. Por la razón anterior entenderé la noción de continuidad física en un sentido más general que es compatible con el materialismo funcionalista.

Con base en lo que hemos dicho podemos definir la teoría de continuidad física:

(Teoría de Continuidad Física): A y B son la *misma* persona sii A y B son continuos físicamente, i.e., sii B tiene una estructura física mínima lo suficientemente compleja y semejante a la de A como para satisfacer sus mismos roles funcionales.

La idea detrás de la definición anterior es la siguiente. De acuerdo con el materialismo funcionalista, podemos argumentar que la continuidad física involucra o implica cierta continuidad funcional y que todo cambio o variación a nivel funcional de una persona requerirá de un cambio o variación a nivel de la estructura física mínima (o realizador funcional físico) de esa persona. Es importante notar que la definición dada no requiere de un realizador físico específico como el cerebro, aunque quizá en la mayoría de casos el cerebro sea un buen candidato para ser el tipo de estructura física mínima (y realizador funcional físico) capaz de satisfacer los roles

funcionales de una determinada persona. La definición en cuestión tampoco requerirá que sólo una y la misma estructura física (o realizador funcional) a través del tiempo sea capaz de satisfacer los mismos roles funcionales, es decir, la definición parece ser compatible con el hecho de que podamos tener distintos realizadores o estructuras físicas que satisfagan un mismo rol funcional. Esta noción de continuidad física no es una noción muy explorada y como veremos a continuación esta caracterización nos permitirá solucionar los casos de fusión y fisión de un modo más natural.

### 4.3. ¿Continuidad Psicológica vs. Continuidad Física?

Hasta este momento contamos ya con dos herramientas teóricas para ver cómo es que podríamos justificar una respuesta determinada a los casos de fusión y fisión. Sin embargo, antes de presentar las soluciones que cada una de las teorías de continuidad nos darían, es de suma importancia decir por qué es que las dos teorías presentadas de la manera que lo hice no colapsan.

Alguien muy astutamente puede pensar que tal y como se describió la teoría de continuidad física presupone la teoría de continuidad psicológica. Esto tiene como consencuencia que las respuestas de cada una de estas teorías a los casos de fusión y fisión sean simétricas: i.e., ambas ofrecen las mismas respuestas en cada uno de los casos. Pero consideremos lo siguiente.

Parece que para poder afirmar que T es una buena teoría de IP, T debe dar cuenta de casos obvios de IP. Si T da cuenta de los casos obvios entonces, parece que T debe rastrear las características físicas y psicológicas que dan lugar a la persistencia de una persona. La idea es la siguiente, si T rastrea las características psicológicas de una persona, T rastreará el cerebro; y si T rastrea las características físicas de una persona, T rastreará el cerebro. La razón de lo anterior es que, si el materialismo al estilo Lewis es el caso, la psicología de una persona debe ser realizada en un cuerpo físico. Parece que hay partes del cuerpo que, por sí mismas, no son capaces de realizar rasgos psicológicos: mi brazo no puede realizar mi psicología, mis piernas tampoco, etc. No obstante, si tuviéramos que ir descartando partes

corporales que puedan realizar propiedades psicológicas, llegaríamos a la conclusión de que el cerebro es la parte física mínima del cuerpo que puede realizar estas propiedades psicológicas. Esto implica que, si rastreamos la psicología, rastreamos al cerebro. Por otro lado, supongamos que queremos rastrear el cuerpo. De nuevo, parece que hay partes del cuerpo de una persona que, por sí mismo, no dan lugar a la persistencia: mi útero es una parte que no parece relevante para dar cuenta de mi IP. Si vamos caso por caso, llegaremos a la conclusión de que el cerebro es la parte física del cuerpo que puede dar cuenta de la persistencia. De esta manera, si rastreamos la continuidad física de una persona para dar cuenta de IP, entonces rastreamos el cerebro.

De las intuiciones mencionadas, parece que si queremos dar cuenta de IP debemos rastrear el cuerpo y la psicología. Pero aparentemente si rastreamos ambas características, llegaremos a la conclusión—totalmente intuitiva—de que hay una conexión muy importante entre la psicología y cerebro: al rastrear el cuerpo o rastrear la psicología nos llevará directo al cerebro. El cerebro será, en casos no tan extraños de IP, la estructura física mínima que da lugar a la persistencia de una persona.

Por lo tanto, no es que la continuidad física presuponga la continuidad psicológica. Más bien, parece que rastrear cualquiera de ambas continuidades nos conducirá a que el cerebro tiene un vínculo importante con la continuidad física y psicología de las personas. Por esta razón, parece que una buena teoría para la IP debe dar cuenta del vínculo anterior para atender a casos obvios y exóticos de IP.

#### 4.3.1. Soluciones al Caso de la Fusión

La pregunta problemática del experimento mental de fusión era la siguiente: ¿quién es C? Es preciso recordar que lo único que sabemos es que C es la persona que resulta de juntar el cerebro del gemelo A y el cuerpo (a excepción del cerebro) de B. A continuación, veremos que, de acuerdo con ambas teorías de continuidad, la respuesta más natural sería que C es la misma persona que A, o sea (R1).

De acuerdo con la teoría de continuidad física, para que dos personas sean las mismas es suficiente con que preserven una estructura física (mínima) lo suficientemente compleja y semejante como para mantener los mismos roles funcionales de una persona. De acuerdo con nuestro caso, tenemos dos opciones para satisfacer aquella estructura física. Por un lado, tenemos el cerebro y por otro lado el resto cuerpo. Parece que el cuerpo con un cerebro satisface el criterio de continuidad física. Sin embargo, sería extraño decir que un cuerpo sin cerebro, por sí solo, podría ser una estructura física mínima lo suficientemente compleja como para mantener o soportar ciertos rasgos funcionales. De esta manera, la opción del cuerpo sin cerebro queda descartada. Como C sólo está formado por el cuerpo de B sin su cerebro, B no es el mejor candidato para ser la misma persona que C. El segundo candidato para satisfacer el criterio de continuidad física sería el cerebro. Plausiblemente el cerebro sí puede ser considerado, en este caso, una estructura física mínima lo suficientemente capaz de mantener o soportar los rasgos funcionales de una persona. Por esta razón, la opción del cerebro (entendido en términos de estructura o realizador funcional físico) parece ser una opción más viable. Consideraremos que el cerebro es la estructura física relevante que (en este caso) nos ayudará a responder la pregunta “¿quién es C?”. Sabemos que el cerebro de A es el mismo que el cerebro de C. Ahora bien, si consideramos que el cerebro de A es la estructura física mínima, lo suficientemente compleja como para mantener sus estados funcionales, A es un continuo físico de C. Por esta razón, diríamos que A es la misma persona que C. La respuesta (R1), analizada desde la base de la teoría de continuidad física sería la más natural para el caso de la fusión.

La teoría de la continuidad psicológica nos dice que es suficiente que dos personas sean continuas psicológicamente para que sean las mismas personas. Desde la base de una teoría materialista, si fijamos lo físico, fijaremos también lo mental. Tenemos buenas razones para pensar que el realizador físico de lo mental es el cerebro. Así, si A y C tienen el mismo cerebro, seguramente tendrán conti-

nidad psicológica. Y si A y C son continuos psicológicamente, serán la misma persona. En consecuencia, (R1) sería la respuesta más natural al caso de la fusión de los gemelos.

Como podemos ver, las dos teorías de continuidad responden de manera simétrica al caso de la fusión.

#### 4.3.2. Soluciones al Caso de la Fisión

A continuación, veremos algunas posibles respuestas al caso de la fisión. Aquí veremos que la respuesta más natural, sobre la base de las teorías de continuidad física y psicológica, es que Dr. Jekyll es la misma persona que ambos duplicados Dr. Jek y Dr. Kyll, o sea, (R3). La información que teníamos en este experimento mental es sólo que los duplicados generados a partir de la información del Dr. Jekyll eran duplicados *simpliciter*.

De acuerdo con la solución física, si ambos duplicados son duplicados *simpliciter*, ambos duplicados tienen una estructura física cualitativamente idéntica. La forma en la que definimos la teoría de la continuidad física no exige que la estructura física capaz de mantener los estados funcionales de una persona tenga que ser exactamente la misma (en términos de la composición de su materia, por ejemplo) que la de otra para que digamos que son la misma persona. De esta manera, los duplicados del Dr. Jekyll no tienen que tener el mismo cerebro (numéricamente) que el del Dr. Jekyll para que haya continuidad física. Es suficiente que los cerebros del Dr. Jek y del Dr. Kyll sean realizadores funcionales físicos (aunque materialmente distintos) lo suficientemente complejos como para mantener los estados funcionales del Dr. Jekyll. La identidad cualitativa entre cerebros cumple con este supuesto. Luego, como la estructura física de los duplicados incluye la estructura física y funcional del cerebro del Dr. Jekyll, parece que ambos duplicados son continuos físicamente con el Dr. Jekyll. Nuestra teoría de continuidad física nos diría, por tanto, que el Dr. Jekyll es la misma persona que el Dr. Jek y el Dr. Kyll. En consecuencia, (R3) es la respuesta más apropiada al caso de la fisión.

De acuerdo con la teoría de continuidad psicológica para IP, si dos personas son continuas psicológicamente, son la misma persona. La respuesta más inmediata al caso de la fisión sería que Dr. Jekyll es la misma persona que Dr. Jek y Dr. Kyll. El argumento para esta respuesta sería el siguiente. Sabemos que ambos duplicados son duplicados *simpliciter* del Dr. Jekyll. Esto implica que ambos duplicados tienen una estructura física y psicológica cualitativamente idéntica a la del Dr. Jekyll. Por esta razón, los duplicados deben mantener sus rasgos psicológicos. Entonces, Dr. Jekyll sería continuo psicológicamente al Dr. Jek y al Dr. Kyll. Luego, de acuerdo con el criterio de la continuidad psicológica, Dr. Jekyll sería la misma persona que el Dr. Jek y el Dr. Kyll. Sobre esta base, por lo tanto, la respuesta más natural al caso de la fisión es (R3).

#### 4.3.3. Solución con Cláusula de no Fisión

Hay otra solución posible al caso de la fisión con continuidad psicológica. Esta respuesta está basada en la cláusula de la no fisión (Parfit 1984, Kagan 2012). A grandes rasgos, esta regla establece que, si hay fisión, no hay IP (incluso si hay continuidad psicológica o continuidad física). Esto implica que sólo en un caso normal (sin fisión) con continuidad psicológica o física, tendremos a la misma persona. Una de las motivaciones para optar la cláusula de la no fisión es mantener la relación de identidad numérica entre los continuantes físicos o psicológicos de una persona. En un caso de fisión con continuidad psicológica o física, como el del Dr. Jekyll, no podremos concluir que él sobrevive al experimento. La respuesta que nos daría alguien que sostenga la cláusula de la no fisión, tendría que ser (R4). No son lo misma persona debido a que éste es un caso de fisión. No obstante, si (por medio de la cláusula de no fisión) decimos que ninguno de ellos es la misma persona que Dr. Jekyll, deberíamos asumir que, a pesar de que haya continuidad física o psicológica entre Dr. Jekyll y sus duplicados, no hay IP.

Esto querría decir que las teorías de continuidad psicológica o física como criterio de posibilidad para la IP fallan en casos exóticos como los de la fisión si se considera la identidad numérica

como criterio para mantenerla. La motivación (Kagan, 2012) para adoptar esta cláusula es que no tengamos que comprometernos con la respuesta (R3) dado que esta solución es problemática. Su adopción, sin embargo, nos llevaría a concluir que las teorías de continuidad psicológica o física no son suficientes para IP.

Si bien puede parecernos atractivo adoptarla, es preciso notar que tiene su lado controvertido (Kagan, 2012). El problema de aceptarla es que la IP se sustentaría sobre ciertos hechos extrínsecos. La cláusula de la no fisión establece que los hechos acerca de mi persistencia dependen de que no pueda haber alguien (en otro lugar) que mantenga continuidad psicológica o física conmigo. Es decir, dependería de hechos externos a mí y no de hechos meramente internos. Pero parece que intuitivamente los hechos acerca de nuestra IP deberían consistir en mantener ciertas relaciones de continuidad o semejanza entre las propiedades intrínsecas de distintos continuantes.

## 5. Tetradimensionalismo e Identidad Personal

Hasta aquí he desarrollado dos teorías de continuidad clásicas en la literatura de IP. También he expuesto dos experimentos mentales y la manera en la que cada una de las teorías de continuidad responden a estos escenarios. Como vimos, las respuestas de las dos teorías son simétricas y no parece haber razones para elegir una sobre la otra. Sin embargo, las teorías de continuidad deben complementarse con alguna teoría de persistencia de objetos. Este artículo no trata el problema de la persistencia y el cambio en los objetos concretos, pero utilizaré algunas herramientas conceptuales de este debate. Sólo me concentraré en una posición específica: la teoría tetradimensional de los estadios temporales [stage view] o, también llamado, exdurantismo [exdurantism].

De acuerdo con la concepción de los estadios temporales, los objetos ordinarios, también llamados particulares concretos, persisten si “exduran”, i. e., si tienen diferentes estadios temporales que están relacionados por medio de la relación de contraparte temporal (“CT” en adelante).

Según el exdurantista, yo soy la misma persona que ayer en virtud de que tengo un estadio temporal pasado que es CT de mi estadio temporal presente entre estos estadios temporales. La manera de entender “misma persona” para el exdurantista no es en el sentido de identidad numérica porque la relación de CT es más débil. Es muy importante notar que, para el exdurantista, las personas, así como los gatos y las mesas, son sólo idénticas a estadios temporales instantáneos, y no a la colección o suma de sus estadios temporales (como en el caso del perdurantismo); cada estadio temporal está sólo en una relación de contraparte. Antes de continuar, es importante contrastar a grandes rasgos el exdurantismo del perdurantismo y el endurantismo (una posición tridimensional sobre la persistencia) y mencionar algunas de las motivaciones por las que en este artículo me enfocaré en el exdurantismo.

Tanto el perdurantismo (Lewis 1976, 1986) como el exdurantismo (Sider 2001, 2010) son teorías tetradimensionales de la persistencia, ya que ambas coinciden en que los objetos materiales se extienden en cuatro dimensiones: tres espaciales y una temporal. En la primera, los objetos no solo tienen partes espaciales, sino también partes temporales; mientras que en la segunda no existe un único objeto tridimensional en el transcurso del tiempo, sino distintos objetos tridimensionales (o estadios temporales) en distintos tiempos. Estas teorías difieren significativamente en su concepción acerca de cómo persisten los objetos a lo largo del tiempo. El perdurantismo, también conocido como “la teoría de gusano” [*worm view*], sostiene que los objetos persisten como una suma de partes temporales conectadas que forman un único objeto continuo: un gusano espacio-temporal. Desde esta perspectiva, un objeto nunca está completamente presente en un instante específico, sino que es idéntico a la totalidad de sus partes temporales, es decir, al gusano espacio-temporal en su totalidad.

Por otro lado, el exdurantismo, o “teoría de estadios” [*stage view*], propone que los objetos son estadios temporales instantáneos: partes temporales únicas que no se extienden en el tiempo, lo que los hace similares a los objetos tridimensionales. Según el



exdurantismo, la persistencia de lo que consideramos “un mismo objeto” se explica mediante relaciones de contrapartes entre estos estadios temporales, relaciones que se fundamentan en la continuidad causal, la similitud y conexiones causales. En cambio, el perdurantismo interpreta la persistencia como un cambio o variación de propiedades intrínsecas entre las partes temporales de un objeto a través del tiempo. Cuando digo algo como “yo fui una niña”, el exdurantismo no hará referencia a un gusano espacio-temporal que abarca toda mi existencia, como sostendría el perdurantismo, sino únicamente a un estadio temporal específico del pasado.

En contraste con el exdurantismo, el endurantismo (que es una teoría tridimensional) argumenta que los objetos persisten en virtud de estar totalmente presentes en cada instante de su existencia. Es decir, un mismo objeto existe totalmente a través del tiempo, independientemente de los cambios que pueda experimentar. Mientras que el exdurantismo interpreta a los objetos como una serie de distintos estadios temporales (objetos tridimensionales), relacionados entre sí.

Para los fines de este capítulo el exdurantismo resulta una opción teórica más conveniente. Esto ya que ofrece una explicación coherente sobre cómo un objeto puede mantener una continuidad en el tiempo a través de relaciones de contrapartes, permitiendo así un análisis en términos de los requisitos de continuidad para IP analizados anteriormente.

### **5.1. Contrapartes Temporales: Tres Condiciones para la Identidad Personal**

La versión original o canónica de la teoría de contrapartes es propuesta por David Lewis (1986). Uno de los objetivos de Lewis con la versión modal de la teoría de contrapartes es que ésta puede servir como alternativa lógica y filosófica para abordar el problema modal acerca de la identidad a través de mundos posibles y al problema de cómo es posible representar *de re* que un individuo tiene distintas propiedades modales. Con base en la teoría de contrapartes, el enunciado “Angélica pudo haber perdido un brazo” se tra-

duce en el lenguaje de contrapartes como “existe una contraparte modal de Angélica que, en otro mundo posible, pierde un brazo”.

No hay nada en la teoría de contraparte modal que impida hacer una versión temporal de ella (Sider, 2010) para construir la teoría exdurantista de la persistencia. El objetivo de usar una teoría de CT en el debate de la persistencia es dar cuenta de enunciados temporales tales como “Angélica fue una niña” o “Angélica será la misma persona que envíe este artículo”. De acuerdo con Lewis (1976), hay por lo menos tres condiciones que los estadios o partes temporales de un objeto deben cumplir para considerarse CT entre sí:

- (1) que  $x$  e  $y$  sean similares;
- (2) que  $x$  e  $y$  estén conectados por una cadena de dependencia causal;
- (3) que  $x$  e  $y$  sean continuos.

Ahora bien, ¿qué tipo de relaciones de similaridad y continuidad son las relevantes para dar cuenta de cómo persisten desde el marco exdurantista, en particular, las personas? Esto depende de qué es lo que consideremos que es una persona y, como vimos anteriormente, los aspectos o rasgos que se quieran rastrear para dar cuenta de la continuidad. Como decía, yo soy similar a Billie, mi gata, en tanto que ambas estamos compuestas de materia, etc. Pero si ésta fuera una relación de similaridad relevante, podríamos llegar a conclusiones muy poco intuitivas, por ejemplo, que un estadio temporal de Billie es CT de algún estadio temporal mío. Cualquier teoría satisfactoria de IP debe descartar fácilmente casos como el anterior y, por ende, clarificar qué relaciones de similaridad o continuidad son las que importan. Si los objetos son estadios temporales, necesitamos que una teoría de IP que asuma esta posición pueda diferenciar mis estadios temporales de los de cualquier otro objeto (por ejemplo, de Billie) y en particular de los de cualquier otra persona.

Como ya he mencionado, de manera intuitiva, para responder la cuestión acerca de cómo persisten las personas se vuelve necesario

decir algo acerca de qué consiste ser una persona. Mencione que ser una persona tiene relación con propiedades o rasgos físicos y psicológicos. De esta manera, las condiciones de similaridad y continuidad de la teoría de CT, para el caso de IP, debe rescatar o incluir estos dos rasgos de manera en la que se suplementen entre sí.

Supongamos que hoy todas y cada una de mis propiedades psicológicas serán transferidas a un cuerpo distinto y, posteriormente, mi cuerpo entero será destruido. Además de lucir distinto, el cuerpo al que será transferida mi psicología, éste cuenta con un cerebro de silicio. Por *mor* del argumento, supongamos que como resultado de la transferencia tenemos a una persona. La pregunta interesante es la siguiente: ¿seré yo la persona que resulte después de la transferencia o no lo seré? Si nos concentramos sólo en las propiedades físicas, no hay forma de sostener (al menos de manera obvia) que yo sobreviví a la transferencia.

Pero podríamos ser más optimistas y decir que si la persona resultante de la transferencia puede pensar como yo, tener las mismas creencias que yo, tener la misma personalidad que yo, las mismas metas, etc., esa nueva persona seré yo. Considerando las propiedades psicológicas, parece tenemos mejores razones para sostener que, en efecto, seré yo. La nueva persona con cerebro de silicio, intuitivamente, va a satisfacer las condiciones que necesitamos para decir que yo seré esa persona: será similar psicológicamente a mí, dependerá causalmente de mí, y será continua a mí, pues no habría cambio radical a nivel psicológico. Este escenario nos ofrece por lo menos una motivación por la cual considerar el asunto de la similaridad y continuidad en términos únicamente de propiedades psicológicas.

La idea de que las propiedades psicológicas son las que nos permitirán hablar de similaridad y continuidad entre estadios temporales es compatible con la postura materialista funcionalista descrita arriba. Esta postura nos compromete, en primer lugar, con la tesis de que las propiedades psicológicas supervienen de las físicas. A saber, cualquier cambio o variación a nivel de las propiedades psicológicas requiere igualmente un cambio o variación a nivel

de las propiedades físicas. Esta tesis es también compatible con la posición funcionalista, según la cual, los estados psicológicos no son más que estados funcionales. En el caso imaginario de arriba, podremos decir que, si la nueva persona es psicológicamente isomorfa a mí, también lo será a nivel funcional. Si el materialismo funcionalista es el caso también tenemos que comprometernos con la idea de que los estados psicológicos deben ser realizados por un realizador funcional que, necesariamente sea físico. Asumir estas tesis es útil para sostener que siempre que mantengamos el realizador funcional, mantendremos la psicológica de una persona. Será importante tener en cuenta esto ya que más adelante usaré estos supuestos para trabajar algunos casos de IP.

### **5.1.1. Similaridad y Dependencia Causal**

La relación de CT es una relación de similaridad o semejanza entre los distintos estadios temporales de un individuo. El estadio temporal de quien comenzó a escribir este artículo es similar al estadio temporal de quien lo escribe ahora en varios sentidos. Los aspectos en los que nos enfocaremos serán los psicológicos, porque me parecen ser más relevantes. Dos estadios temporales  $x$  e  $y$  son similares si comparten propiedades psicológicas.

Que dos estadios temporales sean similares en virtud de compartir propiedades psicológicas no quiere decir que tales estadios deban tener exactamente las mismas propiedades psicológicas. Mi estadio temporal de ayer tiene la mayoría de propiedades psicológicas en común con el de hoy. Pese a ese cambio, mis estadios siguen siendo similares dado que ambos tienen la mayoría de propiedades psicológicas en común.

Aquí algo importante a destacar es que la noción de CT es sensible a contextos de manera que un estadio  $x$  e  $y$  sean CT dependerá del contexto de similaridad seleccionado. Así, tendremos distintas CT dependiendo de las similaridades que consideremos relevantes (Divers, 2002). Para IP éstas son similitudes en un contexto de rasgos psicológicos.

Por otro lado, la dependencia causal es otra condición que cualesquiera dos estadios temporales deben cumplir para ser tomados como CT entre sí (Lewis 1976, 1986). Esta característica consiste en que los estadios temporales de una persona deben tener poderes causales sobre sus otros estadios. Por ende, hay relaciones de causa y efecto entre estadios temporales. Por ejemplo, supongamos que, en  $t_1$ , estoy escribiendo este texto con la intención de completar un capítulo. Este estadio está causalmente conectado con un estadio futuro,  $t_2$ , en el que gracias a la confianza de poder publicarlo lo entrego ya listo para su evaluación. A grandes rasgos, para que un estadio temporal futuro sea CT de mi estadio temporal presente, no es suficiente que el estadio futuro sea similar y continuo con el presente; también es necesario que existan cadenas de dependencia causal entre los estadios, lo cual implica que sus estados psicológicos, i.e., que las cosas que uno crea, piense, sienta, etc., estén conectadas causalmente.

### **5.1.2. Continuidad**

En la cuarta sección hablé sobre teorías de continuidad psicológica y mencioné que una manera de entender esta noción es de la siguiente manera: una persona es continua psicológicamente si sus estados psicológicos son continuos. Esta definición es imprecisa. Para reforzarla, en esta sección la definiremos en términos de cambio gradual de propiedades psicológicas. Esta forma de entender la noción de continuidad es compatible con la intuición detrás de la noción que ya habíamos dado en el capítulo anterior, pues si recordamos, en algún momento mencionamos que la continuidad psicológica podía entenderse como la existencia de patrones psicológicos que se mantienen de un tiempo al otro. Sin embargo, con el objetivo de no confundir la noción de continuidad con la de similaridad, la noción de continuidad entendida como cambio gradual de propiedades nos puede servir para clarificar en qué sentido los patrones psicológicos de los que hablábamos anteriormente se mantienen. El sentido de “mantener”, en el caso de la continuidad, no es el mismo de “compartir” que usamos para definir la noción de similaridad. La

continuidad requiere una conexión temporal y causal que explique cómo una propiedad P se conserva o cambia de manera gradual entre los estadios, mientras que la similitud solo implica la posesión de propiedades comunes sin esa conexión temporal. Por lo tanto, sería posible considerar la continuidad como una relación primitiva no reducible a la similitud, ya que su definición incluye condiciones temporales y causales que van más allá de la mera comparación de propiedades. Considerando lo anterior, diremos que dos estadios temporales son continuos a través del tiempo si o bien el cambio de las propiedades psicológicas de ambos estadios es *gradual*, o bien no hay cambio de propiedades psicológicas.

Alguien podría pensar que en realidad la continuidad entre estadios y la similaridad son una y la misma cosa. Pero esto no es así, consideremos por ejemplo el caso de Matusalén, una persona que vive cientos de años. Supongamos que a los cien años Matusalén era una buena persona y gradualmente se convierte en una mala persona. A los cuatrocientos años Matusalén será el ser más perverso y despreciable de todo el mundo. En este escenario, el cambio de Matusalén es gradual, o sea, no se hace perverso de un momento a otro. Esto último implica que Matusalén es continuo a través del tiempo. Pero, si tomamos el estadio temporal de Matusalén a los 100 años y a su estadio temporal de 400 años, nos daremos cuenta que no hay similaridad alguna. De esta manera, la condición de continuidad es lógicamente independiente de la de similaridad y *viceversa*.

## 6. Tratamiento Exdurantista

El objetivo de la sección anterior fue explicar en qué consiste la relación de CT como herramienta teórica para el problema de IP. Las condiciones que los estadios temporales de una persona deben cumplir para ser CT son: similitud, dependencia causal y continuidad. En esta sección el objetivo será revisar cómo es que podemos aplicar el modelo exdurantista a los casos que hemos visto en la Sección 3.

## 6.1. Tetradimensionalismo: Fusión y Fisión

Antes de presentar las soluciones, es necesario decir algo acerca de una constricción que es deseable respetar: cualquier modelo de persistencia de personas debe poder dar las mismas soluciones que se dieron sobre la base de las teorías de continuidad, las cuales son capaces de explicar las intuiciones que tenemos acerca de IP. La idea que a lo largo de este artículo se ha tratado de motivar es que asumiendo, por *mor* del argumento, que el exdurantismo tiene algunas ventajas sobre otras teorías de persistencia (como el perdurantismo o el endurantismo), otro de los beneficios de la teoría exdurantista es que logra compatibilizar el materialismo funcionalista con la noción de continuidad psicológica que es fundamental para la IP.

### 6.1.1. El Caso de la Fusión

Recordemos que la solución que las teorías de continuidad habían dado a este experimento fue la opción (R1), i.e., que C es la misma persona que A debido a que el cerebro de A es trasplantado a C. Dentro del modelo tetradimensional para la IP, la relación de CT es la que nos permitirá dar cuenta de alguna solución para este caso. Es importante resaltar que para fijar una relación de CT, es necesario satisfacer tres condiciones: similaridad, dependencia causal y continuidad. Teniendo estas condiciones en cuenta, veamos qué solución podemos extraer del modelo exdurantista.

Antes que nada, tenemos que decir por qué algunas soluciones no pueden ser el caso a la luz de nuestro modelo. Hemos dicho arriba que el tipo de propiedades que consideraremos relevantes para dar cuenta de IP, en términos de CT, son las propiedades psicológicas. Esto excluye dos casos. Por un lado, el caso en el que C es la misma persona que B y, por otro lado, el caso en el que C es la misma persona que A y B.

El estadio temporal C no puede ser la misma persona que B debido a que no satisface las condiciones para ser CT futura de B. Esto quiere decir que C no es similar a B en el sentido que consideramos relevante porque las únicas propiedades que podrían compartir son físicas (en relación con su cuerpo) y esto no es su-

ficiente para garantizar la IP ya que sería controvertido decir que un cuerpo sin cerebro es capaz de realizar la psicología de una persona, lo cual es justo lo que necesitamos. Además, C y B no están conectados causalmente. Las propiedades psicológicas o estados psicológicos de C no dependen causalmente de las de B. Finalmente, C no es continuo psicológicamente con B, y esto es así porque, en un caso como éste, no es claro cómo rastrear un cambio gradual de las propiedades psicológicas de B hasta las de C. De esto sigue que C no puede ser contraparte temporal futura de B y, por ende, B y C no pueden ser la misma persona. Esto implica, a su vez, que C tampoco puede ser la misma persona que A y B, dado que, como acabamos de ver, es falso que C sea la misma persona que B.

Ahora nos queda por cubrir el caso en el que C sería la misma persona que A. Este caso es el que habíamos privilegiado anteriormente. La cuestión ahora es revisar cómo podríamos explicarlo en términos del modelo tetradimensional exdurantista para IP. Si decimos que C es la misma persona que A, debemos comprometernos con que C es CT futura de A. De ser esto el caso, necesitaremos mostrar que C es similar, continuo y depende causalmente de A.

Suponiendo que la operación fue exitosa y que todas las propiedades psicológicas de A (o por lo menos la mayoría de las propiedades relevantes: sus creencias, deseos, recuerdos, personalidad, etc.) fueron transferidas junto con su cerebro al cuerpo de B. Así, C compartirá las mismas propiedades psicológicas de A y, por tanto, diremos que ambos son similares psicológicamente.

Otra razón a favor de la similitud entre C y A es la siguiente. Dado que asumimos una postura materialista funcionalista, sabemos que, si mantenemos el realizador funcional de los estados psicológicos de A, mantenemos la misma psicología. El supuesto es que el realizador puede ser diferente, pero la psicología se mantiene si tienen las mismas funciones; sin embargo, una manera de mantener las funciones es precisamente mantener el realizador. Dado que A y C tienen el mismo realizador funcional (el cerebro), ambos deben tener la misma psicología y las mismas propiedades psicológicas.



Si C y A comparten el mismo cerebro y, en virtud de ello, las mismas propiedades psicológicas, parece ser el caso que las propiedades psicológicas de C, dependen causalmente de las de A. El hecho de que C piense como piense, actúe como actúe, crea lo que crea, recuerde lo que recuerde, etc. parece ser causa directa del hecho de que A pensaba de tal forma, actuaba de tal manera, etc. Cuando hablamos de la noción de continuidad mencioné que tenía que ser interpretada en términos de un “cambio gradual de propiedades psicológicas”. Interpretando caritativamente lo anterior, hay continuidad psicológica si o bien no hay cambio o bien las propiedades psicológicas cambian gradualmente a través del tiempo. El caso de C sería un caso en el que plausiblemente, no hay cambio ni radical ni gradual de un momento a otro. Así, si todo va bien, como en una cirugía ordinaria con anestesia y exitosa, lo que podemos esperar es que C despierte y piense, crea, desee, actúe, etc. justo como lo hacía A antes de ser sometido a la operación.

Hemos visto que tenemos buenas razones para decir por qué C es similar, continuo y causalmente conectado con A. Si tales razones son buenas, habremos mostrado que C satisface las condiciones que debería satisfacer para ser considerado CT futura de A. Luego, si C es CT futura de A, podemos concluir que A fue la misma persona que C. Hemos logrado rescatar la misma respuesta a la que habíamos llegado antes de manera consistente. Hasta el momento, parece que tenemos buenas razones para considerar el exdurantismo como un buen modelo para la IP.

### 6.1.2. El Caso de la Fisión

La solución que las teorías de continuidad habían dado al caso de la fisión fue la (R3): que el Dr. Jekyll es la misma persona que el Dr. Jek y que el Dr. Jekyll es la misma persona que el Dr. Kyll. Esta solución podía ser problemática pues levanta algunas inconsistencias: si el Dr. Jekyll es la misma persona que sus dos duplicados, el Dr. Jek y el Dr. Kyll, y “misma persona” es entendida en términos de identidad numérica, entonces, Dr. Jek y Dr. Kyll (por transitividad, que es una propiedad de la identidad numérica), deberían ser la misma persona.

Pero no es intuitivo, por ejemplo, pensar que una misma persona puede ocupar dos regiones espaciotemporales distintas.

Una forma de evitar este problema, si queremos quedarnos con las teorías de continuidad, es agregar una clausula de no fisión. Pero, en esta sección, veremos cómo es que agregando una teoría de contrapartes tenemos otra alternativa que no implica aceptar la clausula de no fisión.

Comencemos con la cuestión acerca de cómo interpretar “ $x$  es la misma persona que  $y$ ”. De acuerdo con el modelo tetradimensional, podemos traducir la proposición anterior de la siguiente forma: “ $x$  e  $y$  son estadios temporales distintos que existen en distintos tiempos, tal que  $y$  será CT futura de  $x$ ”. Si desde el modelo tetradimensional queremos decir que “Dr. Jekyll será la misma persona que Dr. Jek y que Dr. Jekyll será la misma persona que Dr. Kyll”, necesitaremos que los duplicados en cuestión sean estadios temporales futuros de Dr. Jekyll y que estos mantengan una relación de CT futura con Dr. Jekyll. Es importante notar que, si bien Dr. Jekyll tiene que ser un estadio temporal que exista en un tiempo distinto al de Dr. Jek y Dr. Kyll, como la relación de contraparte temporal puede ser uno a muchos, no hay problema con que los duplicados del Dr. Jekyll existan en el mismo tiempo.

En este caso, las propiedades psicológicas que son usadas para generar los duplicados en el proceso de teletransportación son las del mismo Dr. Jekyll. Esta situación, en la que se comparte los mismos estados psicológicos entre individuos para distintos tiempos es la misma ya analizada para el caso de fusión. De aquí que, la condición de dependencia causal entre el Dr. Jekyll y sus duplicados se satisface. Así mismo la manera en la que se dio la teletransportación sugiere que no debería haber ningún cambio, ni gradual ni radical, entre las propiedades de Dr. Jekyll y la de sus duplicados en el futuro, podemos decir que los duplicados resultantes son continuos con Dr. Jekyll.

Hemos llegado a la respuesta que necesitábamos: los duplicados son similares, continuos y dependen causalmente del Dr. Jekyll. De esto se sigue que son igualmente buenos candidatos para ser

CT futuras de Dr. Jekyll y, por ende, diremos que Dr. Jekyll es la misma persona que Dr. Jek y que el Dr. Jekyll será la misma persona que el Dr. Kyll.

Es importante notar que esta solución está libre de inconsistencias. Desde el modelo exdurantista, interpretar “misma persona” en términos de CT y no en términos de identidad numérica nos permite evitar inconsistencias en el caso de la fisión. Esto se debe a que la relación de CT, a diferencia de la identidad numérica, no es necesariamente simétrica ni transitiva. Por lo tanto, el Dr. Jekyll (un estadio temporal) puede estar en una relación de CT con un estadio futuro que es idéntico al Dr. Jek, y también con un estadio futuro que es idéntico al Dr. Kyll, sin que los estadios “Dr. Jek” y “Dr. Kyll” sean CT entre sí. Interpretar la noción de “misma persona” en términos de CT evita comprometerse con la necesidad de que los duplicados sean idénticos entre sí, lo que sería requerido bajo una interpretación de identidad numérica. Así, el modelo exdurantista permite resolver el caso de la fisión sin necesidad de recurrir a la cláusula de la no fisión, preservando la coherencia del análisis.

En resumen, asumir el modelo exdurantista implica que el Dr. Jekyll continua en dos personas: el Dr. Jek y el Dr. Kyll. Lo importante es evitar que el Dr. Jek y el Dr. Kyll sean la misma persona (lo cual es una inconsistencia). El modelo exdurantista no nos obliga a aplicar una cláusula de no-fisión y a desechar la posibilidad de la fisión desde el principio. Si tomamos la posibilidad del escenario de fisión en serio, el modelo propuesto permite una salida que respeta tanto su posibilidad como el resultado que hemos considerado intuitivo (dentro de las reglas que impone el escenario de fisión) de que el Dr. Jekyll continúa en dos personas.

## 7. Conclusión

A lo largo de este artículo hemos tratado de esbozar un modelo tetradimensional para la IP. Este modelo sostiene que “ $x$  en  $t_1$  es la misma persona que  $x$  en  $t_2$ ” debe ser analizado en términos de la relación de contraparte temporal. Esta relación requiere tres condiciones: similaridad, dependencia causal y continuidad entre

estadios temporales. Las propiedades que este modelo considera relevantes para dar cuenta de IP serán las propiedades psicológicas. La motivación por la cual estoy tratando de dar razones para quedarnos con este modelo es que puede acomodarse con el materialismo funcionalista y hacerlo compatible con nuestro análisis de continuidad psicológica a la luz de algunas de las posibilidades genuinas (fusión y fisión) acerca de IP sin atenerse a una cláusula de no fisión.

## Referencias:

- Copenhaver, R. (2009). Reid on memory and personal identity. *Stanford Encyclopedia of Philosophy*.
- Divers, J. (2002). *Possible Worlds*. Routledge.
- Hume, D. [1739] (1978). *Treatise of Human Nature*, Oxford: Clarendon Press, 1978; partly reprinted in Perry 1975.
- Levin, Janet. (2008). Functionalism. *Stanford Encyclopedia of Philosophy*.
- Lewis, D. (1976). Survival and identity. In Amelie Oksenberg Rorty (ed.), *The Identities of Persons*. University of California Press, pp. 17-40.
- Lewis, D. (1976). The Paradoxes of Time Travel. *American Philosophical Quarterly*, 13 (2):145-152.
- Lewis, D. (1986). *On the Plurality of Worlds*. Malden, Mass.: Wiley-Blackwell.
- Locke, J. (1975). *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press (original work, 2nd ed., first published 1694); partly reprinted in Perry 1975.
- Parfit, D. (1971). Personal identity. *Philosophical Review*, 80: 3-27.
- Parfit, D. (1984). *Reasons and Persons*. Oxford, GB: Oxford University Press.
- Reid, T. (1785). *Essays on the Intellectual Powers of Man*. University Park, Pa.: Cambridge University Press. Edited by Derek R. Brookes & Knud Haakonssen.
- Shoemaker, S., (ed.). (1963). *Self-Knowledge and Self-Identity*. Ithaca, N.Y.: Cornell University Press.
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7 (4):269-85.
- Shoemaker, S. (1984). *Personal identity*. Oxford, England: Blackwell. Edited by Richard Swinburne.

- Shoemaker, S. (1997). Self and Substance. *Philosophical Perspectives*. En *Mind, Causation and World*, 283–304.
- Shoemaker, S. (1999). Self, Body, and Coincidence. *Aristotelian Society Supplementary Volume*, 73 (1):287-306.
- Sider, T. (1996). All the World's a Stage. *Australasian Journal of Philosophy*, 74 (3):433 – 453.
- Sider, T. (1997). Four Dimensionalism. *Philosophical Review*, 106 (2):197-231.
- Sider, T. (2001). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford, GB: Oxford University Press.
- Sider, T. (2004). Four Dimensionalism: An Ontology of Persistence and Time. *Philosophy and Phenomenological Research* 68 (3):642-647.
- Sider, T. (2008). Temporal Parts. In Theodore Sider, John Hawthorne & Dean W. Zimmerman (eds.), *Contemporary Debates in Metaphysics*. Blackwell, pp. 241--262.
- Sider, T. (2018). Asymmetric Personal Identity. *Journal of the American Philosophical Association* 4 (2):127-146.



**SEGUNDA SECCIÓN**

**FILOSOFÍA DE LA MENTE**





## CAPÍTULO IV

### LOS ESTADOS MENTALES COMO ESTADOS DE ACTITUD PROPOSICIONAL

*Elberto A. Plazas*<sup>1</sup>

Dentro de la tradición analítica de la filosofía el estudio de los estados mentales se ha realizado principalmente a través del análisis de los reportes de actitud proposicional, que son enunciados que informan acerca de algún tipo de relación psicológica entre un sujeto y un contenido proposicional. A partir de este análisis se ha concebido a los estados mentales como estados de actitud proposicional. Este capítulo pretende introducir al lector en el estudio de los estados de actitud proposicional, como un tema de intersección entre las áreas de la filosofía del lenguaje, la filosofía de la acción, la filosofía de la mente y la psicología científica. El capítulo se divide en seis secciones. En la primera se hace una exposición general

---

1 Elberto A. Plazas es doctor en Lógica y Filosofía de la Ciencia por la Universidad de Salamanca (España) y profesor asociado en la Fundación Universitaria Konrad Lorenz (Bogotá, Colombia), donde dirige el Laboratorio de Percepción y Cognición. Es investigador en las áreas de conducta simbólica, semiótica y filosofía del lenguaje y la mente.

de los reportes de actitud proposicional (RAPs) y los estados de actitud proposicional (EAPs) y se presenta su relación con la explicación intencional de la acción. En la segunda sección se hace una introducción a la semántica de condiciones de verdad y en la tercera se explican las propiedades intensionales (con-s) de los RAPs. En la cuarta sección se habla de las propiedades intencionales (con-c) de los EAPs. La quinta sección revisa las principales teorías semánticas sobre los RAPs. En la última sección se ofrecen algunas conclusiones generales.

## **1. Los estados mentales como estados de actitud proposicional**

A grandes rasgos, se puede decir que la Filosofía Analítica es la tradición que se ha caracterizado por resolver (o disolver) los problemas filosóficos a partir de un análisis lógico o semántico de los enunciados lingüísticos en las cuales dichos problemas son expresados. El supuesto de fondo es que los problemas filosóficos son, en gran medida, confusiones producidas por las múltiples ambigüedades del lenguaje cotidiano en el cual se discute sobre ellos, y una teoría semántica que regimiente el lenguaje cotidiano podría ser de ayuda para aclarar esos problemas (Urmson, 1956/1978). La teoría tomada como canónica para realizar estos análisis es la *Semántica de Condiciones de Verdad* (SCV), que se origina en la lógica de Frege (1972) y luego va a ser desarrollada en los *Principia Mathematica* de Whitehead y Russell (1927) y el *Tractatus* de Wittgenstein (1922). De acuerdo con el método del análisis lógico del lenguaje, los enunciados lingüísticos tienen una estructura gramatical profunda que refleja la estructura lógica de los hechos que describen. El análisis lógico-semántico de esta estructura gramatical permite develar ambigüedades, que serían el origen de los problemas filosóficos. Uno de los temas que más interés ha generado dentro de la tradición analítica ha sido el de los estados mentales, especialmente por los retos que ha involucrado su tratamiento dentro de los límites de la SCV.

La filosofía de la mente surge con el *problema mente-cuerpo* heredado del dualismo interaccionista de Descartes. Según Descartes (1637, 1642), el hombre está compuesto de dos sustancias irreductibles entre sí, el cuerpo y la mente. El cuerpo es una sustancia material, con extensión y cuyo movimiento en el espacio se explica por las leyes de la física mecánica. En cambio, la mente es una sustancia inextensa, pensante, cuya acción es voluntaria y no está sujeta a leyes. Además, los estados mentales se caracterizan por ser privados y auto-conscientes, a los cuales solo cada mente tiene un acceso directo privilegiado a través de la introspección.

Sin embargo, en el marco cartesiano resultaba imposible explicar cómo una sustancia extensa y otra inextensa podían ejercer poder causal mutuo. Esto motivó diversas soluciones en términos de qué sustancias constituyen la realidad. Algunas eran dualistas, como el ocasionalismo de Malebranche (1712/2009), el doble aspectualismo de Spinoza (1677/2000), o el paralelismo psicofísico de Leibniz (1695). Otras eran monismos materialistas, como el mecanicismo de La Mettrie (1747/1961) o el epifenomenalismo de Huxley (1874). Otros eran monismos idealistas como el de Berkeley (1710/1992), o como el de Hegel (1807/1966). Pero a inicio del siglo XX predominó un monismo neutralista, fundado en un fenomenalismo neutral<sup>2</sup> que era escéptico respecto a la posibilidad de definir sustancias últimas de la realidad (p. ej., Mach, 1906/1914; James, 1912/2020; Russell, 1921). En la primera mitad del siglo XX, gracias a la influencia del Positivismo Lógico, se impuso un *monismo fisicalista*<sup>3</sup>, que ha sido dominante hasta hoy. En este marco, el problema mente-cuerpo se ha transformado en el problema de cómo dar cuenta de las propiedades mentales en un mundo físico. Actualmente se considera que los estados menta-

2 El *fenomenalismo neutral* es la posición asociada principalmente al empirismo y al positivismo, según la cual nuestro conocimiento del mundo está basado en fenómenos, sin que estos nos autoricen a realizar inferencias últimas acerca de la realidad fuera del hecho de ser experimentados.

3 El *fisicalismo* es la posición de que todo en el universo es físico o no requiere para su existencia de nada por encima o por debajo de las entidades y procesos reconocidos por las ciencias físicas (Rakova, 2006, p. 143)

les se caracterizan por dos propiedades: la *intencionalidad* y la *conciencia* (p. ej. Dennett, 1969/1996). En este capítulo me dedicaré principalmente a la primera, que es aquella en la cual la tradición analítica realizó mayores desarrollos.

A finales del siglo XIX, el filósofo austro-germano Franz Brentano (1874/2020) afirmó que los fenómenos mentales se diferencian de los fenómenos físicos en que, además de ser conscientes, representan un contenido intencional; es decir, siempre *se refieren a* o son *acerca de* algo. Por ejemplo, cuando pienso, pienso en algo; cuando deseo, deseo algo; cuando espero, espero algo, etc. Los fenómenos mentales siempre están dirigidos a algún objeto, que es su contenido. Pero dicho contenido es *intencionalmente in-existente*, es decir, es interno al acto mental y no requiere contrapartida en la realidad. Por eso, pueden ser contenido de nuestros actos mentales objetos inexistentes o también llamados *objetos intencionales*. Se llama *Tesis de Brentano* a la idea de que todos y solo los estados mentales son estados intencionales. Por lo tanto, nada en el mundo, excepto los estados mentales, manifiestan algo como la intencionalidad<sup>4</sup>.

En la tradición analítica, el aspecto *intencional* de los estados mentales ha sido abordado en la conceptualización de los estados mentales como *estados de actitud proposicional*. A inicios del siglo XX, el filósofo británico Bertrand Russell, uno de los padres de la tradición analítica, propuso que los enunciados sobre estados mentales son *Reportes de Actitud Proposicional* (en adelante, RAPs) (Russell, 1921). Los siguientes son ejemplos de RAPs:

- (1) Tomás quiere ganar el torneo de ajedrez,
- (2) Juan cree que los extraterrestres nos visitan,
- (3) María teme perder el examen,

---

4 Brentano tuvo un círculo inmediato de discípulos destacados que debatieron su teoría de la intencionalidad, tales como Twardowski (1894/1977), Meinong (1904/1960), y en particular Edmund Husserl (1913/1982), el padre de la fenomenología. La escuela fenomenológica ha sido la más influyente dentro de la llamada tradición continental en filosofía, que por razones geográficas y temáticas generalmente ha sido contrapuesta a la tradición analítica en el pensamiento contemporáneo.

(4) Natalia espera que la llame Jorge.

En los ejemplos de RAPs (1)-(4) podemos identificar cuatro componentes: Primero, un sujeto psicológico, que en estos casos son Tomás, Juan, María y Natalia. Segundo, un verbo psicológico, como *querer*, *creer*, *temer* y *esperar*. Tercero, la conjunción ‘que’. Cuarto, una oración subordinada que sigue a la conjunción ‘que’, y en los ejemplos corresponden a “yo ganaré el torneo de ajedrez”, “los extraterrestres nos visitan”, “yo voy a perder el examen” y “me llamará Jorge”. A este enunciado subordinado se le denomina también *cláusula-que*, y lo que dice dicho enunciado constituye el *contenido* o aquello de lo que trata el estado de actitud proposicional del sujeto que se reporta en el RAP. Se puede decir en general (aunque esto es debatible, como veremos en § 5) que el contenido de una *cláusula-que* corresponde a un estado de cosas posible que el sujeto está considerando. Dado que los RAPs describen a los estados mentales de un sujeto en relación a un contenido, presentan los estados mentales como intencionales, en el sentido de Brentano que vimos anteriormente<sup>5</sup>.

Pero además de un contenido, ya sea proposicional u objetual, el verbo psicológico determina la *modalidad* en que ocurre el estado proposicional para el sujeto. Si comparamos a continuación (5) y (6),

- (5) Tomás cree que su equipo favorito ganará el campeonato
- (6) Tomás desea que su equipo favorito gane el campeonato

observamos que ambos RAPs comparten el mismo sujeto y el mismo contenido proposicional, pero cambia el verbo psicológico. Se dice entonces que en cada caso el mismo sujeto se encuentra en un modo o actitud psicológica diferente respecto al mismo estado de cosas que constituye el contenido de la cláusula-que.

5 Sin embargo, en la tradición de Brentano y sus discípulos el contenido del estado mental es objetual más que proposicional (para una revisión ver Plazas, 2023, §2.2). Pero como vamos a centrarnos en los estados mentales dentro de la tradición analítica, nos limitaremos al contenido proposicional.

De acuerdo con Russell (1918, 1921, 1940), los estados mentales descritos por RAPs son *Estados de Actitud Proposicional* (EAPs), que consisten en que un sujeto se encuentra en una relación de actitud psicológica hacia un contenido proposicional particular. El ejemplo (2) expresaría que Juan está en una actitud de creencia hacia el posible hecho de que los extraterrestres nos visitan. Verbos como *creer, desear, considerar, pensar, suponer, esperar, temer, dudar, querer*, etc., son verbos transitivos, que requieren un objeto, que en este caso es el contenido proposicional dado en la cláusula-que, respecto al cual se tiene una actitud psicológica.

John Searle (1983/1986) señaló que los EAPs además de un contenido representacional y un modo psicológico tienen unas *condiciones de satisfacción*. Así, por ejemplo, las creencias pueden ser verdaderas o falsas y los deseos pueden ser cumplidos o incumplidos. Esto involucra que los estados mentales tienen un carácter normativo, por el cual son valorados según un criterio. Además, las condiciones de satisfacción tienen una *dirección de ajuste*. Para las creencias, esa dirección es mente-mundo; mientras que para los deseos la dirección es mundo-mente. Por ejemplo, la creencia de Tomás en (5) se evalúa según lo que ocurre en el mundo; mientras que el deseo de Tomás en (6) se valora según el mundo se corresponda con el contenido mental de Tomás<sup>6</sup>.

Aparte de los EAPs, los filósofos de la mente reconocen otros tipos de estados mentales, en especial los llamados *estados fenoménicos*, que consisten en experimentar sensaciones, como ver, oír, oler, sentir dolor, hambre, sed, inquietud, o sentir alegría o tristeza. Estos estados fenoménicos se caracterizan porque tienen un aspecto cualitativo (*qualia*) que es vivido por cada individuo de manera subjetiva (Pineda, 2012, § 2.4). Una diferencia importante entre los EAPs y los estados fenoménicos es que los primeros tienen un carácter más permanente, y su posesión por parte de un

---

6 Este carácter normativo de los estados mentales había sido reconocido inicialmente por Brentano (1874/2020, Caps. 6-9). Sin embargo, la introducción de las nociones de condiciones de satisfacción y dirección de ajuste vienen de la *teoría de los actos de habla*, del propio Searle (1969/2009, 1979).

sujeto no requiere que se sea consciente de él. Si tomamos (6), podemos decir que Tomás tiene el deseo de que su equipo gane el campeonato, a pesar de que esté pensando en otras cosas en este momento y que siga deseando lo mismo luego de dormir y soñar con cosas diferentes. En cambio, la sensación del sonido de la sirena de una ambulancia se experimenta en el presente, mientras la ambulancia esté pasando cerca de nosotros. En § 6 trataré el asunto de la relación entre los EAPs y los estados fenoménicos. Por ahora, nos centraremos solo en los EAPs.

Los RAPs cumplen una función muy importante en nuestras interacciones sociales cotidianas. Los RAPs nos permiten atribuir estados mentales con contenido a otros individuos para explicar y racionalizar sus acciones. Uno de los principales criterios de mentalidad es la *agentividad*. Ser *agente* es ser capaz de actuar sobre el mundo de manera intencional, más que ser un objeto pasivo al que le ocurren cosas. Además, nuestras acciones intencionales están motivadas o son puestas en marcha por nuestros estados mentales (Anscombe, 1957/1991; Davidson, 1963). Así, al atribuir a un agente un EAP a través de un RAP, afirmamos cuál es el estado mental que motivó su acción. A esta se le llama una *explicación intencional*, porque está basada en los contenidos intencionales de los EAPs atribuidos al agente.

La explicación intencional se realiza a través de un tipo de argumento denominado *silogismo práctico*, que de manera simplificada tiene la siguiente forma:

- (7) El sujeto S desea que O.  
       S cree que si hace la acción A obtendrá O.  
       Por lo tanto, S se dispone a realizar A<sup>7</sup>.

Por ejemplo:

- (8) Miguel desea pasar el examen

---

7 Elizabeth Anscombe (1957/1991) recuperó el silogismo práctico de las obras éticas de Aristóteles. Esta es una versión simplificada del argumento, pero su aplicación es por lo general más compleja (ver von Wright, 1971/1979, Cap. 3)

Miguel cree que si estudia pasará el examen  
Por lo tanto, Miguel se dispone a estudiar

Lo interesante de esta estructura es que las dos premisas del argumento son RAPs, la primera acerca de un estado de deseo y la segunda de un estado de creencia, y los contenidos de estos estados son los que determinan el tipo de acción que realizará el agente y lo que se expresa en la conclusión del argumento. El silogismo práctico es la base de nuestro razonamiento práctico, el cual es un *razonamiento medios-fines*, en el que la creencia establece una condición para el cumplimiento del deseo, y esto determina la acción a realizarse. Los EAP se consideran *razones* para la acción; así que al ofrecer una explicación intencional de las acciones de otros, *racionalizamos* al agente, y dichas razones pueden ser tomadas también como *justificación* de lo que el agente hace.

La explicación intencional de las acciones de un agente a partir de sus actitudes proposicionales no se aplica solo a los seres humanos. También las usamos para dar cuenta de la acción de animales, como cuando se dice que un perro busca su correa y ladra porque quiere que su dueño lo saque al parque. Ocasionalmente se extiende a la naturaleza en general, como en el pensamiento animista cuando se dice que las catástrofes naturales son un castigo a nuestras malas acciones. Daniel Dennett (1978, 1987/1998) llamó *Actitud Intencional* (*Intentional Stance*) a la estrategia de invocar EAPs para explicar la conducta de otros. La atribución de estados mentales a otros para dar cuenta de lo que hacen es la forma en que funciona nuestra *Psicología Popular* o psicología del sentido común (Baker, 1999). A través de la atribución de EAPs a otros, “leemos sus mentes”, y esta habilidad constituye tener una *Teoría de la Mente* de los otros, y de nosotros mismos, la cual se reconoce como una habilidad interpersonal muy importante que se adquiere en el desarrollo humano (Gopnik, 1999; Tomasello, 1999). Incluso, algunos han sugerido que nos resulta más natural explicar el mundo y la realidad de manera intencional que causalmente (p. ej., Schütz, 1954).



En la tradición analítica, el estudio de los estados mentales como EAPs ha pasado por el análisis lógico-semántico de los RAPs. Para entender las peculiaridades semánticas de los RAPs y sus implicaciones respecto a los estados mentales, debemos primero introducirnos a algunos elementos de la semántica de condiciones de verdad, que veremos a continuación.

## 2. La semántica de condiciones de verdad

La Semántica de Condiciones de Verdad (SCV) tiene su origen en la lógica simbólica moderna formulada por Frege en su *Conceptografía* (1879/1972) y que es la base de todas las aproximaciones lógicas contemporáneas. El desarrollo de esta lógica fue parte del proyecto logicista de Frege de fundamentar la aritmética en la lógica, en oposición al psicologismo que fue predominante en la primera mitad del siglo XIX<sup>8</sup>. Frege se enfrentó contra los formalistas en matemáticas, quienes creían que enunciados como  $2 \cdot 2^3 + 2$ , “18” y  $3 \cdot 6$  son expresiones diferentes, pues para ellos no habría diferencia entre forma y contenido. Frege creía, en cambio, que estas expresiones compartían el mismo *contenido conceptual*, a pesar de tener diferente forma. Rechazó la construcción sujeto-predicado de la vieja lógica aristotélica, porque en ella dos proposiciones como “Julio Cesar conquistó las Galias” y “Las Galias fueron conquistadas por Julio Cesar” pasan por proposiciones diferentes, a pesar de que tienen el mismo contenido. Así que desarrolló un sistema de lógica que permitiera capturar el *contenido conceptual* de los enunciados. Para ello, reemplazó la estructura sujeto-predicado por la construcción matemática de *función-argumento*. Una *función* es una expresión incompleta, insaturada, que necesita de compleción por parte de otras expresiones para tener sentido; mientras que un argumento es una expresión saturada y completa. Un nombre de función tiene uno o varios lugares vacíos para ser llenados por uno o más argu-

<sup>8</sup> Se llamó *psicologismo* a la tendencia a reducir o explicar las reglas de la lógica y la matemática a procesos mentales de pensamiento, y que puede ilustrarse en el *System of Logic* de J. S. Mill (1886).

mentos. La función expresa un *concepto*, que es la referencia de un predicado, y según la cantidad de argumentos para completar su sentido podrá ser monádica, diádica, triádica, etc. En cambio, un argumento refiere a un *objeto*, que es algún ente concreto individual expresado por un nombre propio o término singular (1891a, 1892a, 1895).

De acuerdo con Frege, la relación lógica fundamental es la de “caer bajo un concepto” (1895, p. 113). Si tenemos el concepto monádico: ‘\_\_\_\_ es alto’ y el argumento ‘Carlos’, el enunciado “Carlos es alto” será verdadero en cuanto Carlos entre dentro de la extensión del conjunto de personas altas, y será falso si Carlos está fuera de dicha extensión. Si tenemos el concepto diádico ‘\_\_\_\_ ama a \_\_\_\_’, el enunciado “Carlos ama a María” será verdadero en cuanto Carlos y María, en ese orden, entran dentro de la extensión del conjunto de pares ordenados por el que el primer elemento del par se encuentra en la relación de amar al segundo elemento del par.

En la concepción de Frege, los enunciados compuestos se pueden entender como funciones de los valores de verdad de sus enunciados simples junto con la estructura lógico-sintáctica del enunciado, definida por los conectores lógicos (negación, conjunción, disyunción, condicional, bicondicional). Por ejemplo, el enunciado

- (9) Ecuador tiene costa en el Océano Atlántico o en el Océano Pacífico

se compone de la disyunción entre los enunciados simples:

- (10) Ecuador tiene costa en el Océano Atlántico  
(11) Ecuador tiene costa en el Océano Pacífico,

Sabemos que (10) es falso y (11) es verdadero, y también sabemos que una disyunción es verdadera si al menos uno de sus disyuntos es verdadero; por tanto, (9) es verdadero. Esto ejemplifica el *Principio de Composicionalidad*, según el cual el significado de un enunciado compuesto depende del significado de sus enunciados componentes y de su estructura sintáctica. Una consecuencia de esto es la idea de

que entender un enunciado consiste en conocer sus *Condiciones de Verdad*, es decir, bajo qué condiciones el enunciado en cuestión es verdadero. Estas condiciones son el valor de verdad de sus enunciados componentes y la estructura lógico-sintáctica del enunciado compuesto. Este es el núcleo de la SCV. Una consecuencia de lo anterior es que dos enunciados con las mismas condiciones de verdad tendrían el mismo significado. Por ejemplo, (12) tiene el mismo significado que (9) porque ambas comparten las mismas condiciones de verdad, dado que se componen de los mismos enunciados simples y tienen formas lógicas equivalentes<sup>9</sup>.

(12) Si no es el caso que Ecuador tiene costa en el Océano Atlántico entonces tiene costa en el Océano Pacífico.

Para la SCV, las condiciones de verdad de los enunciados simples dependen de cómo es el mundo real extra-lógico, de tal manera que un enunciado simple será verdadero si los objetos nombrados entran dentro de la extensión del concepto expresado en la función del enunciado. Como consecuencia, la SCV respeta el llamado *millianismo*, o la opinión de John Stuart Mill (1886) de que el significado de los nombres propios se agota en su referente, o el objeto que es su portador. De acuerdo con la SCV el aporte significativo que hace cualquier expresión singular que actúe como argumento de un enunciado es su referente objetual. Entre las expresiones singulares se incluyen los nombres propios (como ‘Aristóteles’), las descripciones definidas (como ‘El maestro de Alejandro Magno’), o expresiones deícticas (como, ‘aquel filósofo griego al que señalo ahora’).

En la SCV se cumple también el principio de *Generalización Existencial*, según el cual de un enunciado simple que afirme (o niegue) algo de un referente con verdad podemos inferir que hay algo de quien se puede predicar lo que afirma (o niega). Así, por ejemplo, de la verdad de (13) se puede inferir (14):

9 En lógica proposicional existe la regla de transformación de *implicación material*, por la cual  $(P \rightarrow Q) \equiv (\sim P \vee Q)$ . El enunciado (12) tiene la forma lógica  $(\sim P \rightarrow Q)$  que es equivalente a  $(\sim \sim P \vee Q)$ , y aplicando la regla de *doble negación*, es equivalente a  $(P \vee Q)$ , que es la forma de (9).

- (13) Aristóteles fue el más grande filósofo de la antigüedad
- (14) Alguien fue el más grande filósofo de la antigüedad.

Finalmente, la SCV respeta otro principio importante, el *Principio de Sustitutibilidad de Términos Correferenciales Salva Veritate* o también *Ley de Leibniz*. Este principio dice que si en un enunciado compuesto reemplazamos una expresión singular por otra que denote el mismo referente, el valor de verdad del enunciado no se altera. En consecuencia, si Aristóteles es el maestro de Alejandro Magno y reemplazamos a 'Aristóteles' en (13) por 'el maestro de Alejandro Magno' para obtener (15), y (13) es verdadero, entonces (15) será igualmente verdadero:

- (15) El maestro de Alejandro Magno fue el más grande filósofo de la antigüedad.

Se considera que la marca de un *lenguaje extensional* es que todas sus expresiones cumplen con el principio de sustitutibilidad de correferenciales *salva veritate* (Carnap, 1947; Russell, 1940). Frege aspiraba a que su lógica sirviera como un lenguaje lógicamente perfecto para las ciencias naturales, y este debería ser un lenguaje extensional, en el que todo nombre tendría un referente y solo habría un nombre para cada referente (Frege, 1891b, 1892b, 1895). La idea de que el significado de un enunciado consiste en sus condiciones de verdad, fue la base de la doctrina *verificacionalista*<sup>10</sup>, que fue muy influyente en el Positivismo Lógico. Los positivistas lógicos plantearon que la unidad de la ciencia se lograría a partir del hecho de que todas las ciencias se expresaran en un *lenguaje fiscalista*, que en el fondo consistía en un lenguaje extensional (p. ej., Carnap, 1935/1998). Fiscalistas prominentes del siglo XX, como W. V. O. Quine (1960/2001), defendieron que el conocimiento científico

---

10 El *verificacionalismo* fue una doctrina introducida por Wittgenstein en su *Tractatus*, según la cual solo son significativas las tautologías o los enunciados empíricos de las ciencias. Los positivistas lógicos lo emplearon como un criterio de demarcación del conocimiento científico frente a lo no científico (ver p. ej., Carnap, 1932; Hempel, 1950).

solo podía expresarse dentro de los límites de la lógica de primer orden. Sin embargo, los enunciados sobre estados mentales resultaron ser un gran desafío, como veremos a continuación.

### 3. El carácter intensional de los reportes de actitud proposicional

Como hemos visto, la substitutibilidad *salva veritate* de términos co-referenciales es la marca distintiva de un lenguaje extensional. Sin embargo, en el lenguaje natural son frecuentes los contextos lingüísticos en los que la substitutibilidad no se sostiene. Uno de estos contextos es el de *habla indirecta*, en el cual se parafrasea lo que alguien dijo. Si, por ejemplo, se reemplaza ‘Aristóteles’ por ‘el maestro de Alejandro Magno’, a pesar de la verdad de (16), no se garantiza la verdad de (17):

- (16) Miguel dijo que Aristóteles fue el más grande filósofo de la antigüedad
- (17) Miguel dijo que el maestro de Alejandro Magno fue el más grande filósofo de la antigüedad.

Rudolf Carnap (1947) distinguió entre *contextos extensionales* e *intensionales* (con-s), siendo los primeros aquellos en los que la substitutibilidad aplica y los segundos aquellos en los que no. Quine (1980/2002; 1960/2001) llamó a los primeros *contextos referencialmente transparentes* y a los segundos *referencialmente opacos*. Además del contexto de habla indirecta, otro contexto intensional u opaco es el de *citación* o entrecomillado. Por ejemplo, mientras (18) es verdadero, no es así (19):

- (18) “Aristóteles” es una palabra de once letras
- (19) “El maestro de Alejandro Magno” es una palabra de once letras.

También están los *contextos modales*, como los que involucran las modalidades aléticas de necesidad y posibilidad. Por ejemplo, de la verdad de (20) no se sigue la de (21):

- (20) Necesariamente Aristóteles es Aristóteles
- (21) Necesariamente Aristóteles es el maestro de Alejandro Magno.

Esto es así porque (21) es una verdad histórica contingente, que podría ser diferente si las circunstancias hubiesen sido otras.

Ahora bien, los reportes de actitud proposicional (RAPs) son casos paradigmáticos de contextos intensionales. Por ejemplo, de la verdad de (22) puede no seguir la verdad de (23):

- (22) Miguel cree que Aristóteles fue el más grande filósofo de la antigüedad
- (23) Miguel cree que el maestro de Alejandro Magno fue el más grande filósofo de la antigüedad,

si es el caso que Miguel ignora que Aristóteles fue el maestro de Alejandro Magno. De la negación de (23) sería cierto (24):

- (24) Miguel no cree que el maestro de Alejandro Magno fue el más grande filósofo de la antigüedad.

Pero si (22) y (24) son ciertos, y aplicamos el principio de sustituibilidad obtendríamos:

- (25) Miguel cree que Aristóteles fue el más grande filósofo de la antigüedad y a la vez cree que Aristóteles no fue el más grande filósofo de la antigüedad,

---

11 Quine (1956) reconoció que los RAPs son estructuras ambiguas y que es posible una lectura *de re* de ellas, en las que la sustituibilidad es posible, en oposición a una lectura *de dicto* en la que la sustituibilidad no es posible. Sin embargo, la lectura *de re* puede llevar a considerar que el agente sostiene creencias contradictorias, como se muestra en (25).

por lo que estaríamos atribuyéndole creencias contradictorias a Miguel y dudando de su racionalidad (Quine, 1956)<sup>11</sup>.

Además de la falla en la sustitutibilidad, los RAPs también presentan problemas con el principio de Generalización Existencial. Por ejemplo, de (26) no se sigue (27):

(26) Hesíodo creía que Pegaso es el caballo alado que vive en el Olimpo

(27) Hay algo que es un caballo alado que vive en el Olimpo.

Además, los verbos psicológicos no parecen actuar como funciones de verdad, dado que la verdad de (26) no depende de la verdad o falsedad de (28):

(28) Pegaso es el caballo alado que vive en el Olimpo.

Por lo tanto, los RAPs tampoco cumplen el principio de composicionalidad de la SCV.

Se han desarrollado sistemas de lógica modal y de semántica de mundos posibles que han tenido éxito con el tratamiento formal de expresiones modales o temporales (ver p. ej., Gamut, 1982/2010); pero los RAPs siguen siendo difíciles de explicar semánticamente, al punto que algunos han afirmado que estos enunciados introducen contextos *hiperintensionales* (Bäuerle & Cresswell, 2003. Cf. Carnap, 1947).

#### **4. El carácter intencional de los estados de actitud proposicional**

A mediados del siglo pasado, el filósofo estadounidense Roderick Chisholm consideró que las fallas en la generalización existencial, la composicionalidad y la sustitutibilidad *salva veritate* de los RAPs son criterios de lo que llamó *lenguaje intencional*, en oposición al lenguaje extensional de la lógica simbólica moderna y la SCV. También afirmó que el empleo de un lenguaje intencional era una manifestación del carácter *intencional* (con-c) de los estados mentales

(Chisholm, 1955-56, 1957, Cap. 11). De esta manera, Chisholm recuperó la *Tesis de Brentano* de que todos y solo los estados mentales son estados intencionales, pero esta vez reformulada en términos lingüísticos. En palabras de Chisholm:

Podemos formular ahora una tesis pareciéndose a la de Brentano refiriéndonos al lenguaje intencional. Digamos (1) que no necesitamos usar lenguaje intencional cuando describimos fenómenos no-psicológicos, o “físicos”; podemos expresar todo lo que sabemos, o creemos, acerca de tales fenómenos en lenguaje que no es intencional. Y digamos (2) que, cuando deseamos describir ciertos fenómenos psicológicos —en particular, cuando deseamos describir, pensar, creer, percibir, ver, saber, esperar, tener la esperanza y semejantes —o bien (a) debemos usar lenguaje que es intencional o (b) debemos usar un vocabulario que no necesitamos usar cuando describimos fenómenos no-psicológicos o “físicos” (Chisholm, 1955-1956, p. 129, traducción mía).

De acuerdo con Edmund Husserl, la intencionalidad de los estados mentales se manifiesta en dos propiedades fundamentales: la *independencia de la existencia del objeto* y la *dependencia de la perspectiva del sujeto* (McIntyre & Woodruff-Smith, 1989; Crane, 2001). En primer lugar, gracias a la propiedad de independencia de la existencia del objeto, es posible que un estado mental sea no solo acerca de objetos existentes presentes (p. ej., la ciudad de Atenas), sino también de objetos pasados (p. ej., el Imperio Griego) o futuros (p. ej., la primera ciudad en la Luna); así como también estados de cosas posibles (p. ej., la reversión del cambio climático), o falsos (p. ej., que Napoleón ganó en Waterloo), objetos ficticios (p. ej., Pegaso), o incluso imposibles (p. ej., el círculo cuadrado); llamados objetos intencionales u objetos de pensamiento (Crane, 2013). La independencia de la existencia del objeto se manifestaría en las propiedades semánticas de la falla en la generalización existencial y el incumplimiento del principio de composicionalidad. En segundo lugar, la dependencia del contenido de la perspectiva del sujeto se manifiesta en la falla en la sustitutibilidad *salva veritate* de términos correferenciales. En los ejemplos anteriores (22) y (23), lo que



Miguel cree de Aristóteles y del maestro de Alejandro Magno es diferente, porque en cada caso la perspectiva del sujeto frente al referente es diferente<sup>12</sup>.

Un lenguaje puramente extensional, como el de la lógica simbólica moderna y la SCV, pretende ser objetivo y no debería reflejar propiedades intensionales. Para ajustar los reportes de actitud intencional (RAPs) a un lenguaje fisicalista, algunos positivistas lógicos promovieron una posición llamada *conductismo lógico* (p. ej., Carnap, 1932-3; Hempel, 1949)<sup>13</sup>, la cual buscaba una reducción de los enunciados mentales, como los EAP, a enunciados sobre conducta observable, bajo el supuesto de que los enunciados sobre conducta observable se expresan en un lenguaje extensional. Chisholm rechazó las pretensiones de los positivistas, y buscó mostrar que no es posible una reducción del lenguaje mental a uno conductual sin pérdida de significado. Por ejemplo, si ante un estímulo de fuego se dice que alguien presenta una “respuesta de fuego”, como huir o gritar, esto involucra sin embargo que el sujeto en cuestión “cree que hay fuego cerca”, lo que es intencional porque no implica necesariamente que hay algo en llamas cerca. Algunos positivistas, como Ayer o Carnap, habían sugerido que creer en algo es una disposición a asentir a una afirmación; pero Chisholm señaló que asentir a una afirmación involucra entender lo que la afirmación significa, y ‘entender’ es una palabra ordinariamente intencional (Chisholm, 1955-1956; 1957). La reformulación lingüística de la tesis de Brentano por parte de Chisholm supone, al menos, un dualismo de propiedades; que tiene como consecuencia la imposibilidad de la psicología como una ciencia natural. Debido a la intensionalidad del lenguaje mental, algunos defensores del fisicalismo sugirieron que los estados de actitud proposicional

12 Algunos autores han cuestionado los criterios intensionales (con-s) de la intencionalidad (con-c), bajo la idea de que los primeros son innecesarios e insuficientes para los segundos (p. ej., Crane, 2009; Searle, 1983/1986, Cap. 1). Pero puede considerarse que más bien el carácter intencional de los estados mentales permite que podamos hacer uso de un lenguaje con propiedades intensionales (Plazas, 2023, § 8.8).

13 Ryle (1949/2005) presentó una versión más sofisticada, en términos de disposiciones de conducta.

(EAPs) deberían ser eliminados del inventario de fenómenos naturales (Quine, 1960/2001) y la psicología no podría ser más que una ciencia intencional (Dennett, 1969/1996) o interpretativa (Davidson, 1970, 1974).

En la tradición analítica, el estudio semántico de los RAPs se ha visto como un medio para estudiar la naturaleza y propiedades de los estados mentales, incluyendo sus aspectos intencionales (Clapp, 2006). Esto parece relevante para la posibilidad del estudio de los estados mentales dentro de las ciencias naturales. Además, las propiedades intensionales (con-s) de los RAPs han motivado muchos desarrollos en la filosofía analítica del lenguaje. En la próxima sección revisaremos algunas de las principales teorías que buscan explicar las peculiaridades semánticas de los RAPs.

## 5. Teorías sobre los reportes de actitud proposicional

### 5.1 Relacionismo proposicionalista

A Russell le debemos no solo las nociones de reporte y estado de actitud proposicional (RAP y EAP), sino también su primera explicación semántica. Según Russell, la estructura gramatical de los RAPs nos muestra la estructura lógica de los hechos mentales que describen (Russell, 1918, Confs. 3 y 4; 1921, Conf. 12; 1940, p. 65). Los RAPs tienen una estructura relacional, en los que el verbo psicológico ( $\Psi$ ) actúa como un relator diádico, que pone en relación a un sujeto psicológico ( $s$ ) y un contenido proposicional ( $p$ ). Si tomamos el verbo *creer* como paradigma de los verbos mentales, el ejemplo (29) puede esquematizarse en la forma gramatical (30), que al final corresponde con la forma lógica (31):

(29) David cree que Alejandro Magno fue el conquistador de los persas

(30)  $S$  cree que  $p$

(31)  $\Psi_{sp}$ ,

De acuerdo con (30) y (31), el sujeto se encuentra en una relación de creencia hacia una *proposición*. Pero, ¿qué es una proposición? Los

lógicos hablan de proposiciones para referirse al contenido conceptual de un enunciado u oración declarativa. Por ejemplo,

- (32) Alejandro Magno fue el conquistador de los persas,
- (33) El conquistador de los persas fue Alejandro Magno,
- (34) El discípulo de Aristóteles venció a los persas y los incorporó a su imperio,
- (35) Alexander the Great was the conqueror of the Persians,

son oraciones diferentes, pero expresan la misma proposición, dicen lo mismo; por lo cual están en las mismas relaciones de compatibilidad e incompatibilidad, antecedencia y consecuencia lógica con otras proposiciones. Dado que una proposición es una entidad común a muchas oraciones diferentes, se considera una entidad *abstracta*. Además, las proposiciones tienen valores de verdad, pueden ser verdaderas o falsas (Crimmins, 1997). También se tiende a considerar que las proposiciones son entidades estructuradas, lo que se refleja en la estructura sintáctica de la oración que las expresa. (32)-(35) corresponden a una proposición simple, que consiste de un término singular ‘Alejandro Magno’ (o su correferencial ‘el discípulo de Aristóteles’) y un predicado ‘conquistador de los persas’, (o su correferencial, ‘vencedor de los persas e incorporador de estos a su imperio’). Podemos expresar su estructura como <Alejandro, conquistador de los persas>. Por tanto, la estructura lógica de (29) sería la siguiente:

- (36)  $Cd <a, c>$ ,

donde *C* está por el verbo *creer*, *d* está por David, *a* por Alejandro Magno y *c* por el predicado conquistador de los persas. Por otra parte, el relacionismo proposicional permite hacer válidas ciertas inferencias que intuitivamente lo parecen: por ejemplo, que de (29) se concluya que “David cree en algo” o que, si “María cree que Alejandro Magno fue un conquistador”, se concluya que “María cree lo mismo que David” (Schiffer, 2008).

En la posición de Russell, las proposiciones se refieren, en general, a *estados de cosas* del mundo, y contienen sus condiciones de verdad. A estas se les llama *proposiciones russellianas*, así que el relacionismo proposicionalista se ajusta a la SCV. Sin embargo, por esa razón el relacionismo proposicionalista tiene problemas con las propiedades intensionales de los RAPs. De acuerdo con el principio de substitutibilidad de términos correferenciales *salva veritate*, si (29) es verdadero, también lo es (37):

(37) David cree que el discípulo de Aristóteles fue el conquistador de los persas.

Pero si David ignora que Alejandro Magno fue el discípulo de Aristóteles, es posible que (37) sea falso. El relacionismo proposicional tiene problemas para explicar la diferencia en el valor de verdad entre (29) y (37), porque para esta posición las cláusulas-que de estos RAPs expresan la misma proposición.

El relacionismo proposicional también tiene problemas con los términos singulares referencialmente vacíos. Supongamos que,

(38) Alejandro creía que Pegaso vive en el Olimpo.

(38) es problemático porque si el caballo mitológico no existe, entonces el término singular ‘Pegaso’ falla en referir, y la proposición completa *Pegaso vive en el Olimpo*, no puede ser acerca de algo. Sin embargo, las relaciones se sostienen dada la existencia de sus correlatos y propiedades. Pero si Pegaso no existe, la relación de creencia de Alejandro no se puede sostener. Además, el relacionismo proposicional debería permitir que fuera aplicable el principio de generalización existencial, pero en (38) no es posible inferir que “hay algo que vive en el Olimpo”.

Frente a las dificultades del relacionismo proposicional, se han propuesto diferentes soluciones que han buscado presentar o bien interpretaciones diferentes de las proposiciones o sus componentes, o abandonar el proposicionalismo o el relacionismo..

## 5.2 El proposicionalismo fregeano

Una alternativa es defender el relacionismo proposicionalista, pero ofreciendo una noción diferente de las proposiciones. Esto es lo que ocurre en la Teoría de Frege sobre los RAPs. Frege reconoció que las lenguas ordinarias están lejos de ser un lenguaje lógicamente perfecto, porque en ellas son frecuentes los giros intensionales, como los RAPs. Para dar cuenta de las propiedades intensionales de los RAPs, Según Frege (1892b), en contextos lingüísticos con oraciones subordinadas, se presenta un *cambio de referente*, así que las oraciones subordinadas no tienen su referente habitual, sino un referente indirecto, que consiste en su sentido habitual. Para Frege, todos los signos o expresiones lingüísticas tienen una *referencia* (*Bedeutung*) y un *sentido* (*Sinn*). En el caso de los términos singulares, su referente es el objeto que denota, mientras que su sentido es su *modo de presentación* o manera de darse. Por ejemplo, ‘Alejandro Magno’ y ‘el discípulo de Aristóteles’ son dos modos de presentación o sentidos del mismo individuo. Las oraciones declarativas también tienen sentido y referencia<sup>14</sup>, siendo su sentido el *pensamiento* (*Gedanke*) que expresan, mientras que su referencia es su valor de verdad, es decir, lo verdadero o lo falso. Se puede decir que los ejemplos (32)-(35) expresan diferentes pensamientos, pero tienen el mismo referente, que en este caso es lo verdadero.

Para Frege (1892b), el sentido del signo media el acceso a la referencia. En el caso de las oraciones declarativas, primero captamos el pensamiento de esta y luego avanzamos hacia su valor de verdad en el juicio (Frege, 1918-9). Pero a pesar de que todo signo tiene un sentido, no todo signo tiene una referencia. Así, por ejemplo, la cláusula subordinada de (38): “Pegaso vive en el Olimpo” tiene un sentido que es captable para cualquier usuario competen-

---

14 De acuerdo con Frege, las expresiones predicativas también tienen una referencia, que consiste en su concepto, como término de función o insaturado en un enunciado; pero él no dejó claro cuál era su sentido (1891a, 1892b, 1895). Para Carnap (1947), la *intensión* (o sentido) de una expresión predicativa es el *atributo* que señala, mientras que su *extensión* (o referencia) es el conjunto de individuos de los que dicho atributo se predica con verdad.

te del lenguaje, pero dado que ‘Pegaso’ es un término referencialmente vacío, no es posible que el predicado “vive en el Olimpo”, le aplique con verdad o falsedad. Y dado que la referencia de una oración declarativa es su valor de verdad, esta sería una oración sin referencia. Si las proposiciones son entendidas como pensamientos fregeanos, es posible que ellas sean el contenido de nuestros EAPs, sin que tengan una referencia. De esta manera se soluciona el problema de los RAPs con términos singulares referencialmente vacíos.

Por otra parte, Frege afirmó que a los sentidos también les aplica el principio de composicionalidad, de tal manera que el pensamiento de una oración está compuesto por los sentidos de sus términos singulares y predicativos. Así, el pensamiento de (30) estaría compuesto del modo de presentación de ‘Alejando Magno’ ( $m_{AM}$ ) y el modo de presentación de ‘conquistador de los persas’ ( $m_{CP}$ ). Por tanto, el pensamiento de (32) tendrían la estructura  $\langle m_{AM}, m_{CP} \rangle$ ; y podemos considerar que la estructura lógica de (29) es:

$$(39) \text{Cd} \langle m_{AM}, m_{CP} \rangle.$$

Ahora bien, puede considerarse que el pensamiento de (37) tiene entre sus componentes el modo de presentación de “el discípulo de Aristóteles” ( $m_{DA}$ ), así que la estructura del contenido proposicional es,  $\langle m_{DA}, m_C \rangle$ ; por tanto, la estructura lógica de (37) es:

$$(40) \text{Cd} \langle m_{AD}, m_C \rangle.$$

En consecuencia, David está en una relación de creencia con pensamientos diferentes en (29) y (37). Y dado que en el contexto de los RAPs la referencia de la cláusula-que es su sentido habitual, es decir, su pensamiento, no se incumple el principio de sustitutibilidad de correferenciales *salva veritate*. Por tanto, el proposicionalismo fregeano tiene éxito en explicar las propiedades intensionales de los RAPs, a la vez de mantener el relacionismo proposicionalista.

Sin embargo, el proposicionalismo fregeano se enfrenta a varios problemas. Aquí me concentraré en dos de los más importantes. El primero es el de la naturaleza de los pensamientos. Según Frege (1918-9), los pensamientos no son representaciones (*Vorstellung*) subjetivas, ni realidades concretas objetivas, sino entidades abstractas, públicas, captables intersubjetivamente, pertenecientes a un tercer reino. Un ejemplo es el Teorema de Pitágoras, el cual es verdadero independiente de que lo captemos o no. Sin embargo, no son claros los criterios de identidad de los pensamientos, es decir, cómo consideramos que dos pensamientos son iguales o diferentes. Frege parecía sugerir que dos pensamientos son idénticos si nadie que los capte puede considerar que tienen distinto valor de verdad (Fernández, 2012, p. 545). Pero entonces, no es claro cómo distinguir entre sí todos los pensamientos considerados por todos como verdaderos (o falsos). Carnap (1947) propuso que dos expresiones tienen la misma intensión (el mismo sentido) si son lógicamente equivalentes (L-equivalentes) en un lenguaje, es decir, si son verdaderas en la misma cantidad de mundos posibles. Sin embargo, él reconoció que en contextos de RAPs podía fallar la sustitutibilidad entre expresiones L-equivalentes. Quine (1980/2002) consideró que los sentidos, así como otras entidades intensionales, son “creaturas de la obscuridad”, porque sus criterios de individuación no son claros: no hay forma de decidir si aquel posible hombre gordo en la puerta y aquel posible hombre calvo en la puerta son el mismo o diferentes. Además, Quine era escéptico respecto a que se pudiera explicar el significado de una proposición apelando a algún criterio de sinonimia, sin incurrir en circularidad. Por otra parte, Davidson (1978) rechazó la idea de que en los contextos de RAPs se presenta un cambio de referente, pues esto involucra una pérdida de nuestra “inocencia semántica” de que las palabras “Alejandro Magno fue el conquistador” tienen un referente diferente porque van antecedidas por “David cree que”.

Una segunda preocupación general respecto a la teoría de Frege es que conduce a un *internalismo* del contenido mental. En un intercambio epistolar con Russell a finales de 1904, Frege escribió

“el Mont-Blanc, con todas sus nieves, no es parte componente del pensamiento de que el Mont-Blanc tiene una altura superior a los cuatro mil metros.” A lo que Russell le responde: “Yo opino que el Mont-Blanc mismo, pese a todas sus nieves, es una parte componente de lo que aseveramos... No se asevera el pensamiento... se asevera el objeto del pensamiento...” (citado por García-Carpintero, 1996, pp. 223-4). El internalismo es la idea de que los contenidos de nuestros estados mentales son independientes de cómo son las cosas en el mundo. La preocupación de Russell tiene que ver con que el fregeanismo hace que las condiciones de verdad no hagan parte de los contenidos de las creencias. Esto hace que los contenidos proposicional sean demasiado arbitrarios y oscurece la relación mente-mundo.

### 5.3 Teoría descriptiva de la referencia

Otra forma de defender el relacionismo proposicionalista es suponer que algunos términos singulares, como los nombres propios, no actúan como expresiones propiamente referenciales, sino más bien predicativas. Esto es lo que planteó Russell (1905) en su famosa *teoría de las descripciones definidas*, según la cual nombres propios como ‘Pegaso’ realmente son descripciones definidas abreviadas, es decir, expresiones predicativas complejas que cuantifican sobre un único elemento, aunque lo dejan indeterminado, como ‘el caballo alado de la mitología griega’, ‘el maestro de Alejandro Magno’, ‘el hombre más alto del mundo’, etc<sup>15</sup>. Según Russell, enunciados como “Pegaso vive en el Olimpo” establece una relación entre predicados, independiente de que refieran a algo, así que pueden llegar a ser significativos, a pesar de que no haya un objeto que satisfaga el sujeto de la oración. Además, si se despoja a los nombres de su carácter referencial, es posible explicar algunos casos de falla en la

---

15 Russell (1912/1995, 1918) consideraba que los *nombres lógicamente propios* son demostrativos deícticos como ‘esto’, ‘eso’, ‘aquel’, que permiten señalar ostensivamente un objeto, y se asocian a lo que puede ser conocido directamente o por familiaridad; en oposición a lo que se conoce por descripción.



sustitutibilidad. Sin embargo, Russell no aplicó sistemáticamente su teoría a los RAPs<sup>16</sup>.

Por otra parte, una forma de superar el primer problema del fregeanismo, el de la identidad de los pensamientos, es considerar que el *sentido* de los nombres propios consiste en un conjunto de descripciones asociadas a ellos. Frege sugirió esta vía, al afirmar que el sentido de un nombre propio como ‘Aristóteles’ podría ser ‘el discípulo de Platón’ o ‘el maestro de Alejandro Magno’ (Frege, 1892b, n. 4). Peter Strawson (1959/1989) sostuvo que es posible identificar y reidentificar particulares a partir de expresiones descriptivas que sean suficientes para su individualización. John Searle (1967) sugirió que los nombres propios tienen sentido y no solo denotación, en tanto que actúan como ganchos para colgar descripciones, correspondientes a la información del individuo que asociamos al nombre. Según Searle, a cada nombre propio se asocia un racimo (*cluster*) de descripciones unidas disyuntivamente, de las cuales ninguna es necesaria, pero aun así desempeñan un papel fundamental en la determinación de la referencia.

Las teorías descriptivistas de la referencia fueron duramente criticadas por los defensores de la teoría de la referencia directa. Saul Kripke (1981/1995) mostró que un nombre propio, como ‘Aristóteles’ es un *designador rígido*, en cuanto que refiere al mismo individuo en todos los mundos posibles, mientras que una descripción como ‘el maestro de Alejandro’ es un *designador no-rígido*, en tanto que si las cosas hubieran sido de otra manera es posible que el maestro de Alejandro no hubiera sido Aristóteles. Por otra parte, Kripke mostró que las descripciones asociadas a un nombre no son ni necesarias ni suficientes para la fijación de la referencia del nombre. Hilary Putnam (1975) realizó consideraciones similares

---

16 La teoría de las descripciones definidas de Russell estaba motivada en una polémica con Alexius Meinong (discípulo de Brentano) respecto a la significabilidad de oraciones con términos referencialmente vacíos, como “Pegaso es un caballo mitológico” y oraciones existenciales negativas, como “Pegaso no existe”, más que a solucionar el problema la intensionalidad de los RAPs. Para una introducción a la polémica de Russell con Meinong, puede consultarse Santamaría (2016, §§ 1.2-1.3).

para los términos de clases naturales (p. ej., ‘agua’, ‘oro’, ‘tigre’, etc.), y David Kaplan (1977) mostró que expresiones indexicales como ‘yo’, ‘el’, ‘suyo’, ‘este’, ‘aquí’, ‘ahora’, etc., tienen un significado exclusivamente referencial, que varía con el contexto. Por otra parte, se ha criticado que la teoría del racimo de descripciones hace que el significado de un nombre propio sea demasiado idiosincrático para cada usuario del nombre, impidiendo la comunicación intersubjetiva (Blasco, et al., 1999, McKay, 2005).

### 5.4 Neo-Proposicionalismo russelliano

Los teóricos de la referencia directa, como Kripke, Putnam, Kaplan, Salmon y Soames, rechazaron la idea fregeana de que el acceso a la referencia estuviera mediado por algún sentido o descripción identificadora. Kripke (1981/1995) propuso una *teoría causal de la referencia*, según la cual la relación de referencia se sostiene por una cadena histórico-causal, desde la asignación inicial del nombre hasta su uso presente. Estos autores son millianos, en cuanto defienden la posición de Mill (1886) de que los nombres propios no tienen connotación y su significado se agota en la denotación. A partir de las consideraciones sobre el significado de los nombres propios, los indexicales y los términos de clases naturales, estos teóricos han revivido la concepción original russelliana de que el contenido de las proposiciones consiste en sus condiciones de verdad. Lo anterior supone un *externalismo* del contenido, por el cual las condiciones de verdad son parte componente de los contenidos de los estados mentales (p. ej., Putnam, 1975). Esto en oposición al *internalismo*, más asociado al fregeanismo y su versión descriptivista, que rechaza que las condiciones de verdad sean parte del contenido mental (p. ej. Searle, 1983/1986).

Sin embargo, un problema inmediato del proposicionalismo russelliano es su dificultad para dar cuenta de la intensionalidad de los RAPs, como vimos en § 5.1. Nathan Salmon (1986, ver también Soames, 1989) argumentó que la diferencia en valores de verdad entre (29) y (37) es solo aparente. Él apeló a la distinción de Paul Grice (1975) entre el *contenido semánticamente expresado* de una

emisión y su *contenido pragmáticamente comunicado*. Según Salmon, el modo en que el objeto es aprehendido por cada nombre hace parte de lo pragmáticamente comunicado por el enunciado, pero no de lo que es semánticamente expresado. Así, en (29) y en (37) David está en una relación de creencia hacia la misma proposición, solo que (29) comunica pragmáticamente que lo está mediante el modo de aprehensión de ‘Alejandro’, mientras que (37) comunica pragmáticamente que lo está mediante el modo de aprehensión ‘el discípulo de Aristóteles’.

A esta propuesta se le han hecho varias críticas. Se ha cuestionado que en la teoría de las implicaturas conversacionales de Grice (1975) los hablantes se dan cuenta de la diferencia entre lo que se dice y lo que se implica, mientras que para el neorusselliano el hablante no sería consciente de esa diferencia. Además, Salmon no parece identificar con claridad cuál es el principio pragmático por el cual se confunde lo que es dicho semánticamente y lo que es comunicado pragmáticamente. Además, esta propuesta no soluciona el problema de las creencias con términos singulares referencialmente vacíos. Otra dificultad es que, a partir de las diferencias, ya sean semánticas o pragmáticas, entre (29) y (37), sería admisible suponer (41),

- (41) David cree que Alejandro fue mejor estrategia que el discípulo de Aristóteles;
- pero el neorusselliano debería aceptar que
- (42) David cree que Alejandro fue mejor estrategia que Alejandro,

lo cual hace pasar a David por irracional. Finalmente, el neorusselliano tendría dificultades para explicar por qué dadas las diferencias entre (29) y (37) David podría actuar de modo diferentes frente a ‘Alejandro’ y a ‘el discípulo de Aristóteles’ (Recanati, 1997, §§17.3-4, McKay, 2005, §5; Nelson, 2019, §5).

## 5.5 Russellianismo contextualista

Mark Richard (1989, 1990, 2006) propuso que los verbos de actitud psicológica expresan relaciones diádicas, pero tienen elementos indexicales que los hacen sensibles al contexto, de tal manera que un RAP podría ser verdadero en ciertos contextos, pero falso en otros. De acuerdo con Richards, la cláusula-que de un RAP es una *Matriz Anotada Russelliana* (MAR) lingüística, que consiste de tuplas o anotaciones de pares de expresiones lingüísticas y sus “interpretaciones russellianas”. También existen MARs cognitivas, que refieren a los contenidos de los estados mentales del sujeto adscrito. Un RAP será verdadero en cuanto se den las correlaciones correctas entre las anotaciones de la MAR lingüística, usada por quien adscribe el RAP, y las anotaciones del MAR cognitivo del sujeto adscrito. No obstante, dicha correlación puede variar contextualmente. Una de las restricciones contextuales que determina las condiciones de verdad de un RAP consiste en el sistema representacional cognitivo del sujeto adscrito. De este modo, puede ser el caso que (29) sea verdadero y (37) falso si, para el sistema cognitivo de David, no hay anotaciones que relacionen a Alejandro con el discípulo de Aristóteles.

Esta teoría supone que el problema de la substitutibilidad se presentará dependiendo del contexto dado por el sistema de creencias del sujeto de atribución. Pero puesto en estos términos parece más una paráfrasis del problema de la intencionalidad que una explicación de éste. Además, los contextos son subjetivos, lo que dificulta establecer la substitutibilidad para atribuciones de RAPs colectivas, como ocurre entre “todos los amigos de David creen que Alejandro Magno fue el conquistador de los persas” y “todos los amigos de David creen que el discípulo de Aristóteles fue el conquistador de los persas”. Finalmente, la noción de MAR cognitiva es bastante oscura y depende de la noción de representación mental, que como veremos (§ 5.8), también trae sus propios problemas.

## 5.6 Teoría del Indexical Oculto

Esta es una teoría intermedia entre el russellianismo y el fregeanismo, propuesta originalmente por Stephen Schiffer y defendida después por Crimmins y Perry (1989; Crimmins, 1992), según la cual los RAPs tienen un constituyente inarticulado (no pronunciado) que se refiere a cómo el sujeto concibe la proposición hacia la que tiene la actitud. El RAP hace referencia implícitamente a una ‘idea’ o ‘forma de captar’, la cual es determinada por el contexto extralingüístico. En esta teoría se considera que los contenidos de las clausulas-que de los RAPs son proposiciones russellianas, involucrando sus condiciones de verdad, pero se reconoce que la relación del sujeto con dicho contenido se da a través de un modo de presentación. Así, (29) y (37) se pueden analizar como

- (43) David cree la proposición russelliana <Alejandro, Conquistador de los persas> bajo el modo de presentación que identifica a Alejandro como el rey macedonio,
- (44) David cree la proposición russelliana <Alejandro, Conquistador de los persas> bajo el modo de presentación que identifica a Alejandro como el discípulo de Aristóteles.

Siendo posible que (43) sea verdadero y (44) falso, si David no reconoce a Alejandro como el discípulo de Aristóteles. De acuerdo con esta teoría, las creencias involucrarían una relación ternaria entre el sujeto, la proposición russelliana y el modo de presentación por el cual el sujeto cree en la proposición en cuestión (Schiffer, 2008, §13.2.1).

Se ha criticado que esta teoría viola el principio de restricción lingüística. Según este principio, lo que se dice siempre se puede rastrear hasta algún elemento sintáctico en la oración. Este principio se considera valioso, porque delimita la información transmitida por un enunciado, pero que no forma parte de lo que dice la oración pronunciada. Sin embargo, la teoría sostiene que ‘creer’ expresa una relación de tres lugares, aunque funciona sintácticamente como una relación de dos lugares (Nelson, 2019, § 6). Por

otra parte, el mismo Schiffer mas tarde criticó que un enunciado de creencia del tipo “ $S$  cree que  $p$ ” sea una relación ternaria de la forma “ $S$  cree que  $p$  bajo  $m$ ”, porque la primera expresión es semejante a la de “Camelia besó a Ralph bajo el muérdago”, pero no se supone que besar sea una relación ternaria (Schiffer, 2008).

### 5.7 Proposicionalismo no estructuralista: La teoría de los mundos posibles

Las proposiciones, sea desde la perspectiva russelliana o fregeana, son tomadas como entidades estructuradas. Pero también es posible considerarlas como entidades sin estructura. La principal propuesta en este sentido es la concepción de las proposiciones como conjuntos de mundos posibles, defendida por autores como Hintikka (1969), Stalnaker (1984) y Lewis (1986). Según esta teoría, el enunciado “Alejandro conquistó a los persas” consiste en el conjunto de todos los mundos posibles en los que Alejandro ejemplifica la propiedad de conquistar a los persas. Los contenidos de las oraciones son funciones (intensiones) que mapean de conjuntos de mundos posibles a valores de verdad. Sin embargo, un problema bien conocido es que la teoría no distingue el significado de oraciones lógicamente equivalentes, exigiendo a los sujetos omnisciencia lógica. Por ejemplo, los enunciados “ $2 + 2 = 4$ ” y “la matemática es incompleta”, expresan proposiciones verdaderas en los mismos mundos posibles, es decir, en todos, y, por tanto, según esta teoría, dicen lo mismo. Pero, puede ser verdad que “Juan cree que  $2 + 2 = 4$ ” mientras que sea falso que “Juan cree que las matemáticas son incompletas”, incumpléndose la sustitutibilidad de correferenciales *salva veritate*.

### 5.8 Relacionismo No-proposicionalista

Las teorías anteriores suponen que el contenido de nuestros estados mentales es proposicional, aunque se tengan diferentes concepciones de las proposiciones. Sin embargo, autores con proclividades nominalistas han rechazado apelar a proposiciones para evitar cualquier compromiso con entidades abstractas (Goodman, 1977;

Quine, 1960/2001 § 43; Davidson, 1978). Una primera solución, esbozada por Carnap (1947) es considerar que en un EAP un sujeto se encuentra en una relación, no con una proposición, sino con la oración en la cláusula-que del RAP que la describe. Así, las anteriores (29) y (37) se tratarían como:

- (45) David cree que “Alejandro Magno fue un conquistador”
- (46) David cree que “el discípulo de Aristóteles fue un conquistador”

donde el entrecomillado encierra la oración que es objeto de la relación de creencia para David. Las oraciones son tomadas como particulares lingüísticos más que lógicos. Ahora bien, es posible que David asienta a la cláusula-que de (45), pero no a la de (46), porque está en relación con oraciones diferentes. De esta forma se explica la falla en la sustitutibilidad de correferenciales *salva veritate*, conservando el relacionismo.

Un problema inmediato de esta propuesta es que si fuera el caso de que David no hablara español, (45) y (46) serían falsas. Carnap sugirió el criterio de *isomorfismo intensional*, por el cual dos oraciones  $O'$  y  $O''$  además de ser lógicamente equivalentes, tienen la misma estructura y componentes intensionalmente equivalentes. Si consideramos las oraciones anteriores (32) y (34) como isomórficamente intensionales, (45) se podría analizar como:

- (47) Hay una oración  $O$  (32) en un sistema semántico  $S$  tal que  $O$  es intensionalmente isomórfico con  $O'$  (34) y David estaría dispuesto a asentir a  $O$ .

Sin embargo, Alonzo Church (1950) mostró que (45) y (47) no tienen las mismas condiciones de verdad cuando se traducen a otros idiomas. Además, es posible que David muestre creencia a la oración (32), pero no a (34), a pesar de que muestren isomorfismo intensional, manteniéndose la falla en la sustitutibilidad (Orlando, 2020).

Donald Davidson (1978) sugirió una forma más sofisticada de oracionalismo con su *análisis paratático*. Este es un análisis que aplicó primero al discurso indirecto y luego lo extendió a verbos psicológicos como ‘creer’. La propuesta es que la palabra ‘que’ es un demostrativo que funciona de manera semejante a ‘esto’. Por tanto, si se tienen

- (48) David dijo esto. Alejandro Magno fue un conquistador,
- (49) David dijo esto. El discípulo de Aristóteles fue un conquistador,

es posible afirmar que (48) y (49) tienen diferentes condiciones de verdad, pues la oración señalada por ‘eso’ en cada caso es diferente. Pero al igual que con Carnap, es posible que (48) y (49) sean falsas si David no habla español. La solución de Davidson fue analizar (48) como “Una emisión de David *dice-lo-mismo* que esta emisión mía: Alejandro Magno fue un conquistador”. Pero se ha cuestionado que la relación de *decir-lo-mismo* introduce un criterio de sinonimia, que invita de nuevo al proposicionalismo que se quería evitar (Orlando, 2020; Schiffer, 2008).

Otra forma de relacionismo no-proposicionalista está presente en la *Teoría Representacional de la Mente* (TRM) de Jerry Fodor (1975/1984, 1978, 1987/1994). De acuerdo con Fodor, un EAP consiste en una relación psicológica de un sujeto con una *representación mental* (RM). Una RM es un estado particular del sistema nervioso, que actúa como un vehículo representacional respecto a cierto estado de cosas del mundo, que constituye su contenido. Así que una RM tiene un contenido intencional, evaluable semánticamente (p. ej., como verdadero o falso), tiene una estructura sintáctica y está instanciado físicamente en el sistema nervioso. Esta teoría permite explicar cómo podemos estar en relación a contenidos que incluyen términos singulares referencialmente vacíos, como en (38), pues se está en relación directa con la representación mental de dicho contenido y no con sus condiciones de satisfacción. Igualmente, es posible dar cuenta de la falla de susti-



tutibilidad de correferenciales *salva veritate* entre (29) y (37), suponiendo que se está en relación con dos representaciones mentales diferentes en cada caso. La TRM ha sido la explicación más popular en ciencias cognitivas respecto a cómo el contenido de nuestros estados mentales tiene un papel causal sobre nuestra conducta (Pylyshyn, 1984; Sterelny, 1990; von Eckardt, 1993).

Sin embargo, la principal dificultad de la TRM es explicar cómo estados subpersonales del sistema nervioso pueden adquirir un contenido representacional. Se han ofrecido explicaciones basadas en la relación de isomorfismo estructural (p. ej. Gładziejewski, 2015; O'Brien & Opie, 2009, 2015, Ramsey, 2007); roles conceptuales (p. ej., Block, 1986; Harman, 1982); causalidad (p. ej., Dretske, 1981/1987; Fodor, 1987/1994); o funciones bioteleológicas (p. ej., Millikan, 1984, 1989; Neander, 1995); pero todas estas teorías han sido objeto de críticas y se considera que no hay una solución suficientemente convincente del asunto hasta el momento (para una revisión del tema, Plazas, 2023, Cap. 3).

## 5.9 No relacionismo

Una última alternativa es considerar que los EAP no consisten en una relación del sujeto con algo. Quine (1960/2001, 1970/1973), uno de los principales críticos del proposicionalismo, avanzó esta idea, considerando que los estados de actitud psicológica se pueden tratar como propiedades monádicas del sujeto. Así, verbos como *creer* dejan de considerarse como relaciones, y más bien unidos al contenido de la cláusula-que conforman una propiedad compleja del sujeto. Por ejemplo, (29) sería analizado como:

(48) David cree-Alejandro-Magno-fue-un-conquistador,

en el que David tiene la propiedad de tener esa creencia con ese contenido particular<sup>17</sup>. Una ventaja de este análisis es que, dado

17 En Plazas (2023, § 2.1) se presenta un análisis semejante de la posición de Brentano sobre la estructura de los estados mentales.

que el estado de creencia es una propiedad no relacional del sujeto, las propiedades intensionales, como las creencias en inexistentes o la falla de sustitutibilidad, no requieren tener que dar cuenta de la naturaleza del contenido. Sin embargo, el no-relacionismo parece conllevar a un *internalismo* muy radical, que hace que los estados mentales estén inconexos con la realidad y sean difícilmente explicables. Por otra parte, la fusión el verbo psicológico con el contenido de la cláusula-que en un todo impide que sea analizado composicionalmente. Pero el rechazo del principio de composicionalidad, supone que cada predicado de creencia debe aprenderse por separado y no composicionalmente, lo cual parece inviable (García, 2011).

## 6. Conclusiones y algunas reflexiones finales

Como puede observarse, todas las teorías presentan problemas, y no hay acuerdo entre los especialistas acerca de la superioridad de una sobre las demás. Uno de los principales desacuerdos es el de si se tiene una actitud realista o antirrealista frente a las proposiciones. La mayoría de las propuestas son realistas, pero aún hay dificultades para ofrecer una explicación convincente de qué es una proposición, que no sea tan austera como el russellianismo, ni tan exótica como el fregeanismo. En cambio, la gran mayoría de propuestas afirma el relacionismo, y parece que su abandono es demasiado costoso. Haciendo un balance general, se podría concluir que encontrarse en un estado de actitud proposicional consiste en estar en una relación con algo, aunque aún es difícil determinar qué es ese algo.

Una tensión importante entre estas teorías es la que hay entre *internalismo* y *externalismo* del contenido. Las propiedades intensionales de los RAPs son más fácilmente explicables por el *internalismo*, pero este hace que sea problemático dar cuenta de la relación mente-mundo. El *externalismo*, en cambio, permite dar cuenta de por qué nuestros contenidos son usualmente acerca del mundo, pero tiene dificultades con las propiedades intensionales. Lo que cada una defiende parece irrenunciable, y al parecer la verdad se debe ubicar en un punto intermedio entre las dos.

Las teorías revisadas son conservadoras, en cuanto que buscan preservar los aspectos centrales de la SCV e intentan acomodar a ella las particularidades semánticas de los RAPs. (Clapp, 2006). La SCV sigue siendo un patrimonio fundamental del análisis filosófico dentro de la tradición analítica. Sin embargo, puede ser que los RAPs más que anomalías a ser regimentadas, simplemente muestren limitaciones fundamentales de la SCV, que exija alternativas radicales frente a la SCV. En la tradición analítica del lenguaje se han realizado desarrollos alternativos a la semántica desde la pragmática (Clapp, 2006), como la *Teoría de la Relevancia* de Sperber y Wilson (1986) o la *Teoría de la Representación del Discurso* de Kamp (Asher, 1993). También se han ofrecido otras interpretaciones de los RAPs desde teorías alternativas del significado, como la teoría semiótica de Charles Peirce (p. ej. Plazas, 2023). Pero no es posible abordar estas propuestas en este escrito.

Finalmente, no debe olvidarse que estos análisis se han limitado a los enunciados sobre estados de actitud proposicional, pero también se reconocen estados mentales fenoménicos (§ 1). El análisis de los RAPs se extiende a los estados fenoménicos, en la medida en que se considere que tales estados también tienen un contenido intencional. Por ejemplo, cuando se dice que la percepción supone un tipo de juicio, este juicio puede ser asimilado a un estado de creencia y entendido como un estado semejante a los de actitud proposicional (ver p. ej., García-Carpintero, 2000). Sin embargo, los defensores de las *Teorías de la Intencionalidad Fenomenal* suponen que los estados fenomenales son más básicos que los EAPs, y más bien los segundos deberían analizarse en términos de los primeros, de tal manera que rechazarían un intento de expresar los estados fenomenales en términos de RAPs (p. ej., Mendelovici & Bouget, 2014, 2020).

## Referencias

- Anscombe, E. (1991). *Intención*. México: UNAM/Paidós. (Publicado originalmente en inglés en 1957)
- Asher, N. (1993). *Reference to abstract objects in discourse*. Kluwer.

- Baker, L. R. (1999). Folk psychology. En R. A. Wilson & F. C. Keil (Eds.). *The MIT Encyclopedia of the Cognitive Sciences*. pp. 319-320. The MIT Press.
- Bäuerle, R. & Cresswell, M. J. (2003). Propositional Attitudes. En D. M. Gabbay & F. Guenther (Eds.), *Handbook of Philosophical Logic*, vol 10. 2nd Edition. (pp. 121-142). Springer.
- Berkeley, G. (1710/1992). *Tratado sobre los Principios del Conocimiento Humano*. Alianza.
- Block, N. (1986). Aviso en favor de una semántica para la psicología. En E. Rabossi (Comp.) (1995), *Filosofía de la Mente y Ciencia Cognitiva* (pp. 289-330). Paidós.
- Brentano, F. (2020). *Psicología desde el Punto de Vista Empírico*. Sigüeme. (Original publicado en alemán en 1874)
- Carnap, R. (1932). La superación de la metafísica mediante el análisis lógico del lenguaje. En A. J. Ayer (comp.), *El Positivismo Lógico* (pp. 66-87). Fondo de Cultura Económica.
- Carnap, R. (1932-3). Psicología en lenguaje fisicalista. En A. J. Ayer (Comp.) (1965). *El Positivismo Lógico* (pp. 171-204). Fondo de Cultura Económica.
- Carnap, R. (1947). *Meaning and Necessity: A study in semantics and modal logic*. The University of Chicago Press.
- Carnap, R. (1998). *Filosofía y Sintaxis Lógica*. UNAM. (Original publicado en inglés en 1935).
- Chisholm, R. M. (1955-1956). Sentences about Believing. *Proceedings of the Aristotelian Society*, 56, 125-148.
- Chisholm, R. M. (1957). *Perceiving: A Philosophical Study*. Cornell University Press.
- Church, A. (1950). On Carnap's analysis of statements of assertion and belief. *Analysis*, 10, 67-99.
- Clapp, L. (2006). Propositional Attitude Ascription: Philosophical Aspects. En A. Barber & R. J. Stainton (Eds.), *Concise Encyclopedia of Philosophy of Language* (pp. 613-617). Elsevier.
- Crane, T. (2001). *Elements of Mind*. Oxford University Press.
- Crane, T. (2009). *La Mente Mecánica: Introducción filosófica a mentes, máquinas y representación mental*. 2da Edición. Fondo de Cultura Económica.
- Crane, T. (2013). *The Objects of Thought*. Oxford University Press.

- Crimmins, M. & Perry, J. (1989). The prince and the phone booth: Reporting puzzling beliefs. *Journal of Philosophy*, 86, 685-711.
- Crimmins, M. (1992). *Talk About Beliefs*. The MIT Press.
- Crimmins, M. (1997). Proposition. En P. Lamarque (Ed.), *Concise Encyclopedia of Philosophy of Language* (pp. 287-291). Pergamon.
- Davidson, D. (1963). Acciones, razones y causas. En D. Davidson (1995). *Ensayos sobre Acciones y Sucesos* (pp. 17-36). UNAM/Crítica.
- Davidson, D. (1970). Sucesos mentales. En D. Davidson (1995). *Ensayos sobre Acciones y Sucesos* (pp. 263-287). UNAM/Crítica.
- Davidson, D. (1974). Psicología como filosofía. En D. Davidson (1995). *Ensayos sobre Acciones y Sucesos* (pp. 289-307). UNAM/Crítica.
- Davidson, D. (1978). Al decir that. En D. Davidson (1990). *De la Verdad y de la Interpretación: Fundamentales contribuciones a la filosofía del lenguaje* (pp. 108-122). Gedisa.
- De la Mettrie, J. O. (1747/1961). *El Hombre Máquina*. Eudeba.
- Dennett, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Bradford Books, Publishers.
- Dennett, D. C. (1996). *Contenido y Conciencia*. Gedisa. (Original publicado en inglés en 1969).
- Dennett, D. C. (1998). *La Actitud Intencional*. Gedisa. (Original publicado en inglés en 1987).
- Descartes, R. (1637). Discurso del método. En R. Descartes (2011). *Obras* (pp. 97-152). Gredos.
- Descartes, R. (1642). Meditaciones metafísicas seguidas de las objeciones y respuestas. En R. Descartes (2011). *Obras* (pp. 153-413). Gredos.
- Dretske, Fred I. (1987). *Conocimiento e Información*. Salvat. (Original publicado en inglés en 1981).
- Fernández, Luis (2012). Sentido / Referencia. En L. Vega & P. Olmos (Eds.), *Compendio de Lógica, Argumentación y Retórica* (pp. 543-547). Trotta.
- Fodor, J. A. (1975/1984). *El Lenguaje del Pensamiento*. Alianza. (Original publicado en inglés en 1975)
- Fodor, J. A. (1978). Las actitudes proposicionales. En E. Rabossi (Comp.) (1995). *Filosofía de la Mente y la Ciencia Cognitiva* (pp. 173-203). Paidós.
- Fodor, J. A. (1994) *Psicosemántica: El problema del significado en la filosofía de la mente*. Tecnos. (Original publicado en inglés en 1987).

- Frege, G. (1891a). Función y Concepto. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 53-79). Tecnos.
- Frege, G. (1891b). Carta a Husserl. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 80-83). Tecnos.
- Frege, G. (1892a). Sobre Concepto y Objeto. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 123-139). Tecnos.
- Frege, G. (1892b) Sobre Sentido y Referencia. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 84-111). Tecnos.
- Frege, G. (1895) Comentarios sobre Sentido y Referencia. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 53-97). Tecnos.
- Frege, G. (1918-9). El Pensamiento: Una investigación lógica. En G. Frege (1998). *Ensayos de Semántica y Filosofía de la Lógica* (pp. 196-225). Tecnos.
- Frege, G. (1972). *La Conceptografía*. UNAM. (Original en alemán, publicado en 1879)
- Gamut, L. T. F. (2010). *Lógica, Lenguaje y Significado: Lógica intensional y gramática lógica*. Editorial Universidad del Rosario. (Original en holandés publicado en 1982).
- García-Carpintero, M. (2000). Las razones para el dualismo. En P. Chacón & M. Rodríguez (Eds.), *Pensando la Mente: Perspectivas en filosofía y psicología* (pp. 27-119). Biblioteca Nueva.
- Gładziejewski, P. (2015). Explaining cognitive phenomena with internal representations: A mechanistic perspective. *Studies in Logic, Grammar and Rhetoric*, 40, 63-90. <https://doi.org/10.1515/slgr-2015-0004>
- Goodman, N. (1977). *The Structure of Appearance*. 3a Ed. Reidel.
- Gopnik, A. (1999). Theory of Mind. En R. A. Wilson & F. C. Keil (Eds.). *The MIT Encyclopedia of the Cognitive Sciences* (pp. 838-340). The MIT Press.
- Grice, P. (1975). Lógica y conversación. En En L. M. Valdés (Comp.) (2012), *La Búsqueda del Significado* (pp. 520-538). Tecnos.
- Harman, G. (1982). Semántica del rol conceptual. En L. M. Valdés (Coomp.) (2012). *La Búsqueda del Significado* (pp. 649-668). Tecnos.
- Hegel, G. W. F. (1807/1966). *Fenomenología del Espíritu*. Fondo de Cultura Económica.
- Hempel, C. (1949). The Logical Analysis of Psychology. En H. Feigl and W. Sellars (Eds.), *Readings in Philosophical Analysis* (pp. 373-84). Appleton-Century-Crofts.

- Hempel, C. G. (1950). Problemas y cambios en el criterio empirista de significado. En A. J. Ayer (comp.), *El Positivismo Lógico* (pp. 115-136). Fondo de Cultura Económica.
- Hintikka, J. (1969). Semantics for propositional attitudes. En J. Davis et al. (Eds.), *Philosophical Logic*. D. Reidel.
- Husserl, E. (1982). *Investigaciones Lógicas*. Vols. 1 y 2. Segunda Edición. Alianza. (Original publicado en su segunda edición en alemán en 1913).
- Huxley, T. H. (1874). On the Hypothesis that Animals are Automata, and its History. *The Fortnightly Review*, 16, 555–580.
- James, W. (2020). *Ensayos de Empirismo Radical*. Cactus. (Original publicado en inglés en 1912).
- Kaplan, D. (1977). Demostrativos: Ensayo sobre la semántica, la lógica y la epistemología de los demostrativos y otros indéxicos. En M. Excurdia (Comp.) (2014). *Los Indéxicos y la Semántica de Kaplan*. (pp. 51-139). UNAM.
- Kripke, S. (1995) *El Nombrar y la Necesidad*, México: UNAM. (Original publicado en inglés en 1981).
- Leibniz, G. W. (1695). Nuevo sistema de la naturaleza y de la comunicación de las sustancias así como de la unión que hay entre el alma y el cuerpo. En E. De Olaso (Editor) (1982). *G.W. Leibniz: Escritos Filosóficos* (pp. 459-471). Charcas.
- Lewis, D. (1986). *On the Plurality of Worlds*. Basil Blackwell.
- Mach, E. (1914). *The Analysis of Sensations and the Relation of the Physical to the Psychological*. Quinta edición. The Open Court Publishing Company. (Original publicado en alemán en 1906).
- Malebranche, N. (2009). *La Investigación de la Verdad*. Sígueme. (Trabajo original, sexta edición, publicada en 1712).
- McIntyre, R. & Woodruff-Smith, D. (1989). Theory of intentionality. En J. N. Mohanty & W. R. McKenna (Eds.), *Husserl's Phenomenology: A Textbook* (pp. 147-149). University Press of America.
- McKay, T. (2005). *Propositional Attitude Reports*. E. N. Zalta (Ed.), Stanford Encyclopedia of Philosophy. <http://plato.stanford.edu/archives/fall2005/entries/prop-attitude-reports/>

- Meinong, A. (1904). The Theory of Objects. En R. M. Chisholm (Ed.) (1960). *Realism and the Blackground of Phenomenology* (pp. 76-117). Free Press.
- Mendelovici, A. & Bourget, D. (2014). Naturalizing intentionality: Tracking theories versus phenomenal intentionality theories. *Philosophy Compass*, 9 (5), 325-337.
- Mendelovici, A. & Bourget, D. (2020). Conciousness and Intentionality. En U. Kriegel (Ed.), *The Oxford Handbook of the Philosophy of Conciousness* (pp. 560-585). Oxford University Press.
- Mill, J. S. (1886). *System of Logic. Ratiocinative and Inductive*. 8a Ed. Longmans Green, and Co.
- Millikan, R. G. (1984). *Language, Thought, and other Biological Categories*. Cambridge, MA: The MIT Press.
- Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, 86, 281-297.
- Neander, K. (1995). Misrepresentating & Malfunctioning. *Philosophical Studies*, 79, pp. 109-141.
- Nelson, M. (2019). *Propositional Attitude Reports*. En Edward N. Zalta (Ed.), Stanford Encyclopedia of Philosophy. <http://plato.stanford.edu/archives/spr2019/entries/prop-attitude-reports/>
- O'Brien, G. & Opie, J. (2009). The role of representation in computation. *Cognitive Processing*, 10, 53-62. <https://doi.org/10.1007/s10339-008-0227-x>
- O'Brien, G. & Opie, J. (2015). Intentionality lite or analog content? *Philosophia*, 43, 723-730. <https://doi.org/10.1007/s11406-015-9623-5>
- Orlando, E. (2020). Atribuciones de actitudes proposicionales. En E. García-Ramírez (Coord.) *La Estructura del Significado* (pp. 235-265). UNAM, Editorial Colofón.
- Pineda, David (2012). *La Mente Humana: Introducción a la filosofía de la psicología*. Cátedra.
- Plazas, E. A. (2023). *Intencionalidad y Contenido Mental: Aproximación desde una Semiótica Naturalista*. Disertación doctoral. Universidad de Salamanca.
- Putnam, H. (1975). El significado del significado. En H. Putnam (2012). *Mente, Lenguaje y Realidad* (pp. 165-241.). Universidad Autónoma de México.



- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a foundation for cognitive science*. The MIT Press.
- Quine, W. V. O. (1956) Cuantificadores y actitudes proposicionales. En L. M. Valdés Villanueva (Comp.) (2012). *La Búsqueda del Significado* (pp. 171-180). Tecnos.
- Quine, W. V. O. (1973). *Filosofía de la Lógica*. Alianza. (Original publicado en inglés en 1970).
- Quine, W. V. O. (2001). *Palabra y Objeto*. Herder. (Original publicado en inglés en 1960).
- Quine, W. V. O. (2002). *Desde un Punto de Vista Lógico*. Paidós. (Original publicado en su segunda edición en inglés en 1980).
- Rakova, M. (2007). *Philosophy of Mind A-Z*. Edinburgh University Press.
- Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge University Press.
- Recanati, F. (1997). *Direct Reference: From Language to Thought*. Oxford: Blackwell.
- Richard, M. (1989). Cómo digo lo que tú piensas. En M. Valdés (Comp.) (1996), *Pensamiento y Lenguaje: Problemas en la atribución de actitudes proposicionales* (pp. 251-280). UNAM.
- Richard, M. (1990). *Propositional Attitudes: An essay on thoughts and how we ascribe them*. Cambridge University Press.
- Richard, M. (2006). Propositional attitude ascription. En M. De Witt & R. Hanley (Eds.) (2006). *The Blackwell Guide to the Philosophy of Language* (pp. 186-211). Blackwell.
- Russell, B. (1918). La Filosofía del Atomismo Lógico. En J. Muguerza (Sel.) (1974). *La Concepción Analítica de la Filosofía* (pp. 139-251). Alianza.
- Russell, B. (1905). On Denoting. *Mind*, 14, pp. 479-493.
- Russell, B. (1918). La Filosofía del Atomismo Lógico. En J. Muguerza (Sel.) (1974). *La Concepción Analítica de la Filosofía* (pp. 139-251). Alianza.
- Russell, B. (1921). *The Analysis of Mind*. George Allen & Unwin Ltd.
- Russell, B. (1940). *An Inquiry into Meaning and Truth*. Routledge.
- Russell, B. (1995). *Los Problemas de la Filosofía*. Labor (Original publicado en inglés en 1912).
- Ryle, G. (2005). *El Concepto de lo Mental*. Paidós. (Original publicado en inglés en 1949).

- Salmon, N. (1986). *Frege's Puzzle*. Ridgeview Publishing Company.
- Santamaría, F. (2016). *Hacer Mundos: El nombrar y la significatividad*. Segunda Edición. Siglo del Hombre.
- Schiffer, S. (2008). Propositional Content. En E. Lepore & B. C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 267-294). Clarendon Press.
- Schütz, A. (1954). Formación de Conceptos y Teorías en las Ciencias Sociales. En J. M. Mardones (Cooomp.) (1991). *Filosofía de las Ciencias Humanas y Sociales: Materiales para una fundamentación Científica*. Anthropolos.
- Searle, J. (1967). Nombres propios y descripciones. En En L. M. Valdés Villanueva (Comp.) (2012), *La Búsqueda del Significado* (pp. 105-114). Madrid: Tecnos.
- Searle, J. (1979). *Expression and Meaning: Studies in the theory of speech acts*. Cambridge Universtiy Press.
- Searle, J. (1986). *Intencionalidad: Un ensayo en filosofía de la mente*. Tecnos. (Original publicado en inglés en 1983).
- Searle, J. (2009). *Actos de Habla*. Cátedra. (Original publicado en inglés en 1969).
- Soames S. (1989). Referencia directa y actitudes proposicionales. En M. Valdés (Comp.) (1996), *Pensamiento y Lenguaje: Problemas en la atribución de actitudes proposicionales* (pp. 219-250). UNAM.
- Sperber, D. & Wilson, D. (1986). *Relevance: communication and cognition*. Segunda Edición Blackwell.
- Spinoza, B. (2000). *Ética Demostrada según el Orden Geométrico*. Trotta. (Trabajo original publicado en 1677).
- Stalnaker, R. (1984). *Inquiry*. The MIT Press.
- Sterelny, K. (1990). *The Representational Theory of Mind: An introduction*. Basil Blackwell.
- Strawson, P. (1989). *Individuos: Ensayo de metafísica descriptiva*. Tecnos. (Original publicado en inglés en 1959).
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.
- Twardowski, K. (1977). *On the Content and Object of Presentations: A Psychological Investigation*. Martinus Nijhoff. (Original publicado en alemán en 1894).

- Urmson, J. O. (1978). *El Análisis Filosófico*. Ariel. (Publicado originalmente en inglés en 1956)
- Von Eckardt, Barbara (1993). *What Is Cognitive Science?* The MIT Press.
- von Wright, G. H. (1979). *Explicación y Comprensión*. Alianza. (Original publicado en inglés en 1971)
- Whitehead, A. N. & Russell, B. (1927) *Principia Mathematica*, Vol. I. Segunda edición. Cambridge University Press.
- Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*. En L. Wittgenstein (2009) *Obras*, pp. 1-137. Gredos.



# CAPITULO V

## THE MIND-BODY PROBLEM IN THE ANALYTIC TRADITION

*Michael Schon*<sup>1</sup>

### RESUMEN

El principal objetivo de este capítulo es trazar ciertos debates acerca de la relación entre la consciencia y el cerebro a partir de los

---

1 Michael Schon es profesor asistente visitante en el Departamento de Filosofía de la Texas Tech University (EE. UU.). Es doctor en Filosofía por la Universidad de Wisconsin-Madison y especialista en filosofía de la mente y metafísica.

*Acknowledgments:* For helpful discussions, I'm indebted to Larry Shapiro, Alan Sidelle, Martha Gibson, Hubert Marciniak, and Greg Nirshberg. Also, a special thanks to the students of my 2023 seminar on the Mind-Body Problem at Texas Tech University whose questions pushed me to further clarify the connections between these two historical periods in my own thinking. Insightful comments by an anonymous referee for the volume pushed me to clarify several important points of the paper. The initial idea for the paper was inspired by discussions in Martha Gibson's 2013 Pro-Seminar and her 2014 seminar on Consciousness. So, thank you to all the participants in those discussions as well.

debates, a inicio del periodo moderno, entre Descartes y Arnauld (entre otros) a lo largo de la tradición analítica. Con el auge del conductismo y la teoría de la identidad mente-cerebro, a mediados del siglo veinte, los filósofos analíticos mayoritariamente pensaron que el argumento Cartesiano en que la concebibilidad es una guía para la posibilidad era irrelevante para resolver el problema mente-cuerpo. Mas bien, la identidad entre la mente y el cerebro fue vista como una hipótesis científica, y, por ende, como un hecho contingente. En la obra *Naming and Necessity*, Kripke muestra que esta posición estaba errada, así, reintroduce el argumento Cartesiano a favor del dualismo. Los materialistas respondieron que la concebibilidad—una noción epistémica—no puede ser una guía confiable a la posibilidad metafísica. Éstos argumentaron que las ciencias físicas explican todas las relaciones causales, y ya que ninguna mente o propiedad fenomenal han sido propuestas por éstas ciencias, tales mentes o propiedades fenomenales deben ser en realidad físicas. Los dualistas tienen muchas respuestas, todas respetando a las ciencias naturales, y, a la vez, dando cabida en el mundo natural a propiedades no-físicas. Este impase ha guiado a los filósofos analíticos a explorar posiciones que ni caen dentro del dualismo, ni del materialismo. El capítulo muestra cómo los argumentos del periodo moderno temprano informaron los debates, de los filósofos analíticos, acerca del problema mente-cuerpo. Así también, muestra como los desarrollos filosóficos del siglo veinte reformularon dichos debates.

## ABSTRACT

The main aim of this chapter is to trace the themes of the debates about the relationship between consciousness and the physical brain from debates in the early modern period between Descartes and Arnauld (among others) through the analytic tradition. With the rise of behaviorism and mind-brain identity theory in the middle of the twentieth century, analytic philosophers largely thought the Cartesian argument that conceivability is a guide to possibility was irrelevant to resolving the mind-body problem. Instead, the identity between the mind and brain was seen as a scientific hypothesis and

therefore a contingent fact. Kripke's *Naming and Necessity* showed this to be misguided, thus reintroducing the Cartesian argument for dualism. Materialists responded by arguing that conceivability – an epistemic notion – cannot be a reliable guide to metaphysical possibility. Instead, they have argued that the physical sciences explain all causal relations and since no minds or phenomenal properties are proposed by the physical sciences, such minds and phenomenal properties must actually be physical. Dualists have many responses at the ready, all of which proclaim to respect the natural sciences while making room in the natural world for distinct, non-physical properties. This has resulted in something of an impasse that has led analytic philosophers to explore views that don't fall into either dualist or materialist categories. The chapter will show how the arguments of the early modern period informed analytic philosophers' debates about the mind-body problem while also demonstrating how the philosophical developments of the twentieth century reshaped the debate.

## 1. Introduction: Descartes and His Critics

The mind-body problem is the problem that results from trying to answer the question "What is the relation between the mind and body (specifically the brain)?" The way the problem gets discussed in analytic philosophy can be seen in a nascent form already in Descartes' arguments for the distinction of the mind and body in his *Meditations on First Philosophy*<sup>2</sup>, first published in Latin in 1641, especially in the sixth meditation and the responses to the arguments he makes therein. The two major questions that arise out of these debates are i) whether merely conceiving of some state of affairs is proof of what is genuinely possible (as Descartes argues) and ii) how the mind, as only a thinking, non-extended thing, is causally related to the body, an extended, non-thinking thing. In what fol-

---

2 I'll be using Michael Moriarty's (2008) translation into English of both Descartes' *Meditations on First Philosophy* and the *Objections and Replies*. I'll be using Lisa Shapiro's (2007) translation of the correspondence between Princess Elisabeth and Descartes.

lows, I'll trace how Descartes attempts to answer these questions and show how both the questions and Descartes' answers get developed in the analytic tradition throughout the twentieth century.

I'll begin here by laying out Descartes' argument, focusing on the way he presents it in his *Meditations on First Philosophy*<sup>3</sup>. Then I'll describe the way in which Antoine Arnauld raises the first major question, followed by a description of how Princess Elisabeth of Bohemia raises the second major question. I'll also present Descartes' responses to these questions. In the next section, I'll recount some trends in philosophy and science in the twentieth century that mark some important departures from the way in which the problem was discussed by Descartes and his contemporaries in the seventeenth century. With these differences in mind, in the third section, I'll demonstrate how the Cartesian argument resurfaces (at least in outline) within the context of the analytic tradition, focusing in particular on Saul Kripke's presentation in *Naming and Necessity*. I'll also show how the responses from Princess Elisabeth and Antoine Arnauld also resurface (again, at least in outline). Finally, I'll conclude by surveying some alternative views that have been a major development in the latter part of the twentieth century and into the twenty-first century due to the growing dissatisfaction with the prevailing solutions to the mind-body problem.

Descartes' argument for the distinction between the mind and body is well-known, influential, and controversial. It is often called his Conceivability Argument. The basic premises of that argument are that:

- 1) The mind can be clearly and distinctly thought of as existing separately from the body.
- 2) If two things can be clearly and distinctly thought of as existing separately, then it is possible that they are distinct entities.

---

3 Descartes makes quite a similar argument in his *Discourse on the Method* and elsewhere. The reason for focusing on his presentation in just one place is to avoid any difficulties that may arise from potential differences in his arguments elsewhere (of which there are several).



- 3) If it is possible that two things are distinct entities, then they are actually distinct, even if we have no idea how that separation might occur.

The conclusion of this is that since the mind could have been created separately from the body, they are in fact two things, not one (Descartes, 2008, p. 55). This is quite a quick presentation of the argument, and each step is difficult to understand, so let's break them down and see why Descartes believes each premise is true. If they are, he has a sound argument, and the mind and body are indeed distinct.

In the first premise, what is crucial to recognize is that we do not merely have *any* kind of thought that the mind and body might exist separately. Instead, our thinking that the mind and body can exist separately is *clear and distinct*. In the Second Meditation (2008, p. 23), Descartes distinguishes between thoughts that are imperfect and confused and those that are clear and distinct. He argues that many of the things we commonly take ourselves to understand “most distinctly” are actually things about which our ideas are quite confused. It is only after a great deal of contemplation – or, perhaps more appropriately given the title of his book, “meditation” – that we can arrive at a genuinely clear and distinct thought about something, even something as ordinary and common as a piece of wax. This is the example that Descartes uses at the end of the Second Meditation to demonstrate just how prone to error we are when thinking about nearly everything. As such, we must carefully contemplate everything to avoid errors.

Descartes begins with a basic description of a piece of wax, “it has only just been removed from the honeycomb; it has not yet lost all the flavour of its honey; it retains some of the scent of the flowers among which it was gathered; its colour, shape, and size are clearly visible; it is hard, cold, easy to touch, and if you tap it with your knuckle, it makes a sound” (2008, p. 22). But then the wax is brought close to a fire. It begins to melt, losing any remaining flavor and smell. The shape, color, and size all change. It is now hot, liquid,

and makes no sound when struck. Nevertheless, Descartes says, it is the same wax. None of its previous qualities remain, yet the wax itself does. We recognize this fact even though, to our senses, everything about the wax is different. Thus, it must not be by the senses that we know what the wax is, even though our initial description is always sensory. Descartes concludes that the wax is “something extended, flexible, mutable: certainly, that is all” (2008, p. 22). Moreover, he is perceiving these properties of the wax with his “mind alone”. Nevertheless, this wax that he perceives with his mind alone is the same wax that he first described using his senses. “But yet – and this is important – the perception of it is not sight, touch, or imagination, and never was, although it seemed to be so at first: it is an inspection by the mind alone, which can be either imperfect and confused, as it was before in this case, or clear and distinct, as it now is, depending on the greater or lesser degree of attention I pay to what it consists of” (2008, p. 23). The point of using a piece of wax is that it is a common object about which our beliefs are usually quite confident. We know that the wax has such-and-such properties by looking at it, smelling it, touching it, and so on. But this confidence does not free us from error.

As he shows us, our normal beliefs about the piece of wax are only about its appearances. What the wax *really* is must be arrived at through a process of careful contemplation. This allows us to think about it clearly and distinctly, which is how we are sure that our thoughts about it are not wrong. To put it simply, clear and distinct thoughts about something are thoughts which are free from error because we have carefully considered what it is we are *really* thinking about rather than merely being confident that the appearances contain no falsehoods. This is why it is important that Descartes claims he can *clearly and distinctly* conceive the mind and body existing separately. He is not merely relying on their initial appearances (which might vary from person to person). He is claiming that the conception he has of the mind and brain are such that they could exist separately from one another, even after deep and careful consideration and even though they always appear together.

The second premise of Descartes's argument is that anything that we can clearly and distinctly conceive of as existing separately from some other thing could have been created separately from it. What he means is that even though some pairs of things might always accompany one another, if we can conceive of them existing separately from one another, it is possible for them to come apart. What is crucial about this premise is the *possibility* of the two things coming apart, not that we could ever actually see this happening. As he states it, "they can at least be produced separately by God" (2008, p. 55, emphasis mine). It may in fact be the case that God did not create these two things separately, but the fact that we can conceive of them existing on their own is enough to prove that it would have been possible for God to create them separately. Of course, since we only exist in the universe as it is, we won't be able to *actually* see this separation. It is something that is merely possible. An example should help illustrate Descartes' meaning.

Let's consider a statue and the stuff that it's made of<sup>4</sup>. Just to be very clear, let's consider a specific statue, say the Statue of Liberty. This statue is made of copper. The specific lump of copper from which it was made was shipped from France in various parts and assembled to create the statue in New York where it stands today. Because it was constructed piece by piece, the whole lump of copper and the statue came together at exactly the same time. So, anywhere and at any time that we have the Statue of Liberty, we have had the exact lump of copper it is made from. We have never (and suppose we never will) seen the two separated. Nevertheless, it seems as if the statue and the copper must be two different things. After all, while the facts of history are such that the lump of copper came to be the Statue of Liberty, it seems perfectly reasonable to suppose that the very same copper could have

---

4 This puzzle arises in many metaphysical arguments in ancient Greek philosophy, most famously in Aristotle's works. They are concerned primarily with an issue concerning material objects (and the example is still prominent in such debates today in analytic philosophy), but the puzzle demonstrates the importance of the idea of deriving possibilities from the way we think of things.

been used to make some other statues – or no statue at all! It also seems quite reasonable that we could melt the copper into a giant indistinct shape, thus preserving the copper as one lump but destroying the Statue of Liberty. Indeed, who could deny such a fact? But we will probably never see any such thing happen. All we have are our thoughts about the statue and the copper and the fact that they *could* exist separately, even though they in fact do not. By similar line of reasoning, Descartes argues that the fact that we can clearly and distinctly conceive of the mind existing separately from the body (and the brain in particular) proves that the mind and the body could have been made such that they would exist separately, even though as a matter of the way the world is we will never see this happen.

The third premise of Descartes' argument states that any two things that could possibly have been created separately are actually distinct things, even if as a matter of fact, they are never separated (and as above can't be separated within the laws of the universe as they currently exist). The basic idea here is pretty simple, but it has important consequences. The idea is that if a pair of things could exist separately, then one of them could exist while the other doesn't exist at all. But then this shows that they are *actually* two distinct things rather than merely being one thing that *could have been* two distinct things. Thus, Descartes concludes his argument by saying that "I am really distinct from my body, and can exist without it" (2008, p. 55). Or, to take up the example from above, there are really two things, the Statue of Liberty and the copper which composes it. The remarkable point about this premise is the move from some truth about what might merely be possible to some truth about the way the world *actually is*.

There is much more to be said on this argument and its plausibility. But, for now let's focus on Antoine Arnauld's objection to the argument from the Fourth Objections in the *Objections and Replies to the Meditations*. The main question he raises is whether it really makes sense to use our conceptions of things as proof for genuine possibility – he challenges the truth of premise 2. He does

this by way of an example (Descartes, 2008, pp. 128–130). Suppose someone understands that a triangle has one right angle and thus that it is a right triangle. But, due to some bad teacher, this person has come under the impression that the sum of the square of the lengths of its two legs does not equal the square of the hypotenuse. In other words, this person thinks that the Pythagorean theorem does not apply to this right triangle. Arnauld says that “Reasoning in the same way as M. Descartes has put forward, it seems he will confirm himself in his incorrect belief” (2008, p. 129). By mathematical proof, we know that the Pythagorean theorem applies to all right triangles, but following Descartes line of reasoning, this mistaken person will only confirm that the Pythagorean theorem does not apply to this right triangle. The argument might look like this:

- 1) I can clearly and distinctly conceive of this right triangle as existing separately from any triangle that is such that the Pythagorean theorem applies to it.
- 2) If two things can be clearly and distinctly thought of as existing separately, then it is possible that they are distinct entities.
- 3) If it is possible that two things are distinct entities, then they are actually distinct, even if we have no idea how that separation might occur.

Thus, this right triangle is separate from any triangle that is such that the Pythagorean theorem applies to it. Arnauld rightly points out that this is false. The Pythagorean theorem necessarily applies to all right triangles. Not even God could create a right triangle such that the Pythagorean theorem doesn’t apply to it<sup>5</sup>. Arnauld identifies the problem as coming from the second premise. The man

---

5 As with everything in philosophy, not everyone agrees with this statement. In particular, philosophers influenced by the metaphysical positions of Alexius Meinong maintain that there is a class of objects that are non-existent which can have contradictory properties. Thus, they believe that the statement in the main text is not true for such non-existent objects. For more on this, see (Marek, 2024).

in the example clearly and distinctly conceives of the right triangle but is wrong in his thinking that it could exist without being subject to the Pythagorean theorem. So, having a clear and distinct idea of something as being a certain way does not guarantee that it exists that way.

For the mind-body case, he concludes, “Likewise why should I not be deceived inasmuch as I think that nothing belongs to my nature, which I clearly and distinctly know to be that of a thinking thing, except my being a thinking thing – when in fact perhaps it is also part of my nature to be an extended thing?” (2008, p. 129). The idea behind premise 2 is that when we can clearly and distinctly conceive of something, we know what that thing *really* is. Thus, we know what it might or might not be. But here Arnauld gives us a good mathematical example where we seem to clearly and distinctly know what something is, and yet this gives us the wrong result about what is possible with respect to that thing. Thus, the second premise is false.

Arnauld briefly considers a response to his objection. He considers that Descartes might say that the man in the example does not really have a clear and distinct idea of the right triangle. But he argues that this is unsatisfactory because if he doesn’t have a clear and distinct idea of the right triangle, how could it be said that anyone has a clear and distinct idea of the human mind? The man knows very well that he is considering a right triangle, and the mind is vastly more complex than the triangle, after all. So, if he can make this kind of mistake with respect to a simple triangle, it seems quite likely that Descartes is making a similar mistake in his thinking about the mind.

Descartes’ response to this is to show that Arnauld’s example is disanalogous in an important way. Not only does Descartes have a clear and distinct idea of the mind as something that is not an extended thing, he also has a clear and distinct idea of the body that it is not a thinking thing. He concludes, “And there can be no greater proof of the distinction between two things, than the fact that, to whichever of them we turn our attention, we discover nothing at

all in it that is not distinct from the other” (2008, p. 146). On the other hand, he says that Arnauld’s triangle example is not analogous because no one can have a clear and distinct idea of a triangle to which the Pythagorean theorem applies without concluding that it is a right triangle. Part of the concept of the theorem is that it applies to all and only right triangles. Thus, we can see that the requirement for a clear and distinct conception of two things existing separately is actually quite a significant requirement.

If this were the only argument that Descartes’ contemporaries had against his dualism, it would be a pretty one-sided affair. Descartes has given us an argument for his dualistic conclusion, but Arnauld has only responded by trying to show that the argument is inadequate. However, this isn’t the only problem that Descartes’ contemporaries had with his position. The second major question for Cartesian dualism is about how it can explain mental causation. In her correspondence with Descartes, Princess Elisabeth of Bohemia raises this question by asking for Descartes to explain “how the soul of a human being (it being only a thinking substance) can determine the bodily spirits, in order to bring about voluntary action”<sup>6</sup>. She gives three suggestions of how physical objects are caused to move. They are due to i) “the impulsion of the thing moved”, ii) “the manner in which it is pushed by that which moves it”, or iii) “by the particular qualities and shape of the surface of the latter” (2007, p. 62). She does not expand upon these further except to mention that the first two require physical contact and extension is required for the third. The first seems to be in reference to an object having some non-zero momentum, already being in motion. This is its “impulsion”. But to change an object’s momentum requires something else coming into contact with it. The same

---

6 In the Fifth Objection to the *Meditations*, Gassendi raises a similar question. However, Descartes’ reply to him is much less helpful than his reply to Princess Elisabeth. Lisa Shapiro, in a footnote to her translation of the correspondence between Princess Elisabeth and Descartes, notes that Descartes’ reply is “less hospitable” to Gassendi. While true, it should also be mentioned that Gassendi’s objection is much less hospitable than that given by Princess Elisabeth, despite some similarities in content. It seems to me that this is a good demonstration of the saying “what goes around comes around”.

goes for the second suggestion, which is in reference to something with zero momentum, at rest. In order to move it, some other object would have to physically push it. And for the third, anything with a shape will require some extension in space. She points out that by Descartes' definition of the mind, it isn't extended, and thus can't meet the third requirement. Furthermore, it is unclear how anything immaterial could make contact with a material object in order to meet either of the first two conditions. Although Princess Elisabeth doesn't press the issue too strongly, we can view this as a version of an argument for materialism, which is sometimes called the Causal Argument. Put simply:

- 1) All causes are physical.
- 2) The mind causes physical motions in the body.
- 3) Thus, the mind is physical.

With this argument, we now have the making for quite a tricky debate. On the one side, dualists like Descartes have what seems to be a strong argument from what is conceivable to what is possible, proving that the mind is distinct from the brain – where this distinctness is supposed to be at the fundamental level. On the other side, materialists have a strong argument on the basis of causation to conclude that the mind is not distinct from the brain, at least, not fundamentally distinct<sup>7</sup>. Of course, Descartes' response to this argument is to deny the truth of the first premise. Denying the truth of the second premise would be to deny something that many take to be self-evident. The way that he denies the truth of the first premise is by reference to the substantial union between mind and body, which he claims is a primitive notion without further explanation (2007, pp. 69–71). In none of Descartes' writing does he give a positive account of what a "substantial union" amounts to<sup>8</sup>.

---

7 There are many materialists that agree that the mind and brain are distinct but where the mind ontologically depends on the brain so that they are not so distinct as dualism argues. This form of materialism will be discussed in more depth later as a response to Kripke's argument.

8 For a corroborating report of this, see Jonathan Bennett's note in his translation of Descartes' replies to objections to the *Meditations* in (Descartes, 2017, p. 59).



He does explain what the relation is *not* by pointing to a metaphor of a sailor in a ship. Like a sailor and his ship, the mind does seem to have some control over the body, but *unlike* the sailor the mind is not capable of coming and going. So, the substantial union of the mind and body is not like a sailor in a ship. Their union is more tightly connected. While it's not exactly clear what this relation is, it seems that he is pointing to something like the constant conjunction between the mind and brain, to borrow a phrase from Hume. Thus, while there is no *necessary* connection between them, the mind and brain always accompany one another in our experience.

This is, admittedly, a somewhat lackluster explanation for why some causes might not be physical – it seems to only explain why we *think* they are non-physical (though, on a Humean account of causation, maybe that's all there is to be said). But however we spell out the notion of a substantial union, it is clear that it is a fundamental relation that admits little to no further explanation. Thus, this “substantial union” is very similar to the notion of a fundamental psychophysical law that many analytic dualists claim unites mental and physical properties<sup>9</sup>. These laws are meant to mirror fundamental laws of physics which explain how different fundamental entities and properties interact with one another. Like the fundamental laws of physics, there is nothing more that we can say about them. The best we can do is describe them. But such laws could have been different, which explains why Descartes and other dualists argue that the mind and brain are fundamentally distinct.

## 2. Some Philosophical and Scientific Trends of the Early Twentieth Century

Early analytic philosophy was largely uninterested in the mind-body problem. Notable exceptions to this are C.D. Broad, Russell's com-

---

<sup>9</sup> The most prominent of these is (Chalmers, 1996), who has been a major figure in recent discussions on the mind-body problem since this book. He discusses psychophysical laws and their importance to his version of dualism in depth beginning on page 242. Current analytic dualists often point to Chalmers' naturalistic dualism as their source of inspiration. For instance, see Amy Kind's “dualism 2.0” in (Kind & Stoljar, 2023).

panion pieces on matter and mind, and later Ryle. Interest seems to be picked up in a more significant way in the 1950's with U.T. Place and J.J.C Smart, but even here we see that they largely argue that the mind should be considered in the same vein as scientific identities – phenomena which have been proven to be identical by the empirical sciences. The success of scientific reductions from phenomena previously thought non-physical was thought to be a mark that everything mental would swiftly follow suit. In particular, the discovery of DNA as the chemical substrate of life – the very fact that something is alive could be explained by the chemical structure of its cells which could in turn be explained by the physical makeup of that chemical structure – seemed to show that nothing was immune from the inevitable march of scientific reductions. This contrasted with the idea that life must be explained in fundamentally different terms from lower sciences, that it couldn't be explained by chemical structures, no matter how complex. The ability to explain life as arising out of complex chemical structures seemed to be evidence that inevitably, everything could be explained in this way, the mind included. But besides this, there are several trends in philosophy and science that set the stage for the debates and help explain why they take such a similar shape to the debate between Descartes and his contemporaries also while marking important differences. In this section, I'll highlight a few significant parts of the landscape of science and philosophy in the twentieth century that are important to understand the mind-body problem in the analytic tradition.

The first is the rise and popularity in analytic circles of Logical Positivism. This movement came out of the Vienna Circle in the 1920's with many of its figures moving to Britain and the United States in the 1930's due to the impending tyranny of the Second World War. There, the ideas that they had developed found a home among the philosophers who had already largely accepted their major tenets. These tenets were defined by conceptual clarity and rigorous, precise argumentation. Both of these were bolstered by the development of modern forms of mathematical logic and

led to the logical positivists attempting to reshape philosophy in the mold of modern science. As a result of this, much scorn was cast upon important figures in the history of philosophy, principle among them being Hegel. His methods of argumentation and literary style were considered a major problem as they made his points ambiguous, obscure, and thus only plausible because no one really understood what he was saying – the idea really being that he was not saying anything at all but only *appeared* to be saying things<sup>10</sup>. In contrast, the logical positivists (and by extension, analytic philosophers of the time) rejected as meaningless any statements that were not provable by either direct observation or rigorous and clear logical argumentation from such statements<sup>11</sup>.

This led to an important similarity between the wider landscapes of both Descartes and analytic philosophy. Namely that idealism<sup>12</sup> is generally not considered a viable solution to the mind-body problem. Of course, the idea that everything might fundamentally be mental rather than physical was known to both Descartes and his critics as well as analytic philosophers, but this solution was considered something to be avoided rather than a genuine possibility to be taken seriously. The fact that the physical sciences were advancing by leaps and bounds in both periods probably had some small part to do with this. But the view was popular in the intervening time between

---

10 This will inevitably touch a nerve for the many people who have found Hegel to be insightful – there is a reason he is a major historical figure in philosophy. In particular, continental philosophers disagree with this statement (and not all analytic philosophers agree with it either). There is good reason to think that (Rorty, 2016, pp. 44–47) is right when he argues that one’s dismissal of Hegel as saying meaningless nothings is a significant part of the analytic/continental divide. It is certainly the case that early analytic philosophers had people like Hegel in mind when critiquing the state of philosophy at the time.

11 For a detailed history of both logical positivism and the Vienna Circle, see (Creath, 2023; Uebel, 2024). The principle in the text, known as the verification principle or just verificationism is discussed at length, including the many challenges to the various formulations of it.

12 There is an important difference between what Descartes would have considered “idealism” and what analytic philosophers tend to think of as “idealism”. The differences are too many to discuss here, but any theory that goes by “idealism” generally states that the physical world is actually made up of ideas.

Descartes and analytic philosophy, during which time the physical sciences were by no means stagnant. So, this can't be the only reason why idealism was not seen as a viable option. It's not entirely clear why the option wasn't considered viable, but in the case of early analytic philosophy, the arguments against British Idealism by G.E. Moore and Bertrand Russell certainly had considerable influence on the broad rejection of idealism as a solution to the mind-body problem. The rejection of British Idealism was largely due to its relation to Hegel's idealism. As previously mentioned, Hegel was strongly associated with obscure methods of argumentation, so the appeal of the idealism for which he advocated was largely considered to be a result of this lack of clarity. Whatever the exact reasons, a strong rejection of idealism is an important constraint on the way the mind-body problem is discussed by Descartes and analytic philosophers. This allows for a striking structural similarity between the two time periods despite the large differences elsewhere in philosophy, science, and life in general. Just as was true for Descartes and his critics, there are considered to be two main solutions that seem capable of solving the mind-body problem, dualism and materialism<sup>13</sup>.

Despite this similarity in what were believed to be the viable positions, another major difference in the two time periods is the way the distinction between mind and matter was formulated. For Descartes and his contemporaries, physical bodies are defined by being extended in space. They are three dimensional objects. On the other hand, minds are non-extended and think. They are thinking things with no spatial dimensions. With the development of modern physics and psychology, this distinction was no longer helpful. While studying psychological states, it became apparent that there are important distinctions within what we call "mental" such that the Cartesian and causal arguments about the mind in general seem to be very plausible if one focuses on some kinds of mental states but not plausible if one focuses on other mental

---

13 I feel the need to emphasize that these are broad generalizations. There has always been a sub-current within analytic philosophy that defends idealism in some form or another. For example, see (Farris & Göcke, 2021; Foster, 1982, 1993; Tse, 2022).

states<sup>14</sup>. The crude distinction between what has a spatial location/extension and what doesn't seems to be too general. For instance, a pain in my toe seems to be located precisely where my toe is and thus seems to have the same rough dimensions as my toe as well. This would make it seem to fit the Cartesian category of a physical state, but pain sensations are paradigm examples of mental states. Moreover, the invention of the computer seemed to be the invention of a purely physical thing that could "think". Nevertheless, computers do not have any kind of sensations. There is nothing it is like to be a standard laptop computer. But there is something it is like for anything with a mind. There is a feeling to a state like pain or drinking a cup of coffee. These properties of the mind are referred to as "phenomenal properties". Additionally, modern physics posits many properties and objects that do not have the kind of spatial dimensions Descartes envisioned. String theory, for instance, proposes that there are fundamental physical objects with only two dimensions. Thus, while they are physical, they do not fit Descartes' criteria for being physical objects. The result of this refinement in our thinking about the mind and matter is that the much of the debate about the mind-body problem is not about objects or substances, as it was with Descartes and his contemporaries. Rather, the mind-body debate is narrowly focused on the relation between phenomenal properties – i.e. those properties of experience that make it something that it is like to have that experience – and physical properties – i.e. those properties used in fundamental physics or which can be derived from such theories, but with a special focus on the physical properties of the brain. In what follows, I will largely use terms like "mental" and "phenomenal" interchangeably.

---

14 Jaegwon Kim's *Physicalism, or Something Near Enough* (2005) makes a strong case for the causal argument being plausible for anything that has an explicit function for animals like us. He also makes a strong case for the Cartesian argument being plausible for things like sensations. While there are many things that one might disagree with on these points, Kim's arguments represent many of the trends in analytic philosophy and the cognitive sciences on these issues.

The last significant difference about the landscape of the twentieth century is the success of psychology and neuroscience as scientific endeavors. Of particular importance in the early twentieth century was the relative success of behaviorism in predicting and explaining states of the mind. Although it has been widely criticized as not delivering on its initial promises, behaviorism was a popular psychological theory in the middle of the 20<sup>th</sup> century that purported to show that mental states can be explained purely in terms of physical behaviors (or tendencies to produce physical behaviors). Obviously, it would seem that if this theory were true (or even something near enough to it), then arguments for dualism are spurious. The early successes of this theory combined with the rejection of idealism and a strong sense of the inevitable march of scientific reductions created an environment that makes it easy to understand the widespread acceptance of the scientific identification of mental and physical states – a position known as identity theory. This is precisely the starting point I'll take to begin my discussion of the analytic mind-body debate.

We'll start with U.T. Place and J.J.C. Smart in the 1950's. They claim that the mind and its mental states are identical to the brain and its physical states despite any appearances to the contrary. As I mentioned at the beginning of this section, the idea is that the mind and brain are analogous to previously proven scientific identities. U.T. Place (1956) and J.J.C. Smart (1959) use an analogy with lightning and static electric discharge<sup>15</sup>. They argue that lightning and static electric discharge have been proven to be identical, but this proof required in-depth scientific research due to the apparent difference between the phenomena. This kind of scientific identity, they claimed, is a contingent identity. Science *could have* proven that lightning is some other kind of thing, but it didn't. Lightning is static electric discharge. The general idea here is that identities in science

---

15 Some qualifications are in order. Obviously, static electricity generated by my sock on the carpet is not lightning. So, "atmospheric" should be added along with similar qualifiers. Nevertheless, the general point stands unchallenged that lightning is (a certain kind of) static electric discharge.

cannot be proven or disproven by the kind of *a priori* argument that Cartesian dualists rely on. Relating it back to the presentation of the argument in the previous section, this is a rejection of the third premise in Descartes argument. That premise states, “If it is possible that two things are distinct entities, then they are actually distinct, even if we have no idea how that separation might occur.” Smart and Place are pointing to many examples where this kind of claim seems to be false, and they are all scientific identities which appear to depend on how the world turns out. Moreover, this is exactly the kind of identity one would suspect the mind and brain to be, if they in fact turn out to be identical. Thus, they argue that what we can conceive about these sorts of scientific claims is not reliable evidence for what the facts are. They may show us some possibilities, but nevertheless the facts in the world point to the identity between the mind (and its mental states) and the brain (and its physical states). The details require psychological and neuroscientific research, but the issue is ready to be handed over to the sciences. Puzzle solved. Or so they would have us believe.

### **3. Kripke’s Revival of the Cartesian Conceivability Argument**

The debate between materialism and dualism was largely seen as won by materialists like Place and Smart. The exact natures of the identities were yet to be discovered, but it seemed pretty clear that the scientists were the ones to clean up the work. The fall of behaviorism was not seen as important to the debate as it was quickly replaced by the classical computational theory of mind – the theory states that the mind is like a Turing machine which is an abstract functional model of a computer, and the brain is like the hardware on which the Turing machine is implemented. This could be seen as a standard scientific development where one theory is supplanted by another, better theory and acted as proof that something about what Place and Smart had argued was correct. Cartesian-style conceivability arguments are not relevant to scientific endeavors. But the 1970’s saw a massive rise in the debate between dualists and

materialists owing to the revival of conceivability arguments within the new scientific context. This was accomplished (in part) by Kripke's arguments against the identity theory in his *Naming and Necessity* – first delivered in three lectures in 1970, the transcripts of which were published in 1972 as part of an edited volume, and finally published in 1980 as a standalone book with added Preface, Addenda, and footnotes.

Remember that the identity theory of Place and Smart insisted that mental states were identical with brain states just like other scientific phenomena like lightning and static electric discharge. And these identities, it was claimed, were contingent because they were proven through scientific research. They are true but could have been false if the world had turned out differently, and we could only know that the world turned out the way it did by studying it. Thus, the third premise of Descartes' argument is false and what is true about the relation between the mind and the brain does not seem to depend on what we can conceive about it. Kripke's response was twofold. First, he reminded everyone that identities must be necessarily true and demonstrated how this works for scientific identities as much as it does for other kinds of identities. Second, he argued that there is a key difference between proven scientific identities, like lightning and static electric discharge, and the proposed identity between phenomenal states and brain states.

Before we get to Kripke's response, let's pause and consider Place and Smart's reasoning for thinking that scientific identities are contingently true. Their thinking is that there are two kinds of identity statements, one that is true by the definition of the terms used and one that is not. The first kind of statement would be a statement like "A square is an equilateral rectangle" or "Bachelors are unmarried men" or to use Arnauld's example "A right triangle is a triangle for which the Pythagorean theorem applies." This kind of identity statement doesn't require knowing anything other than what their terms mean. To understand what a square is just is to know that it is a rectangle with sides of equal length. Thus, these statements are said to be analytically true. They can also be prov-



en to be true without ever investigating the nature of anything in the world. You don't need to measure all the squares to know that they all have sides of equal length. You just need to know the concept. Thus, they are said to be true *a priori*. And, of course, they couldn't have possibly turned out false and so they are necessarily true. By common agreement of the time, this trio of properties always go together.

On the other hand, Place (1956) pointed out that scientific identities do not fit this model. Let's take his example of lightning and static electric discharge. One can very well understand what it means for something to be lightning without knowing anything about electricity. In fact, at some point in our lives, we were all probably in this situation. Young children know what lightning is without knowing anything about electricity. And pre-scientific people certainly knew what lightning was without having any knowledge of electricity. Notice that this is different from the square. To know what a square is requires knowing that it is an equilateral rectangle, but this is known simply by knowing the proper meaning of the word "square". With lightning, on the other hand, you can (and probably did) know perfectly well what the phenomenon was without also knowing that it is static electric discharge. These kinds of statements are said to be synthetic. Additionally, it seems obvious that to prove that lightning is static electric discharge, you must go into the world and observe lightning and static electric discharges in such a way as to show that whenever one occurs so does the other. This, Place points out, requires two different kinds of observation which is what gives the appearance of the two phenomena being different when they are in fact identical. Such statements that can only be proven by empirical investigation are called true *a posteriori*. Finally, Place argues that because the world itself could be different from the way it actually is, any statement that is true *a posteriori* must be only contingently true. Again, by common agreement, the trio of synthetic, *a posteriori*, and contingent were thought to always go together.

When Place and Smart argue that the identity between consciousness and brain states are scientific identities, they mean that the identity fits in this latter category. Scientific truths require empirical investigation and so are true *a posteriori*. So, if this trio of statements always goes together, then scientific truths will also be contingently true. Thus, when Descartes argues that what could be true about the nature of the mind proves what is actually true (that the mind is actually distinct from the brain), he must have been mistaken. Place and Smart argue that the identity between mental states and brain states is a scientific truth and so what might have been the case is irrelevant. Their point is that the identity between mental states and brain states is contingent. So, of course it is possible that these states could not have been identical. Nevertheless, they argued that, like other contingent, scientific identities, what is possibly true has no bearing on what is actually the case. We had to go into the world to determine how the actual world is to determine what lightning is, after all. They argue that mental states and brain states are actually identical but could have been distinct. And so, Descartes' third premise – which states that two things which are possibly distinct are actually distinct – is false. In this way, Place and Smart reject Descartes' argument by denying the truth of the third premise.

A large part of Kripke's project in *Naming and Necessity* is to prove that these opposing trios of properties do not always go together. In particular, he attempts to prove that some statements that are necessary nevertheless require empirical investigation of the world. These are what he calls *a posteriori* necessities<sup>16</sup>. He begins by arguing that identity statements are *always* necessarily true if true at all. He cites Ruth Barcan Marcus as having proven this many years earlier (1963). The reasoning is drawn from examples with proper names of individuals. Let's take Clark Kent for example. This is the name of a certain (fictional) individual who works

---

16 He also argues for the converse, *a priori* contingencies. But these are not central to the mind-body debate, so I omit any discussion of them here.

as a journalist with his (also fictional) co-worker Lois Lane. The idea is that when Lois Lane learns that Clark Kent is Superman, she does not learn about a new person. She just learns that her co-worker who she knew as Clark Kent is also called Superman. Obviously, it's necessary that Superman be identical to himself. He could not have been someone else and still be himself. So, the identity statement Clark Kent = Superman is necessarily true. But it seems just as true that Lois Lane had to learn that Clark Kent is Superman by empirical investigation<sup>17</sup>. This fact can't simply be derived from reasoning about what "Clark Kent" means. The way that Marcus describes the situation is that the names act as "mere tags" or pointers to an individual. They don't have much in the way of meaning at all. But even if names do have some meaning associated with them, there's no meaning to "Clark Kent" such that one might be able to derive that Clark Kent is Superman. Kripke concludes that "it seems that we do not necessarily know *a priori* that an identity statement between names is true. It doesn't follow from this that the statement so expressed is a contingent one if true" (1980, p. 101). Clark Kent = Superman is an example of a statement that Kripke calls an *a posteriori* necessity. The example shows that at least *some* statements can be necessarily true while at the same time require *a posteriori* investigation to prove that they are true.

Thus, the assumption that Place and Smart make – that analytic, *a priori*, and necessary always go together while synthetic, *a posteriori*, and contingent also always go together – is proven false. This raises a puzzle about what to do with the scientific identities they say are analogous with the case of mental states and brain states. Place and Smart say that scientific identities are genuine identities. At the same time, they correctly recognize that they require empirical investigation. From this, they conclude that they are contingent identities. But what Kripke has shown is that this

---

17 Kripke gives Quine the credit for making this kind of point in his dispute with Marcus. The point is made in Quine's discussion and comments in (Marcus, 1963).

last step is a mistake. Simply from knowing that a statement is *a posteriori*, we cannot know that it is contingent. It very well may be an *a posteriori* necessity, like the statement ‘Clark Kent = Superman’. So, we have two options. If scientific identities are genuine identity statements like those that use proper names, then they must be necessarily true. The other option is to insist that they are contingently true. But then it must be false that they are identities at all. Kripke argues that scientific identities are just like identities with proper names. That is, they are genuine identities that must be discovered *a posteriori*. This matches with the way we use the terms after all. Since all such identities<sup>18</sup> are necessarily true, the statement that lightning is static electric discharge is either a genuine identity and thus necessary or else it is really contingent and thus not an identity at all. Thus, scientific identities of the sort that Place and Smart have in mind are examples of *a posteriori* necessities. Some other examples of these are water and H<sub>2</sub>O, heat and mean molecular kinetic energy, and clouds and collections of water droplets.

It’s important to note how this defends Descartes’ third premise from Place and Smart. The third premise states, “If it is possible that two things are distinct entities, then they are actually distinct, even if we have no idea how that separation might occur.” This is simply a different way of stating Kripke’s claim that identity statements must be necessary. A few logically equivalent transformations of the premise will make this clear. First, take the contrapositive of the conditional, so that the statement reads, “If two things are not actually distinct, then they cannot possibly be distinct”. Then, recognize that “not possibly distinct” is the same as

---

18 There is a minor complication here concerning what Kripke calls “non-rigid designators” and identity statements made using them. Non-rigid designators are typically definite descriptions like “the inventor of the bifocals”. Statements that make use of these are not necessarily true. For example, the statement that Ben Franklin is the inventor of the bifocals is a true identity statement, but not necessarily true. So, when I say all “such” statements, I mean to refer to those that exclusively use rigid designators like names and natural kind terms. Going into detail on rigid vs non-rigid designation would take us too far afield though.

“necessarily not distinct” and further that “not distinct” is the same as “identical”, we can rewrite the premise again as “If two things are actually identical, then they are necessarily identical” which is exactly what Kripke has taken himself to prove. The idea is that if Clark Kent could have been separated from Superman, then they would, as a matter of fact, not be the same person. But since they are identical, they must necessarily be inseparable. In fact, there is no “they” to be found. There are just two names for the same person! So, if the identity theorist wants to argue that mental and physical states are in fact inseparable, they have to show that they are really two names for the same thing. But then merely the possibility of them coming apart would be enough to prove that they are in fact distinct.

Of course, all that we’ve said thus far allows for the identity theorist to shift their response. They might agree that because the mental and physical are identical, this statement must be true necessarily. Nevertheless, they might argue that the analogy with scientific identities is still a good one. We are just in the position that our ancestors were in with respect to other scientific identities, like lightning and static electric discharge, before we had the empirical evidence to prove that they are true. That is, mind/brain identity could be an *a posteriori* necessarily true statement. Thus, preserving the spirit of Place and Smart’s position.

In essence, this shifts the response from being an objection to premise 3 to being a response to premise 2 in Descartes’ argument. This premise states that if we can clearly and distinctly conceive of two things existing separately, then it is possible they are distinct entities. The idea goes that at some point in the past, we clearly and distinctly thought that lightning could be different from static electric discharge, but it turned out to be false that they could have existed separately. They are identical. This seems to be true for any other *a posteriori* necessity. Since it’s *a posteriori*, we might not have done enough of the empirical investigation to reveal to us the truth of the matter. After all, it seems like Lois Lane can clearly and distinctly conceive of Clark Kent not being Super-

man, but further empirical facts about them will show that they are the same. As you'll recall, Arnauld's objection to Descartes' argument was that someone might clearly and distinctly conceive of a right triangle that does not conform to the Pythagorean theorem. Nevertheless, it does not prove that any such right triangle exists. Similarly, just because Lois Lane can clearly and distinctly conceive of Clark Kent not being Superman, this does not prove that they are actually different people. As you can see, this objection is a variation of Arnauld's objection against Descartes.

Kripke's response is to reveal a disanalogy between something like lightning and something like pain. He argues that in the case of lightning there was an *illusion of contingency* in the statement that lightning is static electric discharge. It appears to us that the statement could have turned out false, but since the statement is a genuine identity, it is necessarily true. Thus, the contingency is illusory. But the case of pain is different. There is no illusion. Pain is really only contingently related to its associated brain states (which he calls c-fibers firing) and since, Kripke argues, identities are necessary, pain cannot be identical to its associated brain states. Thus, dualism is true. It is this step in Kripke's argument for dualism that is most controversial, so it's worth spending some time looking at how he tries to establish this disanalogy<sup>19</sup>.

Let's begin with Kripke's description of a genuine scientific identity like lightning and static electric discharge. When we begin investigating the nature of lightning, we first identify it by its appearance and other attributes that tend to go along with it. Lightning is that phenomenon which is a bright flash typically followed by a rumbling sound we call thunder. It sometimes accompanies intense rain, dark clouds, and other more intense weather phenomena like tornadoes and hurricanes, but sometimes there is

---

19 Kripke makes extensive use of his quasi-technical apparatus of rigid designation to make this point. I omit making use of it here for the sake of clarity. However, this comes at the potential cost to precision in an argument where precision is crucial. The details of Kripke's argument are found on pages 144-155 in his (1980). His discussion of the lightning example in particular is quite brief but appears on page 132.

lightning without these other things. If one is unfortunate enough to be hit by lightning, it is extremely painful and can be deadly. But the main way we identify lightning is by the way it looks to us. It's the bright flash that is often followed by thunder. In other words, we identify it by a certain sensation (or set of sensations) *S* that it produces in us. This sensation can vary, of course, depending on the kind of lightning we are witnessing, but we know a flash of lightning when we see it. Through our scientific investigations, we discovered that this flash we call lightning is static electric discharge. But it certainly appears to be the case that the flash could have been something else. For instance, our scientific investigations could have revealed the flash to be a spear thrown by Zeus or a discharge of some kind of energy other than electricity. So, it appears as if lightning could have been something other than what it is. This is the illusion of contingency.

Kripke argues that what *really* could have been different is the thing that causes in us the sensation *S*. That is, someone might have been in the same position we were with respect to the fact that they had sensation *S*, a bright flash of light that is often followed by thunder, often accompanies rain, etc. But that sensation could have been caused by something other than static electricity. But since lightning is static electricity, in this situation, what is causing *S* is also *not lightning*, despite all qualitative appearances to the contrary. The opposite, of course, is also true. It might have turned out that static electric discharge did not produce in us sensation *S*. Perhaps because it didn't light up the sky or we had different optic nerves or something. In this situation, there would really be lightning, but it does not cause sensation *S* in us. So, what is contingent is not lightning being static electric discharge but simply the fact that lightning causes in us the sensation *S*.

The situation is different in the case of phenomenal states. Kripke focuses on pain, so let's use his example. Before any neuroscience is done, we identify pain in a similar way to lightning. We identify it by a certain sensation that is painful. It often comes along with winces, saying "ouch, that hurts", trying to avoid the

pain, etc. But primarily, we identify pain by a sensation that is painful. When neuroscientists begin studying pain and come up with a proposed identity for pain, for example the stimulation of c-fibers in the nervous system, one might think that we ought to say a similar thing about the sensation of pain that we said for the sensation S produced by lightning. However, Kripke argues that the difference is that pain *just is* the sensation that is painful. And if c-fibers being stimulated *produce* this sensation, then the pain is not identical to c-fiber stimulation. An effect cannot be identical to its cause. When we say that lightning produces a sensation S in us, there is a distinction between the way we identify lightning (by using sensation S) and what lightning is (static electric discharge). But for phenomenal states like pain, there is no difference between the sensation and the way we identify the phenomenon. Again, pain just is the sensation of pain. So, when we think that pain (the phenomenal state) might exist without our c-fibers being stimulated (the physical state), we are thinking of a genuine possibility where there is a mental state without the corresponding physical state. The idea is that while pain actually is caused by c-fibers being stimulated, it could be produced by the stimulation of something else in the nervous system or by something else entirely. In the same way, the sensation S could have been produced by something other than lightning. This contingent relation in the vicinity is the reason why there is an illusion of contingency for the identity statement that lightning is static electric discharge. Lightning and sensation S are clearly distinct, and so we should conclude that pain and c-fiber stimulation are distinct.

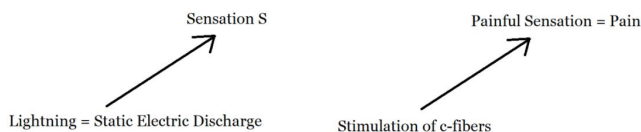


Diagram of Kripke's explanation of the difference between a scientific identity where there is an illusion of contingency (left) and the purported identity between c-fiber stimulation and pain (right). Note that the arrow represents causation (or any other contingent relation) whereas the identity denotes a necessary relation.



Basically, what Kripke is arguing is that in the cases of genuine scientific identities, when we think that they could have turned out otherwise, we are not conceiving of the situation clearly and distinctly. We only *think* that we are doing so, but we need to pay closer attention to what it is that we are actually saying when we say that something like lightning could have been different from static electric discharge<sup>20</sup>. On the other hand, when we do the exact same process for phenomenal states like pain, it is revealed that we *are* clearly and distinctly conceiving of pain being different from c-fibers firing. This is remarkably similar to Descartes' response to Arnauld.

Arnauld's objection to Descartes rejected the premise that from a merely conceivable situation, we could derive something genuinely possible. His example was a right triangle that one might conceive as distinct from a triangle which conforms to the Pythagorean theorem. This is an *a priori* case, but it is one where we seem to be conceiving of a scenario that is not possible. On reflection, Kripke would simply say that we are not clearly and distinctly conceiving of the relevant scenario. The illusion of contingency arises because of a poor description or from simply not understanding the terms one is using to describe the scenario. The case of lightning and static electric discharge is an *a posteriori* corollary to Arnauld's response. And Kripke's reply is the same, to specify the details of the scenario more clearly to reveal that there is merely an illusion of contingency. But once those same details are spelled out for the case of pain and c-fibers firing, the contingency is revealed to be genuine.

One popular response has been to develop versions of materialism that argue against any form of identity theory. These positions argue that the relation between mental states and physical states is still one where the mental states are necessitated by the physical states, but that they are not identical. The exact relation

---

20 He also gives us a method to make sure that we have conceived of a situation clearly and distinctly, but going into detail on what this method entails is too far removed from the discussion here.

differs between the theories, but they are all relations of ontological dependence<sup>21</sup>. One thing ontologically depends on another when the existence of the latter necessitates the existence of the former. The scientific analogy used is between properties like liquidity and the microphysical properties of H<sub>2</sub>O. Liquidity is a physical property of water, and it is not identical to the microphysical properties that give rise to it. However, liquidity arises necessarily whenever you have H<sub>2</sub>O in various conditions (like certain temperature and pressure ranges). Moreover, many things can be liquid, not just H<sub>2</sub>O.

Similarly, it is argued that phenomenal states arise necessarily whenever certain configurations of physical properties are instantiated together, like those in our brains when we experience a pain. Thus, while pain is not identical to any brain state (like c-fibers firing), it is ontologically dependent upon it. It is claimed that this would also help explain how other things without our brains can experience pain (like an octopus, a dog, or if we can make one complex enough, a computer). Such a position seems to offer a way around Kripke's argument as he explicitly targets identity theory and nothing else. Since this kind of materialism (typically called functionalism) does not posit any identity relation between phenomenal states and brain states (in fact, proponents often offer their own arguments *against* identity theory), it might be thought that it can avoid Kripke's argument altogether.

However, this is a spurious response to Kripke. Although he only explicitly addresses identity theory, the way he argues against the view involves showing that the relation between states like being in pain and states like having stimulated c-fibers is only a contingent relation. This has always been the dualist position. Descartes emphasizes the "substantial unity" of the mind and body, thus recognizing that they are intimately tied to one another. But by arguing that this is merely one possible way that things could

---

21 Some of the more notable examples of ontological dependence relations are supervenience and realization.

have turned out, he attempts to prove that the mind and body are fundamentally distinct where there is no necessary connection between phenomenal and physical properties. They are different ontological kinds of thing. Kripke's argument also applies to non-identity forms of materialism because they posit that phenomenal properties ontologically depend on physical properties and ontological dependence holds necessarily. Identity is merely one relation among many that is a necessitation relation; any kind of ontological dependence relation that a materialist might propose is a necessary relation between brain states and the phenomenal states that they give rise to. Since Kripke argues against identity theory by attempting to prove that the *necessity* does not obtain, the argument is just as problematic for non-identity materialists as it is for identity theorists. Thus, we can see that Kripke's argument for dualism is just the same as Descartes' argument for dualism. We can update it for the more recent analytic context:

- 1) Phenomenal properties can be correctly described as existing independently from physical properties
- 2) If two things can be correctly described as existing independently, then it is possible that they are distinct.
- 3) If two things are possibly distinct (and therefore not necessarily connected), then they are actually distinct entities.

Thus, phenomenal properties are distinct from physical properties.

Since Kripke's defense of premise three, much of the debate over the mind-body problem concerns the truth of premise two. Denying the first premise would be to say that pain (for instance) must be described in terms of physical properties for it to be a proper description of pain. But painfulness is not a physical property, so denying the first premise would be to deny the fact that we are describing our pains correctly when we say that they are necessarily painful. This is related to Descartes' first premise because clearly and distinctly conceiving of the mind involves being able to describe it in the right terms. If those terms are entirely physical (as the materialist would say they are), then it would not involve painfulness. So, pain wouldn't necessarily be painful, it

might merely be c-fibers firing. It's hard to see how one could sensibly endorse the position that pain might not be painful, but there is a small minority of analytic philosophers who have tried. On the other hand, it has seemed reasonable to more philosophers to attack premise two, precisely as Arnauld did in the seventeenth century<sup>22</sup>. While we can recognize the general form of the debate is the same, the changes in the philosophical and scientific landscape have allowed philosophers to delve deeper into the issue. As such, the arguments over premise two often involve disputes over the very nature of necessity and possibility and our epistemic access to such facts, disputes over how our phenomenal concepts refer to phenomenal properties, and whether humans can ever answer the mind-body problem<sup>23</sup>.

#### **4. The Causal Argument in Analytic Philosophy**

If this were the only argument in the analytic repertoire, it would be a pretty one-sided affair. Kripke has given us an argument for his dualistic conclusion, but materialists have only responded by trying to show that the argument is inadequate. However, this isn't the only problem that materialists have with dualism. They, like Princess Elisabeth, raise some serious questions for any dualistic position. In particular, it is mysterious how phenomenal properties could play the kinds of causal roles that they seem to play. The same basic datum that pain causes me to shrink away from a flame and say "ow" is even harder for a dualist to account for given the developments in modern cognitive science.

---

22 See (Churchland, 1996; Lewis, 1988) for two philosophers who continue to dispute the truth of premise three. For arguments against premise one, see (Dennett, 1988; Frankish, 2016; Humphrey, 2006; Rey, 1982, 1998). Arguments against premise two include (Block & Stalnaker, 1999; Papineau, 1993; Perry, 2001; Polger, 2004).

23 For the first, see discussions of strong necessities, like (Chalmers, 2014; Cleeveley, 2024; Goff & Papineau, 2014). For the second, see discussions of the phenomenal concepts strategy in (Levin, 2007; Loar, 1990, 1997; McLaughlin, 2001). For the third, see (McGinn, 1989; Nagel, 1986).

Work by figures like Jaegwon Kim (1998, 2005) expands upon Princess Elisabeth's ideas and develops the argument in much greater detail. The basic idea that all physical states have physical causes is still the same, but the idea now has much greater empirical support from the development of the sciences themselves. As we saw above, Princess Elisabeth's argument primarily argues from *a priori* principles about how causation works to the conclusion that the mind must be physical. But one might question whether these principles are true. Since Hume's arguments concerning causation, the idea that there is some *a priori* notion about what causation looks like has been treated as dubious. And just as we saw with Place, Smart, and Kripke, the analytical approach relies heavily on empirical evidence and analogies with science. So, while the argument remains the same, the *a priori* reasons are relied upon less heavily. Instead, the idea that all causes are physical is defended by appeal to physics itself. Every time we are curious about the cause of some physical event, physicists have successfully found the cause to also be physical. They've never needed to rely on mental causes to explain physical events. This is said to give strong inductive support to the premise that all causes of physical events are physical causes<sup>24</sup>. With different support for the first premise, the shape of the argument remains the same.

- 1) All physical events have physical causes.
- 2) Phenomenal properties are the cause of some physical events.

Therefore, phenomenal properties are physical.

---

24 There's an important question here about just what counts as physical in the contemporary discussion. As mentioned previously, contemporary physicists and philosophers have abandoned the Cartesian definition of physical objects as extended in space. But no agreed upon understanding of the physical has adequately replaced it. For in depth discussion and argumentation on this point, see (Montero, 1999, 2001; Stoljar, 2001). For the classic formulation of this problem, known as Hempel's dilemma, see (Hempel, 1969). Without some answer to this question, it's not clear what materialism (or physicalism as it's now often called) is claiming to be true, and thus it's not easy to see whether the causal argument supports the thesis or not.

The two natural responses to this argument are to reject either premise. Rejecting the first premise is similar to Descartes' response and requires positing some irreducible causal role for phenomenal states. Such a response is widely held to not respect the physical sciences, though, given that the support for the premise relies on empirical evidence from the success of those sciences. The second option is to reject the second premise. This yields the very counterintuitive view that while it *appears to us* that pain is what causes us to flinch, pain actually plays no causal role in our actions at all. This view is called epiphenomenalism. Despite its counterintuitive nature, it has the benefit of more clearly respecting both scientific explanations and Cartesian conceivability arguments. Nevertheless, the idea that it isn't my pain that causes me to flinch seems to go against the direct evidence that we all have of the world. This reason alone is enough for many philosophers to reject the view<sup>25</sup>.

## 5. Conclusion: Dissatisfaction with Major Views and Developing Alternatives

The debate between materialists and dualists in analytic philosophy is at something of an impasse. On the one hand, the dualist argument against materialism seems quite strong. Of course, philosophers have been debating Kripke's argument in the decades since he wrote it, but none of the responses have seemed to get at the heart of the argument. Of particular note is work by David Chalmers (Chalmers, 1996, 2010) which develops further details on the conceivability arguments of Descartes and Kripke. On the other hand, the materialist argument against dualism *also* seems quite strong.

---

25 See (Chalmers, 2002) for an interesting discussion on whether physics really doesn't ever propose mental causes for physical events, specifically focusing on collapse theories of quantum mechanics. Chalmers admits that this flies in the face of the received view, but it is an interpretation that is available and would allow someone to both respect the authority of the sciences as well as reject the first premise of the argument. Also, see (Jackson, 1982; Kim, 2005) for interesting defenses of epiphenomenalism. For arguments against epiphenomenalism, see (Robinson, 2023).

Again, philosophers have various responses, but it does seem like materialism takes the success of the physical sciences more seriously than dualism. The debate rages on, but many analytic philosophers have begun reaching for a way to break the impasse. In the past twenty years or so, two views have become quite popular as attempts to solve the mind-body problem while satisfying both the Cartesian arguments and a respect for the success of the physical sciences. So, in closing, I'd like to briefly mention these views and how they attempt to break out of the argumentative impasse that I've described above.

The first view is often called Russellian monism, named after Bertrand Russell, and draws inspiration from a view that he proposed in his (1921, 1927) where he called the view neutral monism (so, sometimes these names are used interchangeably). The view begins by arguing for a certain view of how science works. The idea goes that science investigates the world by noting its structural features and how these features interact with one another. In other words, science studies the causal, structural, and dispositional nature of the world. But what it does *not* do is investigate the *intrinsic* nature of anything in the world. Science can tell us what things *do* but can't tell us what things *are*.

The view further argues that any causal or dispositional properties must be based on the intrinsic properties of the object that has them. There must be some reason why certain things act one way while others act another way, after all. And the idea is that this reason is found in what these things are. For example, the fact that a glass breaks easily when struck or dropped must be explained by some intrinsic properties of the glass itself. Science can tell us of the glass's chemical microstructure, but this structure itself is really just another, more fine-grained dispositional property. If carried on far enough, scientific investigation will reveal fundamental particles, but even these are only defined by the way that they'll behave. An electron is defined (in part) by having negative charge. But having negative charge just means being repelled by similarly charged objects and attracted to objects with an opposite charge.

Russellian monism argues that there must be something intrinsic to an electron such that it reacts this way. We don't know what these are and can't investigate them scientifically. In fact, Russell argues that we have no access to the intrinsic nature of anything in the world, with one notable exception. Phenomenal features of our experience are the intrinsic natures of those experiences. In essence, we have access to the intrinsic natures of our brains whenever we have a phenomenal experience. Thus, phenomenal properties respect the scientific worldview while at the same time respecting the Cartesian arguments that these properties are distinct. Russell himself remained neutral as to what the intrinsic properties of non-minded objects (like electrons) were. But the speculative idea is that phenomenal properties are just those intrinsic properties that physical objects like our brains have when experiencing something. Other physical objects also have such intrinsic properties even if they are very different to phenomenal properties. Since science doesn't investigate the intrinsic nature of the world, this is seen as a supplement to science rather than as an attempt to replace it<sup>26</sup>.

The second view that attempts to break the materialist/dualist impasse is sometimes presented as a variant of Russellian monism. The view is called panpsychism. The etymology of the name would make it seem like the view says that everything (pan-) has phenomenal properties (-psychism). And while there are some historical variants that seem to accept this, the view that has become popular in recent analytic philosophy is one where all fundamental objects have phenomenal properties in addition to their fundamental physical properties<sup>27</sup>. It's easy to see how this can fit well with the Russellian view mentioned above. While Russell's version of his

---

26 For several examples of non-panpsychist versions of Russellian monism, see (Chalmers, 2002; Lockwood, 1981, 1989; Stoljar, 2014). For neutral monist views that differ in important ways from Russellian monism, see (Banks, 2014; Griffin, 1998; Schon, 2023; Stoljar, 2015).

27 For an overview of panpsychism, including a discussion of historical examples in Western philosophy, see (Goff et al., 2022). Contemporary views are found in (Chalmers, 2013; Goff, 2017; Lewtas, 2013; Roelofs, 2021; Seager, 2019).



view remains neutral on the intrinsic properties of objects like the electron, a Russellian panpsychism claims that the intrinsic properties of electrons are phenomenal. Panpsychists are quick to point out that the phenomenal properties of electrons are *much* different from ours. This has to be the case since an electron's experience of the world would necessarily be much different from ours if it has such experiences. The view is often met with a deep suspicion, but adherents argue that the view isn't incoherent, respects scientific progress (especially in its Russellian form), and has no problems with Cartesian arguments, so it should at least be explored. Besides such suspicions, the view does have genuine philosophical problems, though.

The main problem is how any phenomenal properties of electrons (and other fundamental objects) could explain *human* consciousness, which is really what we want explained. A similar problem besets any form of Russellian monism too as it's not clear how intrinsic properties of fundamental objects could provide an explanation for phenomenal properties of experience. However, given the other alternatives, these two views seem to offer a way forward on one of the most difficult philosophical puzzles in human history. And even if they are both misguided, they offer a model for how analytic philosophers think about the mind-body problem. They pay close attention to, and respect the success of, scientific progress while simultaneously trying to navigate the puzzles raised by Cartesian conceivability arguments.

## References

- Banks, E. C. (2014). *The Realistic Empiricism of Mach, James, and Russell: Neutral Monism Reconceived*. Cambridge University Press.}
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, 108(1), 1–46.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

- Chalmers, D. J. (2002). Consciousness and Its Place in Nature. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 247–272). Oxford University Press.
- Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford University Press.
- Chalmers, D. J. (2013). Panpsychism and Panprotopsychism. *Amherst Lecture in Philosophy*, 8.
- Chalmers, D. J. (2014). Strong necessities and the mind–body problem: A reply. *Philosophical Studies*, 167(3), 785–800.
- Churchland, P. M. (1996). The Rediscovery of Light. *Journal of Philosophy*, 93, 211–228.
- Cleeveley, H. (2024). The deep incoherence of strong necessities. *Inquiry: An Interdisciplinary Journal of Philosophy*.
- Creath, R. (2023). Logical Empiricism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2023/entries/logical-empiricism/>
- Dennett, D. C. (1988). Quining qualia. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science* (pp. 42–77). Oxford University Press.
- Descartes, R. (2008). *Meditations on First Philosophy with Selections from the Objections and Replies* (Moriarty, Michael, Trans.). Oxford University Press.
- Descartes, R. (2017). *Objections to the Meditations and Descartes' Replies* (J. Bennett, Ed. & Trans.). [www.earlymoderntexts.com](http://www.earlymoderntexts.com)
- Farris, J. R., & Göcke, B. P. (2021). *The Routledge Handbook of Idealism and Immaterialism*. Routledge.
- Foster, J. (1982). *The Case for Idealism*. Routledge.
- Foster, J. (1993). The Succinct Case for Idealism. In H. Robinson (Ed.), *Objections to Physicalism* (pp. 293–313). Clarendon Press.
- Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, 23, 11–39.
- Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford University Press.
- Goff, P., & Papineau, D. (2014). What's wrong with strong necessities. *Philosophical Studies*, 167(3), 749–762.
- Goff, P., Seager, W., & Allen-Hermanson, S. (2022). Panpsychism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Summer 2022 Edition). <https://plato.stanford.edu/archives/sum2022/entries/panpsychism/>

- Griffin, D. R. (1998). *Unsnarling the World Knot: Consciousness, Freedom, and the Mind-Body Problem*. University of California Press.
- Hempel, C. (1969). Reduction: Ontological and Linguistic Factes. In S. Morgenbesser (Ed.), *Essays in Honor of Ernest Nagel* (pp. 179–199). St Martin's Press.
- Humphrey, N. (2006). *Seeing Red: A Study in Consciousness*. Cambridge, MA: Harvard University Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(April), 127–136.
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind/Body Problem and Mental Causation*. MIT Press.
- Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton University Press.
- Kind, A., & Stoljar, D. (2023). *What Is Consciousness? A Debate*. Routledge.
- Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.
- Levin, J. (2007). Nagel vs. Nagel on the nature of phenomenal concepts. *Ratio*, 20(3), 293–307.
- Lewis, D. (1988). What Experience Teaches. *Proceedings of the Russellian Society, Sydney, AUS: University of Sydney*.
- Lewtas, P. (2013). What It Is Like to Be a Quark. *Journal of Consciousness Studies*, 20(9–10), 9–10.
- Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108.
- Loar, B. (1997). Phenomenal states II. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press.
- Lockwood, M. (1981). What was Russell's neutral monism? *Midwest Studies in Philosophy*, 6(1), 143–158.
- Lockwood, M. (1989). *Mind, Brain and the Quantum: The Compound "I."* Oxford University Press.
- Marcus, R. B. (1963). Modalities and Intensional Languages. *Boston Studies in the Philosophy of Science*, I, 77–116.
- Marek, J. (2024). Alexius Meinong. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Spring 2024 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2024/entries/meinong/>

- McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98, 349–366.
- McLaughlin, B. P. (2001). In defense of new wave materialism: A response to Horgan and Tienson. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents*. Cambridge University Press.
- Montero, B. (1999). The Body Problem. *Nous*, 33(2), 183–200.
- Montero, B. (2001). Post-Physicalism. *Journal of Consciousness Studies*, 8(2), 61–80.
- Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.
- Papineau, D. (1993). *Philosophical Naturalism*. Blackwell.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. MIT Press.
- Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 44–50. <https://doi.org/10.1111/j.2044-8295.1956.tb00560.x>
- Polger, T. W. (2004). *Natural Minds*. Bradford.
- Princess Elisabeth of Bohemia, & Descartes, René. (2007). *The Correspondence Between Princess Elisabeth of Bohemia and René Descartes* (L. Shapiro, Ed. & Trans.). University of Chicago Press.
- Rey, G. (1982). A reason for doubting the existence of consciousness. In R. J. Davidson, G. E. Schwartz, & D. H. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–39). Plenum.
- Rey, G. (1998). A narrow representationalist account of qualitative experience. In J. E. Tomberlin (Ed.), *Philosophical Perspectives* (Vols. 12, Language, Mind, and Ontology, pp. 435–457). Boston, MA: Blackwell.
- Robinson, W. (2023). Epiphenomenalism. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). <https://plato.stanford.edu/archives/sum2023/entries/epiphenomenalism/>
- Roelofs, L. (2021). Is Panpsychism at Odds with Science? *Journal of Consciousness Studies*, 28(9–10), 116–128.
- Rorty, R. (2016). *Philosophy as Poetry*. University of Virginia Press.
- Russell, B. (1921). *The Analysis of Mind*. Allen & Unwin.
- Russell, B. (1927). *The Analysis of Matter*. Kegan Paul.
- Schon, M. (2023). *Neutral Monism Beyond Russell* [Doctoral Dissertation]. University of Wisconsin-Madison.
- Seager, W. (2019). *Routledge Handbook of Panpsychism*. Routledge.

- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68(April), 141–156.
- Stoljar, D. (2001). Two Conceptions of the Physical. *Philosophical and Phenomenological Research*, 62(2), 253–281.
- Stoljar, D. (2014). Four Kinds of Russellian Monism. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind*. Routledge.
- Stoljar, D. (2015). Russellian Monism or Nagelian Monism? In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*.
- Tse, C. Y. P. (2022). Metaphysical idealism revisited. *Philosophy Compass*, 17, 1–21.
- Uebel, T. (2024). Vienna Circle. In E. N. Zalta & U. Nodelman (Eds.), *Stanford Encyclopedia of Philosophy* (Summer 2024 Edition). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2024/entries/vienna-circle/>



## CAPÍTULO VI

### EL CARÁCTER SUBJETIVO DE LA EXPERIENCIA: AUTOCONSCIENCIA E INFORMACIÓN DE PRIMERA PERSONA

*Miguel Ángel Sebastián*<sup>1</sup>

#### 1. Consciencia y Subjetividad

Este capítulo busca presentar, problematizar y ayudarnos a entender el carácter subjetivo que tiene nuestra experiencia consciente. La noción de consciencia que nos interesa aquí es la de *consciencia fenoménica*; el tipo de consciencia que tiene una criatura cuando *se siente de cierta manera* —“hay algo cómo lo que es” (*there is something it is like*) ser una criatura (Farrell 1950; Nagel 1974); es decir, el tipo de consciencia presente en una criatura que tiene *vivencias o*

---

<sup>1</sup> Miguel Ángel Sebastián es doctor en filosofía por la Universidad de Barcelona donde realizó su tesis dentro del grupo Logos. Se incorporó a la UNAM en el 2012 como postdoc y es investigador del Instituto de Investigaciones Filosóficas desde el 2014. Sus principales áreas de interés son la filosofía de la mente y las ciencias cognitivas, especialmente en temas relacionados con la consciencia, la subjetividad, la atención y teorías naturalistas el contenido mental. El Dr. Sebastián también tiene intereses en metafísica y filosofía del lenguaje.

*experiencias conscientes*. Aunque resulta complicado ofrecer una definición o caracterización formal de estos estados o episodios, la experiencia consciente que tenemos del mundo y de nosotros mismos es el aspecto más familiar de nuestra mente. Se siente de cierta manera para una tener una experiencia consciente como pueda ser oler el café en la mañana, saborear un mole, tener un orgasmo o sentir un dolor punzante en la cabeza.

Las expresiones ‘carácter fenoménico’ o ‘fenomenología’ se usa para referirse a la manera en que se siente para una criatura tener una experiencia consciente. De esta manera, asumiendo que las experiencias conscientes son un tipo de estado mental (estados fenoménicamente conscientes), las teorías de la consciencia buscan entender qué hace falta para que un estado tenga fenomenología y así comprender qué es necesario para que una criatura sea fenoménicamente consciente.

Para abordar este desafío, es común hacer una distinción conceptual entre dos aspectos o componentes de la fenomenología y las cuestiones que los acompañan: el aspecto cualitativo y el aspecto subjetivo (Kriegel, 2009; Levine, 2001) Por un lado, debemos entender qué hace falta para que estar en ese estado se sienta de la manera particular en que se siente; esto es, entender en virtud de qué se siente distinto, por ejemplo, tener un orgasmo que oler café. Una teoría del aspecto cualitativo se centra entonces en entender las cualidades que caracterizan a las experiencias conscientes, como la rojez de la experiencia que tenemos al ver una rosa o la particular sensación de dolor de una jaqueca.

Por otro lado, debemos entender, sin apelar a la propia fenomenología, qué distingue un estado consciente de uno que no lo es. Llamemos a este problema el problema de la delimitación (Sebastián, 2022a). Uno podría sugerir que tal problema se reduce al problema de las cualidades, y de esta manera sugerir que basta con que un estado instancie una propiedad cualitativa para que se sienta de cierta manera para el sujeto estar en ese estado. El problema de esta aproximación es que deja abiertas dos cuestiones relevantes. Primero, parece razonable preguntarse qué es lo que distingue a



las propiedades cualitativas de las que no lo son —de manera que pudiéramos tener una respuesta al problema de la delimitación—, donde la respuesta no puede ser que son tales que se siente de cierta manera estar en un estado que las instancia, pues la explicación sería viciosamente circular. Segundo, falta por explicar cómo ese tipo particular de propiedades (cualitativas) da lugar a que surja la perspectiva de primera persona o punto de vista que caracteriza a la consciencia fenoménica. Es por esto que diversos autores han vinculado el problema de la delimitación al problema de la subjetividad y de entender qué hace falta para que aparezca la perspectiva o punto de vista que caracteriza nuestra experiencia. De esta manera, buscan desglosar la idea de que un estado consciente no es un estado que meramente ocurre *en mi* interior —como pueda ocurrir en el caso de la digestión— sino que, de alguna manera que trataremos de iluminar, se trata de un estado que es *para mí*, que se me presenta o que yo vivo.

La discusión filosófica reciente en torno al problema de la consciencia se ha centrado principalmente en el aspecto cualitativo, donde no es extraño postular propiedades intrínsecas, privadas e inefables (*qualia*). Este capítulo busca contribuir a ofrecer nuevos horizontes de investigación, centrándose en el entendimiento del aspecto subjetivo inherente a la experiencia consciente, resaltando su relevancia para entender la consciencia fenoménica, el papel que esta juega en la cognición y el problema que el entendimiento científico de la consciencia parece presentar.

Con ese objetivo, el capítulo está organizado de la siguiente manera: La siguiente sección presenta la íntima relación entre consciencia (fenoménica) e información. La sección 3 introduce la noción de autoconsciencia y su relación con la consciencia fenoménica, diseccionando las principales maneras en que ha sido entendida en la literatura; a saber, como consciencia de la experiencia misma y como consciencia de uno mismo. La sección 4 ilustra, con base en la discusión reciente sobre el pensamiento en primera persona, la manera en que la autoconsciencia debe ser entendida. Para ello, presenta diversas maneras en que la información puede ser

considerada subjetiva y argumenta en favor de una en particular (información perspectiva). Además, discute cómo esta manera de entender la información que la experiencia nos ofrece ayuda a responder a ciertas objeciones. La sección 5 discute la relación entre el aspecto subjetivo entendido en términos de información perspectiva y el problema de la consciencia ilustrado por el argumento del conocimiento. Finalmente, la conclusión discute las implicaciones para el análisis de nuestras capacidades cognitivas y la naturaleza de la consciencia.

## 2. Consciencia e Información

Un artículo en el periódico o una publicación en Instagram transmiten información, reduciendo la incertidumbre que podamos tener sobre el estado de la economía o el lugar donde alguien estuvo disfrutando de sus vacaciones. En nuestro caso, las experiencias conscientes son una fuente primaria de información. Por ejemplo, si no sabemos si hay fruta en la mesa, tal incertidumbre puede disiparse gracias a la información proporcionada por la experiencia consciente que uno tiene al mirar la mesa. El contenido de tal información impone condiciones en la realidad que reflejan lo que es necesario para que la experiencia resulte adecuada. Por ejemplo, el tipo de experiencia que tenemos al ver una piña rugosa encima de la mesa no es adecuada si el objeto que está encima de la mesa, en vez de una piña, resulta ser una banana amarilla. En tal caso, la información proporcionada por la experiencia sería falsa o inadecuada.

Existe, además, una relación íntima entre la información que la experiencia consciente transmite y su fenomenología. En particular, hay cierta información común que distintos ejemplares de experiencias del mismo tipo transmiten a sus sujetos cuando tienen la misma fenomenología (Horgan y Tierson, 2002; Siewert, 1998). Llamemos a esta información “información fenoménica”. Es gracias a la naturaleza informacional de la experiencia que podemos saber de ella, preguntarnos por su naturaleza, etc. De hecho, resulta difícil encontrar razones convincentes para pensar que haya aspectos de la fenomenología que trasciendan la información pro-

porcionada por la experiencia, pues no está claro cómo podríamos tener conocimiento de tales aspectos y llegar a preguntarnos por ellos como inmediatamente uno se pregunta acerca de la naturaleza de la experiencia que tiene (cf. Papineau, 2021). Si bien esta idea se alinea fácilmente con la visión predominante de la cognición, la teoría representacional de la mente, en la cual los procesos cognitivos dependen de la transformación y el almacenamiento de estructuras que llevan información (representaciones), no está en absoluto comprometida con ella.

Resulta natural hablar del contenido de la información de manera análoga a cuando decimos que un artículo o una publicación tiene contenido (Siegel, 2010; 2011). Ahora bien, el término “contenido” es un tecnicismo y distintos autores lo emplean de manera distinta. A menudo, este término se usa para referirse al objeto al que la experiencia está dirigida, el objeto del cual nos percatamos al tener una experiencia —digamos, una piña y su textura en el ejemplo anterior. Si ahora tocamos la piña, tendremos una experiencia consciente distinta. Y no resulta insensato pensar que aquello de lo que nos percatamos al tocarla es también la piña y su textura. Por ello, algunos autores distinguen el *contenido*, a saber, aquello de lo que nos percatamos, de la *manera* en que nos percatamos de ese contenido (p.ej., Kriegel, en prensa). Así, podríamos pensar que la experiencia visual y táctil de la áspera piña tienen el mismo contenido, pero difieren en la manera —visual vs. táctil— en que nos percatamos de ese contenido.

Claramente, ambas experiencias difieren en términos informativos: al tocar la piña, mi experiencia me ofrece la evidencia requerida para saber que la estoy tocando, algo que no necesariamente ocurre cuando la miro. Al hablar del contenido de la información, estamos incluyendo tanto lo que algunos autores denominan “contenido” como la *manera* en que nos percatamos de ese contenido. Para evitar confusión, continuaré hablando en términos de la información que la experiencia provee, entendiendo que tanto el objeto o propiedad al que se dirige como la manera en que lo hace contribuyen a restringir nuestra incertidumbre.

Hablar en términos de información también nos permite circunvalar algunas cuestiones que suscitan experiencias, como pueden ser los estados de ánimo, dolores u orgasmos, que no parecen estar dirigidos a ningún objeto, como más claramente ocurre en experiencias perceptivas. A pesar de ello, estos estados también nos proporcionan información acerca del estado en que nos encontramos o hacia el que debemos tender o evitar.

Una aclaración más antes de continuar: para poder hablar de la información de la experiencia, debemos utilizar oraciones. Eso no implica que tener esa información requiera tener capacidades lingüísticas o poseer los conceptos involucrados en tales oraciones. Por ejemplo, podemos tratar de expresar la información que una fotografía transmite mediante una oración, tal vez una demasiado larga, y es razonable pensar que esa fotografía le transmita cierta información a mi perro, a pesar de que carezca de las capacidades lingüísticas que entender la oración requeriría.

Hechas estas aclaraciones, pasemos a analizar cuál es la información fenoménica que ofrece nuestra experiencia.

### **3. Autopercatación (Self-awareness)**

Si abordamos el problema de la consciencia desde una perspectiva informacional, resulta tentador buscar entender la subjetividad de la experiencia en términos de información subjetiva o de primera persona. La idea de que la experiencia consciente nos informa también acerca de nosotros mismos (autoconsciencia o autopercatación) está comúnmente asociada a Descartes, pero ha sido un tema de debate activo a lo largo de la historia de la filosofía, tanto antigua como medieval (Sorabji, 2006). La noción de que en la experiencia hay una percatación de uno mismo ya se encuentra en Aristóteles, quien argumentaba que al percibir cualquier cosa, también percibimos nuestra propia existencia (De Sensu 7.448a). Por su parte, Descartes sostiene que existe una conciencia de uno mismo en el pensamiento, el famoso “cogito”, anticipado por Agustín, y del cual podemos estar absolutamente seguros (1998). Además, autores de la época moderna como Locke (1690/1994) y Berkeley

(1713/1998) pueden ser interpretados de manera natural como defensores de la idea de que la experiencia consciente involucra percatación de uno mismo.

Por otro lado, Hume se opuso célebremente a la idea de la autoconsciencia que involucra a uno mismo cuando señala:

Hay algunos filósofos que imaginan que estamos cada momento íntimamente conscientes de lo que llamamos nuestro yo [...] En mi caso, cuando entro más íntimamente en lo que llamo yo mismo, siempre tropiezo con alguna percepción particular u otra, de calor o frío, luz u oscuridad, amor u odio, dolor o placer. Nunca puedo atraparme en ningún momento sin una percepción, y nunca puedo observar nada más que la percepción. (1739, p. 252)

La observación de Hume representa un punto de inflexión en la comprensión de la autoconsciencia y, posiblemente, una de las principales motivaciones para examinar la idea de que la experiencia consierne a sí misma. Sin embargo, muchos autores después de él han insistido en puntos de vista que involucran a uno mismo. Por ejemplo, Kant argumenta que cierta forma de autoconsciencia, la “apercepción trascendental”, que parece ser compatible con la observación de Hume, es necesaria para dar cuenta de la unidad de la experiencia consciente a lo largo del tiempo (Brook, 2008).

En la tradición postkantiana, Fichte (1982) argumenta que puntos de vista sobre la autoconsciencia como los ofrecidos por Descartes, Locke e incluso Kant, consideran al yo como un objeto en lugar de un sujeto. Fichte sostiene que esta forma “reflexiva” de autoconsciencia presupone una forma más primitiva de autoconsciencia, previa a cualquier reflexión acerca de quién es el sujeto de la experiencia. Esta forma primitiva permitiría concluir que es de hecho uno mismo, una idea que posteriormente recogen autores como Shoemaker (1968) y Wittgenstein (1958) dentro de la tradición analítica.

La idea de que la conciencia implica la autoconsciencia se puede entender de dos maneras diferentes (Sebastián, 2012; 2022a): (i) la conciencia implica la percatación de uno mismo, el sujeto de la

experiencia —la información fenoménica *conciérne al sujeto* (CS), por ejemplo Rosenthal, (2005); Brown (2015); Sebastián (2012, 2022a)— o (ii) la conciencia implica la percatación del propio estado consciente —la información fenoménica *conciérne al estado mental* (CEM), por ejemplo, Carruthers (2000); Gennaro (1996); Krigel (2009); Williford (2006)—. No es infrecuente encontrar a autores que alternan entre una idea y la otra, como se observa en el trabajo de Brentano (1874/1973).

En línea con esta distinción, diferentes autores recurren a diferentes oraciones para caracterizar, de manera más o menos aproximada, la información fenoménica. Por ejemplo, oraciones en la línea de “esta experiencia presenta tal objeto o cualidad” o “tal entidad está causando esta experiencia” ilustran la idea de que la experiencia transmite información acerca de ella misma (CEM). La idea de que la experiencia transmite algo acerca del sujeto (CS) se ejemplifica en información expresada mediante oraciones como “Estoy relacionado con tal entidad” o “tal entidad está dispuesta a causar tal estado en mí” o “estoy en tal o cual estado”. Finalmente, frases como “Estoy teniendo esta experiencia”, “esta experiencia de tal entidad me pertenece” o “tal entidad está dispuesta a causar esta experiencia en mí” ilustran posiciones en las que la información fenoménica concierne tanto al sujeto como a la experiencia misma. Pasemos a ver las razones y desafíos que enfrenta cada una de las posiciones.

### **3.1. Autopercatación que concierne al propio estado mental (CEM)**

Una motivación inicial para favorecer la posición según la cual la experiencia concierne a la propia experiencia (CEM) y no al sujeto (CS) se deriva de la observación de Hume y del interés en tratar de evitar un compromiso con una entidad ontológicamente controvertida como el Yo. Sin embargo, para abordar estas cuestiones, es esencial analizar cómo caracterizar la información fenoménica y, al mismo tiempo, evaluar los compromisos ontológicos que implica que un estado porte tal información. Si bien no entraremos en detalles sobre esta segunda cuestión, sí exploraremos a fondo la

caracterización de la información fenoménica y regresaremos a la objeción de Hume. A su vez, este análisis ayudará a disipar el temor a posibles compromisos ontológicos de la posición CS.

La idea de que la información fenoménica consciente concierne a la propia experiencia encuentra cierto soporte en la psicología de sentido común, según la cual un estado consciente es un estado del cual nos percatamos —acerca del cual tenemos información— y que un estado del cual no nos percatamos no contaría en ningún sentido como consciente. Este ‘principio de transitividad’ (Rosenthal, 2005) es explotado por las teorías de orden superior (Carruthers, 2000; Gennaro, 1996; Brown, 2015) que sostienen que los estados conscientes son aquellos que están adecuadamente representados por un estado de orden superior.

El problema para esta aproximación surge en casos en los que hay un fallo en la representación (la información no es verídica) y el sujeto no se encuentra en el estado en que la experiencia transmite que se encuentra (Block, 2011). En tal situación, el sujeto experimenta conciencia y, por definición, se halla en un estado consciente pero, *ex hypothesi*, no está en el estado que la experiencia indica (el estado consciente), conduciéndonos a una contradicción. El principio de transitividad podría salvarse si adoptamos una posición CEM y sostenemos que un estado consciente es un estado que se representa a sí mismo (Kriegel, 2009). De esta manera, cuando la representación falla, el estado no se está representando a sí mismo y, por lo tanto, no se consideraría una experiencia consciente.

Sin embargo, un análisis más cuidadoso revela que hay dos sentidos en los que un estado puede decirse que es consciente, y que la aparente contradicción no es tal, pudiéndose rescatar el principio de transitividad sin adoptar una posición CEM (Brown, 2015; Sebastián, 2022a; 2023). Por un lado, un estado puede llamarse ‘consciente’ si es el estado en el que estamos cuando tenemos una experiencia, es decir, el estado *con* el que somos conscientes. Por otro lado, un estado puede llamarse ‘consciente’ si es el estado del cual nos percatamos al tener la experiencia, es decir, si es el estado *del* que somos conscientes. En caso de fallo en la representación, el

sujeto no se encuentra en el estado que su experiencia indica que está, y, por tanto, no está en el estado *del* que es consciente. No obstante, esto es compatible con que de hecho sí esté en el estado *con* el que es consciente y tenga una experiencia. La posición CEM sostiene que el estado *con* el que soy consciente y *del* que soy consciente son el mismo, pero los casos de fallo en la representación no son suficientes para justificar la adopción de esa posición.

En favor de CEM también se encuentran consideraciones epistemológicas. Parece que siempre que tenemos una experiencia, sabemos que la estamos teniendo. Imaginemos que estamos experimentando el olor del café recién tostado. Si alguien nos pregunta qué estamos haciendo, podemos responder de inmediato: “Oliendo café”. Según Sartre (1956), esto se debe a que la experiencia transmite su propia presencia, ya que somos pre-reflectivamente conscientes de nuestras experiencias. Esta pre-reflexividad nos permite responder de inmediato cuando alguien nos interroga sobre nuestras acciones (Sartre, 1956; Zahavi, 2005).

Sin embargo, Schear (2009) ha objetado que los datos fenomenológicos —que mi estado consciente sea transmitido por mi experiencia— no pueden deducirse lógicamente a partir de relaciones epistemológicas —el saber que estoy haciendo lo que estoy haciendo—. Schear propone una explicación de la relación epistemológica que prescinde de los datos fenomenológicos (Thomasson, 2006). A esto, autores como Zahavi y Kriegel (2016) han respondido que el argumento de Sartre no es deductivo sino abductivo: la mejor explicación para lo que uno está en posición de creer de inmediato y sin inferencia es aquello que es transmitido por la experiencia.

Si este argumento es sólido, entonces también apoya la idea de que la información contenida en la experiencia concierne al propio sujeto (CS), ya que se puede declarar de inmediato y sin necesidad de inferencia lo que uno mismo está haciendo. Sin embargo, el hecho de que pueda reportar de manera inmediata lo que estoy haciendo cuando lo estoy haciendo y que la mejor explicación para ello sea la información que ofrece la experiencia, no nos compro-



mete con una posición CEM. La razón es, de nuevo, que podemos distinguir el estado *con* el que soy consciente del estado *del* que soy consciente. Es cierto que cuando tengo una experiencia puedo reportar de manera inmediata el estado *del* que soy consciente (el estado del que me percato), en este caso el de oler café. Esto es suficiente para explicar la inmediatez de mi reporte. Sin embargo, de esto no se deduce que el estado que mi experiencia indica que estoy experimentando, y que me permite realizar un reporte inmediato, deba ser la experiencia misma (Sebastián, 2022a).

### 3.2. Autopercatación que Conciene al Sujeto (CS)

En la discusión contemporánea, diversos filósofos han defendido la idea de que la experiencia conciene al sujeto, basándose en razonamientos *a priori* (por ejemplo, Nida-Rümelin, 2017), en la psicología del sentido común (Rosenthal, 2005), o en premisas epistemológicas (Sebastián, en prensa). Sin embargo, una de las razones más populares, aunque controvertida, proviene directamente de la fenomenología directa. Consideramos a las experiencias conscientes como episodios subjetivos o de primera persona, y parece que lo hacemos a la luz de lo que la propia experiencia presenta —a la luz de la información que nos proporciona—, en lugar de como resultado de un proceso inferencial. En este sentido, el carácter de primera persona es pre-reflexivo como habíamos visto. Este carácter de primera persona es capturado en términos de información de primera persona. Su conexión con la conciencia es enfatizada por Zahavi y Kriegel (2016, p. 38) cuando señalan que el mundo se nos presenta en la experiencia de “una manera distintamente de primera persona”, y remarcan que negar que tal característica esté presente en la experiencia es “no reconocer la propia subjetividad de la experiencia”.

Las experiencias conscientes no solo nos informan sobre cómo es el mundo —en caso de que la experiencia sea verídica—, sino también sobre la relación que mantenemos con este y el hecho de que somos nosotros quienes estamos en esa relación. Consideremos, por ejemplo, el tipo de experiencia que se tiene al mirar una

piña áspera. Esta experiencia consciente no se limita a comunicarnos que en el mundo existe un objeto con cierta aspereza y forma, sino que revela que uno mantiene una relación específica con dicho objeto; en este caso, una relación perceptual que podríamos denominar ‘ver’, la cual se diferencia de la relación en que nos encontramos cuando, por ejemplo, lo tocamos. Si intentáramos expresar en español lo que la experiencia nos transmite, usaríamos naturalmente el pronombre de primera persona, algo así como ‘Yo estoy viendo una piña áspera’. En general, la noción de que la autoconsciencia implica al sujeto mismo puede caracterizarse diciendo que las experiencias transmiten constitutivamente que uno mismo está en cierto estado, en lugar de simplemente indicar que tal propiedad o relación existe en el mundo.

La siguiente sección articula cómo debemos entender esa información presentando formalmente la noción de ‘información perspectiva’, un tipo particular de información de primera persona o información *de se* (Castañeda, 1966; Chisholm, 1981; Lewis, 1979; Perry, 1979), en línea con la idea de que las experiencias conscientes son subjetivas, como señala O’Conaill (2019).

#### **4. Cómo entender la Autopercatación: Información Perspectiva**

Hemos observado que la información de primera persona se expresa utilizando una expresión índice, en el caso de la conciencia del sujeto (CS), mediante el pronombre de primera persona (‘yo’). Sin embargo, al interpretar esta información, es crucial reconocer que la experiencia no especifica el tipo de entidad que el sujeto realmente es. En este sentido, no determina que el tipo de entidad con la que uno se identifica —ya sea un cuerpo específico, una persona o algún otro tipo de entidad metafísicamente exótica como el yo— esté en un determinado estado. La autoconsciencia no implica consciencia de la entidad específica que uno es, sino, como se ha señalado frecuentemente, consciencia de *uno mismo en tanto que uno mismo* (Perry, 1979, p. 11). Apoyándonos en discusiones recientes sobre la naturaleza de los pensamientos índice (aquellos que el sujeto

expresa sinceramente utilizando una expresión indéxica como ‘yo’, ‘aquí’ o ‘ahora’), podemos detallar cómo entender la información fenoménica en términos de lo que denomino información perspectiva (Sebastián, 2022c, 2024).

La idea de que hay algo especial en pensarse a *uno mismo como uno mismo* y que las teorías tradicionales no logran dar cuenta de estos pensamientos, se apoya en experimentos mentales clásicos como los del comprador descuidado de Perry (1979) o los dos dioses de Lewis (1979). No obstante, discusiones contemporáneas han desafiado esta premisa, solicitando argumentos más formales para sustentarla.

Las teorías tradicionales consideran que los pensamientos son el tipo de entidades que pueden ser verdaderas o falsas y además cumplen estas dos condiciones:

**Globales:** Los pensamientos tienen un valor de verdad global, en contraposición a tener valores de verdad solo relativos a sujetos o tiempos. Esto es, dada una descripción completa del mundo, podemos asignarles un valor de verdad. En otras palabras, se les asigna el mismo valor de verdad para cualquier ubicación o punto de evaluación (por ejemplo, espacio, tiempo o sujeto) dentro de un mundo.

**Compatibles:** Son compatibles en el sentido de que diferentes sujetos, en lugares y momentos diferentes, pueden tener el mismo pensamiento. De manera más amplia, no es necesario ocupar una posición determinada en el espacio lógico para poder formarse el pensamiento.

Pensemos por ejemplo en las creencias que Emma y Noa expresarían mediante (1):

(1) Noa mide un metro y sesenta y cinco centímetros.

Este parece un ejemplo claro en el que Noa y Emma tienen la misma creencia y por tanto es *Compatible*. Además, una vez que fijamos cómo es el mundo le podemos asignar un valor de verdad, resultan-

do verdadero si y sólo si Noa mide un metro y sesenta y cinco centímetros en el mundo real. Por lo tanto *Global* también se satisface.

Constrastemos estas creencias con una que Noa expresaría mediante (2):

(2) Yo mido un metro y sesenta y cinco centímetros.

Este ejemplo ilustra un caso en el que Noa piensa en sí misma como ella misma. Noa puede aceptar racionalmente la proposición (2) sin aceptar la (1), lo cual justifica considerar que las creencias en (1) y (2) son distintas. Sin embargo, surge la pregunta: ¿en qué sentido (2) supone un desafío a las teorías tradicionales?

Los escépticos, como Cappelen y Dever (2013) y Magidor (2015), niegan que las creencias en (1) y (2) sean fundamentalmente diferentes y argumentan que estos casos son análogos a otros en los que creencias con la misma referencia varían en significatividad cognitiva. Si esta analogía es correcta, y podemos explicar por qué Hammurabi cree que Hespero es hermoso pero no Fósforo, a pesar de que ambos se refieren al mismo planeta, sin abandonar las teorías tradicionales, entonces no necesitamos abandonar ni *Compatible* ni *Global* para explicar el caso de los pensamientos indéxicos.

Los que ven un desafío en (2) a las teorías tradicionales se dividen en dos grupos, según la condición que estén dispuestos a rechazar. Los globalistas están de acuerdo en que la diferencia en importancia cognitiva en las creencias de Noa es análoga a otros casos que involucran términos co-referenciales. Sin embargo, niegan que los pensamientos indéxicos sean *Compatibles*. Por ejemplo, basándose en la sugerencia de Frege de que el pensamiento en primera persona depende de la forma en que cada uno se presenta a sí mismo: ‘una manera particular y primitiva, en la que se presenta a sí mismo, y de la que no se presenta a nadie más’ (Frege, 1956, p. 298). Por ende, Emma no puede tener la misma creencia que Noa en (2), ya que Emma no puede pensar en Noa de la manera en que Noa se piensa a sí misma.

Por otro lado, los portabilistas niegan que los pensamientos indéxicos sean *Globales*. Lewis (1979) defiende esta posición proponiendo que las creencias son autoatribuciones de propiedades, como tener cierta altura. Así, cuando Noa y Emma expresan la creencia en (2), tendrían la misma creencia al atribuirse la misma propiedad. Consecuentemente la creencia es *Compatible*. Sin embargo, dado que Emma tiene otra altura, la creencia es verdadera para Noa pero no para Emma, y por lo tanto, la creencia no es *Global*, ya que su valor de verdad cambia dentro del mismo mundo.

Pensando en términos de información podemos llamar *privilegiada* a la información que solo puede ser capturada desde una determinada posición en el espacio lógico (por ejemplo, siendo cierto individuo, como Frege sugiere) y que, por lo tanto, no es *Compatible* pese a ser *Global*. Por otro lado, la información *perspectiva* se refiere a aquella que *Compatible*, pero que es de grano más fino que la información acerca de cómo es el mundo y por lo tanto no *Global* (Sebastián, 2024)

La diferencia entre la información privilegiada y la perspectiva puede ilustrarse mediante la analogía de un mapa (Egan, 2006; Sebastián, en prensa), considerando las ciudades como análogas a mundos posibles. Tomemos, por ejemplo, un mapa que muestra la relación espacial entre los tres edificios principales de una ciudad. Esta información es *Global* y *Compatible*, ya que una vez que fijamos la ciudad en la que estamos, podemos determinar si el mapa es adecuado; es decir, el mapa resulta adecuado para algunas ciudades (aquellas cuyos tres edificios principales se encuentran en la relación que el mapa señala) y no para otras, al igual que la información *Global* es verdadera en algunos mundos pero no en otros. Además, la información es la misma independientemente de dónde nos encontremos en la ciudad (*Compatible*).

También podemos considerar la posibilidad de que detalles específicos de la ciudad solo estén disponibles para aquellos en ubicaciones específicas (no *Compatible*); por ejemplo, porque hay mapas especiales que solo se pueden comprar en lugares particulares y

solo se pueden leer allí. Esto sería análogo a la información privilegiada.

Finalmente, podemos pensar en la información perspectiva como un mapa con una flecha que dice “usted está aquí”. La validez del mapa sin flecha depende únicamente de cómo sea realmente la ciudad que se está visitando, mientras que evaluar la adecuación del mapa con la flecha requiere también conocer la ubicación propia dentro de la ciudad. Este último tipo de mapa no solo informa sobre cómo es la ciudad, sino también sobre nuestra ubicación específica dentro de ella.

La información fenoménica se expresa empleando una expresión índice ('yo', 'este estado'); digamos algo así como 'yo estoy en tal estado'. Sin embargo, como acabamos de ver, esta oración es ambigua pues admite al menos tres lecturas posibles dependiendo de si aceptamos *Global* y *Compatible* y debemos evaluar cómo debe ser entendida tal información. Sebastián (2022c, 2024) muestra que si (i) dos individuos diferentes pueden tener experiencias con la misma fenomenología y (ii) dos experiencias de sujetos ordinarios con la misma fenomenología brindan la misma información, entonces esa información es perspectiva (*Compatible* y *no Global*). Veámoslo:

Nuestra comprensión intuitiva y previa a cualquier teorización parece reconocer que dos experiencias numéricamente distintas pueden compartir la misma fenomenología. Consideremos, por ejemplo, las experiencias visuales antes y después de parpadear. Esta idea también se extiende a experiencias entre distintos individuos (Stalnaker, 2000; Shoemaker, 1996). Desde niños, es probable que nos hayamos cuestionado si otras personas perciben las cosas de la misma manera que nosotros, especialmente en lo que respecta a experiencias sensoriales, como al observar un color específico.

Aunque diferentes teorías establecen condiciones variadas para que dos individuos tengan experiencias con la misma fenomenología, es improbable que alguna teoría descarte por completo esta posibilidad. Consideremos, por ejemplo, el caso extremo de dos sujetos que son duplicados físicos. Esto apoya la primera premi-

sa necesaria para el argumento: dos sujetos distintos pueden tener experiencias con la misma fenomenología.

Además, parece evidente que cualquier persona que experimente lo que yo experimento al ver una piña áspera tendría acceso a la misma información que yo. Hemos definido la información fenoménica como aquella que un estado aporta en virtud de su fenomenología específica. Por lo tanto, experiencias con la misma fenomenología transmiten la misma información fenoménica.

Imaginemos que Emma está mirando una piña y que la información que obtiene es <yo veo una piña áspera>. Para determinar si esta información es *Compatible*, debemos considerar si otro individuo, distinto de Emma, puede poseer la misma información. Podemos demostrar que es posible si admitimos que Noa puede tener una experiencia con la misma fenomenología que Emma. En este caso, ambas experiencias contienen la misma información fenoménica, y tanto Emma como Noa tendrían acceso a la misma información, lo que la hace *Compatible*. Por lo tanto, podemos concluir que la información fenoménica no constituye información privilegiada.

Aún queda por evaluar si la información es *Global*. Para esto, necesitamos considerar si, una vez establecido cómo es el mundo en realidad, podemos asignar un valor de verdad a dicha información. En el contexto específico de nuestro ejemplo, deberíamos examinar si dos individuos del mismo mundo pueden poseer la misma información, y si esta puede ser verdadera para uno, pero falsa para el otro.

Si las ilusiones o alucinaciones pueden compartir la misma fenomenología que una experiencia verídica, entonces la respuesta es afirmativa. Imaginemos una situación en la que el sistema visual de Noa falla, resultando en una ilusión de una piña áspera cuando, en realidad, está frente a otro objeto. Si esta experiencia ilusoria comparte la misma fenomenología que la experiencia verídica de Emma, entonces ambas transmiten la misma información fenoménica. No obstante, esta información es verdadera en el caso de Emma, quien efectivamente está en la relación indicada por su ex-

perencia, pero no en el caso de Noa. Por lo tanto, la información fenoménica no es *Global* y, en cambio, se clasifica como **información perspectiva**.

La información fenoménica, entonces, no solo nos informa sobre cómo es el mundo, sino que también nos ayuda a localizarnos dentro de él en el espacio lógico; es decir, no solamente revela cómo es el mundo que habitamos, sino también quiénes, cuándo y dónde somos dentro de él. En este sentido específico, la experiencia es subjetiva, ofreciéndonos, o mejor dicho, constituyendo nuestro punto de vista. Con esta clarificación sobre la naturaleza de la información fenoménica, podemos volver a las objeciones planteadas.

#### 4.1. Objeciones

Autores como Hume parecen poner en duda la idea de que nuestra experiencia consciente nos presenta el mundo desde una perspectiva de primera persona, al menos el desglose de esta aparente obviedad en términos informacionales. Además, aún si concedemos el análisis para nuestras experiencias (de humanos adultos), se puede poner en duda que sea un análisis válido para *todas* las experiencias, y con ello, que este análisis de la subjetividad contribuya a resolver el problema de la delimitación. Esta sección esboza este tipo de debates.

Una primera razón para cuestionar la idea de que la experiencia transmite información sobre el sujeto de la experiencia radica en la preocupación de que esto podría implicar consecuencias ontológicas problemáticas, como la necesidad de postular entidades cuestionables, tales como el yo. Sin embargo, la evaluación de los compromisos ontológicos derivados de poseer cierta información depende de los detalles específicos de una teoría que explique lo que se requiere de un sistema para que pueda tener esa información. Además, la información perspectiva no especifica la naturaleza precisa de la entidad concernida más allá de indicar que se encuentra en un determinado tipo de estado.

Otra razón independiente surge de la introspección, pues, como Hume señala, no parece que encontremos ninguna entidad como el yo al introspectar nuestra experiencia. Esta observación



parece desafiar la idea de que la autoconsciencia concierne directamente al yo. Sin embargo, si la información fenoménica es perspectiva, podemos entender claramente por qué tal conclusión no es necesariamente correcta (Sebastián, 2022a).

Para ilustrar la idea podemos pensar en la información que obtenemos al mirar al ratón a la derecha del teclado. Dejando la autoconsciencia a un lado por el momento, esta experiencia transmite que el ratón está a la derecha del teclado desde cierta ubicación. La experiencia visual transmite algo acerca de la ubicación de la misma manera que la autoconsciencia lo hace sobre uno mismo. Para entenderlo, basta notar que la experiencia no revela la identidad de esta ubicación; mi experiencia en estos momentos no transmite que la taza está a la derecha de la computadora en cierto lugar en la Ciudad de México donde me encuentro en estos momentos. De hecho, si me presentaran una disposición visualmente idéntica en una ubicación muy diferente, tendría una experiencia con la misma fenomenología y ambas experiencias transmitirían la misma información. Para expresar la información que la experiencia transmite emplearíamos la expresión índice *‘aquí’*, algo así como (dejando de lado la autoconsciencia) *que el ratón está a la derecha del teclado desde aquí*.

Como hemos observado, distintos teóricos proponen diversas interpretaciones semánticas de la información expresada a través de una expresión índice. Siguiendo las mismas razones discutidas previamente, la información debe considerarse perspectiva (es decir, *Compatible* y no *Global*). Nuevamente, si es posible tener experiencias con idéntica fenomenología en distintas ubicaciones, entonces la información fenoménica en tales situaciones es la misma, y por ende, *Compatible*. Sin embargo, esta información no puede ser *Global*, ya que puedo tener dos experiencias del mismo tipo (con la misma fenomenología) en lugares distintos que resulten ser verídica en un caso e ilusoria en otro.

Ahora bien, si le pido al lector que introspecciones la ubicación, que examine el “aquí”, probablemente me mirará con extrañeza y no encontrará ninguna ubicación en la introspección. Pero

de ello no concluimos que esta experiencia no transmita cómo están dispuestos los objetos en el entorno de una ubicación, sino que la ubicación en sí no forma parte de la información y por ello no puede ser introspectado. Como afirma Hume, no encontramos en la introspección ninguna entidad con que identificar el “yo” que aparece en la expresión de la información transmitida. El caso es análogo al de la experiencia visual donde no encontramos en la introspección ninguna ubicación que identificar como la referencia del “aquí” en la oración empleada para expresar la información, simplemente porque la referencia no es parte de la información transmitida —lo que a su vez permite que la información sea *Compartible*. Quienes estén dispuesto a seguir a Hume estarán equivocados. Es un error concluir, a partir del hecho de que no logramos encontrar en la introspección ninguna entidad que identificar con la referencia de la expresión índice que la expresión de la información empleando el índice no sea adecuada.

En la filosofía analítica actual, se ha planteado la posibilidad de que, a pesar de que las experiencias de los adultos humanos puedan transmitir información sobre el sujeto que las experimenta, otros seres podrían tener experiencias sin esta característica (Kriegel, 2009; Peacocke, 2014; cf. Kriegel, 2003). Sin embargo, esta posibilidad se plantea sin respaldo argumentativo sólido, y existen dudas sobre su coherencia. La idea de que un estado fenoménicamente consciente no sea subjetivo en el sentido relevante plantea cuestionamientos. Además, dado que no podemos acceder a las experiencias de seres más simples, resulta difícil justificar afirmaciones sobre lo que transmiten sus experiencias, si es que realmente tienen experiencias. En términos metodológicos, parece más sensato utilizar nuestras propias experiencias conscientes como punto de partida para desarrollar teorías y luego, una vez que tengamos una teoría sólida, considerar qué tipo de estados podrían experimentar otros seres.

Ahora bien, aunque algunos de nosotros no podamos dar sentido a la idea de una experiencia que no sea subjetiva en el sentido discutido, enriquece intelectualmente mantener la mente abierta.

Diversos estados alterados, como aquellos que se dan en la meditación, bajo el uso de drogas psicodélicas o en algunas condiciones psiquiátricas, parecen diferir de nuestra experiencia ordinaria de vigilia, y diversos autores sostienen que ilustran casos de experiencias que carecen de aspecto subjetivo.

Si bien no es posible revisar todas las consideraciones psiquiátricas que son susceptibles de afectar al aspecto subjetivo de la experiencia (véase Gennaro, 2015), algunos ejemplos en los que el sentido del yo se ve, de un modo u otro, afectado nos pueden ayudar a entender el desafío. Por ejemplo, la somatoparafrenia es un tipo de delirio corporal en el que el sujeto niega que una extremidad o un lado completo del cuerpo sea suyo (Vallar y Ronchi, 2009; Liang y Lane, 2009; Rosenthal, 2010). Sin embargo, aunque estos pacientes niegan que cierta parte de su cuerpo sea suya, no niegan que la sensación que tienen lo sea. Por lo tanto, no parece haber una tensión con la idea de que la experiencia informa de que uno está en cierto estado de manera constitutiva.

La cuestión está algo menos clara en otros casos, como los casos de despersonalización, donde en situaciones radicales como el síndrome de Cotard (Berrios y Luque, 1995), el sujeto niega incluso estar vivo. En estos casos, se ha sugerido que los pacientes carecen de un sentido del yo, lo que parece desafiar la posibilidad de que se autotribuyan cierto estado. La evaluación de la racionalidad en estos casos resulta complicada, pero la evidencia sugiere que hay una tensión entre la creencia que tienen los sujetos y la información que, supuestamente, otorga la experiencia.

Ahora bien, no está claro que esa tensión revele que en las experiencias no haya información de primera persona, pues en muchas ocasiones rechazamos en el pensamiento la información que nos brinda la experiencia. Por ejemplo, en la conocida ilusión de Müller-Lyer (1889), la experiencia nos dice que una línea es más larga que otra, pero no aceptamos esa información y creemos que ambas tienen la misma longitud.

Un caso patológico ampliamente discutido es el de la esquizofrenia (Sawa et al., 2016). En este caso, los sujetos llegan a repor-

tar que ciertos estados se sienten como si no fueran suyos y pertenecieran a alguien más. Esto supone un claro desafío para quienes articulan la subjetividad en la experiencia como información que dicta que cierto estado le pertenece a uno (sentido de posesión; Billion y Kriegel, 2015; Gallagher, 2000).

En respuesta, estos autores tratan de ofrecer una explicación de los reportes de pacientes con esquizofrenia en términos de otro tipo de alteración, ya sea por defecto o por exceso. Por ejemplo, Gallagher sostiene que lo que les falta a estos pacientes es información acerca de quién genera el estado mental (sentido de agencia) y no acerca de la posesión del estado.

Sin embargo, si el aspecto subjetivo se explica en términos de información acerca del estado en que uno se encuentra, el reporte de que el estado no le pertenece a uno no parece suponer ningún desafío. Para entenderlo mediante un ejemplo sencillo, basta notar que la información de que la casa en la que estoy no es mía no está en tensión con la información de que yo estoy en esa casa.

Otro caso interesante lo presentan las experiencias meditativas y las experiencias con drogas psicodélicas. Se ha argumentado que uno de los mecanismos de acción en la meditación está relacionado con una alteración en el sentido del yo (Hoelzel et al., 2011). De hecho, existen evidencias que muestran que la actividad de las redes cerebrales vinculadas al yo narrativo y la actividad relacionada con el yo se ven influenciadas (Dor-Ziderman, Berkovich-Ohana, Glicksohn y Goldstein, 2013).

En estos experimentos, algunos sujetos llegan a reportar que no hay un sentido de agencia o de propiedad en la experiencia, lo que desafía algunas maneras de entender la subjetividad. Sin embargo, en esos mismos reportes, los sujetos informan estar experimentando lo que está sucediendo, en clara consistencia con la idea de que la experiencia informa del estado en el que se encuentra el sujeto. Además, hay razones para poner en duda estos reportes (Sebastián, 2020).

Los estudios de meditación se han centrado en las prácticas budistas (Lutz, Dunne y Davidson, 2007). En las tradiciones budistas,

no existe tal cosa como el yo, que es una mera ilusión de nuestras mentes. La raíz del sufrimiento humano proviene de nuestra identificación con una entidad ilusoria, y la meditación es una forma de aprender esto al alcanzar un estado de experiencia sin ego en el que se abandona la identificación con el yo (Dalai Lama, 1991). En este sentido, la descripción del estado que se alcanza difiere en su descripción con respecto a otras escuelas en el hinduismo que lo describen de un modo y otro como la unión del alma o yo con lo divino. Sin favorecer una descripción y otra, esta diferencia nos da buenas razones para pensar que la información almacenada y recuperada de sus experiencias, así como la interpretación de estos recuerdos, reflejada en sus reportes, esté probablemente influenciada por las creencias del meditador.

Algo similar ocurre en las experiencias con psicodélicos, que a menudo se describen como alteraciones de la autoconsciencia, desafiando la presencia constitutiva de información perspectiva en la experiencia (Letheby and Gerrans, Milliere, 2017). De hecho, varios estudios que utilizan drogas psicodélicas clásicas (agonistas parciales del receptor 5-HT<sub>2A</sub>), como la psilocibina, la mescalina, el LSD o el DMT, así como otras sustancias psicoactivas como la ketamina, han observado una reducción de la autoconsciencia en comparación con experiencias ordinarias. Sin embargo, hay buenas razones que sugieren que los relatos sobre experiencias psicodélicas pueden estar sujetos al mismo tipo de sesgo mencionado en el caso de las prácticas meditativas (Sebastián, 2020).

El término “psicodélicos” proviene de Humphrey Osmond en diálogo con Aldous Huxley, describiendo un estado de conciencia alterado por sustancias donde se manifiesta (delos) la mente (psique). Huxley fue clave en la difusión de los psicodélicos, tanto por su literatura como por su promoción entre círculos de intelectuales y artistas, influenciado por las religiones de Oriente y viendo en estas sustancias un camino espiritual similar al de la meditación (Stevens, 1998). En “Las puertas de la percepción”, describe la sensación de unidad con el entorno como el “estado final de la falta de ego”, idea que evolucionaría posteriormente. Huxley introdujo

a Timothy Leary al Libro Tibetano de los Muertos, que tanta influencia tuvo en el manual que este escribiría para las experiencias con LSD (Leary et al., 1964). Leary veía paralelos entre las fases de muerte y renacimiento con las identificadas en sus estudios sobre LSD, conceptualizando la pérdida del ego como una trascendencia total, sin visiones, autoconciencia ni pensamientos, sino como una conciencia pura y libertad (Gould, 2007). De esta manera, lo que explicaría los reportes es una experiencia anómala bajo el uso de psicodélicos que es conceptualizada según el estado de disolución del yo que se espera encontrar.

De hecho, la hipótesis de que lo que explica el reporte acerca de la disolución de la subjetividad en la experiencia con psicodélicos sea la propia creencia del sujeto, más que ciertos rasgos de la experiencia, encuentra respaldo en estudios antropológicos. Las drogas psicodélicas tradicionalmente han sido consumidas por comunidades con diferentes puntos de vista cosmológicos fuera de la cultura occidental. Por ejemplo, los hongos psicobios se han utilizado tradicionalmente en Mesoamérica (Wasson & Wasson, 2011), al igual que el peyote que contiene mescalina (La Barre, 2011); la ayahuasca se ha utilizado en la Amazonia (Chaumeil, 1998) y las semillas de anadenanthera en América del Sur (Torres & Repke, 2006), ambas contienen DMT. Los estudios antropológicos muestran que las experiencias con drogas en estas comunidades se centran en el encuentro con espíritus y la obtención de poder y conocimiento de ellos, sin hacer referencia a la disolución del ego o la unión mística con el universo.

## **5. Cualidades, Subjetividad y el Problema de la Consciencia**

Al comienzo de este capítulo, veíamos que el desafío que la experiencia consciente supone para una explicación científica se puede descomponer en dos aspectos: uno asociado al aspecto cualitativo y otro al subjetivo. Una pregunta que surge es si hay alguna relación de dependencia entre ambas cuestiones, de manera que el núcleo

del problema radique en uno de ellos, de modo que, una vez resuelto, el misterio que el otro problema parece albergar se disuelva.

Nagel (1974), por ejemplo, parece apuntar al aspecto subjetivo cuando señala que la experiencia consciente está ligada a un punto de vista que el objetivismo de la ciencia parece dejar de lado. De esta manera, el hecho de que no podamos adoptar el punto de vista del murciélago hace que nos sea imposible saber cómo se siente ser un murciélago. Por otro lado, el argumento del conocimiento de Jackson (1982, 1986) apunta claramente a las cualidades como la fuente del problema. En el experimento mental que Jackson propone, Mary, una neurocientífica con conocimiento físico completo del sistema visual pero que ha pasado toda su vida sin ver colores y estudiando desde el interior de una habitación en blanco y negro, se pregunta cómo se sentirá ver colores como el rojo. Esta incertidumbre se resuelve cuando finalmente es liberada y ve una rosa por primera vez.

En el ámbito de la filosofía analítica contemporánea, la perspectiva más comúnmente aceptada implica abrazar una posición cualitativa que postula que las cualidades son la raíz del problema de la conciencia. No obstante, existen autores que respaldan un enfoque subjetivista, como Kriegel (2009), argumentando que el aspecto subjetivo es la verdadera fuente del problema. Además, algunos autores han defendido que estos dos aspectos son problemas independientes (McClelland, 2016). Considerando que la principal motivación para pensar que son las cualidades de la experiencia la fuente del problema de explicar la conciencia fenoménica se deriva en buena parte del famoso experimento mental de Jackson (1982; 1986), esta sección revisará, a la luz de lo que hemos explorado, una respuesta al Argumento del Conocimiento presentada por Sebastián (en prensa) Esto nos permitirá ampliar las posibilidades para afrontar el desafío que la conciencia ofrece.

La intuición que subyace al argumento del conocimiento es que uno no puede aprender cómo se siente una experiencia que no ha tenido, por mucho que le cuenten acerca de ella (Lewis, 1988). Dicho de otro modo, conocer la distribución de cualquiera que sea

el tipo de propiedades que la ciencia nos diga que son fundamentales —de tal manera que el mundo quede exhaustivamente descrito indicando qué cosas ejemplifican qué propiedades fundamentales (Lewis, 1986)— resulta insuficiente para saber cómo se siente tener una experiencia que no hemos tenido. La conclusión del argumento es que ha de haber otras propiedades (*qualia*) que vayan más allá de las propiedades que la ciencia postula —propiedades que actualmente se asume que son físicas—.

Dentro de su habitación en blanco y negro, Mary sabe qué hay objetos que son de color rojo y entiende que se siente de cierta manera ver esos objetos, al igual que se siente de alguna manera ver los objetos en su habitación, pero se pregunta acerca del tipo de experiencia que la gente tiene al verlos. Esta incertidumbre se resuelve cuando es liberada y ve el color rojo por primera vez. La experiencia que tiene le brinda cierta información y es esa información la que le permite resolver la incertidumbre que tenía. El desafío, entonces, reside en explicar por qué esa información no puede ser derivada del conocimiento científico completo que ella posee, según el experimento mental.

El tipo de respuestas que encontramos en la literatura (aceptando la intuición acerca de la ganancia epistémica de Mary al tener la experiencia) se basan en pensar que la información que da la experiencia es información que sólo puede ser obtenida al tener la experiencia (Howell, 2013). Mary no podía saber cómo se siente ver rojo porque ese conocimiento requiere estar en cierta relación de familiaridad (*acquaintance*) con la propia experiencia (Conee, 1994) —digamos para poder adquirir ciertos conceptos (Balog, 2009; Block, 2006; Loar, 1990; Papineau, 2007) o habilidades (Lewis, 1988; Nemirow, 1990). De esta forma, este tipo de respuestas se basan en pensar que la información fenoménica es **privilegiada** y no es *Compartible*. Se trata de información que solo puede ser obtenida desde una posición particular en el espacio lógico, a saber, aquella que otorga la relación de familiaridad y esto explica porque no puede ser derivada del conocimiento que Mary posee en su interior.



El problema que enfrentan este tipo de respuestas radica en mostrar cómo ese conocimiento puede ser sustantivo, en el sentido de que resuelve una incertidumbre que Mary tenía, de manera compatible con su conocimiento exhaustivo del mundo. Si bien es cierto que, cuando Hammurabi descubre que Héspero es Fósforo, aprende algo a pesar de saber que el hecho de que Héspero sea Héspero era algo que ya sabía, ese nuevo conocimiento implica otros hechos contingentes de los que Hammurabi era ignorante (por ejemplo, que Héspero aparece al amanecer). El cambio en la posición epistémica al conocer un hecho bajo un modo distinto de presentación o concepto requiere la falta de conocimiento de algún detalle de cómo es el mundo. Sin embargo, Mary es omnisciente en este sentido.

Ahora bien, la idea de que estados de incertidumbre son compatibles con el conocimiento exhaustivo del mundo es familiar para los lectores de Lewis (1979): uno puede tener conocimiento exhaustivo del mundo que habita y aún así preguntarse acerca de dónde, cuándo o quién es al interior de ese mundo. Y si la información fenoménica es **perspectiva**, esta es precisamente el tipo de información que provee la experiencia. De esta manera, la información que provee la experiencia no podría ser derivada del conocimiento científico no porque no sea *Compatible* (privilegiada) sino por no ser *Global* (perspectiva). Mary sabe perfectamente quién es y dónde está localizada en el espacio-tiempo en el mundo, pero lo que no sabe es cuál es la información que las experiencias de color aportan, una información más fina que, al no ser *Global*, no puede ser derivada del conocimiento exhaustivo del mundo.

Esta respuesta permite conciliar el conocimiento científico exhaustivo de Mary con el hecho de que aprenda algo nuevo al tener la experiencia, sin comprometernos con la idea de que nueva información del mundo físico pueda ser obtenida a través de la experiencia. Ofrece una solución al argumento del conocimiento. Si las diferencias en la fenomenología entre distintas experiencias se derivan de los distintos estados en que la experiencia nos informa que estamos, el problema de la consciencia resulta ser funda-

mentalmente el problema de explicar cómo es posible que un estado porte información perspectiva. Si bien el antecedente de este condicional está lejos de haber sido establecido, la aproximación subjetivista ofrece una interesante y novedosa perspectiva sobre el problema de la consciencia.

## 6. Conclusión

Este capítulo proporciona un análisis de las diversas formas en que se ha entendido el carácter subjetivo de la consciencia en términos informacionales, presentando razones que sugieren que la experiencia consciente tiene un carácter de primera persona, es decir, que concierne al propio sujeto. Además, ofrece un marco para abordar críticas y objeciones potenciales a esta idea.

La discusión ha evidenciado que la información de primera persona, caracterizada por su naturaleza perspectiva, es compartible, pero se diferencia de la información objetiva proporcionada por las ciencias en que no es *Global*. Esta información, más detallada, no solo nos permite entender cómo es el mundo, sino también nuestro lugar dentro de él, ofreciendo una riqueza descriptiva que las representaciones científicas objetivas no logran capturar completamente.

Mientras que el paradigma científico ha evolucionado significativamente a lo largo de la historia, el desafío de integrar la visión científica del mundo con la riqueza de nuestras experiencias conscientes sigue siendo un obstáculo prominente. En las discusiones contemporáneas, el aspecto cualitativo se percibe como el principal obstáculo para explicar la consciencia, lo que conduce a una exploración limitada de soluciones posibles. Pensar en el problema en términos informacionales y adoptar un enfoque centrado en la información perspectiva propone una vía prometedora para superar este obstáculo, sugiriendo que una comprensión profunda de la consciencia requiere considerar cómo los estados fenomenológicos encapsulan información específica desde el punto de vista del sujeto.

El objetivo de nuestras teorías científicas es elucidar la naturaleza del mundo que habitamos. Sin embargo, no es obvio que esta información sea la más fructífera para que un organismo actúe y

sobreviva. Una arquitectura cognitiva que pueda explotar información sobre uno mismo y su entorno —es decir, sobre “quién y cuándo en el mundo es uno”— podría tener ventajas competitivas. Esta propuesta insta a considerar la información perspectiva como el núcleo para comprender la cognición a nivel personal y sugiere que esta perspectiva abre nuevas direcciones de investigación en la conciencia, particularmente en relación con los problemas cognitivos que esta información está especialmente equipada para abordar. Identificar y explorar estas áreas de investigación, incluyendo cómo la información perspectiva facilita la navegación, la toma de decisiones, y la interacción social, puede ofrecer valiosos insights para el avance de nuestra comprensión de la conciencia y su integración en el marco más amplio de las ciencias cognitivas.

## Referencias:

- Aristotle. (1906). *De Sensu and De Memoria*, Cambridge: Cambridge University Press.
- Balog, K. (2009) Phenomenal concepts. En B. McLaughlin, A. Beckermann and S. Walter (eds), *Oxford Handbook in the Philosophy of Mind*, Oxford University Press.
- Balog, K. (2012). In defence of the phenomenal concept strategy, *Philosophy and Phenomenological Research*, 84, 1-23.
- Berkeley, G. (1713/1998). *Three Dialogues Between Hylas and Philonous*, Oxford: Oxford University Press.
- Berrios, G. and Luque, R. (1995). Cotard's delusion or syndrome?. *Comprehensive Psychiatry*. 36 (3): 218-223.
- Billon, A. and Kriegel, U. (2015). Jaspers dilemma: The psychopathological challenge to subjectivity theories of consciousness. En R. Gennaro (ed.), *Disturbed Consciousness*, Cambridge: Cambridge University Press.
- Block, N. (2006). Max black's objection to mind-body identity, *Oxford Review of Metaphysics* 3.
- Block, N. (2011). The higher order approach to consciousness is defunct, *Analysis*, 71(3).
- Brentano, F. (1874/1973). *Psychology from an Empirical Standpoint*, International Library of Philosophy.

- Brook, A. (2008). Kant's view of the mind and consciousness of self, <http://plato.stanford.edu/entries/kant-mind/>. URL: <http://plato.stanford.edu/entries/kant-mind/3>
- Brown, R. (2015). The HOROR theory of phenomenal consciousness, *Philosophical Studies*, 172, 1783-1794.
- Carruthers, P. (2000). *Phenomenal Consciousness: a naturalistic theory*, Cambridge: Cambridge University Press.
- Castañeda, H. (1966). 'he': A study in the logic of self-consciousness, *Ratio*, 8, 130-157.
- Cappelen, H. and Dever, J. (2013). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*, Oxford: Oxford University Press.
- Chaumeil, J. P. (1998). *Ver, saber, poder. Chamanismo de los yagua de la Amazonía peruana*, Lima: IFEA/CAAAP/CAEA-CONICET.
- Chisholm, R. M. (1981). *The First Person: An Essay on Reference and Intentionality*. Minneapolis: University of Minnesota Press.
- Conee, E. (1994). Phenomenal knowledge. *Australasian Journal of Philosophy*, 72, 136-150.
- Dalai Lama, X. (1991). *Path to Bliss*, NY: Snow Lion.
- Descartes, R. (1998). *Descartes: Selected Philosophical Writings*. Cambridge: Cambridge University Press.
- Dittrich, A. (1998). The standardized psychometric assessment of altered states of consciousness (ascs) in humans. *Pharmacopsychiatry*, 31(2), 80-84.
- Dittrich, A., Lamparter, D. and Maurer, M. (2010). *5d-asc: Questionnaire for the assessment of altered states of consciousness. a short introduction*. Zurich, Switzerland: PSIN PLUS.
- Dor-Ziderman, Y., Berkovich-Ohana, A., Glicksohn, J. and Goldstein, A. (2013). Mindfulness-induced selfess: A MEG neurophenomenological study. *Frontiers in Human Neuroscience*, 7(582).
- Egan, A. (2006). Appearance properties?, *Noûs*, 40(3), 495-521.
- Farrell, B.A. (1950) Experience, *Mind*, 59, pp. 170-198
- Fichte, J. G. (1794/1982). *Science of Knowledge*, 2nd edition, Cambridge: Cambridge University Press.
- Frege, G. (1956). The thought: a logical inquiry, *Mind* 65(259), 287-311.

- Gallagher, S. (2000). Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. En D. Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*, Amsterdam: John Benjamins.
- García-Carpintero, M. (2017). *The philosophical significance of the de se*, Inquiry.
- Gennaro, R. J. (1996). *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, John Benjamins.
- Gennaro (2015). *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness*. Cambridge: The MIT Press.
- Gould, J. (2007). *Can't Buy Me Love: The Beatles, Britain, and America*. Crown Publishing Group.
- Gurwitsch, A. (1941). A non-egological conception of consciousness. *Philosophy and Phenomenological Research*, 1, 325-338.
- Heidegger, M. (1989). *Die Grundprobleme der Phaenomenologie*, Gesamtausgabe Band 24. Frankfurt am Main: Vittorio Klostermann; English translation: A. Hofstadter. *The Basic Problems of Phenomenology*, Bloomington, IN: Indiana University Press.
- Hoelzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R. and Ott, U. (2011). How does mindfulness meditation work? proposing mechanisms of action from a conceptual and neural perspective, *Perspectives on Psychological Science*, 6(6), 537-559.
- Horgan, T. (1984). Jackson on physical information and qualia, *Philosophical Quarterly*, 32, 127-136.
- Horgan, T. and Tierson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. En D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.
- Howell, R. J. (2013). *Consciousness and the Limits of Objectivity*, Oxford: Oxford University Press.
- Hume, D. (1739). *A Treatise of Human Nature*, Oxford University Press, USA.
- Jackson, F. (1982). Epiphenomenal qualia, *Philosophical Quarterly*, 32(April), 127-136.5
- Jackson, F. (1986). Whay mary didn't know, *The Journal of Philosophy*, 83(5), 291-295.
- Kriegel, U. (2003). Consciousness, higher-order content, and the individuation of vehicles, *Synthese*, 134(3), 477-504.

- Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press, USA.
- Kriegel (en prensa) What Is Inner Awareness? Uriah Kriegel For D. Bordini, A. Dewalque, and A. Giustina (eds.), *Consciousness and Inner Awareness*. Cambridge University Press.
- La Barre, W.: 2011, *The peyote cult*. University of Oklahoma Press.
- Leary, T., Metzner, R. and Alpert, R. (1964). *The psychedelic Experience. A manual based on the tibetan book of the dead*. Citadel.
- Lethaby, C. and Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness*, pp. 1-17.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*, Oxford University Press.
- Lewis, D. (1979). Attitudes de dicto and de se, *Philosophical Review*, 88(4), 513-543.
- Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell Publishers.
- Lewis, D. (1988). What experience teaches, *Proceedings of the Russellian Society*, 13(29-57).
- Liang, C and Lane, T. (2009). Higher-order thought and pathological self: the case of somatoparaphrenia, *Analysis*, 69 (4):661-668
- Loar, B.: (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81-108.
- Locke, J. (1690/1994). *An Essay Concerning Human Understanding*, Prometheus Books.
- Lutz, A., Dunne, J. D. and Davidson, R. J. (2007). *Meditation and the neuroscience of consciousness*. Cambridge Handbook of Consciousness, NY: Cambridge University Press.
- Lycan, W. (1996). *Consciousness and Experience*, Cambridge MA: MIT Press.
- Magidor, O. (2015). The myth of the de se, *Philosophical Perspectives*, 29(1), 249-283.
- McClelland, T. (2016). Can self-representationalism explain away the apparent irreducibility of consciousness?, *Synthese*, 193(6), 1-22.
- McMahan, D. L. (2008). The Making of Buddhist Modernism.
- Millière, R. (2017). Looking for the self: phenomenology, neurophysiology and philosophical significance of drug-induced ego dissolution. *Frontiers in Human Neuroscience*, 11(245).

- Millière, R., Carhart-Harris, Berkovich-Ohana, A.: R. L., Roseman, L., Trautwein, F.-M. (2018). Psychedelics, meditation and self-consciousness, *Frontiers in Psychology*, 9(1475), 1-29.
- Müller-Lyer, FC. (1889). Optische Urteilstäuschungen. *Archiv für Physiologie Suppl.* 1889: 263-270.
- Nagel, T. (1974/2002). What is it like to be a bat?, En D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. *Mind and Cognition*.
- Nida-Rümelin, M. (2017). Self-awareness, *Review of Philosophy and Psychology*, 8, 55-82.
- O'Conaill, D. (2019). Subjectivity and mineness, *Erkenntnis*, 84(2), 325-341.
- Owen MJ, Sawa A, Mortensen P.B. (2016). Schizophrenia. *The Lancet*, 388 (10039): 86-97.
- Papineau, D. (2007). Phenomenal and perceptual concepts. En T. Alter and S. Walter (eds), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, Oxford University Press, pp. 111-144.
- Papineau, D. (2021). *The Metaphysics of Sensory Experience*, Oxford: Oxford University Press.
- Peacocke, C. (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford University Press, Oxford.
- Perry, J. (1979). The problem of the essential indexical, *Noûs*, 13, 3-21.
- Perry, J. (2001). *Knowledge, Possibility and Consciousness*. Cambridge, MIT Press.
- Prinz, J. (2012). *The Conscious Brain*. Oxford University Press.
- Rosenthal, D. (2005). *Consciousness and mind*. Oxford University Press.
- Rosenthal, D. (2010). Consciousness, the self and bodily location. *Analysis*, 70 (2):270-276
- Sartre, J.-P. (1936/1960). *The Transcendence of the Ego: An Existentialist theory of Consciousness*. New York: Hill and Wang. 7
- Sartre, J.-P. (1956). *Being and Nothingness*, trans. Hazel E. Barnes New York: Philosophical Library.

- Scheur, J. K. (2009). Experience and self-consciousness. *Philosophical Studies*, 144(95-105).
- Sebastián, M. Á. (2012). Experiential awareness: Do you prefer “it” to “me”?, *Philosophical Topics* 40(2), 155-177.
- Sebastián, M. Á. (2018). Consciencia, *Enciclopedia de la Sociedad Española de Filosofía Analítica*. URL: <http://www.sefaweb.es/consciencia/>
- Sebastián, M. Á. (2019). Subjective character, the ego and de se representation: Phenomenological, metaphysical and representational considerations on pre-reflective self-awareness. En M. F. Marc Borner and K. Williford (eds), *Senses of Self: approaches to pre-reflective self-awareness*, ProtoSociology.
- Sebastián, M. Á. (2020). Perspectival self-consciousness and ego-dissolution: A theoretical analysis of (some) altered states of consciousness. *Philosophy and the Mind Sciences*, 1(1), 9.
- Sebastián, M. Á. (2022a). First-person perspective in experience: Perspectival de se representation as an explanation of the delimitation problem, *Erkenntnis*.
- Sebastián, M. Á. (2022b). *El problema de la consciencia: una introducción crítica a la discusión filosófica actual*. Cátedra 9788437642727.
- Sebastián, M. Á. (2022c). Conscious Perception in favour of Essential Indexicality. *Belgrade Philosophical Annual*, 35(2):13-30
- Sebastián, M. Á. (2023). Being self-involved without thinking about it: Confusions, virtues and challenges of higher-order theories, in J. Weisberg (ed.), *Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*, Cambridge: Cambridge University Press.
- Sebastián, M.A. (2024) Consciousness and perspectival de se content. *Synthese* 203, 189
- Sebastián (en prensa) Subjectivity as Perspectival Information: A Novel Approach to the Knowledge Argument
- Shoemaker, S. (1968). Self-reference and self-awareness, *The Journal of Philosophy*, 65, 555-567.
- Shoemaker, S. (1996). *Intrasubjective/ intersubjective, The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press, pp. 141-154.



- Siegel, S.: 2010, *The Contents of Visual Experience*, New York: Oxford University Press.
- Siegel, S. (2011). Perceptual content, in E. N. . Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition).
- Siewert, C. (1998). *The Significance of Consciousness*, Princeton University Press.
- Sorabji, R. (2006). *Self: Ancient and Modern Insights about Individuality, Life, and Death*. Oxford: Clarendon Press.
- Stalnaker, R. C. (2000). Comparing qualia across persons. *Philosophical Topics*, 26, 385-405.
- Stevens, J. (1998). *Storming Heaven: LSD and the American Dream*.
- Thomasson, A. L. (2006). Self-awareness and self-knowledge, *Psyche*, 12(2), 1-15.
- Torres, C. and Repke, D. (2006). *Anadenanthera: Visionary Plant of South America*. Routledge.
- Tye, M. (1997). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. The MIT Press.
- Vallar G., Ronchi R. (2009). Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research*, 192 (3): 533-551.
- Wasson, V. and Wasson, G. (2011). *Mushrooms, Russia and history (volume II) (A Complete Book on Mushrooms)*. Pantheon books New York.
- Williford, K. (2006). The self-representational structure of consciousness. En U. Kriegel and K. Williford (eds), *Self-Representational Approaches to Consciousness*, The MIT Press.
- Wittgenstein, L. (1958). *The Blue and Brown Books*, New York: Oxford.
- Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. MIT Press.
- Zahavi, D. and Kriegel, U. (2016). For-me-ness: what it is and what it is not. En Dahlstrom (ed.), *Philosophy of mind and phenomenology: Conceptual and Empirical Approaches*, London: Routledge.



## CAPÍTULO VII

### EL DEBATE ACTUAL SOBRE LA LECTURA DE MENTES EN BEBÉS

*Ányerson Stiths Gómez-Tabares<sup>1</sup> & Santiago Arango-Muñoz<sup>2</sup>*

#### RESUMEN

La capacidad para comprender los estados mentales de los demás se conoce como teoría de la mente o capacidad de lectura de mentes. Estudios recientes han mostrado que los bebés de alrededor de 15 meses comprenden la mente del otro. Esto contradice la tradición

---

1 Ányerson Stiths Gómez-Tabares: Doctor en Filosofía por la Universidad de Antioquia y Doctor en Psicología por la Universidad San Buenaventura. Profesor Asistente, Universidad Católica Luis Amigó, Medellín Colombia.

2 Santiago Arango-Muñoz: Doctor en Filosofía por la Universidad de Bochum. Profesor Asociado, Instituto de Filosofía, Universidad de Antioquia, Medellín Colombia. Este texto es un producto vinculado al grupo de investigación Neurociencias Básicas y Aplicadas de la Universidad Católica Luis Amigó, Medellín Colombia, y al proyecto de investigación “Mental agency” (CODI 2022-52270) del Grupo de investigación Conocimiento, Filosofía, Ciencia, Historia y Sociedad del Instituto de Filosofía, Universidad de Antioquia, Medellín Colombia.

que sostenía que la lectura de mentes emerge con la adquisición del lenguaje alrededor de los cuatro años. Así pues, el debate actual se enfoca en cómo los bebés pueden entender la mente del otro de manera no-lingüística. El objetivo de este capítulo es presentar el debate actual y las controversias más relevantes sobre la atribución de estados mentales en la infancia desde una perspectiva filosófica. Para ello, en este capítulo primero presentamos dos teorías filosóficas que explican las capacidades representacionales de los bebés para comprender los estados mentales de los otros: primero, la teoría de Apperly y Butterfil (2009) y luego la de Carruthers (2016). Posteriormente, analizamos las principales diferencias y limitaciones de estas teorías para caracterizar de manera precisa cómo los bebés comprenden los estados mentales de los demás. Finalmente, exploramos una línea de explicación alternativa según la cual los bebés comprenden los estados mentales mediante representaciones diagramáticas, y determinaremos cómo esta perspectiva puede incidir en el debate entre las dos teorías presentadas.

## ABSTRACT

The ability to understand the mental states of others is known as theory of mind or mind-reading. Recent studies have shown that infants understand others' minds as early as 15 months of age, in particular, by measuring this ability with the implicit false belief task. This contradicts the tradition that mindreading emerges with language acquisition at around four years of age. Thus, the current debate focuses on how infants can understand each other's minds in non-linguistic ways. The aim of this chapter is to present the current debate and the main controversies concerning the attribution of mental states in infancy from a philosophical perspective. To this end, in this chapter, we first present two philosophical theories that explain infants' representational abilities to understand the mental states of others: first, we present Apperly and Butterfil's theory (2009) and then Carruthers' theory (2016). Subsequently, we analyze the main differences and limitations of these theories in order to accurately characterize how infants understand the mental states

of others. Finally, we explore an alternative line of explanation, according to which infants understand mental states through diagrammatic representations, and we will determine how this perspective may affect the debate between the two theories presented.

## 1. Introducción

La capacidad para comprender que el comportamiento está guiado causalmente por estados mentales es fundamental para el éxito en las interacciones sociales a lo largo de la vida. Comprender los estados mentales implica la capacidad para atribuir dichos estados (por ejemplo, intenciones, creencias y emociones) a los demás para explicar su comportamiento (Gangopadhyay, 2017). A esta capacidad cognitiva se le conoce como “lectura de mentes” (Bialecka-Pikul et al., 2021) y ha sido objeto de estudio por parte de filósofos y psicólogos durante varias décadas (Wellman, 2017).

Los investigadores han prestado especial atención al desarrollo de esta capacidad en los niños y han utilizado el test de falsas creencias como una herramienta para evaluarla y comprender los mecanismos sociocognitivos involucrados (Rakoczy, 2021; Wimmer & Perner, 1983). La evidencia sobre la adquisición de esta capacidad en la infancia se encuentra mayormente en el campo de la psicología del desarrollo. Sin embargo, los hallazgos empíricos han sido bastante heterogéneos y aparentemente contradictorios. Mientras que los niños de 18 meses aprueban la versión implícita, no-verbal, del test de falsas creencias, los de 3 años tienen dificultades con la versión explícita-verbal, y solo hasta los 4 años logran aprobarla (Para una revisión, ver Baillargeon et al., 2010; Scott & Baillargeon, 2017; Sodian, 2011; Wellman et al., 2001). Por ello, tradicionalmente se consideraba que sólo a los 4 años los niños lograban desarrollar su capacidad de lectura de mentes. Antes de 2005, sólo se contaba con la versión verbal explícita del test de la falsa creencia, y sólo a partir de este año se inició el estudio de esta capacidad cognitiva usando la versión implícita.

En los test de falsa creencia verbal-explícita, se simula una situación donde hay dos personajes: Sally y Anne. El personaje principal,

Sally, guarda un objeto en un cajón y sale de la habitación. Durante su ausencia, Anne cambia la ubicación del objeto a un cajón contiguo. Posteriormente, se le pregunta al niño dónde cree que el personaje principal, Sally, buscará el objeto al regresar. Si el niño comprende que el personaje principal no tiene conocimiento del cambio y, por lo tanto, atribuye una creencia falsa sobre la ubicación real del objeto a Sally (creyendo que el objeto aún se encuentra en su lugar inicial), se considera que el niño tiene la capacidad de lectura de mentes (Egawa et al., 2022; Wellman et al., 2001).

Los hallazgos de este paradigma verbal-explicito han mostrado que a partir de los 4 años los niños aprueban este test (Wellman, 2014; Wellman et al., 2001; Wimmer & Perner, 1983), lo cual condujo a la interpretación de que la lectura de mentes emerge a esta edad, requiere del aprovechamiento del lenguaje y el uso de conceptos psicológicos para razonar sobre los estados mentales de los demás (Oktay-Gür et al., 2018; Rakoczy et al., 2015; Rakoczy, 2017; Ruffman, 2014).

Sin embargo, el estudio de Onishi y Baillargeon (2005) no solo contravirtió los hallazgos e interpretaciones anteriores, sino que también renovó el debate respecto al origen de esta capacidad. Este estudio modificó el test verbal de la falsa creencia y lo adaptó a una versión no-verbal e implícita, en la que los bebés no necesitan comprender ni proferir ninguna expresión verbal. Los investigadores diseñaron una tarea no-verbal, utilizando el método de violación de expectativas, con el propósito de examinar la capacidad de 56 bebés ( $M = 15$  meses) para predecir el comportamiento de un actor sobre la base de su creencia verdadera o falsa relativa al lugar donde se encuentra un objeto.

El método de violación de expectativas consiste en medir el tiempo de fijación de la mirada del bebé sobre un evento particular: si el evento viola la expectativa del bebé, este fijará su mirada en el evento por más tiempo. El resultado principal de este estudio es que los bebés fijan la mirada más tiempo cuando Sally busca el objeto en el lugar al que fue desplazado, lo que indica que el bebé esperaba que ella buscara el objeto en lugar donde originalmen-

te lo puso. En otras palabras, el bebé esperaba que Sally tuviera una falsa creencia sobre la ubicación del objeto. Al mostrar que los bebés lograban aprobar una versión modificada del test de falsa creencia con baja demanda ejecutiva y de formato no-verbal, este estudio demostró que bebés de 18 meses tienen una comprensión de los estados mentales de los demás (Onishi & Baillargeon, 2005; Scott & Baillargeon, 2017).

Adicionalmente, varios estudios han encontrado que los bebés entre los 15 y 18 meses anticipan el comportamiento de un agente de forma diferencial dependiendo de si el agente había percibido o no un cambio inesperado de un objeto, sugiriendo que los bebés tienen una comprensión implícita de los estados mentales del agente y distinguen entre sus propias percepciones y las de los demás (Burnside et al., 2020; Buttelmann & Kovács, 2019; Baillargeon et al., 2010, 2016; Onishi & Baillargeon, 2005; Siu & Cheung, 2019; Scott et al., 2010). Estos hallazgos se han interpretado a favor de la idea de que los bebés poseen una capacidad de lectura mental implícita que no requiere del lenguaje ni el uso de conceptos mentales, una capacidad que emerge temprano en la ontogenia y les permite reconocer que las acciones de los demás están orientadas intencionalmente (Baillargeon et al., 2016; Fiebich & Coltheart, 2015; Woodward, 2009).

Esta discrepancia en cuanto a la evidencia empírica al utilizar los test de falsa creencia implícita y explícita en bebés y niños se le conoce como el “rompecabezas del desarrollo” (en inglés *development puzzle*) y ha dado lugar a debates importantes en el campo de la filosofía de la mente sobre el origen y desarrollo de la lectura de mentes. En este sentido, el problema sobre la lectura de mentes en la infancia no es un asunto meramente empírico, sino también conceptual y conduce al debate filosófico sobre la naturaleza de la atribución de estados mentales y los mecanismos o sistemas cognitivos que la integran. Así, el debate actual se enfoca en explicar cómo los bebés pueden entender la mente del otro de una manera no-lingüística, con un interés particular por los mecanismos cognitivos involucrados.

El objetivo de este capítulo es presentar el debate actual y las controversias más relevantes sobre la atribución de estados mentales en la infancia desde una perspectiva filosófica. Dada la amplitud de la discusión filosófica y la diversidad de posturas sobre la atribución de estados mentales, es imposible hacer una reconstrucción completa de todas o casi todas las discusiones y posturas filosóficas que han participado en el debate a lo largo de más de las últimas cuatro décadas. Por tal motivo, en este capítulo se abordarán únicamente dos teorías filosóficas que explican las capacidades cognitivas de los bebés para comprender los estados mentales de los demás y se posicionan como posibilidades de explicación del rompecabezas del desarrollo. Estas teorías son la teoría de dos sistemas (en adelante T2S) (Apperly & Butterfill, 2009) y la teoría de un sistema (en adelante T1S) (Carruthers, 2016, 2017; Westra & Carruthers, 2018). Ambas se preguntan por la base cognitiva de la lectura de mentes y la manera cómo funciona en etapas tempranas del desarrollo. Tanto la T1S como la T2S intentan dar cuenta del rompecabezas del desarrollo, lo que conduce al debate de si son uno o dos sistemas cognitivos los encargados de la atribución de estados mentales en la infancia.

En la primera sección de este capítulo se abordará el debate general sobre la base cognitiva de la lectura de mentes y se presentarán los supuestos centrales de la T1S y la T2S para explicar el rompecabezas del desarrollo. En la segunda sección, se analizarán las principales limitaciones de estas teorías para caracterizar de manera precisa cómo los bebés comprenden los estados mentales de los demás, y los medios representacionales que utilizan para lograrlo. En la tercera sección, se brindará una alternativa teórica según la cual los bebés comprenden los estados mentales mediante un formato diagramático. Finalmente, se plantean los alcances de esta perspectiva y cómo puede incidir en el debate entre las dos teorías presentadas.



## 2. Debates sobre la base cognitiva de la lectura de mentes

El problema de la atribución de estados mentales no es un asunto reciente. Las teorías más tradicionales, conocidas como Teoría-Teoría (TT) (Para una revisión ver Hutto & Ravenscroft, 2021), Teoría de la Simulación (TS) (Para una revisión ver Barlassina & Gordon, 2017) y las versiones híbridas que integran tanto TT y TS (Currie & Ravenscroft, 2002), fueron las primeras en analizar la base cognitiva de esta capacidad y en proponer estrategias específicas por las que se atribuyen estados mentales. Según la TT, la atribución de estados mentales sería el resultado de la capacidad humana general para construir teorías para interpretar el comportamiento de los demás (Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997), mientras que para la TS, sería la capacidad para simular o imaginar los estados mentales de los demás sin requerir una teoría de lo mental (Gordon, 1986, 1995; Goldman, 2006).

En la actualidad, se han planteado otros relatos teóricos diferentes a la discusión TT y TS que intentan resolver el rompecabezas del desarrollo y ofrecen un marco de interpretación consistente con los datos en psicología del desarrollo sobre el rendimiento de los bebés en los test implícitos (no-verbales) de falsa creencia. Estos relatos actuales se pueden dividir entre los que ofrecen una solución negativa al rompecabezas y los que ofrecen una solución positiva.

Los relatos negativos son aquellos que parten de la evidencia de estudios que implementan los test de falsa creencia implícita para rechazar la interpretación de que los bebés tienen lectura mental. Esta suposición tiene dos vertientes. La primera es metodológica, en la cual se sostiene que los test de falsa creencia implícita evalúan algo diferente a la lectura de mentes (Dörrenberg et al., 2018). La segunda es filosófica, y versa sobre la tesis de que los bebés submentalizan en lugar de mentalizar (Heyes, 2014a, 2014b). Según esta interpretación, el desempeño de los bebés en los estudios psicológicos refleja capacidades cognitivas de bajo nivel derivadas de mecanismos de anticipación de la conducta (Ruffman & Perner, 2005), el funcionamiento de un esquema de dominio general

de detección de la acción (Burge, 2018) o el contagio de intereses (Falck et al., 2014). Este punto de vista resuelve el rompecabezas del desarrollo al considerar que no existe propiamente la capacidad de lectura de mentes en bebés. Por esa razón decimos que es una solución negativa.

Los relatos positivos son aquellos que parten de los datos del desarrollo para defender la tesis de que los bebés están dotados de una capacidad temprana de lectura de mentes, sea esta innata o adquirida, según el marco de interpretación adoptado. De allí que los relatos positivos se enfocan en caracterizar los mecanismos cognitivos que explican cómo los bebés tienen una comprensión implícita y no lingüística de los estados mentales de los demás (Apperly & Butterfill, 2009; Baillargeon et al., 2010, 2016; Carruthers, 2016; Gómez-Tabares, 2023; Kloo et al., 2020; Scott & Baillargeon, 2017; Sodian et al., 2020; Ting & Baillargeon, 2021).

Sin desconocer la importancia de los relatos negativos, este capítulo se enfoca exclusivamente en analizar los relatos positivos. Aunque existen muchas propuestas positivas, en la literatura filosófica reciente se destacan la T1S y la T2S. La T2S plantea que existen dos sistemas cognitivos de lectura de mentes, uno que es rápido y automático que opera en etapas tempranas del desarrollo y otro que es reflexivo y controlado que emerge tardíamente y está asociado con el lenguaje y el razonamiento (Apperly & Butterfill, 2009). En contraste, la T1S defiende la existencia de un único sistema modular que experimenta un enriquecimiento gradual durante el desarrollo y alcanza capacidades de mayor sofisticación conceptual para la atribución de estados mentales (Carruthers, 2016). Dado que ambas teorías pretenden resolver el rompecabezas del desarrollo, el debate es si la lectura de mentes en la infancia es el resultado de uno o dos sistemas cognitivos.

A continuación, se presentan ambas propuestas, primero la T2S y luego la T1S, que emergió como rechazo a la postura de los dos sistemas.

## 2.1 Teoría de dos sistemas (T2S)

Apperly y Butterfill (2009, 2013a, 2013b) analizaron las diferencias en el procesamiento cognitivo de los niños en tareas implícitas y explícitas de falsa creencia, con un interés específico en desarrollar una teoría computacional en el sentido de Marr (1982) para explicar la lectura de mentes de una manera cognitivamente eficiente, sin requerir, en todos los casos, la sofisticación conceptual.

Los autores propusieron que los humanos cuentan con dos sistemas cognitivos de lectura de mentes, el sistema 1 o Teoría Mínima de la Mente (en inglés *Minimal Theory of Mind Cognition*) y el sistema 2 o Teoría Completa de la Mente (en inglés *Full-blown Theory of Mind Cognition*) (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013b). Esta propuesta deriva del enfoque en psicología de los procesos duales que hacen una distinción entre dos tipos de procesos psicológicos que funcionan a partir de sistemas o mecanismos psicológicos distintos, que son denominados Sistema 1 (en adelante S1) y Sistema 2 (en adelante S2) (Frith & Frith, 2008; Kahneman, 2011). Apperly y Butterfill (2009) aplicaron la teoría de los procesos duales para analizar la lectura de mentes en la infancia, y propusieron que el S1 y el S2 son diferentes en cuanto a la forma de procesar los estímulos sociales y los alcances que tienen en diferentes actividades cognitivas de comprensión de estados mentales.

El S1 es responsable de los procesos implícitos, que son procedimientos cognitivos rápidos, automáticos e inflexibles. Las operaciones de S1 permiten procesar estímulos sociales rápidos que son indicativos de estados psicológicos y utilizan señales socio-perceptivas como expresiones faciales, dirección de la mirada y movimientos corporales para rastrear estados similares a las creencias (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013b; Meinhardt-Injac et al., 2018). La capacidad de los bebés para identificar señales socio-perceptivas es relevante porque les permite tener una comprensión intencional a partir del foco perceptivo o atencional de la otra persona (Woodward, 2009).

El S1 emerge temprano en la ontogenia y está encapsulado. Es decir, el S1 está restringido para participar en la operación del S2

que corresponden a actividades cognitivas más complejas, vinculadas al razonamiento de creencias y que usualmente requieren del lenguaje. Otro rasgo del S1 derivado de su encapsulamiento es su imposibilidad de reclutar otros procesos cognitivos para su funcionamiento (p. ej., las funciones ejecutivas) (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013b).

Por el contrario, el S2 es responsable de los procesos explícitos de lectura de mentes, que son procedimientos cognitivos lentos, exigentes, reflexivos y flexibles. Por tanto, el S2 funciona para la lectura de mentes “en toda regla”, y sus operaciones se basan en el razonamiento psicológico para realizar inferencias, explicaciones y predicciones sobre los estados mentales de los demás (Apperly & Butterfill, 2009). Este sistema presenta un desarrollo tardío que es paralelo y depende del desarrollo del lenguaje y la maduración del funcionamiento ejecutivo en la infancia. Además, el S2 no presenta las limitaciones de procesamiento del S1. Las características del S1 y el S2 se detallan en la tabla 1.

Tabla 1.

| <i>Características del S1 y el S2 de lectura de mentes</i>                                                    |                                                                                                           |
|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| <b>Sistema 1</b>                                                                                              | <b>Sistema 2</b>                                                                                          |
| <b>Teoría de la mente mínima (Minimal ToM Cognition).</b>                                                     | <b>Teoría de la mente completa (Full-blown ToM Cognition).</b>                                            |
| <b><i>Características generales</i></b>                                                                       |                                                                                                           |
| Procesos rápidos, eficientes, inflexibles y automáticos, con gran capacidad de procesamiento y poco esfuerzo. | Procesos lentos, poco eficientes, flexibles, lentos y controlados, de capacidad limitada y gran esfuerzo. |
| <b><i>Propiedades particulares</i></b>                                                                        |                                                                                                           |
| Evolutivamente antiguo                                                                                        | Evolutivamente reciente                                                                                   |
| Compartido con los animales                                                                                   | Distintivamente humano                                                                                    |
| Independiente del lenguaje                                                                                    | Asociado al lenguaje                                                                                      |
| Desarrollo temprano                                                                                           | Desarrollo tardío                                                                                         |
| Asociativo                                                                                                    | Basado en reglas                                                                                          |
| Holístico, perceptivo                                                                                         | Analítico, reflexivo                                                                                      |
| Cognición modular                                                                                             | Inteligencia fluida/flexible                                                                              |
| Pragmático                                                                                                    | Lógico                                                                                                    |
| Independiente de la capacidad cognitiva                                                                       | Dependiente de la capacidad cognitiva                                                                     |
| Subpersonal                                                                                                   | Personal                                                                                                  |
| Independiente de la memoria de trabajo y la función ejecutiva                                                 | Limitado por la capacidad de la memoria de trabajo y la función ejecutiva                                 |

Fuente: Tomado de Gómez-Tabares, 2022, p. 336

Según los autores que defienden la T2S, esta teoría resuelve el rompecabezas del desarrollo al mostrar que el desempeño de bebés y niños en los test implícitos (no-verbales) y explícitos (verbales) de falsas creencias es el resultado del funcionamiento de dos sistemas cognitivos que maduran y funcionan de manera distinta durante el desarrollo infantil. Así, la T2S aporta una solución positiva al explicar las capacidades tempranas de los bebés para tener una comprensión intencional implícita del comportamiento del agente y su disociación con las capacidades de comprensión explícita de estados mentales que requieren del lenguaje y mayor sofisticación conceptual (Poulin-Dubois et al., 2020).

## **2.2. Teoría de un sistema (T1S)**

Carruthers (2013, 2016, 2017) rechaza la tesis de que los procesos cognitivos rápidos, eficientes y automáticos, por un lado, y los procesos lentos, reflexivos y controlados, por el otro, sean el resultado del funcionamiento de dos sistemas como lo plantea la T2S. Esto no significa negar que en ocasiones se llevan a cabo operaciones cognitivas rápidas y eficientes y en otras ocasiones lentas y reflexivas para comprender los estados mentales de los demás, pues la evidencia señala que hay diferentes medios y estrategias atributivas para comprender los estados mentales. La crítica radica en que no se requiere una teoría que postule la existencia de dos sistemas para explicar el desarrollo y funcionamiento de la lectura de mentes, y, por tanto, es innecesario un modelo dual para explicar el rompecabezas.

La T1S propone una teoría más parsimoniosa, según la cual la capacidad de lectura de mentes es el resultado del funcionamiento de un único sistema que experimenta un enriquecimiento gradual en la infancia (Carruthers, 2016). Se trata de un sistema modular (Fodor, 1983) de dominio específico que opera dentro de la arquitectura cognitiva para comprender los estados mentales de los demás (Carruthers, 2013; Baillargeon et al., 2010, 2016). Este sistema opera desde la infancia hasta la adultez, pero sus operaciones cognitivas varían durante el desarrollo humano debido a las experiencias sociales compartidas y la maduración de otras funciones

cognitivas durante los primeros años de vida, lo cual incide directamente en el enriquecimiento conceptual del sistema de lectura de mentes.

Así, la transición de la lectura de mentes implícita a la explícita que experimentan los niños durante sus primeros cinco años de vida es el resultado de un proceso de enriqueciendo conceptual progresivo del sistema, mas no del funcionamiento de dos sistemas disociados (Carruthers, 2013, 2016, 2017; Westra, 2017). Este proceso de enriquecimiento está conectado con la exposición temprana del bebé a experiencias sociales compartidas y prácticas de intencionalidad compartida como lo son la alta disponibilidad emocional y el uso del habla mentalista de la madre (como principal cuidadora) (Sodian et al., 2020). En este proceso también juegan un rol fundamental la maduración gradual del lenguaje, de la función ejecutiva y de la memoria de trabajo que se dan en los primeros años de vida. Todos estos procesos psicológicos son reclutados por el sistema de lectura de mentes para que sus operaciones alcancen progresivamente una mayor sofisticación conceptual en la comprensión de las creencias de los demás (Carruthers, 2016, 2017; Kloos et al., 2020; Sodian et al., 2020; Westra & Carruthers, 2018).

La influencia de estos factores sociales y cognitivos explicaría tanto los cambios progresivos del sistema de lectura de mentes como la razón de por qué los bebés y niños exhiben diferentes grados de éxito para comprender los estados mentales de los demás cuando se someten a los test implícitos y explícitos de falsa creencia, lo cuales presentan demandas disímiles en cuanto a la función ejecutiva, la memoria de trabajo y el uso del lenguaje. Así, la T1S resuelve el rompecabezas del desarrollo al proponer que los niños experimentan un enriquecimiento conceptual progresivo que explica la aparente disociación en su rendimiento en los test de falsa creencia (Carruthers, 2013, 2016, 2017).

Recientemente se pusieron a prueba las predicciones de la T1S y la T2S a partir de una revisión sistemática de estudios longitudinales y se analizó si el rendimiento en las tareas implícitas es un precursor cognitivo de continuidad con el rendimiento en las ta-

reas explícitas en diferentes momentos del desarrollo (Gómez-Ta-  
bares, 2023). Se incluyeron estudios que evaluaran los cambios en  
la capacidad de lectura de mentes en varios periodos del desarro-  
llo en una misma población. Se encontró un patrón consistente de  
continuidad de las capacidades mentalistas en la infancia que inicia  
con el reconocimiento de acciones intencionales a los 6 meses, la  
sincronía del foco atencional con un agente a los 12 meses, la toma  
de perspectiva visual a los 15 meses, la anticipación de la mirada a  
los 18 meses, hasta alcanzar un razonamiento explícito de los esta-  
dos mentales de los demás a los 48, 50, 60 y 70 meses (Gómez-Ta-  
bares, 2023).

Estos hallazgos presentados en la revisión sistemática son con-  
sistentes con las predicciones de la T1S al señalar que los bebés  
disponen de un sistema de lectura de mentes que experimenta un  
enriquecimiento gradual a lo largo de la primera infancia. Este en-  
riquecimiento implica cambios cualitativos y cuantitativos en las  
capacidades mentalistas en la infancia, las cuales van alcanzando  
una mayor sofisticación conceptual durante el desarrollo humano,  
hasta llegar a una etapa madura para realizar inferencias y predic-  
ciones de comportamiento y estados mentales.

### **3. Limitaciones de la T1S y la T2S para caracterizar la estructura de las representacionales de la lectura de mentes en bebés**

Tanto la T1S como la T2S son alternativas teóricas consistentes  
con la evidencia para explicar la lectura de mentes en bebés. Si la  
lectura de mentes funciona a partir de un sistema modular o dos  
sistemas cognitivos es un debate actual que persiste en el campo  
de la filosofía y las ciencias cognitivas. Sin embargo, ambas teorías  
comparten dos aspectos. El primero es que resuelven el rompecabezas  
del desarrollo a partir de un relato positivo, y, segundo, no  
se comprometen con la idea de que la comprensión intencional en  
bebés requiera de la posesión de conceptos mentales o capacidades  
de meta-representación.

Este segundo aspecto deriva en un problema fundamental para comprender el funcionamiento y naturaleza de la lectura de mentes en la infancia, el cual versa sobre la identificación del medio o mecanismo cognitivo a través del cual los bebés representan algún aspecto de las intenciones, deseos o creencias de los demás de una manera no-lingüística. En el caso de los niños de cuatro años que aprueban el test explícito de falsa creencia, el medio representacional son los conceptos lingüísticos que les permite razonar sobre las creencias de los demás. Sin embargo, en el caso de los bebés que aprueban la versión implícita (no-verbal) del test de falsa creencia no es tan claro cuál es el medio representacional o mecanismo que explica la comprensión intencional de los bebés, dada las limitaciones conceptuales y lingüísticas que presentan. Así, tanto la T1S como la T2S tienen el reto de caracterizar los medios o vehículos representacionales de la capacidad de lectura de mentes implícita en bebés. A continuación, se analiza este problema desde ambas teorías, primero la T2S y luego la TS1.

De acuerdo con la T2S (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013b), el vehículo representacional<sup>3</sup> de la lectura de mentes es diferente para las operaciones del S1 y el S2. En el S2 es la capacidad humana de razonar a partir de conceptos psicológicos para explicar y predecir el comportamiento de los demás. Por esta razón, las operaciones que realiza el S2 son lentas, exigentes y reflexivas (Apperly & Butterfill, 2009). En cuanto al S1, el medio representacional no parece tan claramente definido. En principio, la T2S plantea que las operaciones del S1 facilitan la comprensión de información social relevante a partir de patrones perceptibles

---

3 Aquí es fundamental explicar la diferencia entre contenido, representación y vehículos representacionales. El contenido se refiere a la información que una representación expresa, es decir, lo que describe sobre el mundo o la realidad. La representación mental es la estructura cognitiva que porta ese contenido dentro de la mente, como una imagen mental, un mapa cognitivo o una proposición. Los vehículos representacionales, por su parte, son los formatos o mecanismos a través de los cuales estas representaciones se expresan, como puede ser el lenguaje, diagramas o mapas mentales, o incluso las interacciones neuronales. Estos formatos influyen en cómo el contenido es procesado e interpretado (Dretske, 1995). Esto se ampliará en la sección 3.1.



en los demás (expresiones faciales, voz y movimientos corporales) y permiten que los bebés puedan identificar las intenciones que acompaña el comportamiento de los agentes. Estas operaciones del S1 no son conceptuales y se vinculan más con procesos atencionales y perceptivos que al razonamiento basado en conceptos<sup>4</sup> (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013b).

Al respecto, Apperly y Butterfill (2009), Butterfill y Apperly (2013b) propusieron dos funciones centrales del S1, las cuales son el encuentro (en inglés *Encountering*) y el registro (en inglés *Registering*). El concepto de encuentro denota una relación entre un agente y un objeto, basándose en indicadores sociales observables. El registro, por otro lado, implica una relación entre la intención del agente, el objeto y un contexto particular, lo que permite que el bebé reconozca que el comportamiento del agente está dirigido intencionalmente a un objetivo. La función del registro en las operaciones del S1 es rastrear estados mentales similares a las creencias (Butterfill & Apperly, 2013b). Así, el S1 no atribuye estados mentales con contenido conceptual, sino registros, que Apperly y Butterfill (2009) definen como estados mentales relacionales.

Para Butterfill y Apperly (2013b) el registro es un estado agencial en la medida que desempeña un papel funcional similar al de la creencia. La creencia vincula causalmente el comportamiento del agente con algún estado mental o epistémico, y el registro vincula causalmente el encuentro (vínculo perceptivo del agente con el objeto) con las acciones intencionales del agente dirigidas a un

---

4 Aquí es necesario distinguir entre contenido conceptual y contenido no-conceptual. El contenido conceptual involucra representaciones que dependen de conceptos, los cuales son estructuras mentales que categorizan el mundo. El contenido no conceptual, en cambio, capta información sin la intervención de conceptos abstractos o clasificaciones. Por ejemplo, los bebés o algunos animales pueden seguir el comportamiento de los objetos sin recurrir a conceptos como “creencia” o “deseo”. Dretske (1981) distingue entre información análoga y digital para explicar esta diferencia: mientras que el contenido análogo ofrece información continua y detallada (como una percepción visual), el contenido digital es discreto y categórico (como una descripción textual). En este sentido, el contenido no conceptual proporciona acceso directo a la percepción, mientras que el contenido conceptual implica una interpretación abstracta y categorizada del mundo.

objetivo. Sin embargo, los registros difieren de la noción canónica de creencia en varios aspectos: su naturaleza es no-conceptual y no requieren de recursos cognitivos sofisticados. En consecuencia, el concepto de registro puede funcionar como un medio cognitivo de rastreo de las intenciones de un agente que son indicativos de un estado epistémico (conocimiento o ignorancia) sobre la ubicación o identidad de un objeto (Herschbach, 2015; Jacob, 2019a, 2019b), sin necesidad de involucrar, en principio, conceptos mentales (Butterfill & Apperly, 2013a, 2013b).

La solución planteada por T2S es novedosa e interesante. Sin embargo, tiene un problema en su núcleo explicativo que presentamos a continuación. La T2S plantea que los registros son representaciones no-conceptuales que rastrean estados epistémicos como las creencias o el conocimiento. Sin embargo, dado que el medio representacional de la lectura de mentes implícita no es conceptual, no es claro cuál es el contenido de dichos registros o cómo éste se articula para orientar la acción del bebé. Por tanto, el concepto de registro que caracteriza las operaciones de S1 es aún muy oscuro y no brinda una caracterización satisfactoria de la estructura cognitiva de los registros o de cómo los bebés logran captar los estados mentales de los demás sin involucrar una estructura predicativa compuesta por conceptos.

Por otra parte, la T1S señala un sistema único que funciona de manera progresiva y ascendente. El rasgo central de este sistema es su capacidad para enriquecerse conceptualmente durante el desarrollo humano (Carruthers, 2016, 2017; Westra & Carruthers, 2018). Carruthers (2016, 2017) propone que las capacidades más básicas de cognición social que exhiben los bebés operan con un conjunto de conceptos primitivos, los cuales son unidades básicas de representación que están predispuestas para facilitar la interacción con agentes y objetos, y que posteriormente se derivan en capacidades conceptualmente enriquecidas (Carruthers, 2016, 2017; Westra & Carruthers, 2018). Estos conceptos primitivos se pueden entender como representaciones preconceptuales que permiten rastrear formas abstractas de comportamiento, por ejemplo, los aspectos intencionales

que orientan la acción de un agente en tareas de búsqueda o identidad de un objeto oculto. En este sentido, los bebés utilizan estos preconceptos para comprender que el comportamiento del agente está dirigido intencionalmente a objetivos.

Estas configuraciones preconceptuales están relacionadas con la exposición social y se utilizan para crear patrones de asociación entre el comportamiento observable y las intenciones que motivan la acción de los agentes. Por ejemplo, incluyen asociaciones entre cambios en la ubicación de los objetos, señales sociales simples como expresiones visuales, faciales o corporales, y una gama limitada de repertorios de acciones intencionales. Así, cuando los bebés anticipan el comportamiento de un agente en una tarea familiar por habituación, activan sólo las representaciones preconceptuales relevantes de la estructura composicional más completa. Esto les permite formarse expectativas sobre las creencias del agente, ya sean verdaderas o falsas, y reaccionar en consecuencia (Carruthers, 2016, 2017).

Desde este punto de vista, los bebés estarían dotados de capacidades preconceptuales que según el momento del desarrollo son rudimentarias o limitadas. Estas capacidades son innatas y se enriquecen progresivamente. En este sentido, la progresión de la lectura de mentes implícitas a la explícita no se da por un cambio conceptual radical en cuanto a los medios representacionales, sino por el enriquecimiento conceptual de los preconceptos o conceptos primitivos (Carruthers, 2013, 2016).

Esta postura es cercana a la perspectiva nativista del razonamiento psicológico (Baillargeon et al., 2016; Carruthers, 2013; Ting et al., 2021). El nativismo parte de la idea de que la lectura de mentes es innata (Litchfield & Lambert, 2011), y, por tanto, los bebés nacen dotados de preconceptos que operan como los medios representacionales para comprender a los demás como dotados de estados mentales, y se van enriqueciendo conceptualmente hasta lograr formas más sofisticadas de razonamiento psicológico (Carruthers, 2016; Kushnir & Gopnik, 2005; Schaafsma et al. 2015; Leslie et al., 2004).

La idea de que los bebés nacen con conceptos primitivos que se enriquecen durante el desarrollo hasta alcanzar representaciones conceptuales en toda regla es una solución novedosa. Sin embargo, no es suficientemente clara la estructura cognitiva de los conceptos primitivos ni de cómo estos representan en la mente de los bebés para captar los estados mentales de los demás.

### **3.1. Un problema filosófico emergente**

Tanto la T1S como la T2S no se comprometen completamente con un marco explicativo que implique meta-representaciones o el uso de conceptos en toda regla por parte de los bebés para explicar la lectura de mentes. Por el contrario, proponen medios cognitivos mucho más rudimentarios. Sin embargo, el problema fundamental de ambas teorías es que no es claro qué es lo que representan los bebés cuando implementan los registros o los preconceptos para captar los estados mentales de los demás sin involucrar una estructura predicativa compuesta por conceptos.

Los registros y los preconceptos son medios representacionales no-conceptuales que rastrean estados epistémicos como las creencias o el conocimiento. Sin embargo, no es claro cuál es el contenido de estos cuando el medio representacional de la lectura de mentes implícita es no-conceptual. Por tanto, los registros y los preconceptos no parecen aclarar lo que pretenden explicar. Si el medio representacional no es conceptual en toda regla, entonces, ¿cómo es el contenido representacional de los bebés al implementar medios representacionales como los registros o los preconceptos para comprender los estados mentales de los demás?

## **4. La lectura de mentes en bebés como configuraciones diagramáticas**

Dadas las limitaciones explicativas que hemos señalado de las teorías revisadas hasta ahora (T1S y T2S), proponemos que los preconceptos en la T1S y los registros del S1 de la T2S son configuraciones cognitivas similares a los mapas mentales, cuyo medio representacional es diagramático. Los bebés procesan la acción social obser-

vale de los otros agentes en patrones diagramáticos que indican intenciones, sin necesidad de usar conceptos o lenguaje. Estos patrones configuran mapas cognitivos que guían las respuestas de los bebés en los test implícitos de falsa creencia en cuanto al procesamiento de la información social para identificar similitudes, patrones y relaciones entre agentes, objetos, sus propiedades y el contexto. Más allá del ámbito de la psicología experimental, es probable que los bebés utilicen estos mapas mentales en sus primeras etapas del desarrollo para interactuar con familiares y el entorno.

#### **4.1. Representaciones diagramáticas y mapas cognitivos**

Antes de desarrollar la idea central de este texto, es fundamental considerar que, al atribuir estados representacionales a otros, representamos que éstos tienen representaciones que expresan un contenido evaluable en términos de precisión. Cualquier representación que exprese algo evaluable en términos de precisión lo hace en algún formato representacional. Por formato representacional nos referimos a algo como una oración, un diagrama, un mapa o una imagen. El mismo contenido puede ser expresado en diferentes formatos. Por ejemplo, “la bola está en la caja de la izquierda” expresa que la bola está en la caja de la izquierda, pero una imagen de un par de cajas y la bola también podría expresar esto. En este sentido, cada formato es un medio representacional que determina la manera cómo operan las capacidades cognitivas en cuanto al procesamiento de la información (Aguilera, 2020, 2021; Beck, 2013; Camp, 2007; Heck, 2007; Velázquez Coccia, 2018). Aquí consideramos que las diferencias entre estos dos tipos de representaciones son diferencias en el formato representacional y no del contenido expresado por ambas representaciones.

En la literatura se destacan dos formatos representacionales: cartográficos y lingüísticos (Aguilera, 2020, 2021; Camp, 2007). Las representaciones cartográficas emplean mapas, imágenes, símbolos y diagramas mentales para abstraer información sobre las propiedades físicas de los objetos con el fin de representar relacio-

nes de conectividad, continuidad y convergencia entre ellos para la toma de decisiones. En contraste, las representaciones lingüísticas emplean una estructura conceptual, lo que permite que el pensamiento se estructure a partir de códigos semánticos (Aguilera, 2021). Esto es crucial, ya que los diferentes formatos representacionales poseen características distintivas que determinan su idoneidad para expresar ciertos contenidos, siendo algunos más adecuados que otros. Además, algunos formatos tienen una capacidad expresiva mayor que otros.

De esta manera, podemos comprender que el pensamiento de una criatura se adapta a un formato representacional específico. Este formato, a su vez, tiene la capacidad de expresar ciertos contenidos, pero no otros. Además, la criatura puede manipular estos contenidos de formas particulares según su nivel de desarrollo y sus capacidades cognitivas. Por ejemplo, una criatura que piensa exclusivamente en imágenes solo podría contemplar contenidos que puedan ser expresados visualmente, lo que resultaría en una capacidad de inferencia extremadamente limitada debido a la ausencia de equivalentes imaginísticos para la mayoría de los operadores lógicos (Camp, 2009, p. 111).

El uso de diferentes principios combinatorios resulta en diferencias distintivas entre estos formatos. En algunos aspectos, los mapas básicos son más eficientes y accesibles que el lenguaje. Primero, son densos en información, lo que dificulta expresar el contenido de un mapa básico en forma de oraciones, ya que requeriría un conjunto extenso y complejo de ellas. Segundo, muestran claramente sus implicaciones; por ejemplo, si un mapa de Colombia indica que Medellín está al norte de Bogotá y al occidente de Bucaramanga, también representará explícitamente que Bogotá está al sur de Bucaramanga y que Bucaramanga está al oriente de Medellín, entre otras relaciones. Esta información es cognitivamente transparente, ya que se encuentra explícitamente representada en el mapa. En contraste, la frase “Medellín está al norte de Bogotá y al occidente de Bucaramanga” implica toda esta información, pero no la representa explícitamente; aquellos que utilicen esta

frase deben hacer un esfuerzo para deducir el resto. Finalmente, como resultado de lo anterior, realizar cambios en un mapa básico es simple: mover cualquier icono en el mapa automáticamente actualiza todas las relaciones representadas entre los objetos, manteniendo la coherencia del mapa sin necesidad de esfuerzo adicional (Camp 2007, pp. 160–162).

Decir que la representación del bebé es diagramática no implica que existan, literalmente, diagramas en su cerebro. Más bien, sugiere que sus pensamientos siguen patrones funcionales específicos que hacen uso de un formato de representación particular. En este contexto de análisis sobre la lectura de mentes en bebés, los diagramas representan una forma primitiva de pensamiento eficiente de la información que permite la representación de intenciones o creencias sin la necesidad de una representación conceptual explícita. La eficiencia radica en la inherente simplicidad de una representación de formato diagramático, lo que posibilita que tanto bebés como animales no humanos codifiquen información de una manera intuitiva y visual (Boyle, 2019). Esta simplicidad facilita el procesamiento eficiente de los estímulos del entorno y permite responder rápidamente a estos. Sin embargo, estas representaciones tienen las limitaciones mencionadas previamente; en particular, la ausencia de los rasgos característicos que constituyen los conceptos.

#### **4.2. Diagramas y lectura de mentes implícita**

La idea que exploramos en este texto es que los bebés utilizan un formato representacional diagramático para construir una imagen mental sobre las características observables del comportamiento de los agentes y la manera cómo reaccionan al interactuar con el entorno y los objetos. Ahora bien, un aspecto que se debe aclarar es cómo los mapas cognitivos que forman los bebés a partir de sus representaciones diagramáticas capturan los rasgos no perceptibles del comportamiento del agente, como lo son los estados epistémicos de ignorancia o el conocimiento, la comprensión de intenciones y creencias. Nuestra respuesta es que las representaciones

diagramáticas desempeñan funciones tanto referenciales como atributivas, sin necesidad de vincular capacidades conceptuales, y les permiten a los bebés tener una comprensión intencional del comportamiento del agente.

Por un lado, la función referencial se encarga de registrar la información de los estímulos perceptibles, lo cual incluye el comportamiento de los agentes (acciones, expresiones, movimientos, entre otros), su interacción con los objetos y las propiedades físicas de los objetos (color, forma, tamaño) para formar un mapa mental en el que los aspectos percibidos se representan diagramáticamente. Por tanto, la representación diagramática depende de los registros sensoriales del bebé.

Por otro lado, la función atributiva se encarga de reclutar, clasificar, caracterizar y combinar la información referencial para generar nueva información sobre atributos o propiedades no perceptibles de los agentes (p.ej. creencias, deseos, conocimiento, ignorancia). La función atributiva juega un papel crucial. Gracias a ella, los bebés pueden registrar que un agente mantiene una creencia falsa con respecto a la ubicación de un objeto, aun cuando ellos mismos han registrado dónde ha sido movido y dónde se encuentra realmente el objeto.

Por tanto, la comprensión intencional del bebé no se basa exclusivamente en la información referencial, es decir, en lo que los bebés pueden ver o percibir directamente. A partir de la función atributiva de las representaciones diagramáticas, los bebés captan y procesan aspectos no observables relacionados con el estado mental del agente, lo que implica una actividad cognitiva que integra y sintetiza tanto la información referencial como la atributiva para generar respuestas rápidas y eficientes que le permitan al bebé anticiparse a las acciones del agente al considerar sus estados epistémicos (p.ej. intenciones, creencias, conocimiento o ignorancia)<sup>5</sup>.

---

5 La función atributiva se encarga de combinar información perceptual para generar inferencias sobre propiedades no perceptibles, como creencias, deseos o intenciones. Un tema de debate es si esta función implica necesariamente el uso de conceptos. Clasificar algo podría implicar que ya se ha individuado una clase abstracta, lo cual



En este sentido, las representaciones atributivas sirven como mecanismos eficientes que simplifican y facilitan la toma de decisiones implícitas en los bebés. Este tipo de representaciones se pueden determinar a partir del seguimiento de la dirección y permanencia de la mirada del bebé (es decir, usando el método de violación de expectativas). Dada la complejidad inherente al procesamiento asociativo de múltiples elementos referenciales, la función atributiva resulta esencial para permitir el procesamiento predictivo y orientar la mirada en consonancia con el comportamiento del agente observado.

Sin embargo, las capacidades representacionales de los bebés son limitadas porque dependen de la experiencia previa, lo que restringe su comprensión a las situaciones específicas que han sido experimentadas. En otras palabras, las representaciones diagramáticas atributivas se forman a partir de los registros referenciales, limitando así la posibilidad de generalización y podrían inducir respuestas incorrectas en los bebés cuando se enfrentan a escenarios o agentes que tienen características muy diferentes a las que dieron origen a las representaciones referenciales. Por ejemplo, en un escenario hipotético, un bebé juega con su madre y un juguete de peluche. La madre oculta el juguete detrás de su espalda y el bebé busca el objeto al dirigir su mirada detrás de ella. En este escenario, el bebé construye una representación mental diagramática del objeto (el juguete) y del agente (la madre), apoyándose en los patrones de comportamiento observables (la madre ocultando el juguete). Esto le permite formar una representación mental de la intención de la madre de ocultar el juguete. Así, en los encuentros de interacción social y atención conjunta se configuran las representaciones referenciales, las cuales son necesarias para que los bebés tengan representaciones atributivas.

---

parecería requerir conceptos. Sin embargo, es posible que esta clasificación no requiera de una individuación conceptual explícita, sino que se base en un procesamiento perceptual. De esta manera, los bebés podrían “clasificar” comportamientos sin recurrir a representaciones conceptuales plenamente desarrolladas.

Es posible considerar situaciones cotidianas en las que los procesos triádicos de atención conjunta (bebés, cuidadores y objetos) son esenciales para la formación de representaciones diagramáticas. Los bebés al interactuar con su madre u otro agente de cuidado se someten a una variedad de estímulos, entre ellos, el juego dirigido y la interacción con objetos, e involucran procesos de sincronización del foco atencional, el uso del habla mentalista y la expresión emocional de la madre. Estos procesos socioemocionales son modelados por las prácticas sociales cotidianas y moldean la capacidad cognitiva del bebé para mapear los estados mentales de los demás. Son justamente estos procesos socioemocionales los que le dan forma a las representaciones diagramáticas atributivas en los bebés.

Esta interpretación se alinea con varios estudios que han mostrado que el uso del lenguaje que hace referencia a estados mentales y la respuesta emocional-sensible, afectuosa y empática, por parte de la madre hacia las necesidades del bebé en los procesos intencionalidad compartida promueven la capacidad del bebé para comprender la acción dirigida a objetivos y sincronizar el foco atencional con un agente, adoptar la perspectiva social, visual y epistémica para comprender y predecir el curso de la acción intencional de los demás (Bialecka-Pikul & Bialek, 2021; Kristen-Antonow et al., 2018; Licata et al., 2016; Sodian et al., 2020). Así, estos escenarios de enriquecimiento social moldean la formación de representaciones diagramáticas en los bebés para comprender los estados mentales de los demás.

De acuerdo con lo planteado, las representaciones diagramáticas referenciales no contienen propiamente información acerca de los estados mentales de los demás, y se limitan al registro de las características o rasgos que el bebé logra percibir del contexto, los objetos y los agentes. Estos registros son reclutados para construir representaciones atributivas que se encargan de extraer información acerca del estado epistémico del agente ante una situación particular. Esto puede ser su creencia falsa o verdadera sobre la ubicación del objeto o la intención de buscar el objeto en el lugar incorrecto según su creencia falsa. En este sentido, los mapas mentales son vehículos re-

presentacionales que tienen un papel esencial en la lectura de mentes implícita en bebés (Velázquez Coccia, 2018).

### 4.3. Lectura de mentes diagramática

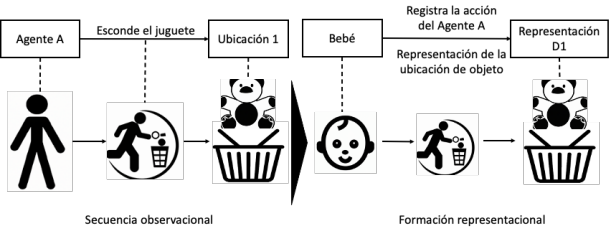
Para analizar más detalladamente la manera como operan las representaciones diagramáticas y su relación con la lectura de mentes implícita, consideremos el siguiente escenario experimental inspirado en el paradigma de violación de expectativas (Onishi & Baillargeon, 2005).

En este escenario, un bebé ( $M^{edad}= 18$  meses) observa a un agente (Agente A) ocultando un juguete en una ubicación (Ubicación 1). Posteriormente, en ausencia del Agente A, el bebé ve a otro agente (Agente C) mover el juguete a una ubicación diferente (Ubicación 2). Cuando el Agente A regresa y busca el juguete en la Ubicación 1, los investigadores miden el tiempo que el bebé mira al agente A. Si el bebé entiende que el agente A tiene una creencia falsa (pensando que el juguete todavía está en la Ubicación 1), entonces no debería sorprenderse y, por lo tanto, no debería mirar más tiempo cuando el Agente A busca en la Ubicación 1. Sin embargo, si el Agente A busca en la Ubicación 2, esto violaría las expectativas del bebé y, por lo tanto, debería llevar a los bebés a mirar más tiempo. Este patrón de mirada más prolongado en respuesta a la búsqueda en la Ubicación 2 se interpreta como evidencia de que el bebé ha formado expectativas sobre dónde debería buscar el Agente A, basándose en su comprensión de la creencia del Agente A sobre la ubicación del juguete.

El mapa cognitivo que el bebé podría haber construido en esta situación podría desglosarse en las siguientes representaciones diagramáticas: El bebé percibe que el Agente A esconde el juguete en la Ubicación 1, lo que da lugar a la representación diagramática referencial 1 (D1) (ver Figura 1). En la figura 1, el Agente A esconde el juguete en la Ubicación 1. El bebé registra esto y forma una representación diagramática del agente A llevando a cabo la acción de ubicar el juguete en la Ubicación 1.

**Figura 1.**

Primera representación diagramática referencial (D1) del bebé de la acción del Agente A y la ubicación del objeto.

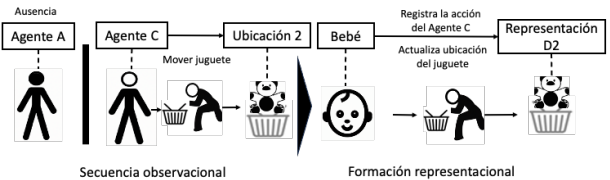


Nota: El Agente A esconde el juguete en la Ubicación 1. El bebé registra esto y forma una representación diagramática del agente A llevando a cabo la acción de ubicar el juguete en la Ubicación 1.

A continuación, en ausencia del Agente A, el Agente C mueve el juguete a la Ubicación 2. Esta secuencia observacional lleva a la formación de la representación D2, en la que se registra la acción del Agente C en ausencia del Agente A y se actualiza la representación de la ubicación del juguete (Ver figura 2).

**Figura 2.**

Segunda representación diagramática referencial (D2) del bebé de la acción de C y el cambio de ubicación del objeto.



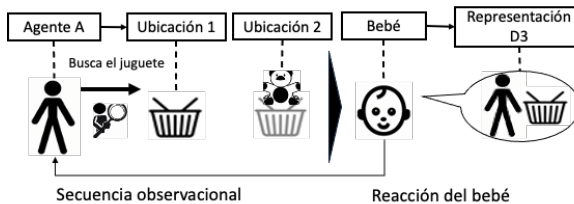
Nota: En ausencia del Agente A, el Agente C mueve el juguete a la Ubicación 2. El bebé registra esto y actualiza su representación diagramática de que el juguete ahora está en la Ubicación 2. Esto implica, por tanto, un cambio del ícono que representa el juguete en la actualización de la representación diagramática que tiene el bebé

Ahora, el Agente A regresa y busca el juguete después de que éste fue cambiado de ubicación. La representación D3 se forma a partir del patrón de observación en D1 y D2 (Ver figura 3). Esto supone que el bebé tiene una representación diagramática que conecta al agente A con el juguete en la ubicación 1 (D1), por un lado, y una representación que conecta al Agente C con el cambio de ubicación del juguete (D2). En otras palabras, D3 recluta el registro observacional de D1 y D2 para actualizar su representación diagramática que determina la reacción del bebé ante el comportamiento del agente al buscar el juguete. A continuación, se muestran dos escenarios para la formación de la representación D3 (Ver figura 3). El escenario A, la representación D3 corresponde con la expectativa del bebé según el registro del D1 y D2, en la cual se forma una representación del agente A buscando el juguete en la ubicación 1, y el escenario B, la representación D3 corresponde a la violación de esta expectativa, y es la del Agente A buscando el juguete en la ubicación real.

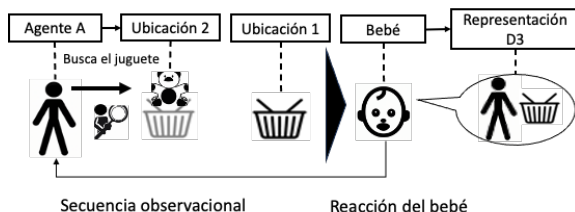
### Figura 3.

Tercera representación diagramática referencial (D3) que indica la expectativa del bebé en el escenario A y la violación de la expectativa en el escenario B acerca del comportamiento del Agente A.

#### Escenario A



## Escenario B



Nota: El Agente A regresa y, dado que el bebé sabe que el Agente A cree que el juguete aún está en la Ubicación 1, el bebé anticipa que el Agente A buscará el juguete en la Ubicación 1 (escenario A). Si el Agente A busca en la Ubicación 2, esto violaría las expectativas del bebé, a saber, la representación diagramática que ha formado (D3) y, por lo tanto, resultaría en una mirada más larga por parte del bebé (escenario B). La flecha que va del bebé al agente en la figura 3, representa que el bebé dirige su atención a observar el comportamiento de A en el escenario propuesto, por lo que es solamente ilustrativa.

En la figura 3, D3 representa la reacción del bebé ante la acción del agente A cuando regresa y busca el juguete en la Ubicación 1. Aunque la creencia del Agente A no ha cambiado desde D1 (todavía cree que el juguete está en la Ubicación 1), la realidad y la representación diagramática del bebé sí han cambiado. Si se considera que la Figura 3 ofrece la explicación más satisfactoria, entonces la reacción del bebé podría atribuirse únicamente a los diagramas referenciales. En otras palabras, la expectativa del bebé sobre dónde buscará el Agente A el juguete (D3), sería el resultado de un patrón asociativo de eventos únicamente observables (D1 y D2).

Sin embargo, hay un aspecto que esta explicación no logra cubrir por completo: el efecto del conocimiento del Agente A (creencia falsa) sobre la ubicación del objeto. El registro de este conocimiento permite al bebé formar una representación que lo lleva a reaccionar cuando el comportamiento de búsqueda del Agente A contradice sus expectativas. El bebé debió tomar en cuenta el conocimiento del agente A sobre la ubicación del objeto, específicamente su creencia falsa, y considerarla para formar una

expectativa sobre dónde el agente buscaría el objeto. Este conocimiento, aunque se deriva o depende de las representaciones referenciales D1 y D2, no es en sí mismo referencial.

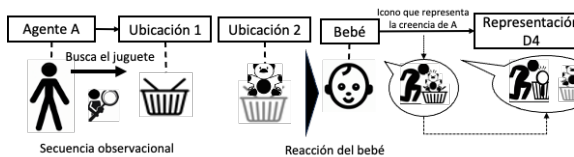
Una forma de abordar este asunto es considerar que la integración de estas representaciones, D1 y D2, podría resultar en una representación diagramática atributiva (D4). Por lo tanto, D3 se descarta y se sustituye por D4.

La figura 4 muestra la secuencia de acción del Agente A en dos escenarios y la representación diagramática atributiva (D4) que indica la expectativa del bebé acerca del comportamiento del Agente A al considerar su creencia falsa sobre la ubicación del objeto. Para que el bebé construya una expectativa sobre el comportamiento de búsqueda del agente A y lo represente buscando el juguete en la ubicación 1 en el escenario A (representación D4), debió tener una comprensión implícita de que el agente tenía una creencia falsa que lo conduce a su búsqueda en dicha ubicación. La figura 4, por tanto, representa de manera iconográfica la creencia del agente A en la mente del bebé para la formación de la presentación diagramática atributiva D4.

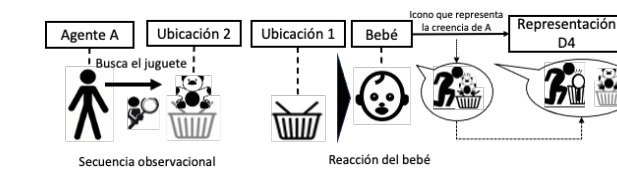
## Figura 4.

Representación diagramática atributiva (D4)

### Escenario A



## Escenario B



Nota: En este punto, el bebé tiene una representación actualizada de la ubicación real del juguete (Ubicación 2) y también de la creencia del agente A (icono que representa la creencia de A) quien, al no estar presente cuando el juguete fue movido, todavía cree que el juguete está en la Ubicación 1. La representación D4 refleja esta discrepancia entre la creencia del Agente A y la realidad, así como la expectativa del bebé de que el Agente A buscará el juguete en la Ubicación 1 (que es donde el Agente A cree que está) (escenario A). Esta discrepancia forma la base del asombro del bebé si el Agente A actúa de una manera que contradice su creencia previa, es decir, buscando en la Ubicación 2 (escenario B).

A pesar de que la representación D4 es similar a la D3 en cuanto a su contenido diagramático, lo que se quiere mostrar es que la reacción del bebé es derivada, por un lado, de la secuencia observada en D1 y D2, y por el otro lado, es afectada por el conocimiento del Agente A (creencia falsa) sobre la ubicación del objeto. En este sentido, la violación de la expectativa del bebé en el escenario B (figura 3 y 4) es producto de la comprensión de que el agente ha formado una creencia falsa sobre la ubicación real del objeto y espera que lo busque en la ubicación 1 (escenario A). Por este motivo, D3 se descarta y se sustituye por D4. La representación D4 está formada tanto de aspectos referenciales como atributivos.

Si se considera que la Figura 4 ofrece una explicación más completa, la reacción del bebé puede ser atribuida a una representación diagramática atributiva, D4, en lugar de la puramente referencial D3. La representación D4 es el producto de la experiencia social del bebé que se deriva de la integración de las representaciones anteriores: D1 (donde el juguete se coloca inicialmente en la Ubicación 1 y el Agente A está presente) y D2 (donde el juguete se mueve a la Ubicación 2 en ausencia del Agente A). D4 tiene en cuenta



la creencia falsa del Agente A con respecto a la ubicación del juguete. Al formar D4, el bebé no sólo se basa en la realidad observable (las ubicaciones del juguete), sino que también incorpora en su representación el conocimiento no observable del Agente A (la creencia falsa). Esta representación atributiva da lugar a una expectativa coherente en el bebé de que el Agente A buscará en la Ubicación 1, a pesar de que el juguete ahora está en la Ubicación 2.

Esta explicación basada en diagramas, en particular la representación D4, captura los aspectos clave del estudio de Onishi y Baillargeon (2005), a saber, la interpretación mentalista de que los bebés tienen una capacidad temprana de lectura de mentes.

En este sentido, el mapa cognitivo que el bebé construye no se limita a aspectos directamente observables (aunque sí depende de estos), sino que también refleja aspectos no observables, como la creencia falsa del Agente A, derivada de la experiencia social y la integración de las representaciones diagramáticas anteriores. Así, el contenido de estas representaciones diagramáticas se expande para incluir aspectos no directamente observables del comportamiento de los agentes. Por lo tanto, D4 es el producto de la experiencia social del bebé y de la integración de D1 y D2.

## 5. Conclusiones

Nuestra propuesta diagramática cuestiona la idea tan arraigada en la filosofía y la psicología de que la cognición humana se estructura a partir de un único formato representacional, el cual privilegia el uso de conceptos psicológicos y el lenguaje para navegar en el mundo social. Por el contrario, defendemos una visión más amplia según la cual las capacidades más básicas de cognición social se desarrollan a partir de un formato diagramático. En el caso de los bebés, estos emplean medios representacionales diagramáticos para comprender el comportamiento de los demás a partir de sus estados mentales.

Esta propuesta ofrece un marco filosófico de explicación del rompecabezas del desarrollo y amplía el contexto de explicación de la T1S y la T2S que revisamos al inicio del capítulo. La pro-

puesta permite caracterizar con mayor precisión los medios representacionales que emplean los bebés para captar los estados mentales de los demás sin involucrar una estructura predicativa compuesta por conceptos.

En los que respecta a la T1S, los bebés pueden emplear un medio representacional diagramático atributivo que funciona como los conceptos primitivos propuestos por los teóricos. Además, este medio representacional establece reglas simples que le permiten al bebé tener una comprensión intencional del comportamiento del agente. El uso de estas reglas simples para rastrear estados similares a las creencias a partir de la aplicación de nociones conceptuales primitivas (Carruthers, 2017) se puede entender a partir de la formación de representaciones diagramáticas referenciales en los bebés a partir de sus experiencias sociales.

Estas experiencias sociales enriquecidas forman un mapa cognitivo en forma de reglas simples para rastrear patrones de comportamiento social, las cuales son reclutadas para dar forma a las representaciones diagramáticas referenciales en los bebés para predecir el curso de acción de un agente al considerar sus estados mentales. Así, al decir que los bebés utilizan conceptos primitivos para pensar o razonar sobre los estados mentales de los demás, no implica decir que utilizan el concepto de creencia, piensan con palabras o tienen meta-representación. Desde nuestra propuesta, quiere decir que piensan y razonan a partir de un formato representacional diagramático.

En lo que respecta a la T2S, la propuesta de lectura de mentes diagramática amplía la comprensión del funcionamiento del S1, en la medida que es plausible la interpretación de que las operaciones cognitivas rápidas, eficientes y espontáneas del sistema emplean un formato diagramático. Las representaciones diagramáticas explicarían con mayor precisión el funcionamiento del registro (registering) y el encuentro (encountering) en las operaciones del S1 de lectura de mentes (Butterfill & Apperly, 2013b).

De acuerdo con Butterfill y Apperly, (2013b), los registros emplean medios representacionales no conceptuales, por lo que es

lógico pensar que el S1 opera a partir de un formato representacional diferente para registrar estados similares a las creencias, el cual es diagramático. Además, las funciones registro y encuentro coinciden con la caracterización de la función referencial y atributiva de los medios representacionales diagramáticos.

Nuestra interpretación es que las funciones de encuentro y registro emplean un medio representacional diagramático, lo que explicaría porque las operaciones del S1 están encapsuladas, presentan límites de procesamiento y su funcionamiento es independiente del S2 (Apperly & Butterfill, 2009). La lectura diagramática de la mente es una propuesta explicativa que brinda una nueva luz sobre la forma en que los bebés (y algunos animales no-humanos como los chimpancés) pueden comprender (al menos parcialmente) la mente de los otros. Seguramente aún hay mucho terreno que recorrer para lograr una explicación completa, pero esta perspectiva nos brinda el bosquejo del mapa cognitivo primitivo a partir del cual los bebés pueden orientarse en el mundo físico y mental que los rodea.

## Referencias

- Aguilera, M. (2020). Mapas, lenguaje y conceptos: hacia una teoría pluralista del formato de los conceptos. *Principia an international journal of epistemology*, 24(1), 121–146. <https://doi.org/10.5007/1808-1711.2020v24n1p121>
- Aguilera, M. (2021). Heterogeneous inferences with maps. *Synthese*, 199(1–2), 3805–3824. <https://doi.org/10.1007/s11229-020-02957-w>
- Andrews, K., Spaulding, S., & Westra, E. (2021). Introduction to folk psychology: Pluralistic approaches. *Synthese*, 199(1–2), 1685–1700. <https://doi.org/10.1007/s11229-020-02837-3>
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953–970. <https://doi.org/10.1037/a0016923>

- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110–118. <http://dx.doi.org/10.1016/j.tics.2009.12.006>
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159–86. <https://doi.org/10.1146/annurev-psych-010213-115033>
- Barlassina, L., & Gordon, R. M. (2017). Folk psychology as mental simulation. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 Ed.). <https://plato.stanford.edu/entries/folkpsych-simulation/>
- Beck, J. (2013). Why we can't say what animals think. *Philosophical Psychology*, 26(4), 520–546. <https://doi.org/10.1080/09515089.2012.670922>
- Bialecka-Pikul, M., & Białek, A. (2021). Developmental continuity in coordination of perspectives as an alternative to the implicit and explicit theory of mind controversy. *Developmental Psychology*, 57(11), 1772–1786. <https://doi.org/10.1037/dev0001254>
- Boyle, A. (2019). Mapping the Minds of Others. *Review of Philosophy and Psychology*, 10(4), 747–767. <https://doi.org/10.1007/s13164-019-00434-z>
- Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, 125(3), 409–434. <https://doi.org/10.1037/rev0000091>
- Burnside, K., Severdija, V., & Poulin-Dubois, D. (2020). Infants Attribute False Beliefs to a Toy Crane. *Developmental Science*, 23(2), e12887. <https://doi.org/10.1111/desc.12887>
- Butterfill, S. A., & Apperly, I. A. (2013a). Replies to three Commentaries on Minimal Theory of Mind. In J. Schwenkler (eds.), *Brains* (philosophyofbrains.com). [https://www.butterfill.com/pdf/minimal\\_brains\\_discussion\\_replies.pdf](https://www.butterfill.com/pdf/minimal_brains_discussion_replies.pdf)
- Butterfill, S. A., & Apperly, I. A. (2013b). How to Construct a Minimal Theory of Mind. *Mind & Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Buttelmann, F., & Kovács, Á. M. (2019). 14-Month-olds anticipate others' actions based on their belief about an object's identity. *Infancy*, 24(5), 738–751. <https://doi.org/10.1111/inf.12303>

- Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, 21(1), 145–182. <https://doi.org/10.1111/j.1520-8583.2007.00124.x>
- Camp, E. (2009). A language of baboon thought? In R. W. Lurz (Ed.), *The philosophy of animal minds* (pp. 108–127). Cambridge University Press.
- Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28(2), 141–172. <https://doi.org/10.1111/mila.12014>
- Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, 7(1), 141–162. <https://doi.org/10.1007/s13164-015-0259-y>
- Carruthers, P. (2017). Mindreading in adults: evaluating two-systems views. *Synthese*, 194(3), 673–688. <https://doi.org/10.1007/s11229-015-0792-3>
- Currie, G. & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press.
- Dennett, D. (1998). *La actitud intencional*. Gedisa.
- Dretske, F. (1981). *Knowledge and the flow of information*. The MIT Press.
- Dretske, F. (1995). *Naturalizing the mind*. The MIT Press.
- Dörrenberg, S., Rakoczy, H., Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, 46, 12–30. <https://psycnet.apa.org/doi/10.1016/j.cogdev.2018.01.001>
- Egawa, J., Kawasaki, K., Hayashi, T., Akikawa, R., Someya, T., & Hasegawa, I. (2022). Theory of mind tested by implicit false belief: a simple and full-fledged mental state attribution. *The FEBS Journal, Early View*, 289 (23), 7343-7358. <https://doi.org/10.1111/febs.16322>
- Falck, A., Brinck, I., &, Lindgren, M. (2014). Interest contagion in violation-of-expectation-based false-belief tasks. *Frontiers in Psychology*, 5, 23. <https://doi.org/10.3389/fpsyg.2014.00023>
- Fiebich, A., & Coltheart, M. (2015). Various Ways to Understand Other Minds: Towards a Pluralistic Approach to the Explanation of Social Understanding. *Mind & Language*, 30(3), 235–258. <https://doi.org/10.1111/mila.12079>
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

- Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron* 60, 503–510. <https://doi.org/10.1016/j.neuron.2008.10.032>
- Gangopadhyay, N. (2017). The future of social cognition: paradigms, concepts and experiments. *Synthese*, 194(3), 655–672. <https://doi.org/10.1007/s11229-016-1162-5>
- Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.
- Gómez-Tabares, A. S. (2022). ¿Dos sistemas de lectura de mentes? Análisis crítico de la teoría de dos sistemas. *Theoria: An International Journal for Theory, History and Foundations of Science*, 37(3), 331–355. <https://www.jstor.org/stable/27191214>
- Gómez-Tabares, A. S. (2023). Is there continuity from implicit recognition of intentional action in infants to explicit mindreading in preschoolers? Systematic review of longitudinal evidence and theoretical implications (¿Hay continuidad entre el reconocimiento implícito de la acción intencional en bebés y la lectura de mente explícita en preescolares? Revisión sistemática de la evidencia longitudinal e implicaciones teóricas). *Infancia y Aprendizaje*, 46(4), 950–982. <https://doi.org/10.1080/02103702.2023.2230044>
- Gopnik, A., & Meltzoff, A. (1997). *Words, Thoughts and Theories*. MIT Press
- Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1-2), 145–171. <https://doi.org/10.1111/j.1468-0017.1992.tb00202.x>
- Gordon, R. M. (1986). Folk Psychology as Simulation. *Mind & Language*, 1(2), 158–171. <https://doi.org/10.1111/j.1468-0017.1986.tb00324.x>
- Gordon, R. M. (1995). Simulation without Introspection or Inference from Me to You. In D. Martin. & T. Stone (eds.) *Mental Simulation* (pp. 53-67). Blackwell
- Heck, R. K. (2007). Are there different kinds of content? In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 117–138). Blackwell.

- Herschbach, M. (2015). Direct social perception and dual process theories of mindreading. *Consciousness and Cognition*, 36, 483–497. <https://doi.org/10.1016/j.concog.2015.04.001>
- Heyes, C. (2014a). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, 9(2), 131–143. <https://doi.org/10.1177/1745691613518076>
- Heyes, C. (2014b). False belief in infancy: a fresh look. *Developmental Science*, 17(5), 647–659. <https://doi.org/10.1111/desc.12148>
- Hutto, D. & Ravenscroft, I. (2021). Folk Psychology as a Theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2016/entries/folkpsych-theory/>
- Jacob, P. (2019a). Challenging the Two-systems Model of Mindreading. In A. Avramides & M. Parrott. *Knowing Other Minds* (pp. 79–106). Oxford Scholarship Online. <https://doi.org/10.1093/oso/9780198794400.001.0001>
- Jacob, P. (2019b). What Do False-Belief Tests Show? *Review of Philosophy and Psychology*, 11(4), 1–20. <https://doi.org/10.1007/s13164-019-00442-z>
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
- Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, 44(2), 107–115. <https://doi.org/10.1177/0165025419850901>
- Kristen-Antonow, S., Licata-Dandel, M., McEuler, M., & Sodian, B. (2018). Maternal cognition talk in the mother–toddler dyad mediates the influence of early maternal emotional availability on preschoolers’ belief reasoning. *Social Development*, 27(4), 841–857. <https://doi.org/10.1111/sode.12301>
- Kushnir, T., & Gopnik, A. (2005). Young Children Infer Causal Strength from Probabilities and Interventions. *Psychological Science*, 16(9), 678–683. <https://doi.org/10.1111/j.1467-9280.2005.01595.x>
- Leslie, A., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences*, 8(12), 528–533. <https://doi.org/10.1016/j.tics.2004.10.001>

- Licata, M., Kristen, S., & Sodian, B. (2016). Mother–child interaction as a cradle of theory of mind: The role of maternal emotional availability. *Social Development*, 25(1), 139–156. <https://doi.org/10.1111/sode.12131>
- Litchfield, K.A., & Lambert, M.C. (2011). Nativist Theory. In Goldstein, S., & Naglieri, J.A. (eds) *Encyclopedia of Child Behavior and Development*. Springer. [https://doi.org/10.1007/978-0-387-79061-9\\_1911](https://doi.org/10.1007/978-0-387-79061-9_1911)
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. The MIT Press.
- Meinhardt-Injac, B., Daum, M. M., Meinhardt, G., & Persike, M. (2018). The Two-Systems Account of Theory of Mind: Testing the Links to Social- Perceptual and Cognitive Abilities. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00025>
- Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, 173, 60–74. <https://doi.org/10.1016/j.cognition.2018.01.001>
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255–258. <https://doi.org/10.1126/science.1107621>
- Poulin-Dubois, D., Azar, N., Elkaim, B., & Burnside, K. (2020). Testing the stability of theory of mind: A longitudinal approach. *PLoS ONE* 15(11), e0241721. <https://doi.org/10.1371/journal.pone.0241721>
- Rakoczy, H. (2017). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese*, 194, 689–707. <https://doi.org/10.1007/s11229-015-0860-8>
- Rakoczy, H. (2021). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. In Bhargava, D, *Theory of Mind: A Psychological Perspective* (pp. 35-50). Vidya Books
- Rakoczy, H., Bergfeld, D., Schwarz, I., & Fiske, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, 86(2), 486–502. <https://doi.org/10.1111/cdev.12311>



- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34(3), 265–293. <https://doi.org/10.1016/j.dr.2014.04.001>
- Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences*, 9(10), 462–463. <https://doi.org/10.1016/j.tics.2005.08.001>
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in cognitive sciences*, 19(2), 65–72. <https://doi.org/10.1016/j.tics.2014.11.007>
- Scott, R. M., Baillargeon, R., Song, H., & Leslie, A. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61 (4), 366–395. <https://psycnet.apa.org/doi/10.1016/j.cogpsych.2010.09.001>
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21(4), 237–249. <https://doi.org/10.1016/j.tics.2017.01.012>
- Siu, T.-S. C., & Cheung, H. (2019). Developmental progression of mental state understandings in infancy. *International Journal of Behavioral Development*, 43(4), 363–368. <https://doi.org/10.1177/0165025419830233>
- Sodian, B. (2011). Theory of mind in infancy: Theory of mind in infancy. *Child Development Perspectives*, 5(1), 39–43. <https://doi.org/10.1111/j.1750-8606.2010.00152.x>
- Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How Does Children's Theory of Mind Become Explicit? A Review of Longitudinal Findings. *Child Development Perspectives*, 14(3), 171-177. <https://doi.org/10.1111/cdep.12381>
- Ting, F., He, Z., & Baillargeon, R. (2021). Five-month-old infants attribute inferences based on general knowledge to agents. *Journal of Experimental Child Psychology*, 208, 105126. <https://doi.org/10.1016/j.jecp.2021.105126>
- Velázquez Coccia, F. (2018). (mind)-reading maps. *Philosophies*, 3(2), 12. <https://doi.org/10.3390/philosophies3020012>
- Wellman, H. M. (2014). *Making Minds, How Theory of Mind Develops*. Oxford Series in Cognitive Development, Oxford University Press.

- Wellman, H. M. (2017). The Development of Theory of Mind: Historical Reflections. *Child Development Perspectives*, 11(3), 207–214. <https://doi.org/10.1111/cdep.12236>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, 72(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Westra, E. (2017). Spontaneous mindreading: a problem for the two-systems account. *Synthese*, 194(11), 4559–4581. <https://doi.org/10.1007/s11229-016-1159-0>
- Westra, E., & Carruthers, P. (2018). Theory of Mind. In T. K. Shackelford., & V. A. Weekes-Shackelford (eds.), *Encyclopedia of Evolutionary Psychological Science*. [https://doi.org/10.1007/978-3-319-16999-6\\_2376-1](https://doi.org/10.1007/978-3-319-16999-6_2376-1)
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining functions of wrong beliefs in young children's understanding of deception. *Cognition*, 13 (1), 103-128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, 18, 53–57. <https://doi.org/10.1111/j.1467-8721.2009.01605.x>





El segundo volumen de Filosofía analítica propone un recorrido por problemas centrales de la metafísica y la filosofía de la mente, desde una tradición caracterizada por la claridad conceptual, la precisión argumentativa y el análisis riguroso. Aborda cuestiones como la fundamentación y la dependencia, el determinismo y el fatalismo, la persistencia y la identidad personal, las actitudes proposicionales, la relación mente-cuerpo, la subjetividad, la consciencia y la comprensión de otras mentes. Estos problemas se examinan a la luz de debates clásicos y contemporáneos, en diálogo con la lógica, la filosofía del lenguaje y las ciencias cognitivas. Más que ofrecer respuestas definitivas, el libro busca proporcionar herramientas conceptuales para analizar con mayor precisión cuestiones complejas de la realidad y de la experiencia humana, mostrando la amplitud de la filosofía analítica y su capacidad para dialogar con distintos campos del conocimiento sin renunciar al rigor de su método.

ISBN: 978-9907-840-05-6



9 789907 840056



**filosófica**  
EDITORIAL