



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Maestría en Ciencias Humanas
Opción Filosofía Contemporánea

Tesis para obtener el Título de Magister en Ciencias Humanas
Opción Filosofía Contemporánea

**“El Problema Difícil de la Conciencia y el Argumento del
Conocimiento de Frank Jackson”**

Autor: Prof. Claudio Lee Lassevich Esperanza

Orientadores: Dr. Álvaro Peláez; Dr. Ricardo Navia

Montevideo, 7 de setiembre de 2020

Aval de los Orientadores de Tesis

México, Ciudad de México, 2 de septiembre de 2020

A quien corresponda

Mediante la presente, quien suscribe, Dr. Álvaro Peláez Cedrés, en mi calidad de asesor de la tesis de maestría del alumno Claudio Lassevich, de título “El problema difícil de la conciencia y el argumento del conocimiento de Frank Jackson”, comunico que la misma se encuentra en condiciones de ser defendida ante el sínodo que se designe para tales efectos.

Sin otro particular, aprovecho la ocasión para enviar un cordial saludo

Dr. Álvaro Peláez Cedrés
Universidad Autónoma Metropolitana – Unidad Cuajimalpa, México

Sres. miembros de la Comisión Académica de Postgrado
Facultad de Humanidades y Cs. De la Educación
De mi mayor consideración:

En mi calidad de cotutor, por la presente dejo constancia que la tesis de maestría de Claudio Lassevich Esperanza titulada “El problema difícil de la conciencia y el argumento del conocimiento de Frank Jackson” está en condiciones de ser presentada para su defensa, a los efectos de la obtención del título de Magister en Humanidades Opción Filosofía Contemporánea.

Sin otro particular:

Prof. Ricardo Navia

A Carolina, Irene y Ariel.

A mi abuelo, Dr. León Esperanza Levi, de Bendita Memoria.

A mi mamá, Virginia Esperanza, de Bendita Memoria, por transmitir ese legado.

Agradecimientos

Esta investigación recibió el apoyo de la Agencia Nacional de Investigación e Innovación (ANII), programa de financiamiento de Maestrías en Uruguay, contrato POS_NAC_2017_1_141302.

La posibilidad de haber realizado cursos de neuroanatomía, neurofisiología, y neurociencia cognitiva y computacional, en la Facultad de Medicina, y en el Espacio Interdisciplinario de la Universidad de la República han posibilitado que esta investigación, siendo filosófica, mantenga un permanente diálogo con las ciencias empíricas. Mis orientadores en la Facultad de Humanidades y Ciencias de la Educación, Dr. Álvaro Peláez, y Dr. Ricardo Navia, han sido un apoyo permanente. Agradezco su paciencia en las interminables lecturas.

La Sociedad Filosófica del Uruguay, la Asociación Filosófica del Uruguay, y la Associação Brasileira de Filosofia e História da Biologia han permitido que presente algunas líneas de la presente tesis en encuentros académicos nacionales e internacionales de diversa índole. También agradezco al Instituto Juan XXIII por ofrecer facilidades en el cumplimiento de mis tareas profesionales que permitieron que me aboque a la culminación de la investigación.

Un par de ideas provenientes de la inquietud de mi hermano, Lic. Daniel Lassevich, hicieron posible culminar esta investigación. También el sostén de mi padre, Jaime Lassevich, generando condiciones sin descanso. El buen sentido hiperlógico de la Prof. Carolina Barboza está presente en cada línea de este texto.

Tabla de Contenidos

AVAL DE LOS ORIENTADORES DE TESIS	III
AGRADECIMIENTOS	VII
TABLA DE CONTENIDOS.....	IX
RESUMEN.....	XIII
ABSTRACT.....	XIV
PALABRAS CLAVE	XV
KEYWORDS	XVI
INTRODUCCIÓN	1
CAPÍTULO 1: LA BRECHA EXPLICATIVA Y EL PROBLEMA DIFÍCIL DE LA CONCIENCIA.....	7
1.1 HACIA UNA EXPLICACIÓN REDUCTIVA EN <i>EL MUNDO</i>	9
1.2 LAS FUNCIONES COGNITIVAS DEL AUTÓMATA CARTESIANO	13
1.2.1 Las sensaciones y la brecha explicativa	20
1.2.1.1 Los niveles de la sensación	22
1.2.1.2 La brecha explicativa	29
1.2 EL PROBLEMA DIFÍCIL DE LA CONCIENCIA	31
1.2.1 Conciencia de acceso y conciencia fenoménica.....	32
1.2.2 Dos problemas respecto a la conciencia.....	35
1.3 LA SUPERVENIENCIA Y SUS FORMAS	41
1.3.1 Superveniencia y reduccionismo.....	44
1.4 EL MECANICISMO COMO TESIS DE SUPERVENIENCIA	46

CAPÍTULO 2. EL PROBLEMA DIFÍCIL DE LA CONCIENCIA: GEOGRAFÍA DE POSTURAS	53
2.1 GEOGRAFÍA DEL PROBLEMA DIFÍCIL DE LA CONCIENCIA.....	54
2.1.1 Materialismos no reduccionistas (Tipo-A).....	55
2.1.2 Materialismos reduccionistas (Tipo-B).....	56
2.1.2.1 Apunte sobre la influencia de este enfoque.....	57
2.1.3 Materialismo Tipo-C e Idealismo Tipo-I	60
2.1.4 Dualismo interaccionista (Tipo-D)	62
2.1.5 Epifenomenismo o Dualismo Tipo-E.....	63
2.1.6 Monismo Tipo-F	65
2.1.7 Funcionalismo	66
2.2 SÍNTESIS	67
CAPÍTULO 3: EL ARGUMENTO DEL CONOCIMIENTO.....	69
3.1 DE LA INTUICIÓN AL ARGUMENTO	70
3.2 LA HABITACIÓN DE MARÍA: EL ARGUMENTO DEL CONOCIMIENTO.....	73
3.2.1 Qué sabemos acerca de María.....	74
3.2.2 La estructura del argumento.....	77
3.3 OBJECIONES AL AC DE FRANK JACKSON.....	80
3.3.1 María no aprende nada: la banana azul	81
3.3.2 María aprende conocimiento no proposicional	86
3.3.2.1 Aprender cómo se siente es adquirir una habilidad	87
3.3.2.2 Dificultad: el conjunto de habilidades de Lewis es muy exigente	91
3.3.2.3 Aprender cómo se siente es familiarizarse	94
3.3.2.4 Un argumento neurobiológico a favor de la hipótesis de la familiaridad	97
3.3.2.5 ¿Se trata de una estrategia efectiva?.....	101
3.3.3 María aprende nuevo conocimiento proposicional sobre hechos ya conocidos	104
3.3.3.1 Críticas a la estrategia “nuevo conocimiento / viejo hecho”.....	112
3.3.3.2 Intensión primaria y secundaria	115

3.3.4 La paridad de razones contra el dualismo	118
3.3.5 Fisicalismo, materialismo, y la premisa (1)	123
3.3.5.1 ¿Qué es una información física completa?.....	124
3.3.5.2 Fisicalismo en sentido estricto y en sentido amplio.....	127
3.3.5.3 Plausibilidad de la interpretación de Dennett.....	130
3.4 VALORACIÓN GLOBAL DEL AC PARA EL DUALISMO	134
CONCLUSIONES	141
BIBLIOGRAFÍA	149

Resumen

A partir de la Revolución Científica de los siglos XVI y XVII, y particularmente de la obra de René Descartes, el problema filosófico que consiste en determinar la naturaleza de la relación entre el cuerpo y el alma cobró una nueva forma. Uno de sus aspectos fundamentales surge de la imposibilidad del modelo mecanicista cartesiano de explicar cómo y por qué hay algunos estados mentales que tienen una cualidad fenoménica.

David Chalmers ha llamado “problema difícil de la conciencia” a dicha interrogante para distinguirla de los llamados “problemas fáciles de la conciencia” que son aquellos que las ciencias cognitivas pueden resolver mediante estrategias reduccionistas. Al problema epistémico le corresponde un problema metafísico vinculado a la determinación del estatuto ontológico de la conciencia. La pertinencia de la distinción ha suscitado posicionamientos materialistas reduccionistas, no reduccionistas, dualistas, y panpsiquistas.

Frank Jackson diseñó un experimento mental consistente en imaginar a María, una neurocientífica que conoce todo lo que hay por conocer sobre la fisiología de la visión en un escenario de aislamiento cromático desde el nacimiento. ¿Se sorprendería María al ver por primera vez algo rojo o su conocimiento físico completo le permitiría prever como se sentiría dicha experiencia? Un análisis detenido del argumento y varias líneas de crítica podría arrojar luz sobre el problema difícil de la conciencia.

Defenderé que el argumento logra fundamentar una postura no reduccionista a nivel epistémico pero fracasa al intentar refutar una postura materialista a nivel metafísico.

Abstract

Since the Scientific Revolution of the 16th and 17th centuries, and remarkably with the works of Rene Descartes, the philosophical problem of determining the nature of the relationship between body and soul took on a new form. One of its fundamental aspects arises from the impossibility of the cartesian mechanistic model to explain how and why there are some mental states that have a phenomenal quality.

David Chalmers has called this “the hard problem of consciousness” to distinguish it from the so-called “easy problems of consciousness”, which are those that the cognitive sciences can solve through reductionist strategies. To the epistemic problem corresponds a metaphysical problem linked to the determination of the ontological status of consciousness. The accuracy of the distinction has given rise to materialistic reductionist, non-reductionist, dualist, and panpsychist positions.

Frank Jackson designed a mental experiment which consists of Mary, a neuroscientist who knows everything there is to know about the physiology of vision in a setting of chromatic isolation from birth. Would Mary be surprised to see something red for the first time, or would her complete physical knowledge allow her to know what it is like to see red? A careful analysis of the argument and several critical strategies could shed light on the hard problem of consciousness.

I will defend that the argument manages to support a non-reductionist position at the epistemic level but fails in trying to refute a materialistic position at the metaphysical level.

Palabras Clave

Conciencia Fenoménica; Problema Difícil de la Conciencia; Argumento del Conocimiento; Superveniencia.

Keywords

Phenomenal Consciousness; Hard Problem of Consciousness; Knowledge Argument; Supervenience.

Introducción

Desde los albores de la historia de la filosofía occidental ha habido reflexión respecto a la naturaleza de lo mental y su relación con lo corporal. En la Modernidad, la obra de Descartes se ha constituido en un verdadero hito fundante disciplinar dado el giro radical que el filósofo le dio a la cuestión, giro que marcó toda la especulación filosófica en el área hasta nuestros días. Rorty (1989) ha señalado que el problema de la relación entre la mente -o el alma- y el cuerpo en la Antigüedad y el Medioevo se enfoca principalmente en la posibilidad de conocer los universales¹. En contraste, la problemática moderna cuerpo-alma se ha forjado en el marco de la búsqueda cartesiana de certeza para la nueva ciencia, y tiene como punto de partida la constatación de la asimetría notoria que existe entre la seguridad aparentemente infalible de nuestras creencias acerca de lo que ocurre en nuestra propia interioridad y las dudas que generan nuestras creencias sobre lo que ocurre en el mundo que nos rodea.

La tradición filosófica contemporánea respecto a la relación entre la mente y el cuerpo continúa signada por la manera como Descartes abordó el problema, por su postura, y por sus notorias fisuras que no tardaron en manifestarse. Quienes se han posicionado en las antípodas del filósofo, lo han hecho atentos a su postura dualista. Más aún, incluso quienes niegan que el problema sea significativo también tienen en vista la ontología dualista del autor, aunque más no sea para intentar superarla señalándola como un caso de confusión lingüística. Su teoría ha sido tan

¹ Cfr. por ejemplo los diálogos clásicos de Platón *Fedón* (1938) y *República* (1999), en donde la inmaterialidad del alma se pone en juego en relación a la posibilidad de conocer las ideas.

influyente que *The concept of mind* de Gilbert Ryle (1949), obra usualmente considerada como la que inaugura a la Filosofía de la Mente entendida como disciplina filosófica autónoma, hace girar su tesis central -incluso en sus aspectos propositivos- alrededor de una crítica a la manera cartesiana de entender lo mental y su relación con lo físico.

Enmarcada en ese contexto disciplinar, la presente investigación ha puesto el foco en los siguientes objetivos. Primero, identificar y formular de manera actualizada el problema de la relación entre la mente y el cuerpo, particularmente en dos dimensiones: la epistemológica y la ontológica. La dimensión epistemológica trata de determinar si es posible la elaboración de una teoría científica acerca de lo mental que proceda con estrategias reduccionistas al estilo de la ciencia moderna que nos legaron Galileo y Descartes. La dimensión ontológica consiste en preguntarse si lo mental y lo corporal conforman dos sustancias diferentes, o si se trata de la misma sustancia, y en ese caso si las propiedades mentales y físicas son de naturaleza diferente o no.

Entiendo preliminarmente “reduccionismo de lo mental” como la tesis epistemológica que afirma que podemos tener un conocimiento completo de las mentes apelando a los conceptos de la Física. Por “Física” me refiero no solo a la disciplina específica “física”, sino al conjunto de las ciencias naturales, entre las que podemos destacar -para nuestros fines- las ciencias cognitivas, la neurobiología, las neurociencias computacionales, la psicología y otras disciplinas afines. Por tanto, entiendo por “fiscalismo” a la afirmación epistemológica que afirma la posibilidad, y necesidad, de reducir todos los enunciados acerca de lo mental a enunciados propios de la Física². Mientras tanto, utilizaré el término

² Para enriquecer la discusión en torno a estas distinciones cfr. p. ej. Carnap (1995), Stoljar (2017), Goff (2017).

“materialismo” para referirme a la postura ontológica que afirma que todo lo que existe es material, por lo que sólo existe una sustancia³.

En los últimos años, el filósofo australiano David Chalmers ha llamado “problema difícil de la conciencia” a la pregunta respecto a cómo y por qué muchos de nuestros estados mentales se sienten de determinada manera peculiar (Chalmers, 1995), distinguiéndolo de los “problemas fáciles de la conciencia” que son los vinculados a la comprensión de las diversas funciones cognitivas. Propuso que el problema difícil no puede resolverse por estrategias reduccionistas, pero más allá de sus tesis y argumentos, interesa en este momento destacar que la terminología ha sido fundante en el sentido de que ha permitido que se desplegara la discusión filosófica respecto a lo mental. En consecuencia, el segundo objetivo de esta investigación consiste en determinar con claridad en qué consiste el problema difícil de la conciencia, y para ello será necesario familiarizarnos con una serie de conceptos técnicos entre los que adelanto la distinción entre “conciencia fenoménica” y “conciencia de acceso”, y el concepto de superveniencia y sus distintas formas. Los dos primeros objetivos serán abordados especialmente en el capítulo 1.

El tercer objetivo se desarrolla en el capítulo 2, en el que exploraré y mapearé las principales posturas que han sido presentadas respecto al problema difícil de la conciencia. Excede a mis propósitos el detenerme en cada teoría para explorarla de forma minuciosa, y me limitaré a presentarlas de forma sintética, brindando las claves para que el lector pueda ubicarse en la geografía lógica del problema⁴. Es un capítulo de corte más expositivo que propositivo o argumentativo,

³ “Idealismo” y “dualismo” se oponen a “materialismo” de dos maneras diferentes: el idealismo -en esta faceta ontológica- es la postura que afirma la existencia de una única sustancia, pero dicha sustancia no es la materia sino el pensamiento; y el dualismo de sustancias es la afirmación de que existen dos sustancias radicalmente diferentes y separadas (materia y pensamiento).

⁴ Parafraseo aquí a Ryle (1949), aunque sin comprometerme con sus tesis de fondo.

cuyo fin es una correcta comprensión del estado del arte y una clara delimitación de las diferentes posiciones, preparación imprescindible para abordar el siguiente objetivo.

El cuarto objetivo de la presente tesis se llevará adelante en el tercer capítulo, en el que analizaré en profundidad uno de los más influyentes argumentos que ha sido presentado con el fin de criticar el fisicalismo y el materialismo. El “argumento del conocimiento”⁵ ha sido propuesto en diversas formulaciones en las últimas décadas. La formulación de Frank Jackson, en su “Epiphenomenal qualia” (1982) ha sido sin duda de las más discutidas, gracias a la utilización de un experimento mental que ha dado lugar a ríos de tinta. Jackson imagina a María, una neurocientífica que posee un conocimiento físico completo de nuestra percepción visual, pero que nunca ha percibido ella misma los colores, pues ha vivido confinada en un *bunker* ambientado de forma monocromática. Cuando es liberada y enfrentada a los objetos del mundo María se sorprendería, por lo que podemos suponer que aprendió algo nuevo que no conocía, a saber, cómo se siente ver algo rojo o azul. Dicho aprendizaje lleva a Jackson a concluir que debe haber algo no físico por saber acerca de la percepción visual de los colores, y que el fisicalismo – como postura epistemológica-, y el materialismo –como postura ontológica-, han de ser falsos.

Analizar detenidamente el AC y sus críticas provenientes de diferentes tiendas es algo que tiene interés por sí mismo, ya que se trata de uno de los argumentos basados en experimentos mentales más discutido de los últimos años, pero también nos permitirá una mejor comprensión del problema difícil de la conciencia y de las variadas posturas epistemológicas y ontológicas que ha suscitado. Defenderé, luego de explorar numerosas objeciones, que el argumento

⁵ Utilizaré de aquí en adelante la abreviación “AC” para referirme al Argumento del Conocimiento. Asimismo llamaré “María” a “Mary”, la protagonista del experimento mental en la formulación de Frank Jackson.

ha sido exitoso en su faceta epistemológica, por lo que efectivamente permite dudar de las afirmaciones del fisicalismo de corte reduccionista. Esta conclusión nos colocará ante la siguiente disyuntiva: o desde un materialismo estricto le negamos entidad a la conciencia fenoménica, o la aceptamos como un ítem de la realidad que no puede explicarse con estrategias reduccionistas. En ninguno de los cuernos del dilema hay espacio para una explicación de la conciencia fenoménica en términos físicos.

También defenderé que el argumento no ha sido exitoso en su faceta ontológica, por lo que no logra totalmente fundamentar una crítica definitiva al materialismo. Asimismo, las críticas al argumento tampoco logran establecer ni la falsedad del dualismo ni la veracidad del materialismo. Tomaré como indicio para esta conclusión el hecho de que, en cada ocasión en que la disputa se torna ontológica, los defensores de las distintas posiciones terminan girando en círculos, o proporcionando evidencia no concluyente, o en un permanente intercambio de la carga de la prueba, situaciones que pueden interpretarse como un estancamiento dialéctico. Por ello el argumento no permite tomar una decisión respecto a la disyuntiva señalada en el párrafo anterior, sin perjuicio de que cada uno pueda volcarse por una u otra alternativa en virtud de otras razones o argumentos complementarios.

Para finalizar, y tomando en cuenta el estancamiento mencionado, cabrá preguntarse cómo se puede vincular el AC con la obra de Descartes, que habría dado inicio al problema mente-cuerpo tal como lo estamos entendiendo hoy día, y si debemos entender ese estancamiento en la disputa metafísica como un indicador de falta de sentido de la actividad filosófica en general.

Como criterio general, he intentado utilizar como fuente traducciones fiables al español. En numerosas ocasiones esto no ha sido posible, por lo que el lector encontrará algunas citas textuales que han sido traducidas por quien escribe directamente del idioma original. Soy optimista respecto a haber preservado el sentido original de dichos fragmentos.

Capítulo 1: La Brecha Explicativa y el Problema Difícil de la Conciencia

Como he adelantado en la Introducción, procederé a delimitar el problema filosófico que nos convoca. Para tal cometido, comenzaré exponiendo cómo, al menos en opinión de quien escribe, el problema difícil de la conciencia surge naturalmente de la manera como Descartes va conformando su proyecto de una física mecanicista para dar una nueva fundamentación metafísica a la naciente ciencia post-revolucionaria. En tal oportunidad el mecanicismo se constituyó como un modelo epistemológico de gran potencial para dar cuenta de los hechos del mundo en general, y el funcionamiento del cuerpo humano en particular. No obstante, al irrumpir el pensamiento, lo mental, se erigió en un auténtico límite epistémico que el filósofo no logró explicar mediante reducción a un funcionamiento mecánico.

A continuación, convendrá apropiarnos de una terminología más actualizada. Para tal fin, primero consideraré la distinción que propone Ned Block (1995) entre dos formas o funciones de la conciencia: la conciencia de acceso y la conciencia fenoménica. Ambas formas aparecen confundidas en la obra de Descartes, y en su distinción hay una clave para avanzar en la comprensión del problema. Mientras que la primera involucra todas las funciones cognitivas que nos permiten adaptarnos al mundo (percepción del entorno, memoria, atención, control motor, reporte de estados internos, etc.); la segunda refiere al hecho de que algunos estados mentales conscientes tienen un contenidos fenoménico, cualitativo, peculiar, que puede o no acompañar a la conciencia de acceso, dependiendo del caso.

A partir de este punto restringiré el foco de atención a la conciencia fenoménica. La dificultad para explicar lo mental en términos mecanicistas, la llamada brecha explicativa (Levine, 1983), podrá enmarcarse en lo que David Chalmers (1995) ha llamado *problema difícil de la conciencia*, terminología que ha sido adoptada tanto por seguidores como por detractores, constituyéndose en eje paradigmático de la discusión contemporánea.

Si bien el problema difícil de la conciencia es, en un primer examen, epistemológico, dado que aborda la cuestión de la explicación de un aspecto de la realidad, el problema también tiene una dimensión ontológica. En este marco, la relación entre la mente y el cuerpo también se ha discutido en términos de *superveniencia* de lo mental a lo material⁶, por lo que una delimitación de dicho concepto y sus variantes nos permitirá entender que la disputa en torno al problema difícil puede asociarse de diversas maneras a una disputa en torno a la forma de superveniencia entre mente y cuerpo.

Repasada la génesis del problema, y precisados los conceptos, quedaremos habilitados para, en el siguiente capítulo, proceder a un mapeo de las principales posiciones al respecto. Las dos nociones teóricas problemáticas -conciencia fenoménica y superveniencia- entran en juego en las intuiciones que suscitan los experimentos mentales que se han esgrimido como argumentos críticos con el fisicalismo y/o el materialismo, entre los cuáles se destaca nuestro foco de atención del tercer capítulo: el experimento de la habitación de María para la conformación del AC.

⁶ En el caso de la excepcional filosofía del obispo Berkeley, cfr. por ejemplo su *Tratado sobre los principios del conocimiento humano* (1992 [1709]), se trataría del caso inverso: superveniencia de lo material a lo mental.

1.1 Hacia una explicación reductiva en *El mundo*

En 1630, Descartes habría abandonado la elaboración de sus *Reglas para la dirección del espíritu* (1984 [1628]), posiblemente por encontrar cierta inadecuación entre el método científico puramente matemático que ahí quería exponer y la pretensión de que dicho método fuera útil para ampliar nuestros conocimientos empíricos acerca del mundo (Hamelin, 1949) (Turró, 1989). Una fundamentación epistemológica de la física tenía que otorgar un justo lugar a los datos de los sentidos. Es en ese punto que comenzó a elaborar *El mundo*⁷, una obra que luego retiraría de la imprenta en 1633 especialmente a raíz del proceso inquisitorio a Galileo, y no por considerar sus contenidos incorrectos (Turró, 1989). En dicha obra se propone una fábula, la invención de un nuevo mundo, cuya materia constituyente es un agregado de partículas, estructurado de manera mecánico-geométrica, que explica el conjunto de apariencias fenoménicas que llegan cotidianamente a nuestros sentidos. Esa “fábula” no es nada más que lo que hoy llamaríamos un modelo.

Los modelos científicos pueden distinguirse groseramente en modelos realistas o instrumentalistas, de acuerdo al grado de compromiso que tengan sus afirmaciones acerca del mundo. Si la aspiración del modelo es simplemente proveer una serie de explicaciones útiles para comprender los fenómenos, pero sin pretender ser descriptivos y literalmente fieles a la realidad, se trata de un modelo de carácter instrumental. Si lo que pretende, además de explicar la realidad, es afirmar que la realidad efectivamente es de esa manera, se trata de un modelo de carácter realista.

⁷ *El mundo [Le monde]* llegó a nosotros publicada en dos partes. El *Tratado de la luz* (Descartes, 1989 [1633]) aborda hasta su capítulo VII la creación de un mundo completo fabulado tal como se viene presentando en el cuerpo de este trabajo, y a partir de ahí una serie de consideraciones de filosofía natural vinculadas a los fenómenos lumínicos. El *Tratado del hombre* (Descartes, 2011 [1633]) posiblemente estuviera pensado para ser la continuación de ese capítulo VII, y consiste en la aplicación de lo fabulado para el mundo en general a la creación de un autómata similar a un cuerpo humano.

El uso que Descartes hace de la palabra “fábula”, en francés “l’ invention d’une Fable” (1989 [1633], págs. 96-97) abona la hipótesis de que se trata de un modelo instrumentalista, pero luego de un giro argumental podrá defenderse que Descartes da un salto y le dota a su mundo fabulado un compromiso más fuerte con la estructura de la realidad.

La fábula supone la creación de un mundo de tal forma que todo lo que en él existiera tendría que ser perfectamente cognoscible. El mundo tiene que estar constituido cumpliendo la condición de que las ideas que podamos tener acerca de sus contenidos sean ideas perfectamente claras y distintas, que por tanto cumplirían con el criterio de justificación y aceptabilidad de creencias que Descartes propone en la regla primera del método. En ese sentido se trata en todos los casos de variables cuantificables, y, por tanto, pasibles de objetivación a través de la matemática.

¿Por qué hablar de un mundo creado, modélico, y no de uno ya existente, como el nuestro? Para asegurarse de que no haya nada incomprensible ni inconcebible en él. Todo lo que hay en ese mundo, todo lo que Descartes va agregando poco a poco, ha de ser perfectamente comprensible. Inventado este mundo, la fábula funciona como una inferencia en la que se parte de la concebibilidad de un objeto para luego concluir su posibilidad fáctica, recurso que el autor ha utilizado en varias ocasiones: “...aun cuando nada haya en él de común con el antiguo mundo, no obstante Dios puede crearlo en uno nuevo, ya que es cierto que puede crear todas las cosas que podemos imaginar.” (1989 [1633], pág. 107).

El mundo fabulado en el *Tratado de la luz* se trata de lo que en la jerga filosófica conocemos como un mundo posible. Será necesario acudir a las *Meditaciones metafísicas* (2011 [1641]) para asistir a la demostración de la existencia de Dios, cuya existencia garantiza que además de ser un mundo posible, es efectivamente nuestro mundo real. Así, el modelo trasciende su carácter meramente instrumental, y se constituye en un proyecto de fundamentación

metafísica de la metodología de la nueva ciencia que tendrá como fin explicar cómo se compone nuestro mundo real y bajo qué leyes ocurren los hechos que le conforman.

Mientras que la física aristotélica explicaba los fenómenos naturales apelando a la determinación de la naturaleza esencial de las cosas que participan en ellos, multiplicando al infinito las explicaciones; Descartes va demostrando, a través de sucesivas aproximaciones, que todos los fenómenos de la naturaleza (físicos, químicos y biológicos) pueden entenderse si postulamos que el mundo físico está conformado por una materia de naturaleza única y homogénea⁸. De ese modo, todo aquello que se muestra a nuestros sentidos, generando una sensación, puede ser explicado en términos de la extensión, figura y movimiento heterónomo de partículas materiales que chocan entre sí. Ese movimiento estaría regido por leyes extrínsecas a la naturaleza de la materia: ley de inercia, de acción y reacción, y de movimiento rectilíneo⁹. Afirma Descartes: "...hay un medio para explicar la

⁸ Cfr. especialmente para profundizar sobre este punto a Clarke (1986) y Turró (1989), además de la propia obra.

⁹ Las tres leyes del movimiento propuestas por Descartes en el Tratado de la luz han constituido un antecedente de la física newtoniana que de hecho generó un intenso debate entre cartesianos y newtonianos, especialmente durante el siglo XVII. La ley de inercia notoriamente ya estaba presente en la obra de Galileo, y posiblemente su germen en el pensamiento de otros filósofos y teólogos del Renacimiento. Las otras dos leyes habrían sido inferidas de la naturaleza de la sustancia extensa y de Dios.

La física aristotélica consideraba al movimiento como pasaje de potencia a acto, y en este contexto el reposo absoluto es el estado natural de cualquier objeto que no es expuesto a alguna causalidad eficiente. Al reducir todo modo de movimiento a cambio de lugar comienza a considerarse el movimiento rectilíneo uniforme como estado natural de los objetos. La posibilidad de un objeto que se mueva indefinidamente en movimiento rectilíneo seguramente llevó a Galileo (y a Giordano Bruno anteriormente) a concebir el cosmos como un espacio infinito, problema que Descartes explícitamente evade en su *Tratado de la luz*, pero que afortunadamente no es relevante para los fines de esta exposición.

Por otra parte, la inmutabilidad de Dios es el fundamento del principio de conservación de la cantidad de movimiento que está en la base de la ley de acción y reacción, así como del movimiento rectilíneo. El movimiento natural fundado en la acción de Dios es el rectilíneo, dado que lo recto es

causa de todos los cambios que ocurren en el mundo y de todas las transformaciones que se dan en la Tierra.” (1989 [1633], pág. 63). Será interesante reconsiderar esta afirmación en la medida que avancemos a través del capítulo.

En el *Tratado del hombre* (2011 [1633]), que supuestamente iría intercalado luego del capítulo VII del *Tratado de la luz*, Descartes aplica los mismos postulados con los que venía trabajando de forma específica al cuerpo humano. Propone un modelo, un autómatas, con las mismas características del cuerpo humano que solemos percibir en nuestra vida cotidiana, y por hipótesis totalmente cognoscible. De manera análoga, luego de la demostración de la existencia de Dios podemos atribuir realismo al modelo. Al igual que con el resto de los fenómenos naturales, se trata de demostrar que es posible explicar toda la fisiología del cuerpo humano en términos de extensión, figura y movilidad de las partículas de sustancia extensa¹⁰. Si bien la explicación reductiva, mecanicista, se ha constituido y ha perdurado como una estrategia fuertemente influyente para la conformación de explicaciones biológicas (Barberis, 2012), veremos que la cuestión puede complicarse cuando ingresamos al funcionamiento de las funciones cognitivas. Pasemos a considerar con mayor detalle ese punto.

más simple que lo curvo, como puede constatarse en la simple comparación de las ecuaciones correspondientes a trayectorias rectas y curvas. El desvío de la trayectoria rectilínea solo es posible por la interferencia de otras partículas, ya que la creación de las partículas y su movilidad inicial es un acto de una sola vez y no un acto continuado de creación.

¹⁰ Se trata de una obra que, al igual que la quinta parte del *Discurso* de 1637, tiene una fuerte inspiración en *De los movimientos del corazón* (1994 [1628]), de William Harvey. En dicha obra, clave para la historia de la revolución científica de los siglos XVI y XVII, Harvey demuestra el movimiento circulatorio de la sangre, y supera la teoría galénica de los fluidos. Se trata de la superación de la medicina de Galeno, autoridad indiscutible en la materia durante toda la Edad Media. Para seguir este interesante derrotero, que está en el germen de la presente tesis, sugiero confrontar (Beaujeu, 1985) (Crombie, 1987) (Garrison, 1966) (Lain Entralgo, 1948) (Mason, 1985) (Papp, 1970).

1.2 Las funciones cognitivas del autómeta cartesiano

En este apartado expondré cómo Descartes provee una explicación mecanicista de las funciones cognitivas del autómeta creado, de manera similar a la que concibió para la naturaleza en general. Intentaré demostrar que se trata de explicaciones que, como resulta evidente, son totalmente anacrónicas en sus detalles, y, sin embargo, conforman una estrategia razonable incluso a los ojos de nuestros días. Vale la pena entrar en mayor detalle, dado que lentamente ingresamos al ámbito de lo mental, foco de atención de la presente investigación.

En el *Tratado del hombre*, y años más tarde en *Las pasiones del alma*, Descartes aborda, entre otras cosas, una explicación mecanicista de diversas funciones que hoy llamaríamos cognitivas. Entre ellas destaco para nuestros fines: sensación, percepción, memoria, imaginación, acción voluntaria, control motor. Todas ellas dependen especialmente de la configuración anatómica del cuerpo (tejidos, tubos, fibras, vasos, poros, concavidades, intersticios, etc.) y de los espíritus animales que se movilizan con mayor o menor facilidad a través de ellos. De manera similar a lo que ya proponía Galeno, Descartes opina que la sangre arterial penetra al cerebro no solo con el fin de alimentar y conservar su sustancia, sino también “para producir en él un determinado hábito muy sutil o, más bien, una llama muy viva y muy pura que llamamos *espíritus animales*” (2011 [1633], pág. 681). El movimiento de dichos espíritus animales no es diferente al movimiento de cualquier otro objeto que haya en el cosmos, son cosas extensas y están regidas por las leyes del movimiento que fueron expuestas en el *Tratado de la luz* y que repasamos más atrás.

Cuando los principales vasos penetran la cavidad encefálica se ramifican y reúnen alrededor de la glándula pineal. Las partículas más grandes y groseras de la sangre tienden hacia la corteza cerebral y la alimentan, mientras que las partículas más pequeñas atraviesan unos poros a través de los cuáles pueden entrar y salir de la glándula, e incorporarse al sistema nervioso. Estas partículas sutiles de sangre, que ahora pasan a circular por los nervios y sus ramificaciones, no son más que los

espíritus animales de cuya acción y movilidad dependen las mencionadas funciones cognitivas tanto sensitivas como motoras¹¹.

Llegados a este punto estamos en condiciones de realizar una reconstrucción sintética de cómo funcionan, en este autómata, algunas funciones cognitivas fundamentales. La sensación es, para Descartes, una forma de pensar. Más aún, la manera como el autor entiende el pensar, la noción representacional de “idea”, parece estar fuertemente atada a la noción de “sensación”. Descartes (2011 [1633]) entiende que la sensación es un proceso que ocurre de la siguiente manera: los objetos del mundo impactan sobre las superficies sensibles de los órganos sensoriales -retina, piel, tímpano, lengua, mucosa olfativa-, y el choque transfiere el movimiento a los espíritus animales que se trasladan a través de los nervios aferentes hasta llegar a la glándula pineal. Algunos de estos espíritus podrán desviarse -gracias a la particular organización del sistema nervioso- hacia vías motoras y accionar fibras musculares para provocar movimientos reflejos, como por ejemplo dirigir la mirada hacia un sonido. La glándula pineal está encargada de recibir los espíritus animales provenientes de las diversas modalidades sensoriales e integrarlas para la formación de un percepto único y consistente, etapa que puede fallar produciendo inconsistencias entre la sensación y el mundo real, tal como establece el autor en su duda metódica de las *Meditaciones*¹².

Salvando las distancias, algunos elementos de la manera como hoy tenemos de comprender el proceso de la sensación ya estaban presentes, en germen, en el pensamiento cartesiano. El autor explica que lo que hoy llamaríamos sistema

¹¹ Lo que diferencia esta concepción de la de Galeno es que no hay solución de continuidad entre sangre y fluido nervioso: este último es simplemente el conjunto de partículas más sutiles de la sangre, mientras que Galeno suponía que existía una cocción y por ende un cambio de naturaleza en la transición entre un fluido y el otro. Es algo coherente con la manera que tiene Descartes de entender de forma unitaria la naturaleza de los elementos de la materia a diferencia de la manera aristotélica.

¹² Cfr. también Clarke (2005).

nervioso del autómatas está conformado de tal manera que la estructura del objeto es reproducida por la estructura del estímulo proximal, por ejemplo, por la forma peculiar como la luz impacta en la superficie retiniana, y que esta estructura es respetada y conservada por las fibras nerviosas que son atravesadas por espíritus animales. Lo que hoy llamaríamos señal nerviosa, el potencial de acción, está representado en la neurofisiología cartesiana por la propagación de los espíritus animales oportunamente impulsados por el estímulo a través de los túbulos de la vía sensitiva para alcanzar la corteza cerebral, y luego a la glándula pineal. Este pasaje respeta la estructura fundamental del objeto¹³. La glándula pineal es capaz de identificar qué túbulos transportaron espíritus animales y cuáles no, logrando identificar la figura del estímulo. También la cantidad de movimiento de los espíritus en cada túbulo representa la intensidad de la luz que impactó en determinada zona de la retina, y lo mismo puede aplicarse para la intensidad de los sonidos y para las demás modalidades. La identificación de bordes para la individuación de objetos podría surgir de identificación de diferencias abruptas de cantidad de movimiento de los espíritus de un sector y de los de un sector (plano, línea, etc.) contiguo. La ubicación de dichos objetos en el espacio es posible en relación a la posición de nuestros miembros, cabeza y la dirección hacia la que apunta el rostro, y esos datos también son provistos por la señal nerviosa que luego la glándula pineal integra, como aporta Clarke (2005, pág. 59) a partir de su lectura de la *Dióptrica* de Descartes¹⁴.

¹³ Cfr. especialmente Descartes (2011, pág. 717).

¹⁴ Puede vincularse la cantidad de movimiento de los espíritus correspondientes con la claridad de la idea, y la nitidez de los bordes con la distinción de las ideas, para encontrar la fundamentación mecanicista de esas variables tan determinantes -claridad y distinción- en la epistemología cartesiana. También podemos intuir que la capacidad que tengamos de procesar a través de las categorías matemáticas dichas intensidades y bordes es la que determina el grado de certeza de las ideas que tenemos, que como sabemos es vulnerable al error para los datos crudos de los sentidos, pero gana en certeza cuando dichos datos de los sentidos son la materia sobre la que se aplica la

La conservación de la estructura del estímulo no significa que la señal nerviosa sea igual al objeto, pero sí que se conserva un patrón estructural del mismo tal que hay una correspondencia uno a uno entre forma del objeto, forma de la imagen impresa en la glándula pineal. Esto último se ratifica en el establecimiento de la distancia idea-cosa, en cuya relación habrá analogía mas no igualdad. Al llegar los espíritus a las estructuras centrales, las figuras que se trazan en la superficie de la glándula pineal serán las que conformen ideas para el alma racional. De ahí que Descartes establezca que “quiero sobreentender normalmente, bajo el nombre *idea*, todas las impresiones que pueden recibir los espíritus al salir de la glándula H [pineal]” (2011 [1633], pág. 718); una idea es esencialmente un patrón cerebral, para cualquiera de las modalidades sensoriales.

Por lo dicho, una sensación es un determinado patrón generado en el sistema nervioso periférico, que al ser transportado conserva su forma, y que se recibe centralmente para ser la base de la conformación de la idea. En esta última etapa cumplen un rol clave para la categorización de los datos sensoriales los conceptos con los que cuenta el sujeto¹⁵.

Hoy sabemos que en las áreas corticales sensitivas de la corteza cerebral hay secciones destinadas específicamente al procesamiento de la información proveniente de zonas precisas de las superficies sensoriales. A diferencia de Descartes, que directamente explica la formación de los patrones por la disposición de las fibras y túbulos del sistema nervioso, los modelos actuales entienden que se

matemática sobre las cualidades que hacen a la sustancia extensa. Tal aplicación de los conceptos matemáticos -que son verdaderamente trascendentales en un sentido kantiano- al procesamiento de la idea sensorial es lo que hace posible nada más y nada menos que la ciencia física moderna que el autor quiere fundamentar.

¹⁵ Qué papel cumplen estos conceptos en la sensación y la percepción es todo un gran problema limítrofe entre lo filosófico y lo neurofisiológico que no abordaré aquí en profundidad. A partir de la página 104 considero una crítica al AC en el que se ponen en juego los conceptos, y en particular los llamados conceptos fenoménicos, es decir, aquellos conceptos indéxicos que permiten la conformación de juicios perceptivos.

trata de mapas corticales auto-organizados gracias a aprendizaje no supervisado¹⁶. Las neuronas y sus conexiones se disponen de forma tal en la corteza cerebral que su disposición topográfica refleja las relaciones intrínsecas de las categorías del mundo real que se representan y procesan en el área correspondiente. Un proceso auto-organizativo puede crear esta analogía entre la topografía de la corteza cerebral y la “topografía semántica” de la información a procesar. Entre los mapas corticales más conocidos se destacan los mapas somatosensoriales o somatomotores, el célebre homúnculo de Penfield, los mapas retinianos, o de distribución de colores en las áreas visuales primarias y subsiguientes. Más allá de la diferencia respecto al mecanismo que permite la conformación de los mapas corticales, es notable que la idea básica de la conservación de la topografía semántica a lo largo de las vías sensoriales ya estuviera presente en la obra de Descartes.

En cualquier caso, lo que viaja a través del sistema nervioso es el patrón, y los espíritus animales no son más que el medio físico de propagación de la señal, como lo es el potencial de acción en las teorías actuales. Aun así, su rol es indispensable siendo que sin tal medio no hay patrón que transmitir. En consecuencia la sensación parece ser una función cognitiva fundada en hechos que ocurren en la sustancia extensa, al menos hasta la etapa de intervención de los conceptos, y por tanto explicable en términos de explicación reductiva mecanicista: explicar la sensación es explicar el mecanismo que la hace posible, esto es, la extensión, figura y movilidad de las partículas que interactúan desde que la luz impacta al objeto (para el caso de la visión) hasta que los espíritus animales llegan a la glándula pineal. Hoy hablaríamos de patrones de disparos de potenciales de acción que se trasladan a través de las vías sensitivas, patrones que están dados por la distribución de la superficie activada del órgano sensorial, así como por la frecuencia de los disparos en cada haz de neuronas; sin embargo, obviando ese

¹⁶ Cfr. por ejemplo Trappenberg (2010), o Kandel y otros (2013).

refinamiento producto de los avances actuales de la neurociencia básica, impresiona como el núcleo duro explicativo ya está presente en la fisiología cartesiana.

Las demás funciones cognitivas pueden ser entendidas, en líneas generales, utilizando las mismas herramientas. Las respuestas motoras voluntarias a los estímulos sensoriales siguen el camino inverso. La glándula pineal impulsa espíritus animales por vías nerviosas eferentes, estos llegan a las fibras musculares adecuadas para provocar que se inflen (contracción muscular) o que se ablanden (relajación muscular). La coordinación a gran escala de contracciones en fibras sinérgicas y relajaciones en fibras de músculos antagonistas, y su alternancia rítmica, permite producir movimientos complejos como el caminar, el masticar, o simplemente girar la cabeza para atender un objeto. Todo el proceso de control motor, incluyendo el de la contracción y relajación muscular puede explicarse bajo el ala del modelo mecanicista.

Para que los espíritus animales puedan ingresar a la glándula, han de atravesar ciertos poros específicos. Este hecho hace que una vez que nuestros sentidos captan tal o cual propiedad de algún objeto, por ejemplo determinada forma cúbica en el cuadrante superior derecho del campo visual, tales poros permanezcan semiabiertos más allá del episodio. Como consecuencia los espíritus animales podrían atravesar esos poros que permanecieron abiertos sin necesidad de que el objeto esté presente. Aquí radica la base mecánica de la memoria, la imaginación y el aprendizaje perceptivo.

En el caso de la imaginación y la memoria, estamos ante la conformación de ideas por mecanismos similares a los de la sensación, pero sin presencia del objeto evocado.

Las huellas de esas ideas van pasando por las arterias hacia el corazón y se irradian así por toda la sangre, (...) me contentaré con deciros también cómo se imprimen en la parte interior del cerebro (...) donde está la *sede de la memoria* (2011 [1633], pág. 719)

Dado que los poros han quedado marcados y han permanecido semiabiertos desde que el sujeto estuvo expuesto al objeto, ahora el sujeto por su propia voluntad puede movilizar espíritus animales hacia esos mismos poros evocando la imagen sin necesidad de que el objeto esté presente. Es por ello que solo podemos imaginar aquello que vimos alguna vez. Asimismo, es por ello que en la Meditación Sexta Descartes entiende que sería imposible imaginar que podría existir sin un cuerpo, aunque sí podría concebirlo (2011 [1641]) (2011 [1649]). La imaginación es un fenómeno enteramente dependiente de la sustancia extensa que compone el cuerpo. Los mecanismos que producen la imaginación son precisamente los mismos que explican el ensueño, las alucinaciones febriles y fenómenos similares.

Entiendo por aprendizaje perceptivo el hecho de que la repetición de la exposición a determinado objeto hace que en cada nueva ocasión podamos tener una percepción más aguda y detallada del mismo. El concepto nos permite entender, por ejemplo, cómo un especialista en un campo puede percibir detalles que un lego no logra ver aunque se esfuerce. Es lo que ocurre cuando un ecografista, que fue entrenado e incorporó la teoría correspondiente, detecta estructuras y movimientos en la imagen que el paciente solo entiende como figuras borrosas y confusas, o cuando un guardavidas o un surfista detecta las corrientes de retorno en el mar cuando un turista tan solo puede ver olas. Estos mecanismos de aprendizaje se aplican en la programación de sistemas de Inteligencia Artificial que pueden aprender de manera no supervisada a individuar e identificar imágenes -caras, malezas en un cultivo, colonias de células tumorales en un órgano, etc.-.

Los espíritus animales impactan “...no desde el primer momento con la misma facilidad ni la misma perfección como en la glándula H, sino mejorando poco a poco, según sea su acción más fuerte y de mayor duración o sea más repetida” (2011 [1633], pág. 719). Hoy le llamamos “plasticidad sináptica” al proceso que puede hallarse en la base de este fenómeno, y consiste esencialmente en que -gracias a una serie de procesos a nivel molecular- la repetición frecuente de una entrada sináptica influye en su peso sináptico, variable que representa la efectividad de la misma. Siguiendo el modelo clásico que debemos a Donald Hebb,

a mayor peso sináptico, mayor es la probabilidad de que un potencial de acción entrante de una neurona determinada provoque que la neurona postsináptica dispare ella misma un potencial de acción saliente. Esta modalidad va conformando verdaderas autopistas para las señales que aparecen con frecuencia, y especialmente para casos en los que lo que se repite es la repetición conjunta de señales, cumpliendo con el lema “*what fires together, wires together*”. Éste último apunte explicaría el mecanismo que reside tras el proceso cognitivo de la asociación de ideas, que como sabemos es una de las bases fundamentales del aprendizaje.

Nuevamente Descartes acierta con su propuesta en lo esencial del proceso, y esto vale para todas las funciones cognitivas reseñadas. Cada proceso cognitivo se puede explicar por el movimiento de ciertas partículas (hoy agregaríamos cargas eléctricas) a través del sistema nervioso, más allá de las novedades que haya tenido que aportar la ciencia contemporánea respecto a la manera como esos procesos ocurren. De las varias funciones cognitivas que repasamos, profundicemos en la sensación, ya que por una parte representa el mecanismo básico que está en la raíz de las demás funciones, y por otra parte es especialmente importante a los fines de nuestra indagación.

1.2.1 Las sensaciones y la brecha explicativa

Las funciones cognitivas comentadas anteriormente reciben por parte de Descartes una explicación, más o menos satisfactoria, en términos mecanicistas. Son fenómenos vinculados a la estructura física del cuerpo humano, y como ha postulado que el mundo cognoscible es sustancia extensa, la ciencia está avalada para explicar lo mental en términos de extensión, figura, movimiento regido según leyes, choques, disposición de los órganos, todas variables pasibles del análisis matemático que requiere la certeza exigible a una explicación científica. Dado que la explicación de la sensación parece ser la que proporciona el modelo sobre el que se apoya la explicación de las otras funciones mencionadas, en lo que sigue me

centraré en ella. Partamos de la siguiente afirmación que hace el autor en la Meditación Sexta¹⁷:

Pero cuando examinaba por qué la tristeza en el espíritu se sigue de un no sé qué sentimiento de dolor, y el gozo nace del sentimiento de placer, o bien, por qué esa no sé qué emoción en el estómago que llamo hambre nos da ganas de comer, y la sequedad en la garganta nos da ganas de beber, y así lo demás, no podía dar ninguna razón de ello, a no ser que la naturaleza me lo enseñaba así; porque en verdad no hay afinidad, ni relación alguna (al menos que yo pueda comprender) entre esa emoción del estómago y el deseo de comer, así como tampoco entre el sentimiento de la cosa que causa el dolor y el pensamiento de tristeza que ese sentimiento hace nacer. (Descartes, 2011 [1641], pág. 209)

El pasaje señala claramente una limitación en el poder explicativo del modelo mecanicista. Da la impresión que el modelo es efectivo para explicar cómo se produce una sensación, pero se detiene en un punto en el que parece reinar cierta arbitrariedad. ¿Por qué un dolor no solo es la detección de un deterioro en los tejidos, sino que también produce una emoción, un malestar, que tiene un carácter fenoménico peculiar? ¿Por qué el hambre, que me lleva a adoptar ciertos comportamientos como el buscar alimentos, se siente de la manera que se siente, como esa “emoción en el estómago”? ¿Por qué la necesidad de líquidos suscita esa sensación que llamamos sed, y que se suma al mero dato de deshidratación y los comportamientos pertinentes para solucionarla?

La terminología es algo oscura para lo que acostumbramos hoy, pero parece legítimo interpretar que lo que Descartes se está preguntando es por qué un estado mental como la sensación, que cumple la función de detectar y reportar hechos que ocurren en el cuerpo, está acompañado de un carácter fenoménico tal que hay una manera peculiar de estar en ese estado. ¿Por qué no es suficiente con el acceso al mero dato? ¿Por qué el acceso al mero dato tiene que sentirse así como se siente? Descartes constata que hay una manera peculiar de sentir el dolor, el hambre y la

¹⁷ Invito al lector, aunque no es imprescindible, a contrastar con el pasaje del *Tratado de la luz* citado en la p. 11.

sed, pero no logra explicar por qué eso es así ni cómo se produce ese estado cualitativo tan característico¹⁸. Para continuar, resultará especialmente esclarecedor revisar una distinción que Descartes hace de diferentes grados o niveles de sensación, distinción que nos permitirá situar con precisión en donde se ubica el límite epistémico señalado.

1.2.1.1 Los niveles de la sensación

En las respuestas a las sextas objeciones a sus *Meditaciones*, Descartes establece que es posible distinguir tres niveles o grados de la sensación. Tal distinción se corresponde punto por punto muy bien con una esquematización contemporánea de las etapas del procesamiento visual, cuyos contenidos iré intercalando con el fin de ir aplicando la terminología propia de la fisiología contemporánea, y de poder valorar en su justa medida las intuiciones que Descartes recogió en su propuesta, tal como lo hice en el apartado anterior.

El primer nivel o grado de la sensación consiste en “...aquello que los objetos exteriores causan inmediatamente en el órgano corporal.” (Descartes, 2011 [1641], pág. 369), e incluye el movimiento de las partículas de ese órgano, que provocadas por el objeto devienen en un cambio en la figura y posición de dichas partículas. También se incorpora a este nivel la transferencia del movimiento a las vías nerviosas aferentes -por ejemplo, los espíritus animales que atraviesan el nervio óptico- y finalmente los movimientos provocados por dichos espíritus en la superficie de la glándula pineal.

¹⁸ Este planteo emerge justo antes de la presentación de su célebre argumentación a favor del dualismo de sustancias en la Meditación Sexta, lo que parece indicar que hay alguna conexión entre el obstáculo epistémico y la adopción de su postura metafísica, más allá de todas las interpretaciones de los críticos que vinculan la adopción del dualismo a su necesidad de acomodar su mecanicismo a la metafísica cristiana.

Es tema de debate actual¹⁹ determinar si en este nivel hay incidencia real de conceptos en el procesamiento sensorial para la conformación del objeto, o si es un proceso enteramente autónomo, de sentido abajo-arriba (desde el sistema nervioso periférico hacia el sistema nervioso central), y cuyo funcionamiento es posible gracias al cableado del organismo que determina ciertas constricciones operacionales útiles para la conformación de un proto-objeto de conocimiento²⁰. Las propiedades del estímulo recogidas en este nivel nunca alcanzan la conciencia, como lo muestra la fenomenología de nuestra experiencia cotidiana: no vemos llanamente bordes moviéndose, sino que las imágenes de las que somos conscientes tienen un mayor grado de complejidad y usualmente están acompañadas de conceptos. Consistentemente con ello, Descartes entiende que en este nivel no hay conciencia sino tan solo la preparación de la materia bruta que conformará el objeto de la conciencia. Menos aún hay conocimiento ni involucramiento de contenidos conceptuales de ningún tipo que pudieran funcionar como moduladores de la sensación.

Una segunda etapa, que también podemos atribuir a los animales no humanos, consiste en

todo lo que resulta inmediatamente en el espíritu por estar unido al órgano corporal así movido y dispuesto por sus objetos [incluyendo] (...) los sentimientos de dolor, de cosquilleo, de hambre, de sed, de los colores, de los sonidos, de los sabores, de los olores, de calor, de frío, y otros semejantes, que en la sexta Meditación dijimos

¹⁹ Cfr. por ejemplo (Pylyshyn, 1999) (Raftopoulos, 2009) (Raftopoulos, 2014) (Macpherson, 2015).

²⁰ Dicho cableado, es decir, la conformación anátomo-funcional del sistema nervioso, permite la ejecución de operaciones tales como la valoración de la intensidad de luces y sombras en distintas posiciones de la retina, identificación de zonas de alto contraste luz-sombra que sugieren presencia de bordes, seguimiento del movimiento, etc., operaciones que participan en la conformación de un esquema primitivo del objeto. Sería un proceso pre-atencional, más allá de algunos procesos automáticos que permitirían enfocar la mirada o dirigir el rostro hacia algún punto del espacio, modificar el diámetro de la pupila, y otros análogos para las demás modalidades, y podemos hallarlo no solo en el ser humano sino también en todos los animales no humanos, cada uno con sus peculiaridades.

que provenían de la unión, por decirlo así, de la mezcla del espíritu con el cuerpo (2011 [1641], pág. 369).

Estamos ante estados mentales que tienen un componente cualitativo que se siente de cierta manera, y que Descartes no puede explicar en términos mecanicistas puros. De ahí la dificultad señalada en el fragmento de la Meditación Sexta, citado en p. 21, que le lleva a postular que este nivel requiere algún tipo de interacción efectiva a través de la glándula pineal entre sucesos físicos que ocurren a nivel de la sustancia extensa (los sucesos del primer nivel) y sucesos que ocurren a nivel de la sustancia pensante que, aún sin participación de conceptos, es en donde acaecen estas proto-ideas.

Raftopoulos (2009) llama a esta etapa “percepción”, y propone que es la encargada de convertir el producto de la etapa anterior en algo que sí pueda ser un estímulo valioso para la cognición. Como este autor se centra en el procesamiento visual, siguiendo a Pylyshyn (1999) llama “visión temprana” al conjunto de los dos primeros niveles. En cualquier caso, estamos ante la conformación de una proto-representación (aún no idea²¹), de origen sensorial, que es no conceptual. Todo o buena parte del producto de esta segunda etapa, es el contenido fenoménico que puede ser evocado, por lo que lo llamaremos de aquí en más “conciencia fenoménica”, y que incluye relaciones espaciales, posición, orientación, movimiento, tamaño, forma, textura y color. Podemos asumir que lo análogo ocurre para otras modalidades sensoriales con las peculiaridades fenoménicas de un sonido, un sabor, aroma o textura.

²¹ Hay una ambigüedad en el uso de “idea” por Descartes, por una parte entendido como patrón cerebral -atributo de la sustancia extensa- y por otro como pensamiento -atributo de la sustancia pensante-. En carta a Mersenne, Descartes define “idea” de una manera que parece volcar la interpretación de dicha noción como pensamiento con contenido conceptual, y por tanto a ser de naturaleza pensante y no extensa: “Por la palabra “idea” entiendo la forma de cualquier pensamiento por cuya aperccepción [inception] soy consciente del pensamiento en sí” (citada en Clarke (2005, pág. 63)). La idea requiere ir más allá de la conformación del patrón cerebral, requiere que tal patrón haya servido para la formulación de un pensamiento con carácter proposicional en el alma racional.

Finalmente, la tercera etapa ocurre enteramente en el pensamiento, y consiste en la elaboración de un juicio perceptivo, en cuya construcción claramente participan conceptos que se aplican, cual categorías kantianas, al contenido no conceptual que resultó del segundo nivel. La proto-idea del segundo nivel de sensación ya tenía algún color, o alguna forma con bordes definidos, se movía (o no), y se distinguía del fondo. Ahora podemos formular una opinión, por ejemplo, que ese objeto que tenemos delante tiene tal o cual color, se mueve hacia determinado lugar, tiene forma esférica, es mayor que ese otro objeto, etc.

Para Descartes, esta es una etapa que depende por entero del entendimiento, y por tanto no deberíamos genuinamente llamarlo “sensación”, aunque es notorio que el autor opina que el entendimiento opera porque hubo previamente una sensación bruta (primer nivel) que se constituyó en una percepción no conceptual con carácter fenoménico (segundo nivel), y que es sobre este resultado intermedio que opera el entendimiento aplicando los conceptos que permiten la elaboración de juicios perceptivos. Solo a esta altura podemos predicar verdad o falsedad, por lo que solo al producto de esta tercera etapa podemos darle algún valor epistémico. Por el contrario, el resultado de la segunda etapa simplemente era algo que se siente, por lo que le cabe el ser indubitable en sus contenidos sin detenernos en la correspondencia o no con los objetos que sean su causa. En palabras de Raftopoulos,

Todos los procesos visuales subsiguientes caen bajo el término *cognición*, e incluyen tanto la interfaz post-sensorial/semántica en la que las unidades de reconocimiento de objetos intervienen, como procesos puramente conceptuales que llevan a la identificación y reconocimiento del arreglo (visión de alto nivel). (Raftopoulos, 2009, pág. 51)

Raftopoulos, que en su propuesta está recogiendo los aportes más actualizados de la neurofisiología de la visión contemporánea, coincide con Descartes en que este nivel indudablemente depende de conceptos que otorgan valor epistémico a la idea resultante, y afirma que tales conceptos pueden aplicarse porque un proceso atencional, ahora dirigido en sentido arriba-abajo, pudo aislar,

identificar y conceptualizar tanto al objeto mismo como a sus propiedades, aplicando significados a los contenidos fenoménicos.

En la Meditación Tercera puede leerse:

Entre mis pensamientos, algunos son como las imágenes de las cosas, y sólo a estos les conviene con propiedad el nombre de ideas: como cuando me represento un hombre, o una Quimera, o el Cielo, o un Ángel, o a Dios mismo. Otros, además, tienen otras formas: como cuando quiero, temo, afirmo o niego; es cierto que concibo entonces algo como sujeto de la acción de mi espíritu, pero con esa acción le añado también alguna otra cosa a la idea que tengo de aquello; y de este género de pensamientos, unos son llamados voluntades o afecciones, y otros juicios. (Descartes, 2011 [1641], pág. 180)

Nótese que persiste la ambigüedad terminológica, pues aquí se refiere por “idea” a la representación del segundo nivel, pero que a pesar de ello sí está muy clara la distinción entre dicha representación de segundo nivel y el pensamiento del tercer nivel, que perfectamente puede comprenderse como una actitud proposicional al que se le agrega el carácter intencional. Mientras que la percepción del segundo nivel no tiene valor epistémico, y se agota en sí misma, el juicio observacional es acerca de algo, y por tanto puede o no coincidir con el caso.

En la Meditación Sexta, la distinción entre estos dos niveles es reforzada cuando se establece que es imposible imaginar un polígono de mil lados, pero sí se lo puede comprender (Descartes, 2011 [1641]). De esta forma, la imaginación parece ser un proceso que se inicia en el primer nivel y, en el ejemplo, se interrumpe. La limitación para imaginar tal polígono es más una limitación técnica, física, de conformación y funcionamiento de nuestro sistema nervioso, que una limitación metafísica. Podemos comprender conceptualmente el miriágono gracias a las operaciones puras del entendimiento, que posiblemente haya obtenido información visual o táctil de polígonos de menor complejidad y puede operar sobre ella para establecer juicios sobre el mismo sin intervención del cuerpo, por lo que es un proceso que nació en el tercer nivel. Es la sustancia pensante actuando en solitario para establecer juicios de carácter matemático con máximo grado de certeza.

El entendimiento puro, en este tercer nivel, puede corregir en sus juicios lo que surge de algunas percepciones engañosas. Por ejemplo, al percibir que una varilla de madera está rota cuando está sumergida en un vaso transparente con agua, el entendimiento es capaz de formular un juicio acertado y teorizar acerca de la refracción de la luz. Tal reformulación no impide que la imagen producida en el segundo nivel de la sensación permanezca igual. Fenómenos como este han sido materia prima de argumentos a favor de la impenetrabilidad cognitiva de la percepción, como ocurre con la persistencia de numerosas ilusiones ópticas aún luego de que comprendamos su funcionamiento^{22 23}.

Establecidas las tres etapas de la sensación por Descartes, y enriquecida su teorización por los aportes de la neurofisiología contemporánea, podemos precisar que es en la transición entre la primera y segunda etapas que el poder explicativo del mecanicismo cartesiano se detiene. Todo lo que sucede entre las superficies receptoras de los órganos sensoriales y la glándula pineal resulta explicable. No parece ser que ocurra nada más en esta etapa que espíritus animales chocando entre sí impulsándose los unos a los otros a través de túbulos cuya ubicación permite el procesamiento sensorial primario. Un borde podría detectarse simplemente por un contraste importante del movimiento de los espíritus a lo largo de una línea. Varias líneas podrían permitir detectar una cosa diferenciando el fondo de la figura. Mas a partir de ahí parece suceder algo más que excede la mera maquinalidad del cuerpo, y que por tanto requiere de explicaciones que van más allá de una identificación lineal entre estado mental y estado cerebral. ¿Cómo se transforma el choque de unas partículas sobre otras en una representación que tiene carácter fenoménico? ¿Por

²² Cfr. Fodor (1986), Pylyshyn (1999) o Raftopoulos (2009), (2014).

²³ En el marco de una discusión de su época respecto a los animales no humanos, y a si estos podrían pensar y por tanto tener un alma, Descartes se pronunció negativamente (Clarke, 2005). Para el autor, un animal no humano, que carece de lenguaje y por tanto de capacidad de abstraer conceptos, es capaz de llevar adelante satisfactoriamente los dos primeros niveles de la sensación. Un animal puede, por ejemplo, detectar deterioro o rotura de ciertos tejidos, dicha detección se siente de alguna manera, pero sin identificarlo como dolor debido a la no participación de conceptos.

qué la segunda etapa del procesamiento sensorial es necesaria? ¿No sería posible establecer juicios sobre el mundo sin necesidad de que los mismos estén acompañados por un carácter fenoménico?

Críticos de las obras de Descartes que dudan de la interpretación tradicional, que privilegia los contenidos de las *Meditaciones metafísicas* por sobre sus otras obras de carácter más científico, como ser Clarke (1986), (2005) y Turró (1989), proponen que el dualismo de sustancias no es tanto un pilar fundamental en el sistema cartesiano sino que debe ser pensando como una solución de compromiso ante un fenómeno para el que el mecanicismo no aporta una explicación satisfactoria, a saber: la conciencia.

Más aún, de la conciencia como hecho resistente a la explicación mecanicista se pueden distinguir dos propiedades. Por una parte, una característica de algunos estados mentales que podemos atribuir a la tercera etapa del procesamiento sensorial, es la capacidad que tienen de referirse a algo distinto y trascendente a ellos mismos. Las creencias son siempre creencias acerca de algo, acerca de un objeto. Incluso las creencias que tienen como objeto uno mismo o la propia conciencia pueden referir, gracias a un proceso reflexivo que en primer lugar es conciencia no reflexiva de objeto, y solo luego es conciencia reflexiva de sí mismo. Sobre la propiedad de la intencionalidad, filósofos de la tradición continental y de la tradición analítica han abundado, y la polémica respecto a si la intencionalidad puede ser explicada en términos puramente físicos aún sigue vigente²⁴.

Por otra parte, lo que ocurre en la segunda etapa, que para Descartes requiere la participación de la glándula pineal como interfaz entre extensión y pensamiento, es lo que ocupará nuestra atención. El estado mental consistente en una proto-idea

²⁴ Cfr. por ejemplo (Searle, 1990) (Churchland & Smith Churchland, 1990).

no conceptual que emerge de esta etapa tiene la especialísima propiedad de que, adoptando la terminología acuñada por Nagel (2000 [1974]), hay una manera cómo se siente estar en ese estado. Estos estados mentales que se sienten de cierta manera, podrían funcionar como transición entre el puro mecanismo de la sensorialidad bruta, y los contenidos conceptuales propios de la cognición. Ese carácter eminentemente transicional los coloca en un lugar muy complicado para las estrategias explicativas reduccionistas, pues parecen ser los responsables del salto entre el mero dato y la representación.

Da la impresión de que en dicho salto ocurre algo más que lo que ocurre a nivel mecánico, y ese algo más se resiste a recibir una explicación científica, como profundizaremos en el siguiente apartado. Aquí es donde ocurre la magia de la metáfora, que convierte el agua en vino, la longitud de onda en color²⁵.

1.2.1.2 La brecha explicativa

Joseph Levine (1983), llama “brecha explicativa” (*explanatory gap*) a esta dificultad para proporcionar explicaciones científicas de los estados mentales, basadas en estrategias que los identifiquen con procesos físicos que ocurren en el sistema nervioso del organismo. Se está refiriendo especialmente a las estrategias explicativas actuales más influyentes: fisicalismo y funcionalismo, pero el razonamiento es análogo al que podríamos hacer respecto al mecanicismo cartesiano, que no guarda diferencias de fondo con las anteriores.

Levine afirma que mientras que en enunciados del tipo “El calor es el movimiento de las moléculas”, nuestros conocimientos de física y química nos permiten entender cómo el movimiento de las moléculas es un proceso que puede

²⁵ Para un enfoque interesante sobre el punto véase Nietzsche (1996 [1896]).

cumplir el papel que en nuestra terminología cotidiana asociamos al calor, no puede afirmarse lo mismo respecto a enunciados del tipo “El dolor es activación de fibras C” o “El dolor es el cumplimiento de tal o cual papel causal” (que podrían ser enunciados propios de una teoría neurobiológica fisicalista o funcionalista, como veremos en el siguiente capítulo). Este último enunciado puede tener éxito para establecer que cierto estado físico del sistema nervioso de un individuo se correlaciona de manera sistemática con la afirmación del individuo de que se siente de cierta manera;

Sin embargo, hay algo más en nuestro concepto de dolor que su rol causal, hay un carácter cualitativo, cómo se siente; y lo que queda sin explicación por el descubrimiento del disparo de las fibras C es *¿por qué el dolor debería sentirse de la manera que se siente!* Dado que no parece haber nada respecto al disparo de las fibras C que “encaje” naturalmente con las propiedades fenoménicas del dolor más de lo que podría encajar con algún otro conjunto de propiedades fenoménicas. A diferencia de su rol causal, la identificación de la faceta cualitativa del dolor con el disparo de las fibras C (o alguna propiedad de los disparos de las fibras C) deja la conexión entre una cosa y la otra completamente misteriosa. (Levine, 1983, pág. 357)

Los fenómenos físicos considerados en sentido amplio, por ejemplo, incluyendo los de la óptica, los meteorológicos, los de la dinámica y cinemática, los de la fisiología humana en general, los astronómicos, etc., pueden explicarse perfectamente en términos de la extensión, figura, movimiento e interacción causal de las partículas de materia que conforman la sustancia extensa en su totalidad. Incluso si bajamos al nivel molecular, fenómenos sutiles y complejos como el de la contracción muscular, la interacción de hormonas con sus correspondientes receptores celulares, la sinapsis neuronal, aún siguen explicándose en términos de la estructura tridimensional de las proteínas en juego. Independientemente de que el propio Descartes haya acertado o no con sus propias explicaciones, conceptualmente el modelo ha tenido un potencial que sin duda ha impulsado el desarrollo científico especialmente entre los siglos XVII y finales del XIX, así como los desarrollos tecnológicos asociados que promovieron la revolución industrial y sus concomitantes transformaciones económicas y sociales del mundo.

Los enunciados del tipo mencionado alcanzan a determinar una correlación que ocurre localmente en un individuo, o a lo sumo determinan una correlación de tipos de estados mentales con tipos de estados físicos o funcionales. No obstante, no explican realmente la razón de tal correlación, esto es, la razón por la que ese estado físico produce ese carácter fenoménico que le acompaña y cómo es que lo hace. Por ello, las afirmaciones de Levine se aplican perfectamente bien al proyecto mecanicista cartesiano. La transformación de extensión en imaginaria mental permanece misteriosa por más precisa que sea la evidencia empírica de la neurobiología, y por más elaboradas que sean las teorías que pretenden conectar una cosa con la otra.

Habiendo revisado cómo explica Descartes las diferentes funciones cognitivas del autómatas, que en los hechos pretende ser el hombre real; y habiendo ubicado en qué nivel de la sensación el poder explicativo de su estrategia se agota, produciendo una brecha explicativa; estamos en condiciones de pasar a definir en términos actualizados y precisos en qué consiste el problema difícil de la conciencia, y cómo se diferencia de los problemas fáciles.

1.2 El problema difícil de la conciencia

Hasta ahora hemos avanzado con un uso confuso del término “conciencia”. Una hipótesis de trabajo con la que tendremos que convivir en estado de permanente incertidumbre a lo largo de la presente investigación, es la que consiste en diferenciar dos conceptos de conciencia que están estrechamente relacionados pero se distinguen tanto por el hecho que connotan como por el concepto que denotan (Block, 1995, pág. 227). Digo que tendremos que convivir con la incertidumbre, y tomar esta distinción como hipótesis de trabajo, porque en cierta manera, determinar si se trata efectivamente de dos formas diferenciadas de conciencia será una de las cuestiones en juego en el experimento mental de María.

1.2.1 Conciencia de acceso y conciencia fenoménica

Por una parte, se puede asignar a la conciencia una función vinculada al procesamiento de información proveniente del entorno, así como de otros estados mentales del sujeto (conscientes o no), con el fin de hacerla accesible al pensamiento racional y al control racional de nuestras acciones (incluyendo para el caso de los seres humanos nuestras preferencias). Por otra parte, se puede entender la conciencia como la experiencia consciente que tenemos desde que nos levantamos hasta que nos acostamos, que solo se ve interrumpida durante el sueño o en casos de enfermedad o trastornos muy específicos como el desmayo o el coma, y que se caracteriza por un permanente flujo de estados que poseen cualidades fenoménicas peculiares. Esas cualidades fenoménicas cumplen una función cognitiva, como ocurre cuando detectamos que un alimento está en mal estado gracias a su olor desagradable. No obstante, podría argüirse, la función cognitiva podría cumplirse sin necesidad de que dichos estados tengan su aspecto cualitativo característico. Podríamos detectar que un alimento está en mal estado de maneras diferentes, a través de datos objetivables como la detección de determinadas poblaciones de microorganismos de cierta magnitud. La posibilidad de la consideración por separado de ambos tipos de conciencia y los conceptos asociados, es una polémica sobre la que el AC podría arrojar luz.

Un estado perceptivo es consciente-A²⁶, en términos groseros, si su contenido -lo que es representado por el estado perceptivo- es procesado por la función de procesamiento de información, esto es, si su contenido llega al Sistema Ejecutivo,

²⁶ Por practicidad utilizaré las abreviaciones que Block utiliza para distinguir ambas formas de conciencia. “Conciencia-A” refiere a “conciencia de acceso” y “conciencia-F” refiere a “conciencia fenoménica”. Chalmers (1999) llama “conciencia psicológica” a la primera forma de las mencionadas y “conciencia fenoménica” a la segunda. En lo que sigue preferiré para la primera “conciencia de acceso” tal como la define Block (1995) con el fin de evitar la polisemia del término “psicología” y sus derivados. En cualquier caso, ambos autores están haciendo referencia a lo mismo cuando utilizan uno u otro término.

en donde puede ser utilizado para controlar el razonamiento y la conducta. (Block, 1995, pág. 229)

Para que un estado sea consciente-A debe cumplir tres condiciones que tomadas en su conjunto son suficientes: (1) ser pasible de uso como premisa en un razonamiento, (2) poder servir para el control racional de nuestras acciones, y (3) poder servir para el control racional de nuestras preferencias. Si bien la tercera condición podría identificarse como una forma particular de la segunda, Block considera deseable su consideración aparte para poder incluir a los animales no humanos como agentes conscientes-A, cumpliendo tan solo las dos primeras, por lo que su peso en la conformación del concepto es menor. Las actitudes proposicionales (creer que *p*; desear que *p*, imaginar que *p*...) son los ejemplos más claros de estados mentales conscientes-A. Asimismo, los aprendizajes en general parecen ser procesos que se juegan más a nivel de la conciencia de acceso que la fenoménica (aunque habría que ver si componentes fenoménicos intensos podrían servir de anclaje para que se desarrollen, lo cual sería tema de otro debate con interesantes aplicaciones prácticas).

Por otra parte, Block solo ofrece una definición ostensiva o por mera sinonimia de lo que es la conciencia fenoménica:

...conciencia-F es experiencia. Las propiedades de la conciencia-F son experienciales, esto es, un estado es consciente-F si tiene propiedades experienciales. La totalidad de las propiedades experienciales de un estado son el “cómo se siente” tener dicho estado. (Block, 1995, pág. 230)

La expresión “cómo se siente”, hace referencia al artículo clásico de Thomas Nagel (2000 [1974]), “*What is it like to be a bat?*”. La tesis de Nagel en dicho artículo, justamente hace referencia a que las neurociencias tienden a omitir el

aspecto fenoménico cuando hacen referencia a los estados conscientes de los individuos²⁷.

Ejemplos típicos de experiencias conscientes-F incluyen eso que sentimos cuando vemos, oímos, tocamos, olemos, degustamos, sentimos dolor, sentimos picazón, tenemos sentimientos o emociones. Muchos de los ejemplos mencionados también tienen aspectos vinculables a la conciencia de acceso. Tomemos por caso el dolor, que como bien sabemos se siente de una manera muy particular al punto que distintos tipos de dolor tienen un carácter fenoménico diferente. Pero además los dolores cumplen la función de informar acerca de, pongamos por caso, daños en los tejidos, o la presencia de una cosa potencialmente peligrosa en el entorno, y permite el direccionamiento racional de la conducta, impulsando comportamientos evitativos y luego comportamientos apropiados para la reparación del daño. Por tanto, los dolores, así como las emociones, y los juicios perceptivos, son estados mentales que parecen tener doble vida -de acceso y fenoménica-.

Lo importante respecto al caso del dolor y los otros mencionados, no es tanto que haya casos de conciencia-F pura o conciencia-A pura, sino el hecho de que ambos aspectos son perfectamente discernibles desde el punto de vista conceptual. Se trata de dos formas diferentes de conciencia que tienen propiedades diferentes. Un estado mental es consciente-A en virtud de su aspecto representacional, esto es, que puede ser utilizado como insumo de razonamientos, mientras que un estado mental es consciente-F en virtud de su aspecto fenoménico. Ambas formas de ser consciente pueden coexistir, como cuando uno piensa en “aquel vestido rojo que regaló a su novia” y a su vez lo visualiza en la imaginación, reviviendo las

²⁷ La consideración del planteo de Nagel perfectamente podría haber sido incluida en el cuerpo de esta tesis: un conocimiento científico completo de la anatomía y fisiología del sistema nervioso de un murciélago no nos podría decir nada respecto a cómo se siente ser un murciélago. La incapacidad de conocer el aspecto experiencial del ser un murciélago es lo que nos permite pensar la conciencia de acceso y la conciencia fenoménica como dos conceptos diferenciados que refieren a dos funciones distintas.

emociones de aquel momento, pero el doble aspecto es conceptualmente discernible al punto que podríamos concebir estados mentales con uno u otro conjunto de propiedades.

1.2.2 Dos problemas respecto a la conciencia

Estamos en condiciones de familiarizarnos con la manera ya clásica de Chalmers de distinguir dos grupos de problemas diferentes respecto a la conciencia. Los *problemas fáciles de la conciencia* consisten en explicar los problemas vinculables a la conciencia de acceso. Preguntas del tipo, ¿cómo se producen las funciones cognitivas?, ¿cómo es posible que un individuo reciba e interprete estímulos de lo que ocurre en su entorno y genere respuestas apropiadas? También podemos incluir algunos problemas que incluyen al lenguaje, como lo es el determinar de qué manera un individuo es capaz de poner en palabras sus estados internos. Otros problemas fáciles pueden ser de aplicación práctica: lograr programar un *software* para que aprenda a jugar al ajedrez, o al *go*, o reconocer caras a partir de fotografías, o a predecir la probabilidad de que un cliente de un banco devuelva un préstamo con el fin de otorgar calificación crediticia, ubicar células tumorales y extirparlas, etc. Las aplicaciones son infinitas.

No se trata de que sean problemas fáciles porque no tengan dificultad. Lo que los problemas llamados fáciles tienen en común es que un poco antes un poco después, podrán ser resueltos gracias al avance de las neurociencias y otras ciencias vinculadas. En contraste, “El impresionante progreso de las ciencias físicas y cognitivas no arrojó ninguna luz significativa sobre la cuestión de cómo y por qué el funcionamiento cognitivo está acompañado de la experiencia consciente” (Chalmers, 1999 [1996], pág. 51). Mientras que ya tenemos alguna idea, desde hace buen tiempo, de cómo un sistema conformado por entidades materiales (extenso en términos cartesianos) puede estar configurado para cumplir funciones cognitivas como recordar, aprender, controlar movimientos, hablar, reportar informes de

estado, etc., no sabemos cómo pueden originar la experiencia consciente. Las neurociencias no han podido, y según Chalmers no pueden con estrategias reduccionistas, explicar cómo y por qué la conciencia de acceso en algunas ocasiones está acompañada por experiencia consciente fenoménica. Block y Chalmers coinciden en que las teorías que han intentado dar cuenta de la conciencia a través de una explicación científica, y que mapearemos en el siguiente capítulo, solo han apuntado a dar una explicación de la conciencia de acceso, perdurando la conciencia fenoménica como un misterio.

Ilustremos la cuestión con un ejemplo simple y esquemático. Caminamos por un bosque hambrientos, nos encontramos ante un árbol y tiene frutos. Utilizamos nuestros sentidos para identificar el fruto y chequear si de acuerdo a nuestros conocimientos previos es o no comestible. Decidimos que es comestible y lo comemos, para lo que llevamos nuestro brazo y mano hacia el fruto, lo arrancamos y llevamos hacia nuestra boca. Estamos procesando información que proviene de nuestros sentidos para hacerla accesible a la consideración racional y el control de nuestras conductas²⁸. Todo esto es perfectamente explicable por las neurociencias tradicionales, al punto que podríamos programar una máquina para que realice las mismas actividades. Tenemos estados mentales que hemos de ubicar en la conciencia de acceso porque se trata de información que se hizo accesible a los centros cognitivos responsables de la toma de decisiones y ejecución de acciones. Lo misterioso es que además de todo eso ocurre algo que no parece ser imprescindible en la explicación: al ver, oler, morder el fruto, nos hallamos en un estado que se siente de determinada manera. Nos hallamos en un estado mental con un carácter fenoménico que parece emerger de las propiedades físicas del fruto, y de nuestro sistema nervioso, pero no entendemos cómo una cosa se conecta con la

²⁸ Debemos abstenernos de entender este proceso desde el ver el árbol hasta la degustación del fruto como si se tratara de acciones voluntarias que pasan por una planificación racional y centralizada a cada momento. Al respecto, cfr. por ejemplo a Ryle (1949) que esclarece –desde su propio marco teórico- la distinción conceptual entre acciones voluntarias e involuntarias.

otra. Hay una brecha explicativa que nos impide pasar de las propiedades físicas del fruto, y de la experiencia de comerlo, a las propiedades fenoménicas que sentimos cuando lo hacemos y lo saboreamos u olemos.

Imaginemos ahora que estamos ante tres manchas uniformes de color anaranjado, rojo y azul sobre fondo blanco. La visión de cada una de esas manchas producirá una experiencia que diferirá de las demás en su fenomenología: el “cómo se siente” al ver la mancha anaranjada es diferente al “cómo se siente” ver la mancha roja y lo mismo respecto a la mancha azul. Lo que sí puede decirse es que el “cómo se siente” al ver la mancha anaranjada es más similar al “cómo se siente” al ver la mancha roja que al “cómo se siente” al ver la mancha azul. Se llama “*quale*” (plural “*qualia*”) a la propiedad que posee la experiencia de ver algo rojo que permite que verlo se sienta de determinada forma, propiedad que no posee la experiencia de ver algo anaranjado ni la experiencia de ver algo azul. Puede afirmarse que la experiencia de ver anaranjado también tiene, en parte, el *quale* propio de la experiencia de ver rojo. La experiencia de ver algo azul carece por completo de ese *quale* (tendrá otro) y por eso su experiencia es más disímil en comparación con la experiencia de ver la mancha anaranjada²⁹.

²⁹ Si bien usualmente la discusión filosófica se ha centrado en la experiencia perceptiva del color, la terminología puede aplicarse a la experiencia visual en general y a otras modalidades sensoriales. Otra de las modalidades que fue ampliamente discutida es la sensación de dolor, y cabe señalar que los dolores pueden cambiar tanto en intensidad como en su carácter cualitativo. Las emociones, como el odio, enamoramiento, miedo, también parecen tener su carácter cualitativo peculiar, se sienten de determinada manera. Por el contrario, no es tan claro que otros estados mentales tengan *qualia* asociados, particularmente las actitudes proposicionales, por ejemplo, la creencia de que hoy es viernes. Quizás la creencia de que hoy es viernes puede provocar en mí una emoción que sí tenga carácter cualitativo, pero la creencia misma no parece tener carácter cualitativo. Es plausible pensar que una actitud diferente ante la misma proposición podría portar un *quale* diferente, v.g. desear que hoy sea viernes parece sentirse de manera diferente que creer que hoy es viernes. En el ejemplo mencionado podría afirmarse que una cosa es la actitud proposicional, y otra cosa es la emoción que ese estado cognitivo suscita, por lo que el *quale* podría adjudicarse a la emoción sin afectar la actitud proposicional. También podría afirmarse que existen actitudes proposicionales con algún componente fenoménico asociado, por lo que se trata de un asunto no dirimido. Para una

Desde una perspectiva de primera persona, parece un hecho indiscutible que gran parte de nuestras experiencias conscientes tienen carácter cualitativo³⁰. Sin embargo, ese carácter cualitativo ha sido un permanente escollo para toda postura filosófica o teoría científica que tuviera alguna pretensión de reducir lo mental a otro tipo de propiedades. Aquí se ubica precisamente la brecha explicativa de Levine, o el problema difícil de la conciencia tal como lo formula Chalmers, que busca determinar la relación entre los procesos físicos vinculados especialmente a la estructura y funcionamiento del sistema nervioso, y la emergencia de la experiencia consciente, en su faceta subjetiva y privada, como experiencia que tiene carácter cualitativo o *qualia*. La distinción de matices de colores, con sus infinitas variaciones posibles, sin duda tiene una gran utilidad dado que es nuestra forma de procesar información fundamental para llevar adelante nuestra vida identificando alimentos, posibles peligros, posibles candidatos para la reproducción, etc. Ahora bien,

Desafortunadamente, el tipo de explicación funcional que tiene tan buen desempeño en los estados psicológicos [cognitivos] no parece funcionar para explicar estados fenoménicos. La razón es simple. Cualquiera sea la concepción funcional de la cognición humana que formulemos, siempre queda una *pregunta ulterior*: ¿Por qué este tipo de funcionamiento está acompañado por la conciencia [fenoménica]? (Chalmers, 1999 [1996], pág. 77).

introducción a la cuestión y compilación de artículos relevantes cfr. (Bayne & Montague, 2011). Nuestro foco estará puesto en los estados mentales que poseen un carácter cualitativo, sean del tipo que sean, por lo que esta cuestión no resulta relevante a nuestra discusión: si hay actitudes proposicionales con carácter fenoménico, nuestra discusión las afectará, en tanto se trata de estados que se sienten de cierta manera. Si no las hubiere, entonces no las afectará. Que las haya o no las haya es una cuestión que corre por una vía diferente, por lo que no es necesario prestarle una atención particular a este debate.

³⁰ Chalmers (1999 [1996], pág. 16) llama “tomar en serio la conciencia” justamente a reconocer como evidente de suyo la existencia del aspecto fenoménico de la experiencia consciente, sin pretender que se trata tan solo de una ilusión. De manera análoga, Goff (2017) le llama a esta postura “constricción de la conciencia”. Al tratarse, según su postura, de algo que se nos manifiesta de forma evidente e inmediata, considera que no es necesario ningún argumento específico para demostrar su existencia. De ahí entiende que cualquier intento por establecer una postura acerca de lo que él llama la “realidad fundamental” ha de incluir a la conciencia-F como uno de sus elementos ineludibles. Es en este punto que los dualistas de propiedades se separan de los llamados eliminativistas que niegan que exista tal cosa llamada conciencia fenoménica.

Volvamos ahora al pasaje ya citado de la Meditación Sexta de Descartes, para constatar que lo que hace es precisamente identificar lo que hoy llamamos problema difícil de la conciencia, salvo por el mantenimiento de una ambigüedad respecto a si nos encontramos en el segundo o el tercer nivel de la sensación.

Pero cuando examinaba por qué la tristeza en el espíritu se sigue de un no sé qué sentimiento de dolor, y el gozo nace del sentimiento de placer, o bien, por qué esa no sé qué emoción en el estómago que llamo hambre nos da ganas de comer, y la sequedad en la garganta nos da ganas de beber, y así lo demás, no podía dar ninguna razón de ello, a no ser que la naturaleza me lo enseñaba así; porque en verdad no hay afinidad, ni relación alguna (al menos que yo pueda comprender) entre esa emoción del estómago y el deseo de comer, así como tampoco entre el sentimiento de la cosa que causa el dolor y el pensamiento de tristeza que ese sentimiento hace nacer. (Descartes, 2011 [1641], pág. 209)

¿Por qué y cómo la detección del vacío del estómago tiene que estar acompañada de esa sensación que se siente de una manera peculiar y llamamos “hambre”? Podría ser suficiente para el mantenimiento de la vida con que el dato sea accesible a los centros cognitivos de toma de decisiones. El mecanicismo cartesiano no es más que una posible estrategia de explicación reductiva de todos los fenómenos del mundo y también de las funciones cognitivas del hombre. Como estrategia, es de un potencial más promisorio que el establecimiento de formas sustanciales para cada cosa y proceso que existe. La sensación (en su primer grado), la imaginación, la memoria, el ensueño, el control motor, la atención, los movimientos reflejos, son todas funciones que se pueden explicar en términos de extensión, figura, movimiento y choque de las partículas que participan del suceso. A ello podemos agregar todas las pasiones que Descartes analiza en *Las pasiones del alma* (2011 [1649]). Más allá de que la mayoría de sus explicaciones tienen buen grado de fantasía completando los huecos naturales del estado de la ciencia de su época, y de los instrumentos de observación que tenía a disposición, aquí lo que interesa es que metodológicamente la estrategia es factible para la enorme mayoría de las funciones. La naciente ciencia moderna camina bien y va cumpliendo con sus objetivos hasta que se topa con lo mental. Lo mental sigue representando un misterio, a tal punto resistente a toda explicación mecánica, que Descartes detiene

su explicación para dar paso a la única forma sustancial que conserva de la física de estilo aristotélico: el alma.

El fenómeno de la vida, como actividad de aquellos seres que mantienen su equilibrio interno y pueden reproducirse, también es pasible de una explicación mecánica, y por tanto ya el alma no necesita ser -como para los antiguos griegos- aquello que da vida al cuerpo. El funcionamiento de la máquina da cuenta perfectamente de qué quiere decir estar vivo. Pero el fenómeno del pensar, la capacidad de tener ideas, en particular ideas con carácter fenoménico, escapa a todo mecanismo por más complejo que sea. Se ha debatido si el objetivo de establecer una metafísica dualista pudo tener que ver con conciliar su ciencia con las posturas centrales de la Iglesia Católica relacionadas a la inmortalidad del alma. En lo que respecta a nuestro foco de interés, impresiona que el dualismo de sustancias aparece en la obra de Descartes cuando éste identifica claramente que no hay explicación posible de la conexión entre extensión y pensamiento.

Es fácil advertir en la cita, y en general en la obra de Descartes, que las nociones hoy distinguibles de conciencia de acceso y conciencia fenoménica se encuentran totalmente identificadas. De esa manera podemos leer que la brecha explicativa afecta por una parte a la conciencia-A, v.g. que de la sequedad de la garganta se siguen las ganas de beber, y de la misma forma a la conciencia-F, v.g. de la sensación de placer se sigue la alegría. En conclusión, para Descartes el problema difícil de la conciencia, imposible de resolver para su modelo mecanicista, incluye toda forma de pensamiento, tanto conceptual como no conceptual, y se trata de explicar cómo pasamos del primer grado de la sensación al segundo y el tercero. Como vimos, ya la transición del primer grado al segundo requería el ingreso de los espíritus animales a la glándula pineal, y el tercer nivel ocurría por entero en su interior, lo que es consistente con la idea de que el dualismo de sustancias es el resultado de la limitación del poder explicativo del mecanicismo que debe detenerse cuando tiene que dar cuenta de la mentalidad.

Hoy, a sabiendas de que los estados conscientes-A, que incluyen las actitudes proposicionales propias del tercer nivel de la sensación descrito, pueden contar con explicaciones científicas reductivas, podemos restringir el alcance del problema a la dimensión fenoménica de la mentalidad. El carácter cualitativo de la experiencia consciente parece ser uno de los últimos reductos en el que el misterio aún reina por sobre nuestra capacidad para entender la realidad. Las propiedades físicas y químicas de las cosas, los hechos meteorológicos, la astronomía, la biología, el funcionamiento del cuerpo humano, las funciones cognitivas de los animales y del hombre, el pensamiento superior, parecen ser todos fenómenos que hallan su fuente en la estructura fundamental de la realidad física, y la experiencia consciente perdura como bastión a conquistar. Revivamos aquella pregunta que quedó abierta a lo largo del capítulo: ¿realmente todos los hechos pueden explicarse, como Descartes pretendía, en términos de extensión, figura y movimiento? Para poder precisar más esta última observación pasaré a desarrollar el concepto de superveniencia y sus formas particulares. Una vez hecho esto estaremos en condiciones de tener delimitado nuestro problema filosófico, mapear posibles respuestas, y pasar al análisis del AC en particular.

1.3 La superveniencia y sus formas

El término anglosajón “*supervenience*” fue usado inicialmente en el ámbito de la biología de inicios de siglo XX, tomando como antecedente a los emergentistas británicos, para referir eventos en los que una propiedad nueva emerge como resultado de una organización compleja de nivel inferior. Tales propiedades habían de considerarse como algo por encima y más allá de lo físico, pero a su vez como contando con eficacia causal propia para incidir sobre los hechos del mundo físico. Un ejemplo clásico es la vida, la capacidad de autosustentarse y reproducirse, entendida como emergente de las propiedades físico-químicas de la materia orgánica. El significado del término ha ido variando en relación al contexto de uso. En filosofía hay tesis de superveniencia en metaética, estética, filosofía de

la ciencia en general, etc. Dada la variabilidad de significados, y el hecho de que en nuestro idioma no hay siquiera un uso cotidiano del cuál servirse para obtener intuiciones que permitan su comprensión, conviene limitarnos a lo que ha sucedido últimamente en el contexto de la filosofía de la mente.

Una definición inicial puede formularse de la siguiente manera: un conjunto de propiedades A superviene a un conjunto de propiedades B si y solo si dos cosas no pueden diferir respecto a A sin diferir respecto a B.

Esto significa que no puede haber diferencias en A sin diferencias en B. No se trata tan solo de que no haya diferencias, se trata de que no puede haberlas. Dicho en otra forma, diferencias en A requieren diferencias en B; igualdad en B garantiza igualdad en A. Por ejemplo, si afirmamos que las propiedades biológicas supervienen a las propiedades físicas, lo que estamos diciendo es que independientemente de cuál sea la teoría física vigente en el momento, no puede haber diferencias biológicas sin que haya diferencias físicas, y a la inversa, igualdad física garantiza igualdad biológica.

Las tesis de superveniencia pueden variar de acuerdo a su ámbito de aplicación. Las propiedades A *supervienen localmente* a las propiedades B si no puede haber diferencia respecto a A en un individuo sin que haya diferencias respecto a B. De manera análoga, la tesis de *superveniencia* será *global* si se aplica a las propiedades A y B de un mundo entero, por lo que no habría dos mundos posibles idénticos respecto a A sin que lo fueran respecto a B. Como caso intermedio respecto de los anteriores, las propiedades A *supervienen regionalmente* a las propiedades B, si en una determinada región de la realidad no puede haber diferencias en A sin que haya diferencias en B. Un ejemplo podría ser aplicable a las teorías institucionalistas del arte, por las que las propiedades relevantes de un objeto -v. g. una lata de sopa Campbell- pueden ser supervenientes a lo físico en un pasillo de supermercado y no serlo en un museo. El ejemplo citado corresponde a regiones delimitadas espacialmente. Algo similar ocurre con la jerarquización aristotélica del cosmos en las regiones sub y supralunar.

En otra dimensión, podemos clasificar las tesis de superveniencia respecto a su grado de fuerza modal. Cuando afirmamos que *no puede* haber diferencias respecto a A sin diferencias respecto a B, ese “no puede” es pasible de interpretación bajo el criterio de la legalidad lógica o de la legalidad física. Algunos admiten la posibilidad de niveles intermedios de fuerza modal en una legalidad metafísica. Consideraré a la necesidad metafísica como similar a la necesidad lógica: a nuestros efectos, un mundo metafísicamente posible es lo mismo que un mundo lógicamente posible.

Las propiedades A *supervienen lógicamente* a las propiedades B si y solo si ningún par de situaciones lógicamente posibles son idénticas respecto a B pero diferentes respecto a A. Esta tesis significa que no es concebible un mundo en el que haya igualdad en B pero diferencia en A. Así, un Dios que hubiera creado un mundo, habría hecho todo el trabajo respecto a las propiedades A al caracterizar las propiedades B. Por ejemplo, si asumimos que las propiedades biológicas supervienen lógicamente a las propiedades físicas, o para ser más breves que lo biológico superviene lógicamente a lo físico, creemos que Dios no podría crear dos mundos físicamente idénticos, pero biológicamente diferentes. Bajo esta tesis, conocer todos los hechos vinculados a propiedades B asegura que ya conocemos todos los hechos vinculados a propiedades A.

Las tesis de superveniencia fáctica o nomológica son más débiles. Las propiedades A *supervienen nomológicamente* a las propiedades B si las leyes de la naturaleza prescriben que no hay diferencias respecto a A sin diferencias respecto a B. Sin embargo, según una tesis con este grado de fuerza modal, sí sería posible otro mundo en el que la configuración de propiedades A no esté fijada por la configuración de propiedades B. Así, sería imposible en este mundo, en virtud de las leyes de la naturaleza, que dos objetos 1 y 2 difieran respecto a A sin diferir respecto a B, pero es lógicamente posible un mundo en el que eso sucediera. Por tal razón, el Dios creador de ese mundo tendría que crear las propiedades B de bajo nivel, y luego para fijar las propiedades A de alto nivel tiene que hacer más trabajo, por ejemplo, determinar las leyes de la naturaleza que regirán ese mundo. En otro

mundo posible podría haber un juego diferente de leyes naturales que permitiera tal desligamiento de las propiedades A respecto a las B.

La superveniencia lógica requiere como condición necesaria pero no suficiente a la superveniencia nomológica, por lo que del establecimiento de la primera se sigue el de la segunda, y no al revés. Las condiciones para la primera son más fuertes e incluyen las condiciones para la segunda. Esto es, A puede ser superveniente nomológicamente a B sin ser lógicamente superveniente a B³¹.

1.3.1 Superveniencia y reduccionismo

Las tesis de superveniencia son tesis ontológicas, y tienen su correlato epistémico en el carácter reduccionista de las teorías científicas. Una explicación reductiva es una explicación de un fenómeno en términos de entidades más simples. Por una parte, es intuitivo plantear que, si las propiedades A supervienen a las propiedades B, podemos explicar por qué algo tiene la propiedad A mostrando su configuración respecto a propiedades B. Hacer ciencia con un enfoque reduccionista es explicar las propiedades supervenientes explicando aquellas sobre las que supervienen. Si en algún sentido falla la superveniencia, entonces las explicaciones reduccionistas dejarán algo sin explicar.

Ahora bien, ¿qué fuerza modal requiere la tesis de superveniencia para poder fundamentar debidamente el reduccionismo epistémico? Este punto se ha debatido profusamente. Por ejemplo, Chalmers (1999 [1996]) entiende que el establecimiento de la superveniencia lógica es condición necesaria para

³¹ Esta distinción será la clave sobre la que operarán los experimentos mentales de concebibilidad, como el argumento modal de la Meditación Sexta de Descartes o los argumentos de *qualia* ausentes (Block, 1980) (Chalmers, 1999). Lo que buscan es establecer que, aunque de hecho la variación de ambos conjuntos de propiedades esté ligado de forma invariable, al ser pensable un escenario de desligamiento de la variación de ambas propiedades, debemos concluir que no son idénticas.

fundamentar una estrategia reduccionista. Por tal razón, demostrar la no superveniencia lógica de la conciencia fenoménica sobre lo físico es su estrategia para demostrar que el problema difícil de la conciencia no es soluble mediante los enfoques reduccionistas habituales. En contraste, Churchland (1985) entiende - usando otra terminología- que una reducción conceptual punto por punto (lo que vendría a ser una tesis de superveniencia lógica) es una exigencia demasiado alta para cualquier intento de hacer ciencia y se conforma con una tesis de alcance nomológico en donde las propiedades de cierto tipo se puedan formular en términos de entidades más simples con el establecimiento de las leyes puente apropiadas.

No es tan claro que la superveniencia sea suficiente para la explicación reductiva, porque dependerá de qué tipo de reducción estamos hablando. Si por reducción entendemos identidad (como en la teoría de la identidad psicofísica), o si requiere una relación de implicancia de forma que una configuración de propiedades de bajo nivel implique con fuerza deductiva una configuración de las de alto nivel, entonces la superveniencia lógica no es suficiente para garantizar la legitimidad de explicaciones reductivas. No es claro que la superveniencia lógica suponga identidad entre ambos conjunto de propiedades. Cuando se da B, se da A, pero ¿eso significa que A no es nada más allá de B? Esta discusión también puede ser central al debate sobre la naturaleza de lo mental, pues incluso en el ámbito del materialismo podemos encontrar alternativas reduccionistas y no reduccionistas dependiendo de la respuesta que se dé a estas preguntas, como veremos en el capítulo siguiente.

Otros autores, ahora aumentando el grado de exigencia, opinan que el establecimiento de una relación de superveniencia no es suficiente para fundamentar estrategias reduccionistas. Lycan, citado en (Horgan T. , 1993), acuñó el término “superduperveniencia” para referirse a aquellos casos en los que la superveniencia es explicable. Horgan (1993) entiende que una postura materialista respecto de lo mental no puede contentarse con establecer que hay una relación de superveniencia lógica de lo mental sobre lo físico, sino que tendría que mostrar que tal relación es explicable de manera robusta como una relación de carácter

ontológico y no como una mera identificación conceptual. A su vez los términos de dicha explicación tienen que ser *kosher* [sic] para los materialistas. Por tal razón, quien quiere aceptar que los estados mentales fenoménicos existen realmente, y que se puede dar cuenta de ellos en términos de una teoría materialista, debe lograr dar cuenta de una relación de superduperveniencia de lo mental sobre lo físico, cumpliendo con todas esas exigencias. Este es el nivel más alto de exigencia para una explicación reductiva, ya que requiere no solo dar cuenta de la manera como ambos conjuntos de propiedades se relacionan, sino que pretende dar cuenta de la naturaleza intrínseca de dicha relación.

En síntesis, de la misma manera que las tesis ontológicas de superveniencia pueden presentarse con distintos niveles de fuerza modal, las tesis epistémicas de reducción teórica o conceptual también pueden pensarse con un diferente grado de compromiso. Ya establecido el concepto de superveniencia, sus tipos y derivados, y su relación con las explicaciones reductivas, apliquemos el mismo al sistema cartesiano. Tal ejercicio nos permitirá, en primer lugar, determinar qué tipo de modelo es el modelo mecanicista cartesiano de acuerdo a la fuerza modal de superveniencia que el modelo postule, y, en segundo lugar, determinar el alcance de tal superveniencia y, por consiguiente, si hay algún ámbito de la realidad para el que no podamos establecer relaciones de superveniencia, que resultarían en la resistencia a ser explicados con teorías reductivas.

1.4 El mecanicismo como tesis de superveniencia

Habiendo desarrollado un marco teórico apropiado, estamos en mejores condiciones para comprender cómo emerge el problema difícil de la conciencia en el sistema cartesiano. Descartes, en su *Tratado de la luz* (1989 [1633]), nos propone un mundo en el que todo lo que hay es perfectamente cognoscible, y que está compuesto por materia extensa y divisible, cuyas partículas se mueven reguladas por ciertas leyes. En este mundo, todas las propiedades de los objetos -químicas,

biológicas, sensoriales- supervienen a las propiedades de la materia entendida como sustancia extensa. Como la bondad de Dios no permitiría que vivamos en un mundo tal que nuestras creencias claras y distintas fueran falsas, y como la existencia de Dios es una verdad innata que conocemos por la infalible luz natural de la razón, entonces Descartes entiende que es necesario que el mundo en el que vivimos sea tal como él lo presenta. Un mundo en el que los fenómenos no tuvieran una infraestructura extensa matematizable que los determine, y por tanto un mundo en el que las explicaciones reduccionistas mecanicistas no tengan un grado absoluto de certeza, sería un mundo imposible dado que su creación contravendría la bondad divina.

Provisionalmente, podemos establecer que Descartes está postulando una tesis de superveniencia lógica y global de cualquier tipo de propiedades a lo físico, entendido como extensión. No es concebible que Dios creara un mundo, en el que haya propiedades ocultas que no pudiéramos conocer a través de explicaciones reduccionistas que apelen a la infraestructura mecánica de la realidad, entendida como sustancia extensa.

Un caso problemático resulta de considerar que eventualmente podríamos concebir un mundo diferente al que concebimos. Primero, ¿podríamos concebir un mundo en el que el movimiento se comporta según leyes diferentes? Segundo, ¿podríamos concebir un mundo en el que los limones saben dulces? Da la impresión que se trata de casos diferentes. Respecto al primer caso, en la medida que las leyes del movimiento son inferidas por Descartes directamente a partir de las características de la sustancia extensa y de la idea de Dios, no parece que Descartes admitiera como posible concebir un mundo en el que las leyes del movimiento fueran diferentes. Uno podría afirmar que algo así puede ser producto de una idea oscura y confusa, y que una aplicación detenida del método analítico-sintético permitiría rectificar la concepción hasta volver a que el único mundo concebible con claridad y distinción, es aquel en el que las leyes del movimiento son las que él propone, deducidas de las propiedades de la sustancia extensa y de la idea misma de Dios. Si restamos fuerza lógica al carácter de dicha inferencia, sí podría ser

concebible un mundo con leyes físicas diferentes, pero no parece que fuera fiel al espíritu del sistema cartesiano. Estamos habilitados a decir que el concepto de Dios es el que permite inferir, partiendo de la superveniencia nomológica establecida en el mundo fabulado, la superveniencia lógica. Solo hay un mundo posible y es precisamente este, y por ello lo fabulado no es más que lo real.

De lo dicho se sigue que solo habrá diferencias en los fenómenos observables de cualquier ámbito de la realidad, cuando haya diferencias en la extensión, la figura y el movimiento de las partículas de materia que conforman el mundo. Efectivamente se trata de que las propiedades de todo fenómeno de la naturaleza supervienen lógicamente sobre las propiedades de la sustancia extensa. Aquí hemos de incluir lo astronómico, lo meteorológico, lo físico (en sentido estricto) y lo químico. Cabe una mención especial a lo biológico -incluyendo lo concerniente al cuerpo humano-. La vida es, para Descartes, una propiedad que superviene a las propiedades de la sustancia extensa. Y por ello no es necesario postular la existencia de la naturaleza esencial “alma” entendida como lo hacían los antiguos griegos, esto es, como principio inmaterial dotador de vida. La vida es una propiedad de la máquina puesta en funcionamiento, y no precisa de ningún otro ingrediente que la sustente. Por tal razón aún hoy entendemos “vida” en términos de *autopoiesis*, de sistemas metabólicos enmarcados en límites de permeabilidad selectiva. En conclusión, desde el comportamiento de los cuerpos inertes, hasta el funcionamiento más refinado del cuerpo humano, incluyendo su sistema nervioso, que está organizado de manera en extremo compleja, son al menos en principio explicables en términos de una teoría reduccionista que consiste en exhibir como *explanans* el mecanismo fundamental del que emerge la propiedad en cuestión.

No obstante, la respuesta a la segunda pregunta no es tan clara. ¿Podríamos concebir un mundo en el que los limones saben dulces? Del modelo cartesiano, incorporando su sustento metafísico en la bondad divina, no surgen elementos que prohíban que un mundo así fuera concebible. El caso hace mella justamente en el pasaje del primer nivel de sensación al segundo. Lo que sucede entre el impacto de las partículas del jugo del limón en las papilas gustativas hasta que la señal nerviosa

llega a las áreas cerebrales responsables del procesamiento de tal señal es perfectamente explicable en términos mecanicistas. Mas cómo esa señal, esos unos y ceros, se transforman en experiencia consciente peculiar, privada, incomunicable, permanece un misterio. Hay entonces al menos una serie de propiedades en el mundo que no supervienen lógicamente a las propiedades de la sustancia extensa en el mundo de Descartes: el “cómo se siente” de las experiencias conscientes.

Por esta razón, y siguiendo la lectura de Clarke (2005), es que Descartes se ve obligado a conservar una única naturaleza esencial: el alma. Mas ya no le comprendemos como principio dador de vida, sino como sustancia pensante. En otras palabras: el ámbito de su eficacia causal se reduce, ya que no se trata de eso que permite que ocurra la magia de la vida, sino que ahora pasa a ser eso que permite la magia del pensamiento (y en particular de la experiencia consciente). De lo dicho se explica que el segundo nivel de la sensación requiera ese impacto de los espíritus animales en la superficie de la glándula pineal -sede del alma-.

Aunque parece sensato vincular la adopción del dualismo de sustancias de Descartes a la necesidad que tuvo de conciliar su ciencia mecanicista con la metafísica cristiana, desde el punto de vista estrictamente interno al desarrollo de la ciencia, resulta pertinente entender el dualismo de sustancias como la solución que Descartes halló al problema difícil de la conciencia. Obsérvese que la crítica de Isabel de Bohemia en carta del 16 de mayo de 1643 (2011 [1642]), en la que la filósofa interroga a Descartes por la manera en cómo la sustancia pensante y la sustancia extensa interactúan causalmente, no hace más que poner en evidencia que la solución dualista al problema difícil de la conciencia no logra salvar con éxito la brecha explicativa. Dada la forma como fueron caracterizadas ambas sustancias - de manera que tienen propiedades completamente antagónicas y excluyentes- hace que sea imposible dar cuenta de una relación de superveniencia que pueda tener una explicación robusta en términos aceptables para el mecanicismo (siguiendo las

condiciones de Horgan (1993) para el establecimiento de una relación significativa de superveniencia o superduperveniencia)³².

En síntesis, y aprovechando la terminología incorporada en este capítulo, Descartes niega que la conciencia fenoménica supervenga sobre lo físico, entendido como sustancia extensa. La no superveniencia es lógica, porque es concebible sin contradicción que el limón sepa dulce, y por tanto no es verdad que en todo mundo posible a iguales propiedades mecánicas se sigan iguales experiencias conscientes. También se está negando la superveniencia nomológica, que tiene una exigencia más débil, dado que nuestro mundo es el único posible y no hay un juego claro de leyes psicofísicas que den cuenta de la conexión entre lo mecánico y lo fenoménico. Es por ello que apela al “milagro” de la glándula pineal como interfaz entre sustancias (Descartes, 2011 [1649]).

La no superveniencia tiene como consecuencia posiblemente deseable para Descartes, el mantener intacta la libertad de la voluntad frente al determinismo ineludible que reina en el mundo físico. Pero deja muchos asuntos sin resolver. Además de las contradicciones señaladas por Isabel, el problema de la comunicación de las sustancias y de la innecesaria multiplicación de los entes, otras tantas. La dificultad que surge de tener que emparejar almas a cuerpos en relaciones estables uno a uno; la dificultad de mantener como cierta la ley de conservación de la cantidad del movimiento, y por tanto el cierre causal del mundo físico, y a su vez suponer que el pensamiento puede tener eficacia causal en el mundo.

La ciencia moderna revolucionaria pretendió sustituir todas las formas esenciales del teleologismo aristotélico, por explicaciones que redujeran la diversidad de fenómenos naturales a hechos de un único tipo. Entender la naturaleza como máquina, con el reloj como metáfora, propició el avance vertiginoso de la

³² Para profundizar sobre este punto cfr. Descartes & Isabel de Bohemia (2011 [1642]).

ciencia y la tecnología que disfrutamos hoy día. Durante el siglo XX las estrategias explicativas reductivas alcanzaron el funcionamiento del sistema nervioso animal y humano, y el siglo XXI ha conocido el desarrollo de sistemas de inteligencia artificial que buscan imitar y mejorar las funciones cognitivas cuyo funcionamiento ya conocemos. No obstante, aquella pregunta que nos habíamos planteado, ¿será efectivamente que hay un medio para explicar la causa de todos los cambios que ocurren en el mundo?, por el momento tiene que llevar una respuesta negativa. La nueva ciencia ha fracasado en su pretensión de dar una explicación reductiva a la experiencia consciente que, cual aquellos irreductibles galos narrados por Goscinn y Uderzo, resiste los embates del conquistador romano.

Hoy reina cierto acuerdo en que la conciencia de acceso es una función, o un conjunto de funciones, en principio explicable a través de las estrategias reduccionistas. Es un hecho que ya existen sistemas artificiales que pueden tener estados internos conscientes-A, que les permiten realizar numerosas actividades con mayor presteza que los seres humanos. La neurociencia cognitiva y computacional se ha constituido en una herramienta eficaz para resolver los problemas fáciles de la conciencia. Pero la conciencia fenoménica es otro cantar. Con el intento cartesiano nace también el problema difícil, ¿cómo y por qué ciertos hechos físicos que ocurren en nuestro sistema nervioso generan estados mentales que tienen un carácter fenoménico -*qualia*- que se siente de cierta manera? ¿Puede un simple mecanismo tener experiencia consciente? Y si así fuera, ¿cómo?, porque hasta ahora no hemos podido dar cuenta del pasaje del primer nivel de la sensación al segundo. Perfectamente podría saltarse hasta el tercero.

En el transcurso de este capítulo he intentado presentar nuestro problema filosófico partiendo de su génesis en la filosofía cartesiana para luego proceder a su delimitación más precisa con utilización de terminología actualizada, a modo de marco teórico. En el siguiente capítulo procederé a realizar un mapeo de las principales maneras como se ha respondido al problema difícil de la conciencia, clasificándolas siguiendo el criterio propuesto por Chalmers (2002), que considero

será de utilidad para el ulterior análisis del argumento del conocimiento en el tercer capítulo.

Capítulo 2. El problema difícil de la conciencia: geografía de posturas

En el capítulo anterior rastreamos los orígenes del denominado *problema difícil de la conciencia*, hasta ubicarlo en las limitaciones del alcance explicativo de la filosofía mecanicista de Descartes, siguiendo la idea de que el problema mente-cuerpo cobra un cariz novedoso en su obra, con la entrada de la modernidad y en el contexto de la denominada revolución científica de los siglos XVI y XVII (Rorty, 1989). Mientras que el problema mente-cuerpo de las épocas antigua y medieval pasaba más por el conocimiento de los universales y la necesidad de postular un alma inmaterial e inmutable que pueda dar cuenta de ello, Descartes resignifica la noción de alma. El fenómeno de la vida es perfectamente explicable en los términos del naciente mecanicismo, y no requiere la postulación de una naturaleza esencial específica para su comprensión. Mientras que “...la muerte no ocurre nunca por ausencia del alma, sino solamente porque alguna de las principales partes del cuerpo se corrompe” (2011 [1649], pág. 465), serán las funciones cognitivas las que sí dependan de la postulación de un alma esencialmente mental, más que espiritual.

Por una parte, la imposibilidad que tiene el mecanicismo de explicar el aspecto fenoménico de las funciones cognitivas; y por otra parte la asimetría epistémica entre los cuerpos (el propio y los otros) y nuestra propia mente, asimetría que genera la necesidad de explicaciones desde dos puntos de vista diferentes -el de la tercera persona y el de la primera persona-; nos inducen a pensar la relación problemática entre el alma (ahora entendida como mente) y el cuerpo, de una manera diferente. Tal brecha epistémica fue capturada por David Chalmers en una

sucesión de artículos y obras de mayor envergadura³³, especialmente en su influyente distinción entre problemas fáciles y el problema difícil de la conciencia, distinción que estableció el escenario para el debate filosófico de las últimas décadas. La obra del filósofo ha sido fundante en el sentido de generar una terminología a través de la cual los distintos actores han podido expresar sus posturas respecto a la naturaleza de lo mental y del conocimiento de lo mental.

De la mencionada sucesión de artículos, “Consciousness and its Place in Nature” (2002) presenta el mapeo más completo y detallado que permite categorizar las múltiples posiciones que han sido adoptadas acerca, no solo del problema difícil, sino a un nivel superior, del estatuto epistémico de dicho problema difícil. El propósito de este breve capítulo es presentar las posturas más relevantes acerca del problema difícil, no para profundizar en cada una de ellas, sino para agruparlas bajo las categorías de Chalmers y generar una terminología útil para cumplir con el objetivo del capítulo siguiente.

2.1 Geografía del problema difícil de la conciencia

Es posible clasificar, a grandes pinceladas, las múltiples posturas acerca del problema difícil de la conciencia en dos grandes grupos. Uno de ellos, de carácter monista y materialista, niega la relevancia de la distinción entre los problemas fáciles y el problema difícil de la conciencia. El otro, por el contrario, subraya la distinción cualitativa de los dos tipos de problemas.

Recordemos que los problemas fáciles son aquellos que refieren a la comprensión de las funciones cognitivas. Explicar cómo se cumple una función es

³³ Para seguir este tema en la obra del autor sugiero cfr. especialmente Chalmers (1995), (1995b), (1999 [1996]), (1997), (2002).

explicar el mecanismo que cumple esa función en el sistema nervioso humano, o eventualmente en un sistema cognitivo de otra naturaleza como ser una inteligencia artificial u otro tipo de organismos vivos. Para su resolución, son perfectamente apropiados los métodos tradicionales de la ciencia moderna en el contexto de la neurobiología, la psicología, la ingeniería inversa de software, las ciencias cognitivas y computacionales, y otras disciplinas afines; porque dan cuenta de las estructuras y dinamismos que constituyen los mecanismos relevantes a la función.

El problema difícil de la conciencia consiste en explicar cómo y por qué el desenvolvimiento de algunas de dichas funciones cognitivas está acompañado por la experiencia consciente cuando todo hace indicar, al menos desde la perspectiva de Chalmers, que las funciones no necesitan de la misma para su cumplimiento. Este problema es difícil porque, según quienes defienden su carácter distintivo respecto de los fáciles, no es posible explicar la conciencia fenoménica mediante su reducción a mecanismos y si se lo quiere resolver hay que recurrir a estrategias no reduccionistas. Pasemos sin más a considerar el mapeo propuesto por Chalmers.

2.1.1 Materialismos no reduccionistas (Tipo-A)

Dentro de las posturas materialistas hay, en primer lugar, una serie de críticos que han negado de manera rotunda que la distinción fácil-difícil tenga algún sentido. A la interna de esta clase se podrá hallar maneras y argumentos diferentes para negarle sentido a la distinción. Por ejemplo, materialistas eliminativistas como Daniel Dennett (1995) (1995b) o Patricia Smith Churchland (1994) (1997), cada uno a su manera y con sus argumentos, entienden que no tiene sentido distinguir ambos grupos de problemas, y que el gran abanico de las ciencias cognitivas es capaz de desentrañar todo lo que hay por desentrañar, que serían justamente las funciones cognitivas que Chalmers y Block adjudican a la Conciencia-A, tal como la he presentado en el capítulo anterior.

Puesto en términos de superveniencia, son posturas que afirman que lo mental superviene lógicamente y nomológicamente a la materia. La afirmación está referida a “lo mental” y no a “la experiencia consciente” porque esta última es una expresión ambigua de la psicología popular. Todas las expresiones cotidianas, de la llamada psicología popular, así como todas las expresiones más técnicas, pero herederas de la terminología popular, han de ser paulatinamente abandonadas para lograr una terminología técnica que se limite a designar aquello que es tangible y público del sistema nervioso, y que por tanto puede ser objeto de la ciencia.

En el mismo grupo podría incluirse una obra fundacional como *El concepto de lo mental* de Gilbert Ryle (1949), en la que se critica duramente el dualismo cartesiano para negar toda interioridad en el psiquismo humano en favor de una teoría disposicional-conductual de la mente.

A grandes trazos, podemos asociar las posturas filosóficas Tipo-A con los enfoques conductistas en psicología (Watson, 1945 [1924]), y con los enfoques computacionales (Churchland & Sejnowski, 1992) en neurobiología.

2.1.2 Materialismos reduccionistas (Tipo-B)

Un segundo grupo de teorías, agrupadas en lo que Chalmers llama materialismo Tipo-B, tienen en común la afirmación de que sí existe ese carácter fenoménico de ciertos estados mentales que requieren explicación, y por tanto que el problema difícil de la conciencia es un problema relevante. Sin negar el contenido del problema difícil de la conciencia, sí hay una negación de la separación tajante entre ambos niveles de problematización, dado que, para este tipo de teorías materialistas de lo mental, la Conciencia-F es nomológicamente superveniente a lo material, haya o no superveniencia lógica.

Esto se puede entender de la siguiente manera: algunas posturas plantean que la identidad entre lo fenoménico y lo material puede establecerse

conceptualmente; otras afirman que solo es posible fundamentarla *a posteriori*, por ejemplo, gracias a numerosas observaciones que registren que los eventos fenoménicos y los eventos físicos se correlacionan de manera sistemática, y por tanto son el mismo evento. En esta última variedad, de identidad establecida *a posteriori*, la misma debe entenderse como identidad en la referencia, mas no en el significado (Place, 1956). Cuando se afirma “El dolor es excitación de fibras-C”, el nexos “es” debe entenderse como constitutivo (“consiste en”) y no como una atribución de significado al término “dolor”.

Se sigue del establecimiento de tal identidad, que los enfoques metodológicos clásicos son perfectamente aptos para explicar cómo y por qué hay experiencia consciente acompañando a los hechos físicos que ocurren en el sistema nervioso.

Encontramos en este grupo las llamadas Teorías de la Identidad Psico-Física, como la que proponen los filósofos australianos U. T. Place (1956) y J. J. C. Smart (1959), que funciona como antecedente clásico a numerosos enfoques que tienen como meta recuperar lo mejor de los dos mundos: afirmar la relevancia de la experiencia consciente, así como afirmar con optimismo que la misma puede ser comprendida a través del estudio de los hechos físicos fundamentales a los que superviene. El Monismo Anómalo (Davidson, 1981), es una variante relevante a dichas teorías que también cabe mencionar, a tal punto que se le atribuye al filósofo el haber traído de otros contextos disciplinares el término “superveniencia” como herramienta para pensar el problema mente-cuerpo.

2.1.2.1 Apunte sobre la influencia de este enfoque

Puede leerse en el Prefacio a uno de los más difundidos manuales de introducción a la neurociencia:

El objetivo último de la neurociencia es entender cómo el flujo de señales eléctricas a través de circuitos neuronales da lugar a la emergencia [*give rise*] de la mente -a cómo percibimos, actuamos, pensamos, aprendemos y recordamos. Si bien estamos

a muchas décadas de lograr ese nivel de comprensión, los neurocientíficos han realizado progresos significativos para lograr una visión en los mecanismos neurales que subyacen la conducta, la salida observable de los sistemas nerviosos de los humanos y otros organismos (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013, pág. xli)

Y en el capítulo introductorio los autores continúan:

Una de las ideas directrices que desarrollaremos en este libro es que la especificidad de las conexiones sinápticas establecidas durante el desarrollo, subyacen a la percepción, la acción, la emoción y el aprendizaje (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013, pág. 55)

En estos pasajes se está explicitando, al menos en la visión de los autores, que sin duda es de las visiones más difundidas en el ámbito académico global, los objetivos de la neurociencia como disciplina. Merece la pena comentar brevemente dichos pasajes. Primeramente, dan como un hecho, como corresponde al rol del científico que no está reflexionando a nivel metafísico, que lo mental es lo que emerge -en algún sentido de “emergencia” no elucidado- del funcionamiento del sistema nervioso. Luego se provee una lista de las facultades que quedan incluidas en lo que los autores entienden por lo mental. Se trata de facultades cognitivas que deben incluirse en lo que hemos llamado Conciencia-A, esto es, la mente está entendida como el cumplimiento de ciertas funciones. Se trata también de facultades cognitivas que estarían presentes por igual -aunque pueda haber diferencias razonables de grado- en humanos y animales, o al menos en animales superiores.

Seguidamente, los autores se inscriben (en cierta forma) en lo que llamaremos materialismo Tipo-C, ya que son optimistas de que el desarrollo lento pero seguro de la neurociencia podrá dar cuenta de manera cada vez más completa de los fenómenos de la vida mental. Como comentaré en el próximo apartado, dicho tipo de materialismo puede no ser considerado como un tipo particular en sí mismo, ya que puede resolverse en un materialismo Tipo-A, o B, o incluso un monismo Tipo-F, dependiendo del caso. Y justamente esto depende de ese sentido no explicitado de “emergencia”. Si entendemos por dicho término la salida a la luz de una nueva realidad, por entero dependiente de la anterior, pero de naturaleza

diferente, entonces estaríamos entrando bajo alguna forma de dualismo, pero no parece ser el caso. Si dicha emergencia ha de entenderse como un proceso causal que solo puede ocurrir en sentido abajo-arriba, posiblemente se trata de una postura epifenomenista (dualismo Tipo-E). Puede haber otras opciones. Lo que es claro, es que la discusión de qué se entienda por dicho proceso de “emergencia” de lo mental por sobre y a partir de lo que ocurre a nivel de lo físico, está en la clave para la dilucidación del problema difícil de la conciencia y para determinar cuál es la postura subyacente en obras canónicas como la mencionada y otras similares.

Hay un detalle que no debe escapar a nuestra vista. Los autores omiten mencionar, justamente, el aspecto fenoménico de la vida mental. Sí mencionan la “percepción” como una de las facultades emergentes, pero perfectamente podría entenderse por dicho término algo que ocurre en el ámbito del acceso de la información del entorno con el fin de generar y modular comportamientos adaptativos, pero sin poner el énfasis en el hecho de que el producto de dicha percepción tenga un carácter cualitativo peculiar. Dicha omisión podría indicar un compromiso explícito con una formulación Tipo-A. Sin embargo, luego puede leerse, en el Prefacio, que:

Ya no estamos constreñidos simplemente a inferir funciones mentales de la conducta observable (...) la neurociencia podría en poco tiempo desarrollar las herramientas necesarias para sondear los misterios biológicos más profundos -las bases biológicas de la conciencia y el libre albedrío” (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013, pág. xliii).

Quedamos ubicados en el reconocimiento de la legitimidad del ámbito de la conciencia, y hay razones para pensar que se trata de la conciencia fenoménica - aunque no esté explicitado- porque se lo menciona como agregado adicional a la lista de funciones cognitivas mencionadas en el fragmento inicial.

Por tal razón es dable admitir, a modo de hipótesis, que la noción metafísica que hay subyacente a la actividad cotidiana de la neurociencia, es la Teoría de la Identidad Psico-Física. Esta idea, que me fue propuesta por el Dr. Rogerio Passos Severo, debe ser sometida oportunamente a un examen mucho más minucioso de

las fuentes, no solo manuales de divulgación o introducción de las neurociencias, sino publicaciones académicas de “ciencia normal”. Aventuro que dicho examen no arrojará resultados puros como el que se puede leer en el Prefacio de la obra citada, especialmente cuando empezamos a ingresar en campos más específicos y que ocasionalmente trabajan con métodos diferentes, como el de la neurociencia cognitiva y computacional, la neurofisiología, la etología animal, etc. En cualquier caso, se trata de una postura materialista Tipo-B, por la que se reconoce la existencia de estados mentales con carácter fenoménico *-qualia-*, y se es optimista respecto a la posibilidad de dar una explicación reductiva a la emergencia de dichos estados, entendiéndolos como el aspecto mental de un hecho físico que ocurre en el sistema nervioso.

El hecho de que, siguiendo con prudencia la hipótesis propuesta, estemos ante el enfoque predominante en la enseñanza y práctica de las neurociencias, hace de este tipo de materialismo uno muy relevante para contrastar con lo que arroje el análisis del argumento del conocimiento en el capítulo siguiente.

2.1.3 Materialismo Tipo-C e Idealismo Tipo-I

El llamado “materialismo Tipo-C” incluye aquellas posturas que incorporan la dimensión temporal en la posibilidad de la resolución del problema difícil. Reconocen que hoy no tenemos una idea clara de cómo podría explicarse la experiencia consciente, pero que, análogamente a como ha ido sucediendo en la historia de la ciencia, es de esperar que el misterio pueda ir resolviéndose cuando podamos construir los conceptos apropiados. Luego, la dificultad de la resolución no es de principio sino de limitación teórica y técnica contingente. Es una cuestión de darle tiempo al desarrollo natural de las ciencias cognitivas y sus derivadas, para que puedan formular los conceptos que permitan entender qué es y cómo surge la experiencia consciente. Lo que hace esta categoría es simplemente agregar la dimensión temporal a lo que proponen los materialismos Tipo-A y Tipo-B. O bien

se trata de que aún no hemos desarrollado los conceptos apropiados para caer en la cuenta de que la Conciencia-F no era más que mecanismos en funcionamiento, y por tanto es un concepto prescindible o carente de significado, y a eliminar; o bien se trata de que la neurociencia no avanzó lo suficiente para dar cuenta de la Conciencia-F, y es tan solo una cuestión de tiempo para que logremos conquistar ese misterio tal como la ciencia ha logrado conquistar otros. Sea cual sea el caso, el materialismo Tipo-C no es, conceptualmente, diferente a alguno de los otros tipos ya mencionados. Otro caso que podría integrarse a esta modalidad es el de aquellos misterianistas (*mysterians*) que, como Colin McGinn (2003), entienden que el problema difícil es soluble en principio, pero no por nosotros los seres humanos, dadas nuestras peculiaridades cognitivas.

Cabe mencionar brevemente un idealismo Tipo-I³⁴, que también le negaría lugar al problema difícil de la conciencia, pero por razones radicalmente opuestas a los grupos anteriores. Se trata de aquellas posturas radicalmente idealistas que afirman que todo lo que existe, existe en forma de ideas en la mente de un sujeto, y que la existencia de un mundo externo, material, es dudosa y debe ser demostrada (ahí hablaríamos de un “problema difícil de la materia”), o más aún, niegan la existencia de un mundo físico material. No abordaré en el transcurso de esta tesis una postura de este tipo, ya que, si bien se muestra como muy atractiva en tanto curiosidad histórica, no ha tenido exponentes relevantes en los últimos tiempos. Podría defenderse, en cierta manera, que hay cierta analogía por simetría entre el idealismo Tipo-I y el materialismo Tipo-A, de forma que se trata de posturas que postulan lo mismo pero de manera invertida.

³⁴ Cfr. por ejemplo (Berkeley, 1992 [1709])

2.1.4 Dualismo interaccionista (Tipo-D)

Del otro lado de la frontera hallamos las posturas que sí reconocen que hay un problema difícil de la conciencia y que, tal como ha sido formulado por Chalmers, afirman que hay algún aspecto de lo mental que no es explicable por los métodos reduccionistas tradicionales, por lo que defienden la pertinencia de la distinción fácil-difícil. Establecida la brecha epistémica, son tres las grandes estrategias disponibles para saltarla: dualismo Tipo-D, dualismo Tipo-E y monismo Tipo-F.

El dualismo Tipo-D es el de inspiración cartesiana, posiblemente el más afín a lo que Kim (2006) llama “dualismo *folk*” o dualismo de sentido común³⁵. Según esta postura es necesario asumir que hay dos niveles ontológicos, claramente delimitados, de forma tal que uno de ellos está integrado por, y por tanto permite entender, los hechos físicos, mientras que el otro nivel está integrado por, y por tanto permite entender, los hechos fenoménicos. La teoría arquetípica de este grupo es la que surge de las *Meditaciones Metafísicas* de Descartes (2011 [1641]), sin perjuicio de otros exponentes que hayan procurado incorporar la postura al debate contemporáneo, como Popper y Eccles (1980) y Foster (2002). Puede sintetizarse como la afirmación de que existen dos sustancias fundamentales -la materia y el pensamiento-, que son irreductibles la una respecto de la otra. Aun asumiendo que

³⁵ Se ha mencionado (Kim, 2006) o incluso explicado y defendido (Bloom, 2004) que el dualismo es la postura habitual, por defecto, de la mayor parte de las personas adultas o incluso infantes; lo cual no implica, claro está, que sea una postura verdadera (Fiala, Arico, & Nichols, 2012). Una de las críticas que se le ha realizado consiste en señalar que dicha afirmación se ha sostenido sobre una vaga caracterización de dicha postura (Balmaceda, 2016). La postura dualista puede adjudicarse tanto a la distinción cuerpo-mente, poniendo el énfasis en la irreductibilidad de algunas funciones cognitivas, como a la distinción cuerpo-alma, remarcando la irreductibilidad de algunas funciones espirituales (Richert & Harris, 2008). A los efectos de esta tesis es la primera la que interesa, mas a los efectos de la postulación del dualismo como postura de sentido común, queda la impresión de que se trata de la segunda modalidad. Asimismo, vimos en el capítulo anterior, siguiendo a Rorty (1989), que hay una transformación en la noción de alma en la obra de Descartes, transformación que justamente opera radicando las funciones cognitivas -en vez de las espirituales- en lo que llama “alma”.

nuestra mente y nuestro cuerpo son dos sustancias diferentes, es notorio *prima facie* que lo que ocurre en nuestro cuerpo influye en nuestros pensamientos (como cuando tenemos sensaciones, por ejemplo), y al mismo tiempo solemos pensar que nuestros pensamientos son capaces de iniciar acciones voluntarias. Este tipo de interacción entre lo mental y lo corporal es un hecho del que hay que dar cuenta si no queremos distanciarnos demasiado del sentido común. Lo que caracteriza a las posturas Tipo-D es que admiten como factible la interacción causal entre lo físico y lo mental³⁶.

2.1.5 Epifenomenismo o Dualismo Tipo-E

Esta postura guarda un interés particular en la presente investigación porque era la posición del propio Frank Jackson cuando postuló su experimento mental de la habitación de María (Jackson, 1982). El epifenomenismo afirma la existencia de dos tipos claramente diferenciables de sustancias, físicas y fenoménicas, pero entiende a la Conciencia-F como un epifenómeno de lo que ocurre en el sistema nervioso. Por epifenómeno debe entenderse una suerte de precipitado que emerge de lo que ocurre a nivel físico, y por tanto halla en el mundo físico su causa, pero él mismo no puede ser la causa de nuevos fenómenos. Funciona como un eslabón final de las cadenas causales en el mundo natural, admitiendo entonces que hay

³⁶ Los dos argumentos que propone Descartes para fundamentar dicha postura pueden hallarse en sus meditaciones segunda (a partir de la asimetría epistémica entre cuerpo y pensamiento) y sexta (a partir de un argumento modal que parte de la concebibilidad de la separación entre ambas sustancias para concluir el carácter real de dicha separación) (Descartes, 2011 [1641]). De las objeciones más agudas vale mencionar la propuesta por Isabel de Bohemia, luego discutida en intercambio epistolar, vinculada a la interacción entre las sustancias (Descartes & Isabel de Bohemia, 2011 [1642]); y la propuesta por Gilbert Ryle (1949), quien sostiene que Descartes comete un error categorial al pensar al pensamiento con las categorías lógicas equivocadas. Quizás se trate del nudo problemático que ha motivado el inicio de esta investigación, hace ya un buen tiempo, pero por razones de extensión he omitido desarrollarlo con mayor detenimiento.

interacción causal en sentido abajo-arriba pero no admitiendo causalidad arriba-abajo.

La posición Tipo-E es una postura dualista y a la vez intenta sortear la disonancia que genera la interacción causal entre las sustancias con la tesis del cierre causal del mundo. Si partimos de la base de que existe algo llamado “mentalidad”, más allá de cómo lo caractericemos, y que está de alguna manera separado o diferenciado de lo corporal, parece a primera vista necesario asumir que han de existir relaciones causales entre eventos mentales y eventos corporales (en ambos sentidos), entre dos o más eventos físicos, y entre dos o más eventos corporales. Es de sentido común asumir entonces que los eventos mentales están tejidos intrincadamente en la red de las relaciones causales de la naturaleza. ¿Cómo puede una cadena de eventos físicos terminar desencadenando una experiencia consciente?, y a la inversa, ¿cómo puede un estado mental generar nuevos estados mentales o incluso provocar acciones voluntarias en los agentes?

La agencia, es decir nuestra capacidad de actuar movilizados por razones, que pueden ser, por ejemplo, creencias o deseos, solo sería posible si asumimos que lo mental tiene eficacia causal. Y en el otro sentido, la percepción, y la cognición, suponen la posibilidad de que eventos del mundo generen estados mentales. Finalmente, que algunos estados mentales puedan generar o influir en otros estados mentales también parece algo bastante intuitivo. Nadie se sorprenderá si proponemos que la creencia de que “hoy es lunes” puede generar el deseo de que la proposición “hoy es viernes” sea verdadera ya mismo. Por lo dicho, se puede afirmar que la causación mental -sea desde o hacia lo físico, o entre estados mentales- es un tipo de relación que habría que salvar si establecemos una postura que diferencie de alguna manera lo mental de lo físico.

Respecto al segundo elemento de nuestra disonancia, cabe afirmar que, si entendemos que el mundo físico es un mundo completo y cerrado en sí mismo, es necesario asumir que está causalmente cerrado. Imaginar que la conciencia es algo que es causado por hechos del mundo físico, y a su vez que lo que ocurre en la

conciencia puede ser causa de eventos físicos, implica afirmar que hay una vía de escape para la cantidad de energía existente. Por otra parte, suponer que un evento mental puede tener por causa tanto un evento físico como otro evento mental, nos coloca ante un posible caso de sobredeterminación. ¿Se trata de una verdadera sobredeterminación, esto es, de un caso genuino de confluencia de dos causas independientes siendo que cada una de ellas es condición suficiente para el efecto? No es una solución intuitiva ni elegante. Impresiona como una confluencia demasiado casual para interacciones cuerpo-mente que están ocurriendo permanentemente. Es más atractivo pensar, que los eventos causa son el mismo evento, como podría proponer un defensor de la teoría de la identidad, o directamente que lo mental no tiene eficacia causal. En conclusión, para que los principios físicos de conservación de la energía y de la materia se sostengan indemnes es necesario negar toda interacción causal entre la conciencia y el cuerpo de la que emerge, y eso es lo que propone la solución epifenomenista³⁷.

2.1.6 Monismo Tipo-F

Bajo esta categoría, Chalmers ha incluido aquellas teorías que, sin llegar a configurar un idealismo Tipo-I, postulan que la conciencia ha de ser una propiedad fundamental del mundo físico, tal como lo es la masa o la energía. El pansiquismo afirma que la conciencia fenoménica es, en sí misma, una sustancia o propiedad fundamental del mundo físico. Otra postura emparentada, el panprotopsiquismo, es la afirmación de que hay algún tipo de entidades o propiedades fundamentales que

³⁷ Se ha objetado al epifenomenismo el entrar en contradicción con los principios fundamentales de la teoría de la evolución, ya que no permite entender el sentido adaptativo de la experiencia consciente si esta no tiene eficacia causal. Para seguir dicha discusión cfr. Popper (1980), Churchland (1999) y Lycan (2009).

es o son el origen común de todas las propiedades físicas y fenoménicas, por lo que entiende que lo material y lo fenoménico tienen una raíz ontológica común.

En ambos casos, el desafío de una ciencia de la conciencia habría de consistir en indagar mediante estrategias no reduccionistas cuáles son las leyes fundamentales que rigen su funcionamiento, leyes análogas a las leyes fundamentales de la física y que deberían tener el mismo estatuto epistémico. En los últimos años una reconsideración de la obra de Bertrand Russell *The Analysis of Matter* de 1927 ha llevado a filósofos como Philip Goff (2017) y el propio Chalmers (2013) a volver a colocar el llamado monismo Tipo-F como una alternativa relevante a considerar en este mapeo³⁸.

2.1.7 Funcionalismo

Finalmente, hay una familia de posturas acerca del problema de la conciencia fenoménica que es difícil de ubicar en alguna de las dos grandes categorías establecidas. El funcionalismo ha sido, especialmente desde el último tercio del siglo XX, un marco teórico que se ha asentado con fuerza en el panorama de la filosofía de la mente, probablemente debido a que logra compatibilizarse con diversas posturas y atravesar todo el territorio mapeado. Su idea fuerza, la identificación de los estados mentales con el cumplimiento de un rol funcional, es decir, con cierta estructura de relaciones causales con el entorno y con otros estados mentales, está acompañada por el principio de la realizabilidad múltiple, según el

³⁸ El obstáculo más discutido de este tipo de posturas ha sido el llamado problema de la combinación, es decir, el establecimiento de cómo la macroexperiencia o experiencia consciente de los seres como nosotros puede conformarse como una combinación de las miríadas de microexperiencias o microprotoexperiencias propias de las partículas fundamentales que nos constituyen. Puede seguirse esta discusión en Chalmers (2016) y Goff, Seager y Allen-Hermanson (2017). Desde numerosas perspectivas se le ha objetado a esta posición el ser contraria al sentido común, aunque no parece tratarse de una objeción interesante para explorar en un contexto académico.

cual la misma estructura funcional (asimilable a un *software*), puede instanciarse en diversos sustratos infraestructurales (asimilables a un *hardware*). Este principio es el que permite que, si bien se puede hallar cierta afinidad con las posiciones materialistas, la teoría resulta perfectamente compatible con diversas formas de dualismo³⁹.

Ahora bien, es conveniente señalar que justamente uno de los aspectos problemáticos de la Conciencia-F que lleva a Chalmers a distinguir ambas familias de problemas, es que pareciera que la misma no cumple ningún rol funcional en particular. Dicho de otra manera, puede defenderse que toda explicación funcional -entendida como una forma refinada y abstracta de explicación mecanicista- yace impotente ante la Conciencia-F como hecho del mundo a explicar. Por tal razón, si bien puede haber dualismos compatibles con el funcionalismo, a la hora de estar mapeando posturas en torno al estatuto del problema difícil, entiendo corresponde ubicar el funcionalismo en la familia de los materialismos, y dependerá del caso si será Tipo-A o Tipo-B.

2.2 Síntesis

En síntesis, se trata de ocho grupos de posturas teóricas, divididas en dos grupos por el reconocimiento o no del carácter “difícil” del problema difícil de la conciencia. Para las tres posturas materialistas, más el funcionalismo, la mentalidad es un fenómeno cuya explicación científica podrá irse resolviendo con los métodos habituales de las ciencias cognitivas, reduciendo las explicaciones a dar cuenta de los mecanismos públicamente observables que logran cumplir las funciones que la constituyen. Para el idealismo la relación entre conciencia y materia es inversa. Y

³⁹ Por obras representativas del funcionalismo, con sus matices, cfr. Armstrong (1981), Putnam (1975), Block (1980).

para los otros tres grupos de posturas las estrategias metodológicas reduccionistas son incapaces de cumplir con ese cometido, por lo que el problema difícil de la conciencia efectivamente es significativo y es difícil. De esta última aseveración se seguiría que la única manera de entender -si fuera posible- por qué y cómo algunos de nuestros estados mentales están acompañados por un aspecto fenoménico, es buscando nuevas estrategias metodológicas que no pretendan ni negar la dimensión ontológica de dicho aspecto, ni reducirlo a hechos físicos fundamentales.

Los múltiples comentarios y análisis del argumento del conocimiento provienen de actores pertenecientes a dichos grupos. Su mapeo nos ha permitido exponer las coordenadas fundamentales de la disputa, y los puntos de referencia en el territorio. Quedamos en buenas condiciones para adentrarnos, en el próximo capítulo, en el estudio del argumento del conocimiento presentado bajo la forma del experimento mental de María y su liberación.

Capítulo 3: El Argumento del Conocimiento

En el primer capítulo de esta tesis he rastreado los orígenes del problema difícil de la conciencia en las limitaciones que presentó la estrategia mecanicista de Descartes, que logró un modelo explicativo exitoso para dar cuenta de casi todos los aspectos de la realidad por conocer, pero que encuentra dificultades insalvables -al menos en principio- a la hora de intentar explicar por qué algunos de nuestros estados mentales están acompañados por un carácter fenoménico que parece resistirse a cualquier explicación reductiva.

En tal ocasión, además de ir siguiendo la manera como el filósofo se topa con esa verdadera brecha explicativa, presenté el problema utilizando una terminología más cercana a la de uso en nuestros tiempos, dejando establecido el lazo inconfundible entre lo que Descartes discutía en su época, y lo que Frank Jackson, David Chalmers y otros filósofos de diversas posturas están discutiendo en la nuestra. Esto ha sido útil para presentar, en el segundo capítulo, un mapeo sintético que diera cuenta de la geografía de respuestas al problema, agrupándolas según un criterio que se ha mostrado útil para procesar el debate, inteligible y aceptable para las diversas partes.

En este tercer capítulo procederé a la exposición, análisis y valoración del argumento del conocimiento, presentado por Frank Jackson (1982) (1986), y luego ampliamente discutido por exponentes de las diversas posturas teóricas presentadas en el capítulo anterior. Comenzaré determinando qué intuición básica pone en juego el experimento mental, luego ensayaré una formulación del argumento que pudiera ser aceptada por actores provenientes de diferentes enfoques. En el apartado 3.3 analizaré algunas de las objeciones más representativas al argumento. Finalmente,

en 3.4 valoraré su alcance, diferenciando una dimensión epistemológica y una dimensión ontológica de la discusión. Pasemos, sin más preámbulos, a la exposición.

3.1 De la intuición al argumento

Los experimentos mentales suelen tener como objetivo generar un escenario que controle adecuadamente ciertas variables, con el fin de provocar una intuición. Luego, la intuición podrá ser usada en un argumento. ¿Cuál es la intuición que desea traer a la luz el AC?

Díganme todo lo físico que se puede decir acerca de lo que ocurre en un cerebro vivo, los tipos de estados, sus papeles funcionales, sus relaciones con lo que ocurre en otros momentos y en otros cerebros, etc., etc., y aunque sea todo lo inteligente que se tiene que ser para unir toda la información, no se me habrá dicho nada acerca de lo doloroso de los dolores, del picor de las picazones, de las punzadas de los celos, o acerca de la experiencia característica de probar un limón, oler una rosa, oír un ruido fuerte o ver el cielo. (Jackson, 2003, pág. 96)

La intuición que desea poner de manifiesto el experimento mental que pasaremos a considerar, ya adoptando la terminología que traemos desde el primer capítulo, consiste en que lo que podemos saber sobre el carácter fenoménico de ciertos estados mentales, en especial las sensaciones, excede la información física. Bajo el concepto “información física” podemos incluir -siguiendo las principales teorías reduccionistas que he presentado en el capítulo anterior-, la información acerca de los comportamientos, la relativa a los estados neurales, o incluso la información relativa al papel causal de los estados mentales de los individuos. Se trata por tanto de una intuición de carácter epistémico -y de ahí el nombre del argumento-: incluso poseer información física completa sobre el funcionamiento

del sistema nervioso no explica el carácter cualitativo de la experiencia consciente⁴⁰.

Es posible identificar numerosos antecedentes, argumentos similares en el sentido de que se trata de experimentos de pensamiento que establecen un escenario en el que la brecha epistémica (Levine, 1983) irrumpe, y de la misma se infiere una brecha ontológica⁴¹. Jackson (1982) opina que ninguno de los argumentos precedentes logra aislar claramente la intuición que está en juego, y por tanto se propone elaborar un experimento mental que sí logre aislar la intuición de manera que pueda ser aceptable para todos, incluso para los materialistas a quienes quiere convencer. Como la intuición sería universalmente aceptada, y el argumento es deductivamente válido, entonces el AC pretendió ser presentado como uno contundente y terminante en contra del materialismo, como postura que determina

⁴⁰ Hay un verdadero nudo terminológico, que se ha hecho presente en los capítulos anteriores, y que vuelve como convidado de piedra. Se trata de la diferenciación entre lo que hemos de considerar “físico”, concepto que a lo largo de este texto he venido asociando a lo epistemológico, y lo que hemos de considerar “material”, concepto que he venido asociando a lo ontológico. Sin embargo, especialmente a partir de la segunda mitad del siglo XX, y con el desdibujamiento de las posturas más nítidas del positivismo lógico, el término “fiscalismo” ha cobrado un cariz ontológico e incluso en algunas ocasiones se ha utilizado de manera intercambiable con “materialismo” (Stoljar, 2017). En aras de la claridad, y considerando que se trata de una distinción relevante en esta tesis, he decidido ser cuidadoso en la elección de los términos y mantener la diferenciación mencionada. Puntualmente, en esta ocasión, es preferible conservar la terminología elegida por Jackson, porque de hecho se trata de información, que ha de ser expresable en términos proposicionales, y no de entidades.

⁴¹ Cfr. por ejemplo “The Country of the Blind” de H. G. Wells, citado por Jackson (1982, pág. 129) en donde alguien con visión normal tiene que convencer a un pueblo de ciegos de que posee un sentido adicional; en “Mechanism and its alternatives” de C. D. Broad, citado por Stoljar y Nagasawa (2004) se propone un arcángel omnisciente que no sabe cómo se siente oler amoníaco. De especial interés es el argumento del molino en #17 de la *Monadología* (Leibniz, 1977 [1714]), que debe considerarse en tándem con pasajes de su *Nuevo tratado sobre el entendimiento humano* (1977 [1765]) y su Carta a Bayle de 1702; ahí Leibniz discute si una máquina compuesta (como un molino) podría tener percepciones unificadas; sobre el tópico cfr. Duncan (2012), y Kulstad y Carlin (2013). Finalmente mencionar el argumento de los ciegos y sordos congénitos en Sección ii de Hume (2003 [1748]) en donde se intenta defender que toda idea proviene de impresiones, argumento que se mencionará en el devenir de este capítulo. Asimismo, sugeriré en la Conclusión a esta tesis que el argumento para el dualismo en la Meditación Segunda (Descartes, 2011 [1641]) también puede considerarse un argumento similar al que propone Jackson.

cuál es la estructura ontológica de la realidad en general, y de la mentalidad en particular.

La primera versión que propone Jackson del AC en el artículo de 1982 nos presenta a Alfredo, un individuo en todo sentido normal e igual a nosotros pero que por alguna razón neurobiológica tiene la capacidad de discernir con claridad dos matices diferentes de rojo, o de lo que para nosotros entraría dentro de lo que adjetivaríamos como rojo. Para Alfredo se trata de dos colores primarios distintos, rojo1 y rojo2. Ahora bien, aun cuando pudiéramos identificar en las retinas de Alfredo conos con pigmentos diferentes, y vías neurales específicas asociadas a dichos conos, y que podamos tener un conocimiento completo de tales atributos anatómo-fisiológicos, así como de sus comportamientos observables cuando señala al nombrar lo rojo1 y lo rojo2, Jackson asevera que no lograríamos entender cómo se siente la diferencia entre percibir rojo1 y rojo2. Más aún, si pudiéramos injertar a algún individuo las estructuras anatómicas necesarias para discernir ambos colores, dicho individuo saldría de la intervención aprendiendo algo nuevo, a saber, cómo se siente ver algo rojo1, ver algo rojo2, y en qué se diferencian ambas experiencias. Ese “algo nuevo” resultaba ajeno a nuestro saber, aunque tuviéramos toda la información física acerca de la estructura del sistema nervioso de Alfredo.

El propio Jackson dedica buena parte de su artículo a discutir lo que ocurre con Alfredo. María, solo aparece en escena cuando Jackson reconoce que Alfredo es una persona extraordinaria, y que tal característica podría afectar la aceptabilidad de la premisa del argumento que planteara el escenario. María nace como un mero comentario al pasar con el fin de solucionar tales obstáculos, pero se convertirá en la protagonista central del AC en su versión más difundida y discutida, y en la protagonista de lo que resta del capítulo. Pasemos a considerar su situación, y el argumento que ésta propicia.

3.2 La habitación de María: el argumento del conocimiento

El siguiente es el pasaje en el que Jackson nos presenta a María en su *Epiphenomenal qualia* de 1982⁴².

María es una brillante científica que, por alguna razón, ha sido forzada a investigar el mundo desde una habitación en blanco y negro, mediante un monitor de televisión en blanco y negro. Se especializa en neurofisiología de la visión y adquiere, supongámoslo, toda la información física que hay que obtener acerca de lo que pasa cuando vemos tomates maduros, o el cielo, y usamos términos como “rojo”, y “azul”, y demás. Descubre, por ejemplo, precisamente qué combinaciones de longitudes de onda del cielo estimulan la retina, y exactamente cómo esto se produce, *mediante* el sistema nervioso central, la contracción de las cuerdas vocales y la expulsión de aire de los pulmones que dan por resultado que se pronuncie la oración “El cielo es azul”. (...)

¿Qué pasará cuando se libere a María de su habitación en blanco y negro o cuando se le dé un monitor de televisión a color? ¿*Aprenderá* algo o no? Parece perfectamente obvio que aprenderá algo acerca del mundo y de la experiencia visual que tenemos de él. Pero entonces es ineludible que su conocimiento previo era incompleto. Sin embargo, tenía *toda* la información física. *Ergo*, hay más conocimiento por obtener que eso, y el fisicalismo es falso. (2003, págs. 99-100)

A diferencia de Fred, María es un ser humano sin diferencias anatómicas o fisiológicas cualitativas respecto a un ser humano habitual. Mientras que Fred tenía en su retina conos con un pigmento alternativo, además de los conos de las retinas habituales sanas, y tal detalle podría hacer que el escenario planteado por el argumento fuera visto como algo implausible o forzado, María es una persona perfectamente corriente. Quizás sí tenga unas notables cualidades espirituales que le han permitido sobrevivir a ese encierro sin enloquecer, y algunas capacidades cognitivas y metacognitivas aumentadas -v.g. memoria, capacidad de procesamiento de datos, etc.- que son las que le permiten tener un conocimiento

⁴² Elijo la traducción realizada por Laura F. Manríquez, perteneciente a la compilación dirigida por Ezcurdia y Hansberg en 2003, bajo la órbita de la Universidad Nacional Autónoma de México.

completo de la información física posible acerca de la fisiología de la visión. Pero esta especialidad puede pensarse como una diferencia cuantitativa que se monta sobre capacidades cognitivas normales para cualquier ser humano.

María sabe mucho más de lo que podría saber una persona corriente, pero no porque ponga en juego habilidades no presentes en las personas corrientes. ¿Qué sabemos nosotros acerca de María?

3.2.1 Qué sabemos acerca de María

En el texto, Jackson nos presenta tres datos relevantes acerca de María. El primero está dado por su hiperespecialización en fisiología de la visión, María sabe todo lo que hay por saber sobre ese tema, por lo que posee un conocimiento físico completo en el área. ¿Qué significa tener “toda la información física” acerca de la percepción visual de los colores? Esta es una idea problemática a la que tendremos que volver en el transcurso de esta discusión, especialmente por el anudamiento conceptual, ya presentado anteriormente, entre lo que entendamos por “físico” y lo que entendamos por “material”. Por “información física” Jackson está considerando el conjunto de comportamientos observables -incluyendo en estos comportamientos las preferencias verbales audibles y no audibles-, así como los fenómenos biológicos, eléctricos, y químicos que ocurren a nivel del tejido nervioso y todos los demás tejidos relevantes del organismo, y también el papel causal que cada uno de estos estados y fenómenos puede tener, esto es como efectos de los estímulos provenientes del mundo, como efectos de otros estados mentales y/o cerebrales, como causas de otros estados mentales y/o cerebrales, y finalmente como causas de comportamientos como los mencionados.

Por lo dicho, es posible desglosar tres dimensiones en ese fisicalismo que pretende refutar Jackson. La dimensión conductual, la dimensión neurobiológica y la dimensión funcional de los estados mentales, cada una de ellas asociable a

distintas posturas reduccionistas que fueron presentadas en el capítulo anterior⁴³. No obstante, no es en vano volver a clarificar la terminología. En un sentido epistémico, la cuestión será si hay algún tipo de información que no pueda reducirse a alguna de esas tres dimensiones. En un sentido ontológico, la cuestión es si de una respuesta afirmativa al planteo epistémico se sigue que en nuestro inventario de cosas reales haya algún lugar para hechos o entidades no materiales. Para ambas problemáticas, relacionadas, pero no iguales, es posible identificar dos versiones del argumento del conocimiento, tal como veremos en el apartado siguiente.

El segundo dato relevante que surge del texto fuente, y que hay que precisar, es que María nunca tuvo una percepción visual de algún color. María solo ve cosas en blanco, negro y escala de grises. Para fortalecer el argumento agreguemos que María se encuentra en esta situación desde su nacimiento. Dennett (2006) ha señalado que es necesario eliminar una serie de posibles situaciones que podrían poner en tela de juicio la virginidad cromática de María. Ella podría tener experiencias de color por mirar su propia piel, cortarse y contemplar su herida y su sangre, o podría percibir colores presionando sus ojos con sus manos. Es posible imaginar miles de situaciones que alteraran esta característica del escenario propuesto.

Considero que este tipo de detalles son triviales. Tan solo basta con idear algún dispositivo por el que tales efectos sean neutralizados para que el experimento pueda seguir adelante. Nuestra científica podría usar unas gafas que filtren los colores de alguna manera -cuyo funcionamiento, María necesariamente comprende gracias a su conocimiento físico completo de todo lo relacionado con la percepción del color-. Resta también considerar las experiencias de visión de colores que puedan surgir de sufrir algún golpe en los ojos, o incluso las experiencias de

⁴³ Vale señalar que el funcionalismo no es estrictamente una postura materialista, como vimos en el apartado 2.1.7. Cfr. p. ej. Hierro-Pescador (2005), Kim (2006), o al propio Chalmers (1999 [1996]).

percepción de colores que María pudiera tener en sueños o en alucinaciones por estados febriles o por consumo de ciertos fármacos. Toda esta gama de posibles obstáculos al establecimiento del escenario del experimento mental, podría sortearse simplemente agregando a nuestro escenario la condición *ad hoc* de que María nunca tuvo por esas vías, o ninguna otra, experiencia significativa de colores. Si bien no soy amigo de interponer condiciones *ad hoc*, como los personajes que aparecen como por arte de magia al final de una película para resolver el conflicto, el carácter tan trivial del problema a sortear para hacer viable el argumento hace que no vea ningún impedimento en realizarlo en esta ocasión.

El tercer dato que nos proporciona Jackson pone en funcionamiento la intuición presentada anteriormente: “parece perfectamente obvio” que al ver por primera vez algo rojo, María aprende algo nuevo que no sabía nadie, a saber, cómo se siente ver algo rojo. Lo mismo puede decirse de cualquier otro color, y lo mismo podría extenderse a propiedades análogas pertenecientes al mundo auditivo, táctil, gustativo u olfativo. Quizás también podría extenderse a cualquiera de las demás formas de la sensación, como ser la propiocepción, la identificación dolorosa de daños en los tejidos, la recepción de datos sobre la temperatura en la piel, etc.

La conclusión del argumento que surge a partir de los datos mencionados es que esa información física completa no es información completa a secas. Esto es: hay más por saber sobre la percepción del color que la información física correspondiente, y eso es lo que puede explicar que María se sorprenda al ver algo rojo y aprenda algo nuevo que no sabía antes a pesar de su conocimiento físico completo sobre la percepción visual de lo rojo.

Establecido qué sabemos acerca de María, y qué pretende concluir Jackson, pasemos a ordenar las ideas ofreciendo una estructuración del argumento que sea aceptable para las distintas perspectivas teóricas.

3.2.2 La estructura del argumento

En respuesta a una serie de críticas, especialmente Churchland (1985), Jackson publica “What Mary didn’t Know” (1986), en el que defiende el AC. Ahí Jackson establece un análisis del argumento que aparentemente conforma a todas las partes, defensores y detractores, por lo que vale la pena ya mismo presentarlo para poder trabajar con él. Según Jackson el argumento tiene la siguiente estructura⁴⁴:

1. María (antes de ser liberada) sabe todo lo físico que hay que saber acerca de otras personas.
2. María (antes de ser liberada), no sabe todo lo que hay que saber acerca de otras personas (porque ella aprende algo acerca de ellos al ser liberada).
3. Existen verdades acerca de los demás (y de ella misma) que escapan a la explicación fisicalista.

Podría agregarse una conclusión ulterior, de carácter ontológico, que se desprende de (3), que notoriamente tiene un carácter epistémico:

4. El materialismo es falso.

Puede observarse que la primera premisa recoge el primero de los datos mencionados anteriormente: María tiene una información física completa, o lo que es lo mismo para Jackson, María conoce todo lo físico que hay por conocer sobre otras personas (podría ajustarse a la letra del argumento planteando que el conocimiento es específicamente sobre los procesos físicos de la percepción visual de otras personas). Quizás se enfatiza que su saber es acerca de otras personas, y se excluye a la propia María en esta premisa, porque hay procesos físicos que en María

⁴⁴ Nuevamente elijo la traducción de Laura E. Manríquez en Jackson (2003b).

nunca ocurrieron, aunque -esto es importante- María habría de saber qué sucedería en ella si esos procesos físicos ocurrieran, por lo que en la práctica hemos de admitir que conoce, al menos en términos disposicionales, todo lo físico que ocurriría en ella en caso de encontrarse ante algo rojo. Para identificarla fácilmente, podemos llamar a esta premisa (1) “premisa del conocimiento físico completo” (Stoljar & Nagasawa, Introduction, 2004).

La segunda premisa establece los otros dos datos mencionados anteriormente. Por una parte, que María se sorprende al ver algo de color cuando es liberada, y que aprende algo nuevo: cómo se siente ver algo de color; por otra parte, que el conocimiento de María previo a su liberación no era completo. Esta premisa podría ser desdoblada, ya que se puede identificar un paso intermedio. Podemos saber que el conocimiento de María era incompleto antes de su liberación porque María se sorprende y aprende algo nuevo al ser liberada. De esta forma, la premisa (2) puede subdividirse en:

2.a. María al ser liberada se sorprende al ver algo rojo.

2.b. (2.a) María, luego de ser liberada, aprende algo nuevo al ver algo rojo.

Si bien en numerosas ocasiones será suficiente trabajar con la formulación de Jackson, puede resultar oportuno conservar en mente este paso intermedio. Está implícito en el argumento de Jackson que de la sorpresa de María al salir se puede inferir que debe haber aprendido algo nuevo.

Ambas premisas transcurren en el ámbito epistémico. Una versa sobre el saber de María, la otra sobre lo que María ignoraba y luego aprende. El rol de las premisas parece ser establecer la intuición básica que presenté anteriormente. Mientras tanto, la conclusión (3) también es epistémica, pero está al servicio de establecer una postura ontológica que puede desprenderse de la anterior, y que quedó reflejada en la proposición (4). Si hay verdades que escapan a toda explicación fisicalista, entonces el materialismo ha de ser falso. Esto es, si hay

verdades no aprehensibles por la física (considerada en sentido amplio), entonces ha de existir una realidad no material. Lo que el argumento defiende, entonces, es que alguna forma de dualismo ha de ser verdadera, aunque no especifique cuál, o incluso -aunque no fuera el propósito de Jackson- una posición panpsiquista.

A partir de esa separación entre (3) de carácter epistémico y (4) de carácter ontológico se puede realizar una distinción entre una versión débil del argumento, que iría hasta (3), demostrando que hay aspectos de la vida mental de las personas que no pueden conocerse a través de estrategias reduccionistas; y una versión fuerte del argumento si incluimos (4), que busca defender que existen hechos que no son materiales en la vida mental de las personas, y por tanto refutar el materialismo como postura metafísica general (Nida-Rümelin & O Conaill, 2019). No hay dudas que la intención de Jackson es proponer el AC en su versión fuerte, ya que en cierta forma la versión débil funciona como la confirmación de la intuición de la que el experimento mental parte, y haría innecesario el argumento en sí.

En síntesis, si María tiene conocimiento físico completo de la fisiología de la visión (premisa (1)), y tenemos la intuición de que aprende algo nuevo cuando ve algo rojo por primera vez (premisa (2)), entonces hay aspectos de la percepción del color que no se pueden conocer mediante estrategias reductivas (conclusión (3)), y finalmente, hemos de incluir entidades o hechos no materiales en nuestro inventario de lo real (conclusión (4)).

Pasaré en el siguiente apartado a considerar algunos de los enfoques más significativos que han sido críticos con el argumento. Enfoques que, en general, es fácil identificar por la línea teórica de la que proceden.

3.3 Objeciones al AC de Frank Jackson

Aunque no lo haré de forma exacta, puede resultar conveniente seguir el orden establecido por Van Gulick (2004), para ir enumerando maneras de decirle que no a María.

Primero nos preguntaremos, ¿realmente María aprende algo al ser liberada?, y exploraremos la posibilidad de que no haya ningún tipo de aprendizaje en ese momento.

Segundo, si concedemos que hay aprendizaje, nos podemos preguntar, ¿es aprendizaje proposicional lo que obtiene María?, ya que podría ser el caso que efectivamente haya un aprendizaje, pero el mismo no sea del tipo proposicional sino de otro tipo.

Tercero, si concedemos que el aprendizaje que obtiene María en su liberación es del tipo proposicional, ¿no es posible que se trate de un nuevo conocimiento de un hecho que María ya conocía?

Cuarto, si concedemos que el AC es un buen argumento, válido y con premisas verdaderas, ¿cuál es el verdadero alcance de ese argumento? Podría ser el caso que lo que el AC demuestre no sea exactamente lo que Jackson pretendía.

Finalmente, de entre todas las estrategias posibles, ¿hay algún nudo terminológico que disolver respecto a la interpretación de “fiscalismo” o “hecho físico” que implique cambios en el argumento y su fuerza dialéctica?

La literatura específica respecto al problema difícil de la conciencia, y particularmente respecto al AC en la versión de Frank Jackson, es vasta. Incluso algunos autores han presentado versiones y reversiones de su postura, por lo que pretendí hacer una selección de las líneas críticas más significativas, aquellas sobre las que más provecho podemos sacar y que han generado más literatura secundaria.

3.3.1 María no aprende nada: la banana azul

Un enfoque, en general vinculable al materialismo Tipo-A⁴⁵, plantea que el argumento subestima todo lo que María podría llegar a saber si la premisa (1) fuera verdadera. Vimos más atrás que el objetivo de Jackson al proponer el experimento mental de María era trabajar con un escenario más plausible que el de Alfredo. Sin embargo, eso nos lleva a juzgar a María con un lente que posiblemente no sea el más adecuado. La premisa (1) establece que el conocimiento que María tiene acerca del mundo físico es completo; la línea argumental que voy a presentar propone que, si la premisa (1) es verdadera, entonces María no podría aprender nada nuevo al salir de su encierro.

Daniel Dennett (2004), (2006), propone un posible final alternativo al experimento con el fin de demostrar que no logra cumplir su cometido. En general, Dennett es escéptico respecto al uso de experimentos mentales como argumentos filosóficos (2015), pero eso no le ha impedido ingresar de lleno a la polémica alrededor de los mismos para intentar desacreditarlos “desde dentro”. Hace un uso frecuente de los mismos, pero en un tono sarcástico. Al salir de su bunker, en vez de ver una manzana o un tomate rojo, María sería enfrentada por sus captores -que no cesan en su sadismo- a una banana azul. Dennett entiende que lejos de ser engañada por el truco, el conocimiento físico completo que posee María le permitiría exclamar “¡Hey! ¡Me habéis querido engañar! ¡Los plátanos son amarillos, pero éste es azul!” (2006, pág. 125). Esto es, el conocimiento físico completo de María tendría que permitirle prever cómo se sentiría ver algo rojo, o para este caso algo amarillo como una banana. Más aún, María podría haber logrado predecir cómo se sentiría ver una banana gracias a un razonamiento deductivo que incluyera miles de pasos, pero como por hipótesis estamos suponiendo que María tiene toda la información física pertinente como para constituir las premisas de

⁴⁵ Cfr. apartado “2.1.1 Materialismos no reduccionistas (Tipo-A)” en la página 55.

dicho razonamiento, y si acompañamos dicha información por un poder de procesamiento superior al habitual, tendríamos que admitir que un razonamiento de ese tipo es concebible, y tan solo depende de tomarse en serio la premisa (1) y seguir sus indicaciones, esto es, realmente imaginar qué significa tener toda la información física.

La dificultad para imaginar el escenario, consistente en María conociendo la totalidad de la información física que hay por saber sobre la percepción de los colores, radica en que se trata de una tarea descomunal. No es relevante el hecho de que sea descomunal tener ese conocimiento, eso es evidente. El punto es que es descomunal el simple hecho de imaginar a María teniendo ese conocimiento. Quienes aceptan que, siguiendo la premisa (2), María aprende algo nuevo al ser liberada, es porque no imaginan correctamente lo que significaría la veracidad de la premisa (1), y esto ocurre porque al considerar el escenario podemos imaginar que María sabe muchísimo sobre el tema, pero no podemos realmente imaginar que sabe *todo* sobre el tema. Por el contrario, si imaginamos correctamente (si fuera posible), que María sabe todo sobre el tema, entonces estamos obligados a admitir que María no aprende nada. Antes de salir de su encierro María conoce perfectamente qué efectos neurofisiológicos producirá cada color en su sistema nervioso, por lo que lo único que necesita es encontrar una manera de identificar esos efectos en ella misma. Podría notar ciertas reacciones específicas que se producen en ella y que solo ocurren para el amarillo, o solo ocurren para el rojo, y así sucesivamente, y eso está garantizado por la correcta comprensión del alcance de su conocimiento físico completo. De aquí Dennett extrae dos conclusiones encadenadas: primero, que las premisas (1) y (2) son incompatibles; y segundo, que eso constituye un vicio propio de los experimentos mentales como el de la habitación de María.

Dennett no está queriendo afirmar que efectivamente si María existiera sin duda razonaría de dicha manera. Se trata de criticar la idea de Jackson de que “...Parece perfectamente obvio que aprenderá algo acerca del mundo y de la experiencia visual que tenemos de él...” (2003, pág. 100). Es decir, Jackson

propone un experimento mental con el propósito de generar el escenario adecuado para que la intuición de que el conocimiento físico es insuficiente para el conocimiento fenoménico, y a ello Dennett replica que la obviedad de tal intuición está, en cierta forma, contrabandeada. Entiendo que Dennett tiene un punto relevante a considerar aquí: es propio de la tradición filosófica occidental el alentar una actitud escéptica incluso, y especialmente, respecto a aquello que consideramos obvio. Por consiguiente, no es de recibo afirmar que “es obvio” que María aprendería algo. Cuando se utiliza con demasiada frecuencia el operador “es obvio” “tiene que encenderse una señal de alarma porque entonces, como por arte de magia, a los censores se les escabulle una premisa falsa con un guiño...” (Dennett, 2006, págs. 131-132). Si salimos del ámbito de lo obvio, para pasar al ámbito de lo concebible, tenemos que admitir que es tan concebible que María se sorprenda al ver algo rojo, como que María no se sorprenda y detecte el engaño de la banana azul.

Una vez establecido que no es legítimo plantear que “es obvio” que la premisa (2) es verdadera dada la correcta caracterización de la premisa (1), se produce una inversión de la carga de la prueba: Jackson y los defensores del AC deberían defender argumentativamente la veracidad de la premisa (2). Para ello habría que demostrar que sería imposible, aún con millones de palabras, poder explicar cómo se siente ver algo rojo. Según Hume (2003 [1748]), un ciego de nacimiento no puede entender cómo se siente ver algo rojo. Dennett eleva la apuesta, yendo incluso contra un referente del eliminativismo Tipo-A, al afirmar que no es del todo claro que con millones de palabras María no pudiera explicar al ciego de nacimiento (que debería tener la misma capacidad de procesamiento que María) cómo se siente ver algo rojo. De ahí, Dennett invierte la carga de la prueba y pide que se demuestre que María no puede calcular cómo se siente ver algo rojo.

Howard Robinson (2004) ha señalado que hay un error en la manera en que Dennett interpreta la premisa (1). Cuando (1) establece el conocimiento físico completo, se refiere a conocimiento de universales, pero Dennett la interpreta como incluyendo conocimiento de particulares. Es ese conocimiento de particulares, que

no ampara la premisa (1), el que le permitiría a María no sorprenderse al ver la rojez del tomate, y por tanto no aprender nada al ser liberada, o el que le permitiría detectar el engaño de la banana azul. María puede saber, por premisa (1), que luz con una longitud de onda del entorno de los 645 a 700 nm, produce la experiencia que habitualmente nos lleva a decir “eso es rojo”. Pero María no puede saber qué longitud de onda tiene la luz que incide en un objeto específico y llega a su retina, salvo que estuviera munida de algún instrumental especial, pero eso no está amparado por la hipótesis. Si bien se puede afirmar que las vías sensitivas visuales son verdaderos detectores de longitud de onda, la información que arrojan es fenoménica y no física. Por dicha razón, María puede conocer aspectos físicos y objetivables de los colores, pero no los aspectos subjetivos de la experiencia consciente que tenemos al ver un color particular, ya que esa vía neural no ha sido recorrida debido a su aislamiento. Esa es justamente la intuición a la que apunta el AC, y que le impediría a María identificar directamente los colores a pesar de su conocimiento físico completo.

Más aún, todo lo que María podría saber antes de su liberación es lo que uno podría decir y cómo uno podría reaccionar a ciertos estímulos de color. Entonces, cuando es enfrentada a la banana azul solo podría llegar a identificar el engaño haciendo uso de un detector de longitud de onda que arroje información física, y no uno que arroje información fenoménica (como lo es el circuito neural iniciado por los conos y los bastones). Al detectar el aspecto físico del estímulo a través de un instrumental apropiado, María sí podría asociar el estímulo con los enunciados que usualmente se profieren al enfrentarse a dicho tipo de estímulo, y al fin poder denunciar que está siendo engañada,

...pero será una revelación el hecho de que el tipo de estímulo físico que ella sabía se llamaba ‘azul’ y derivaba en conductas ‘apropiadas para lo azul’ *se sentía de esa manera*, y ese hecho fenoménico es lo que ella logra aprender (Robinson, 2004, pág. 71).

Para el lector que no es parte en esta disputa argumental queda la impresión, sobre la que habrá que volver, de que la cuestión se queda en un problema

terminológico. Desde la perspectiva del materialismo Tipo-A se afirmará que todo lo que existe es materia, y esto vale especialmente para la conciencia, por lo que las propiedades fenoménicas que se pueda adjudicar a los estados mentales no son más que ilusiones. La tesis materialista de Dennett implica que la conciencia fenoménica superviene lógicamente a lo físico, de lo que se sigue que si tenemos toda la información física (por premisa (1)), tendremos toda la información fenoménica, que en los hechos es física. Más aún, sería conveniente ir progresivamente eliminando la terminología mentalista que provoca la irrupción de problemas filosóficos donde solo hay problemas terminológicos. Por ello es que María puede, en principio, deducir cómo se vería la banana antes de salir, y denunciar el engaño.

Para Dennett, la experiencia consciente no tiene estatuto ontológico, y lo que se siente al ver algo rojo es un hecho físico (que involucra exclusivamente entidades materiales, que -por otra parte- es todo lo que existe). No hay una porción extra de información, que sería no física, sobre la que hubiera que dar cuenta, y por ello Dennett siente que no tiene la carga de probar nada. Desde su perspectiva, quien no propone la existencia de ninguna entidad adicional en el mundo es él, y por tanto no tiene que ofrecer ninguna explicación respecto a por qué algo no existe. ¿Cómo sería una explicación de ese tipo?

Por el contrario, Robinson sigue la línea que ya he explorado en otros exponentes del no reduccionismo como David Chalmers o Philip Goff. La objeción de Robinson parte de la certeza de que hay algo que se siente de cierta manera, la experiencia consciente, que es la novedad que sorprende a María. Como el no reduccionismo parte del “tomarse en serio la conciencia” (Chalmers, 1999 [1996]), o lo que es lo mismo, de la llamada “constricción de la conciencia” (Goff, 2017), según lo cual la experiencia consciente es un hecho real al que accedemos de manera inmediata y evidente (en términos cartesianos), entonces Dennett debería dar cuenta de ello. Dennett debería brindar una muestra de cómo sería esa inferencia que permita a María deducir, a partir de su conocimiento físico absoluto (de universales), cómo se sentiría ver algo rojo. La carga de la prueba la debería tener Dennett porque es el que propone la hipótesis contraria al sentido común, es el que

quiere desmontar el operador “es obvio”. ¿Cómo hemos de interpretar este ir y venir de cargas de la prueba? Por el momento pasemos a considerar otra línea crítica⁴⁶.

3.3.2 María aprende conocimiento no proposicional

Otra estrategia muy difundida para afrontar el AC, asociable por lo general a los materialismos Tipo-B⁴⁷, consiste en afirmar que, si bien es verdad que María aprende algo al ser liberada, ese aprendizaje no es aprendizaje de un nuevo saber proposicional que refiera a hechos, y por tanto que tenga condiciones de verificación de algún tipo. Para los defensores de esta estrategia, la premisa (1) es verdadera y legítima: efectivamente podemos imaginar un escenario en el que alguien conozca toda la información física que hay por conocer acerca de la visión de los colores. La premisa (2.a) también es verdadera, ya que coinciden en la intuición que consiste en que María podría sorprenderse al ser liberada, indicando que está aprendiendo algo nuevo. No obstante, no es tan claro lo que ocurre en (2.b). No es tan claro que la sorpresa sea un indicador de que María está ganando conocimiento proposicional. Cuestionada (2.b), no se sigue la conclusión que propone Jackson (3), porque no hay razones suficientes aún para señalar que hay verdades sobre la visión de los colores que escapen a la explicación fisicalista. Lo que María aprende en (2.b) cuando por fin sabe lo que se siente ver algo rojo, no es un nuevo conocimiento proposicional sino algún otro tipo de conocimiento. De ahí se sigue que el razonamiento aparentemente válido, es inválido por motivo de un equívoco en el uso de un término central: “saber acerca”, que tiene significados diferentes en (1) y (2).

⁴⁶ Volveré sobre este punto en los apartados “3.3.2.3 Aprender cómo se siente es familiarizarse” y “3.3.5 Fisicalismo, materialismo, y la premisa (1)”, así como en las conclusiones.

⁴⁷ Cfr. apartado “2.1.2 Materialismos reduccionistas (Tipo-B)” en la pág. 56.

Esta estrategia ha tenido dos presentaciones de acuerdo a qué tipo de conocimiento nuevo se postula para la premisa (2.b): la hipótesis de la habilidad (Nemirow, 1980) (Lewis, 2003), y la hipótesis de la familiaridad (Churchland P. M., 1985) (2004) (Bigelow & Pargetter, 2004) (Conee, 2004). Pasemos a analizarlas con sus consiguientes consideraciones críticas.

3.3.2.1 Aprender cómo se siente es adquirir una habilidad

Los defensores de la hipótesis de la habilidad suelen mencionar el siguiente pasaje de una reseña de Lawrence Nemirow como antecedente principal:

...algunos modos de conocimiento⁴⁸ consisten, no en aprehensión de hechos, sino en la adquisición de habilidades (...). De modo que el conocimiento de una experiencia puede entenderse como la habilidad para colocarse uno, a voluntad, en un estado representativo de tal experiencia. Puedo conocer la experiencia de estar mirando algo rojo si puedo visualizar algo rojo por mi propia voluntad (Nemirow, 1980, pág. 475)

Si bien Nemirow no está apuntando explícitamente al AC -es notorio que la publicación del artículo es anterior a la de los de Jackson-, sí se trata de una crítica a la idea de que la experiencia consciente no puede ser explicada totalmente desde el punto de vista de la tercera persona⁴⁹. Tal es así que dicha estrategia ha sido adoptada y elaborada con mayor detalle de manera específica para atacar el AC.

A Nemirow le llama la atención que una característica usualmente señalada de la experiencia consciente es la incomunicabilidad, que es lo que hace imposible, por ejemplo siguiendo nuevamente al propio Hume (2003 [1748]), que podamos explicar a alguien ciego de nacimiento cómo se siente ver algo rojo, o a un sordo

⁴⁸ Traduzco “*understanding*” por “conocimiento” en lugar de el más literal “entendimiento” por considerarlo apropiado dado el contexto.

⁴⁹ Por ejemplo el artículo ya mencionado de Nagel (2000 [1974]), que va en sinergia con el argumento que aquí nos convoca.

de nacimiento cómo se siente el sonido de un clarinete. Nemirow señala que la misma característica puede atribuirse a las habilidades. Para ilustrar, propongo pensar en un biofísico, que puede explicar teóricamente la sucesión de movimientos que ocurren cuando una persona nada, corre, o anda en bicicleta, pero es difícil para alguien explicar cómo es que puede hacer cada una de esas cosas. De la misma manera, no podrá explicar cómo se siente cada una de esas experiencias salvo usando metáforas que pueden tener valor poético pero que no pueden tener valor epistémico. Dicha analogía le lleva a Nemirow a proponer que saber cómo se siente determinada experiencia consciente no es más que tener la habilidad de poder colocarse en ese estado.

Si llevamos este planteo a la discusión específica del AC, tenemos que preguntarnos en qué consiste eso que María no tiene en la conclusión (3), para determinar si se trata algo que pueda explicarse en términos físicos o no. En caso de encontrar que eso que María no tiene es algo que admite una descripción física, es necesario dar cuenta de cómo puede compatibilizarse dicho hallazgo con la premisa (1), según la cual María tiene un conocimiento completo de la información física disponible acerca de la visión de los colores. Según la lectura de Jackson, eso que María adquiere es información fenoménica, no pasible de una explicación física (3), y de ello se sigue la conclusión ontológica (4) por la que el materialismo es falso.

David Lewis (2003), quien ha sido el exponente más influyente de la hipótesis de la habilidad, afirma que María efectivamente ha aprendido algo al salir de su encierro, gracias a la novedosa experiencia de ver algo rojo. Sin embargo, el hecho de que María aprenda algo no afecta en nada la tesis fisicalista, porque lo que aprende es una habilidad que puede ser perfectamente explicada por las estrategias reduccionistas habituales de la neurociencia. Lo que aprendemos a través de la experiencia no es información fenoménica, por lo que el materialista no tiene nada que temer.

La idea original de Jackson respecto a su AC consiste en que lo que María adquiere, que no tenía a pesar de tener el conocimiento físico completo, es información fenoménica. La información fenoménica sería irreductible a, y por tanto no deducible de, información física, razón por la cual María se sorprendería al ver algo rojo. Así, tener una experiencia supone incorporar dos tipos de información: información física (que María ya tendría de antemano por sus conocimientos previos), e información fenoménica (que María no tendría de antemano a pesar de sus conocimientos previos). El objeto de esa información fenoménica sería el carácter intrínseco de la experiencia, que es capaz de suscitar estados mentales con caracteres fenoménicos o *qualia*. Sin embargo, para el materialista Tipo-B esos estados mentales con *qualia* no son nada más (y nada menos) que patrones de activación de redes neuronales, por lo que María habría de saber acerca de dichos patrones.

Quedamos ante un dilema. Dado que la estructura del argumento funciona con validez formal, Lewis tiene que demostrar, por un lado, que eso nuevo que adquiere María no es información fenoménica, sino alguna otra cosa que sea admisible para el materialista, y que no contradiga la premisa (1). Por otro lado, deberá proponer una teoría razonable acerca de qué es eso que aprende María. Para argumentar que no se trata de información fenoménica, Lewis propone dos argumentos. El primero, que podríamos llamar “argumento boomerang” o “argumento de la paridad de razones”, merece una consideración aparte a partir de la página 118. El segundo consiste en atacar la postura epifenomenista, que es la que Jackson busca defender mediante la presentación del AC.

Según Lewis, no se puede defender el AC sin arribar a una postura de ese tipo, mas tal postura es insostenible, por lo que algo del argumento ha de fallar. Como lo que falla no es la estructura formal del argumento, entonces tiene que ser la hipótesis de que lo que María aprehende al liberarse es información fenoménica. Para el epifenomenismo, tal como fue presentado en el capítulo anterior, una característica peculiar que tiene la información fenoménica es que está totalmente aislada de todo otro tipo de información. Por otra parte, la información novedosa a

la que accede María al salir de su encierro habría de permitirle, en tanto información, eliminar posibilidades. Mientras que carecer de información significa tener ante uno muchas posibilidades, la información sobre cualquiera que sea el caso permite ir eliminando posibilidades de acción o de juicio. Pero una información fenoménica completamente aislada del mundo físico, difícilmente podría servir a María para eliminar posibilidades a la hora de actuar o juzgar que algo es el caso. Precisamente la idea del epifenomenismo es que los *qualia* serían el eslabón final de una cadena causal, pero que ellos mismos no son causa de ninguna otra modificación en el mundo físico. Así, esa información fenoménica que María adquiere por una parte ha de permitirle juzgar, por ejemplo, exclamando “¡Así se siente ver un tomate maduro!”, pero contradictoriamente, esa información fenoménica no debería ser capaz de suscitar ella misma hechos del mundo físico como esa preferencia, y mucho menos funcionar como criterio de eliminación de posibilidades.

En sustitución de la idea original de Jackson, Lewis sugiere, a modo de hipótesis, que lo que María adquiere es una habilidad. Más aún, que la expresión “saber cómo se siente” refiere a un saber del tipo saber-como (*know-how*) y no saber-qué (*know-that*). Saber cómo se siente una experiencia, que es lo mismo que tener la experiencia, consiste en tener un conjunto de tres habilidades: la habilidad de reconocer que uno está teniendo la experiencia (nuevamente), la de recordar que uno tuvo la experiencia, y la de poder imaginarse a uno mismo teniendo esa experiencia aun cuando la misma no esté presente. Bajo esta hipótesis, María no adquiere las habilidades gracias a que aprehende un *quantum* de información fenoménica nueva, ya que aprender cómo se siente ver algo rojo no es nada por encima o más allá que tener el conjunto de habilidades de reconocer, recordar e imaginar lo rojo. Dichas habilidades notoriamente admiten una explicación físico-funcional, así que se trata de una hipótesis compatible con el fisicalismo.

Como Lewis no niega que dichos estados mentales tengan un carácter fenoménico, sino que lo que niega es que dicho carácter se deba a la incorporación de información no física, entonces se trata de una postura Tipo-B que puede

formularse de la siguiente manera: la experiencia consciente superviene nomológicamente a lo físico, de manera tal que no puede haber diferencias en las propiedades fenoménicas de un estado mental sin que también haya diferencias en las propiedades físicas. Esto tiene que ser así porque las propiedades fenoménicas quedan explicadas como habilidades cognitivas. No podemos adscribirle a Lewis una tesis de superveniencia lógica porque él mismo se encarga de explicar que la posición fisicalista, que él mismo defiende, es una tesis contingente acerca de cómo es nuestro mundo actual, pero que no descarta que en otros mundos posibles la relación pudiera ser diferente (Lewis, 2003). La identificación del conjunto de habilidades con el “saber como se siente...” ha de ser establecida *a posteriori*.

En síntesis, Lewis admite que la experiencia es capaz de suscitar aprendizajes, pero que hay mejores razones para pensar que se trata de aprendizaje de las habilidades mencionadas y no de nuevas proposiciones acerca de información fenoménica no física del mundo. La hipótesis de la habilidad sería preferible porque no implica una posición insostenible como el epifenomenismo, y porque es inmune al argumento de la paridad de razones de la página 118, mientras que logra explicar por qué la experiencia efectivamente permite suscitar algún tipo de aprendizaje de forma compatible con una estrategia reduccionista.

3.3.2.2 Dificultad: el conjunto de habilidades de Lewis es muy exigente

La identificación del conjunto de habilidades de Lewis (reconocer, recordar e imaginar una experiencia) con la expresión “saber cómo se siente tener la experiencia”, puede expresarse en términos de condiciones necesarias y suficientes. Dado que Lewis propone que es la conjunción de las tres habilidades lo que equivale a “saber cómo se siente...”, entonces puede establecerse que cada una de ellas es condición necesaria, y la conjunción de las tres ha de ser condición suficiente para saber cómo se siente una experiencia. A la inversa, saber cómo se siente tener una experiencia es suficiente para tener cada una de esas habilidades, y necesaria para tener el conjunto de las tres. Una estrategia para defender el AC de la hipótesis de

la habilidad ha consistido en demostrar que esa equivalencia, por diferentes razones, no funciona. Se puede mostrar, habilidad por habilidad, que cada una de ella ni es necesaria ni es suficiente para “saber cómo se siente una experiencia”.

En esa línea, Raymont (1999) propuso un argumento que puede sintetizarse de la siguiente manera, sobre la base de que puede haber contraejemplos para refutar cada bicondicional entre habilidad y experiencia:

1. Puede haber habilidad de reconocimiento de las propiedades de una experiencia sin experiencia.
 2. Se puede saber cómo se siente una experiencia sin que se ponga en juego la habilidad de reconocer una experiencia.
 3. Puede haber habilidad de recordar una experiencia sin haberla experimentado.
 4. Se puede saber cómo se siente una experiencia y luego no tener la habilidad de recordarla.
 5. Puede haber habilidad de imaginar una experiencia sin haberla experimentado.
 6. Se puede saber cómo se siente una experiencia sin por ello tener la habilidad de imaginarla.
-
7. Saber cómo se siente una experiencia no es condición necesaria ni suficiente para tener las habilidades de reconocer, recordar e imaginar una experiencia, ni en conjunto, ni de forma disyuntiva.

Cada una de las premisas puede respaldarse en algún ejemplo o caso emblemático. Aquí es donde el argumento pierde peso, ya que en varias ocasiones se recurre a ejemplos que son, al menos, discutibles. Para sostener la premisa (1), Raymont hace referencia al célebre problema de Molyneux, tomándolo como algo

ya resuelto a partir de cierta evidencia empírica⁵⁰, para afirmar que es posible el reconocimiento de un objeto por un sentido al que se conoció por otro. La idea es que en casos de ese tipo hay reconocimiento de objetos sin experiencia previa, ya que el sujeto podría reconocer a través de la vista un cubo o una esfera que previamente solo había conocido a través del tacto.

Además del defecto de fundar la premisa en un experimento mental para el que no existe evidencia concluyente, hay una disanalogía entre lo que se propone en el contexto del experimento de Molyneux y lo que está proponiendo Lewis en el contexto del experimento de María. Mientras que en el primero la discusión tiene que ver con la posibilidad de reconocimiento multimodal, esto es, de que el reconocimiento de un objeto por una modalidad sensorial permita el reconocimiento del objeto por otra modalidad; en el caso de la liberación de María lo que ocurre compromete propiedades que, como el color, son unimodales. De esa forma, aún si consideráramos dirimido el debate en torno a la intermodalidad sensorial, que no lo está, no es correcta la utilización de ese tipo de casos como contraejemplo de la hipótesis de la habilidad.

De forma análoga, otros contraejemplos que Raymont propone, u otros que podríamos sugerir alternativamente, no están exentos del mismo tipo de dificultades, ya sea porque carecen de evidencia concluyente, o porque no son aplicables al caso de la liberación de María. Sería posible proceder a ampliar el respaldo a la estrategia de Raymont, buscando respaldo académico empírico o especulativo para cada una de las premisas⁵¹, pero considero que cualquier intento

⁵⁰ Cfr. Locke (L. II, Cap. IX, 8, 1956 [1689]); Leibniz (1977 [1765]) ; Raymont (1999); Zakeri & Ghasemi (2016).

⁵¹ Para fundamentar la premisa (2) puede pensarse en cómo puede haber experiencia consciente sin que ello implique reconocimiento, como en la experimentación con ilusión de Kanizsa que puede seguirse en (Vanderbroucke, Sligte, Fahrenfort, Ambroziak, & Lamme, 2012). La premisa (3) podría

de ampliación de ese tipo caería en dificultades similares. A modo de conjetura, sugiero que este tipo de dificultades surge del intento aplicar lo que resulta de la evidencia experimental para defender u objetar una propuesta que es conceptual. La hipótesis de la habilidad es la propuesta de definir de determinada manera a la experiencia consciente, y, si hemos de considerar alguna crítica a dicha hipótesis, la misma ha de ser de la misma naturaleza. Por tal razón remito al lector al apartado “3.3.2.5 ¿Se trata de una estrategia efectiva?” en el que desarrollo una crítica más viable, en opinión de quien escribe, que afecta tanto a la hipótesis de la habilidad como a la hipótesis del conocimiento por familiaridad, otra propuesta que sugiere entender la experiencia consciente como conocimiento no proposicional.

3.3.2.3 Aprender cómo se siente es familiarizarse

El AC ha sido objetado siguiendo la estrategia del conocimiento no proposicional, pero sin concordar con la hipótesis de la habilidad, sugiriéndose como alternativa la llamada “hipótesis de la familiaridad”⁵². La estructura básica de la crítica es similar a la de la hipótesis de la habilidad: María, efectivamente aprende cómo se siente ver algo rojo al ser liberada, pero ese aprendizaje no es aprendizaje de un nuevo conocimiento proposicional, por lo que el argumento no refutaría al fisicalismo y menos aún al materialismo. Lo que los defensores de esta variante afirman, es que “saber cómo se siente una experiencia” no es adquirir habilidades

respaldarse en los experimentos de creación de memorias artificiales en ratas (Ramirez, y otros, 2013) (Liu, Ramirez, & Tonegawa, 2014). La premisa (4) podría hallar su sustento en el curioso experimento mental que propone Hume en la sección (ii) de sus *Investigaciones* (2003 [1748]) en el que se afirma que un sujeto que nunca experimentó un matiz de azul podría imaginarlo a partir del conocimiento previo de los matices adyacentes en una sucesión exhaustiva y ordenada de matices de azul. Es notorio que este experimento mental es sumamente endeble, al punto que va en contra de la posición empirista del propio Hume. Se puede ampliar la lista de contraejemplos al bicondicional que surge de la hipótesis de la habilidad, el punto es que se trata de contraejemplos tan discutibles como la hipótesis misma.

⁵² Traduzco “*acquaintance*” por “familiaridad”.

sino familiarizarse con dicha experiencia. Pasemos a explorar la variante y algunas de las razones por las que podría ser preferible al enfoque de la habilidad.

Según Bigelow y Pargetter (2004), dos creencias pueden ser muy similares entre sí, y sólo diferenciarse por la manera como el sujeto se vincula con los objetos de dichas creencias. Las dos creencias pueden tener las mismas condiciones de verdad y ser verdaderas en los mismos mundos posibles, pero diferenciarse en relación a si el sujeto que cree se encuentra en una relación tal con la estructura del mundo que es en virtud de esa relación que sus creencias son verdaderas en ese mundo. Estas relaciones relevantes con la estructura del mundo son las que pueden catalogarse de relaciones de familiaridad.

Un ejemplo simple puede ser la creencia que tengo de que “me gusta cómo se siente comer chocolate”, que obtiene respaldo simplemente por el hecho de que he comido chocolate, me he familiarizado con su sabor, aroma, textura, etc., y todo ello me ha resultado agradable. Diferente sería el caso de un niño muy pequeño que nunca haya probado chocolate, y que crea que “me va a gustar comer chocolate” porque confía en sus padres que le dicen que es delicioso. La creencia del niño se respalda en la confianza en sus padres y no en que esté familiarizado con el sabor del chocolate. Uno se familiariza con un estado de cosas cuando entra en una forma nueva de relacionamiento con dicho estado de cosas, esta novedad hace que la relación del sujeto con el mundo sea especialmente apropiada para enunciar creencias verdaderas, habilitando a la justificación de afirmaciones como “me gusta cómo se siente comer chocolate”.

Podemos aplicar la categoría de “familiarización” al conocimiento que María tiene respecto a saber cómo se siente ver algo rojo. El nuevo conocimiento de María solo difiere en este sentido muy específico: en la forma de familiarizarse con el *quale* relevante. El escenario planteado por Jackson logra aislar de manera muy precisa el factor familiaridad, colocándolo en el lugar de la novedad que irrumpe en la liberación de María. El nuevo conocimiento de María solo difiere del anterior en la forma que tiene de familiarizarse con la experiencia de ver los colores.

No se trata de un nuevo conocimiento proposicional de hechos desconocidos hasta el momento. Se trata de un conocimiento proposicional que ya se había adquirido, pero que se confirma por una vía nueva, más directa. Mientras que en el caso del niño, seguramente su conocimiento sobre el chocolate sea superior luego de probarlo que antes, en el caso de María, el conocimiento físico completo del que parte podría garantizar que, a pesar del cambio de vía de acceso al conocimiento, este no es mayor que antes de la liberación.

Aprender cómo se siente una experiencia es algo idéntico a familiarizarse con ella, particularmente con sus cualidades fenoménicas. Ahora, desde tiendas fisicalistas ha de entenderse que la cualidad experimentada es una propiedad física. Si el rojo fenoménico es una propiedad física, las premisas (1) y (2)⁵³ pasan a ser premisas incompatibles. La forma de solucionar el conflicto es admitir que María conocía todos los hechos físicos sobre la percepción del color, y que al salir de su encierro no está aprendiendo un nuevo hecho sino familiarizándose de forma directa con alguna propiedad de un hecho ya conocido. No hay aprendizaje de hechos, hay un familiarizarse con propiedades de hechos ya conocidos (Conee, 2004). Esto solo es posible admitiendo que no hay otra forma de familiarizarse con una cualidad que experimentándola, oponiéndose a la sugerencia de Dennett (2006), y que también insinuó Lewis (2003), de que podríamos adquirir este tipo de familiarización a través de otros medios, incluso a través de una deducción, contrariando un núcleo duro del empirismo clásico⁵⁴. Conee entiende que cualquiera de los tipos de conocimiento habitualmente considerados (conocimiento fáctico, conocimiento de habilidad, y conocimiento por familiaridad) ha de implicar que el sujeto se halle en una posición óptima para alcanzar un logro epistémico. Respecto al conocimiento de cómo se sienten las propiedades fenoménicas de los objetos ordinarios que nos

⁵³ Puede ser útil a lector volver a la pág. 77 para recordar cuál es cada premisa del AC, en esta y otras ocasiones.

⁵⁴ Cfr. pág. 81.

rodean, la posición óptima de logro epistémico no puede ser otra que la de familiarizarse con dichas propiedades a través de la experiencia directa. Para esto no es suficiente conocer los hechos, es necesario entrar en una relación particular con los hechos.

Estamos en condiciones de entender, según esta línea, dónde queda ubicado el equívoco terminológico que impide concluir que hay hechos no físicos sobre la percepción de los colores. María entra en un nuevo estado de familiaridad, pero no cree una nueva verdad: ya sabía que los tomates son rojos. Dicha familiaridad solo puede adquirirse por introspección de lo que ocurre cuando se está ante el estímulo apropiado. La familiarización introspectiva de cómo se siente ver el tomate rojo es lo que permite la adquisición de la habilidad de discriminar los tomates rojos de los verdes, por ejemplo. La familiarización es previa y es condición de la adquisición de la habilidad, y por ello sus proponentes defienden que esta hipótesis es mejor que la hipótesis de la habilidad (Bigelow & Pargetter, 2004) (Conee, 2004). Pasemos a considerar un argumento neurobiológico en favor de esta alternativa.

3.3.2.4 Un argumento neurobiológico a favor de la hipótesis de la familiaridad

Paul Churchland (2004) entiende que el enfoque de la hipótesis de la habilidad va en el sentido correcto, dado que Lewis y Nemirow logran dar en el clavo al hallar el equívoco entre las dos formas de saber referidas por “saber acerca” en las premisas (1) y (2) del AC, pero falla en reducir el equívoco a una dualidad saber-qué/saber-como, cuando el caso es que la neurobiología ha identificado muchas otras maneras de conocer. Coherente con su postura eliminativista Tipo-A, entiende que la distinción mencionada pertenece al lenguaje *folk*, y que un enfoque científico no debería ceñirse a una distinción binaria no fundamentada como esa.

Obsérvese, por ejemplo, que en la literatura se distinguen dos grandes vías neurales paralelas de procesamiento de la información visual (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013). Simplificadamente, la vía ventral es la

encargada del reconocimiento de objetos, mientras que la vía dorsal es la que conecta la visión con el control motor. De esta manera, la vía dorsal es preminente para la utilización de estímulos visuales en conexión con el desarrollo de habilidades motrices, como por ejemplo esquivar objetos, apuntar hacia un blanco, caminar manteniendo el equilibrio, etc. En contraste, las habilidades de Lewis parecen ser procesadas por la vía ventral. Hay aquí una primera división binaria.

Sin embargo, la cuestión es más compleja, ya que ambas vías están interconectadas en varios puntos y “colaboran” una con la otra, volviéndose esa distinción binaria menos clara (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013). Esto es, la distinción sigue vigente pero no debe considerarse como representando dos formas radicalmente diferentes de conocer a través de la vía visual. La distinción saber-qué/saber-como no permite identificación de tipos puros, ni es exhaustiva.

¿En qué consistiría la familiarización en términos neurobiológicos? “Estar familiarizado” puede entenderse como haber desarrollado una representación del rango de colores familiares, gracias a la configuración apropiada de ciertas redes neuronales en las áreas corticales apropiadas. Es un proceso que ocurre cuando el estímulo se presenta de forma directa al sujeto, lo que permite que en el sujeto se dispare el proceso de configuración de las redes neuronales apropiadas. El reconocimiento de los colores depende de que esa modulación previa de las redes de la corteza visual haya tenido lugar, especialmente en V4 (Churchland P. M., 2004). No puede haber habilidad de reconocimiento si el individuo no está familiarizado con determinado color. El detalle histológico y anatómico, si se trata de tal o cual área, o en qué consiste la regulación de los pesos sinápticos necesaria para dicha configuración, es de importancia secundaria para nuestros propósitos. Lo medular radica en que no hay habilidad de discriminación de un color si no hay una previa familiarización con dicho color, y, más importante aún, con un esquema general ordenador del mundo del color. Dicha familiarización es un proceso del que se puede dar cuenta -según Churchland- en términos estrictamente materialistas.

Este reconocimiento depende de que la criatura posea una representación previa - una configuración aprendida de sinapsis que conecte las poblaciones relevantes de células- que particione con anterioridad la taxonomía visual de la criatura para que pueda responder selectiva y apropiadamente al flujo de la estimulación visual que llega desde la retina y el núcleo geniculado lateral. (Churchland P. M., 2004, pág. 166)

La representación previa que funciona como condición de toda habilidad de discriminación, recuerdo e imaginación de colores es no discursiva, por lo que puede ser perfectamente un candidato para ocupar el rol que en la propuesta de Lewis-Nemirow ocupa la habilidad misma. Ese marco cognitivo que posibilita dichas habilidades es de lo que María carece durante su cautiverio, y es lo que aprende al ser liberada. Efectivamente es un aprendizaje no proposicional, dejando indemne la idea contenida en la premisa (1) del argumento, según la cual María tiene un conocimiento completo de todo lo que hay que saber (proposicionalmente) acerca de la fisiología de la percepción del color. Ahora sí, el equívoco terminológico puede señalarse en la dualidad en el uso de “saber acerca”, en donde “saber” en (1) refiere a un conocimiento discursivo mientras que en (2) refiere a un conocimiento no discursivo por el que María se familiariza con la percepción de los colores posibilitando la modulación de las redes neuronales adecuadas para el posterior desarrollo de las habilidades de Lewis-Nemirow.

Si se pretende solucionar el equívoco, y asignar a “saber acerca” de (1) el sentido explicado de “saber acerca” en (2), la primera se volvería tan fuerte que sería implausible, definitivamente falsa. En este punto el análisis de Churchland se toca con el de Dennett estudiado anteriormente⁵⁵. Recordemos que Dennett planteaba que los defensores del AC subestiman la fuerza de la premisa (1), esto es subestiman qué implicaría que María sepa todo lo físico que hay que saber sobre la percepción del color. Dennett realiza dicha afirmación porque está entendiendo

⁵⁵ Cfr. el apartado “3.3.1 María no aprende nada: la banana azul”, pág. 81.

“saber acerca” en el sentido no discursivo que propone Churchland para la premisa (2).

El encuentro de las dos líneas de objeciones pone de manifiesto una dificultad para el sostenimiento de la postura de Dennett. Este no podría sugerir que María podría deducir cómo se siente ver algo rojo a partir de su conocimiento no discursivo establecido en la premisa (1)⁵⁶. Para sostener la posibilidad de una deducción de ese tipo, Dennett tendría que asumir que una verdad intuitiva, no discursiva, puede funcionar como punto de partida de una deducción que luego concluya un conocimiento discursivo. La discusión sobre si un razonamiento deductivo de este tipo, que parta de una premisa no discursiva y arribe a una conclusión discursiva, rebasa los objetivos de esta tesis, y, en cualquier caso, es interesante haber encontrado un punto en el que ambas estrategias se cruzan⁵⁷.

Una segunda dificultad irrumpe si nos preguntamos si la premisa (1) interpretada de esta manera sería plausible dado el escenario de confinamiento y asilamiento cromático en el que ha vivido siempre María. Dejo esta última idea inconclusa para retomarla más adelante en la sección “3.3.5 Fisicalismo, materialismo, y la premisa (1)”.

En síntesis, la evidencia neurobiológica no sustenta la distinción binaria entre saber-qué y saber-como, y por el contrario asume que familiarizarse con las cualidades sensoriales de los objetos es condición para luego poder adquirir alguna de las habilidades de Lewis y Nemirov. Presentada la hipótesis de la habilidad, y luego la de la familiaridad, y desarrollados los argumentos para sopesarlas, pasemos a discutir si la estrategia del conocimiento no proposicional (que incluye a ambas

⁵⁶ Cfr. pág. 81 y más adelante el apartado “3.3.5.3 Plausibilidad de la interpretación de Dennett”.

⁵⁷ Es oportuno recordar que, al presentar la distinción cartesiana de tres niveles de sensación, afín a la propuesta de Raftopoulos, vimos que es el tercer nivel de la sensación el que ofrece ideas que a las que se le puede predicar algún valor de verdad (cfr. pág. 25).

hipótesis como alternativas viables) efectivamente responde al AC, y si son suficientes para sostener una posición fisicalista respecto al conocimiento de las cualidades fenoménicas de nuestra experiencia, y materialista respecto a su estatuto ontológico.

3.3.2.5 ¿Se trata de una estrategia efectiva?

Respecto a la hipótesis de la habilidad, Sam Coleman (2009) entiende que, aun siendo correcta, no hace nada para afectar el argumento del conocimiento. Considero que esta postura es perfectamente extensible a la hipótesis de la familiaridad, por lo que puede generalizarse al enfoque que incluye a las dos, y que establece que lo que María adquiere al salir de su encierro es conocimiento no proposicional. De ninguna de las dos variantes se sigue que María no gane conocimiento fáctico, y tampoco se sigue que haya o no haya propiedades o hechos no físicos involucrados en su progreso epistémico.

Discutir si María aprende habilidades, o se familiariza con propiedades, no sirve para decidir si María aprende algo fáctico o no. Una cuestión es si lo que María aprende es reductible a habilidades o no. Otra diferente es si la adquisición de habilidades, o la familiarización, involucra o no hechos. Ambas discusiones son ortogonales, corren por carriles diferentes (Coleman, 2009). De la respuesta que demos a una de ellas no se sigue nada respecto a la otra. Pensar lo opuesto configuraría un error categorial. Cualquiera de los cuatro escenarios es posible: María aprende una habilidad nueva sin aprender sobre nuevos hechos, o María aprende una habilidad nueva gracias a que aprende nuevos hechos, o no hay

habilidades involucradas pero sí nuevos hechos, o finalmente no hay habilidades involucradas ni nuevos hechos⁵⁸.

La habilidad de andar en bicicleta requiere el monitoreo del carácter fenoménico de los estados de equilibrio del sujeto, y no parece disparatado proponer que la información que ofrece la propiocepción tiene un carácter fenoménico. Es plausible afirmar que cuando uno se cae, la experiencia se siente de determinada manera. De lo contrario sería inexplicable la cantidad de personas que desean tener experiencias de caída libre como en el paracaidismo, en los juegos mecánicos de los parques de atracciones, o en el salto *bungee*. Esas personas están sintiendo algo, la caída, la excitación, el miedo, etc. El conocimiento de la biofísica del movimiento y la fisiología de la propiocepción y el control motor, explica la habilidad, pero no es suficiente para su adquisición.

...es la falta de *conocimiento fenoménico* del agente -esto es, más precisamente, su falta de familiaridad con la cualidad de los estados específicos de sensación- lo que implica la imposibilidad de adquirir habilidades en un salón de clases (Coleman, 2009, pág. 11).

Como ya vimos anteriormente, el conocimiento fenoménico es epistémicamente previo a la adquisición de la habilidad, y la cuestión de la naturaleza del conocimiento fenoménico permanece intacta por la crítica de Lewis-Nemirow al AC. Sería posible asentir a sus argumentos y aún mantener que María se sorprende al liberarse porque aprende conocimiento de hechos no físicos, y de ahí defender una postura dualista.

Coleman se limita, en la fuente citada, a defender la inutilidad dialéctica de la hipótesis de la habilidad, y el pasaje citado parece insinuar que podría adoptarse como alternativa la hipótesis de la familiaridad. Sin embargo, como he afirmado al comienzo de este apartado, sobre ésta última es posible realizar aseveraciones

⁵⁸ Vale reiterar que lo mismo que en este apartado comento respecto a la hipótesis de la habilidad vale para la hipótesis de la familiaridad.

análogas. Tal es así que el propio Conee (2004), se encarga de dejar establecido que su propuesta no implica una toma de postura respecto a si hay, o no hay, información no física involucrada en el proceso de familiarización con las propiedades fenoménicas. Aprender cómo se siente una experiencia es familiarizarse con las cualidades fenoménicas de dicha experiencia, si ese proceso involucra eventos no físicos, o solamente físicos, y lo mismo respecto a las propiedades, es algo que escapa a su enfoque. Conee es enfático y lo establece en varios pasajes de su artículo. Con más razón se podría afirmar que la hipótesis de la familiaridad no afirma nada respecto a la estructura fundamental de la realidad.

La respuesta al argumento del conocimiento que ha sido defendida aquí pretende ser compatible con la veracidad del fisicalismo. La propuesta tiene la intención de ser neutral respecto a la pregunta sobre si el evento de una persona experimentando una cualidad fenoménica es un evento enteramente físico o no (Conee, 2004, pág. 206).

La hipótesis de la familiaridad respalda la intuición básica tras el AC, ya que coincide con Jackson en que las cualidades fenoménicas quedan afuera de cualquier teoría física completa. Sin embargo, Conee entiende que esto no implica un desafío para el fisicalismo porque no se sigue de ahí su falsedad.

Los hechos físicos podrían incluir todo hecho acerca de los *qualia*. Aun así, la explicación fisicalista todavía “deja afuera a los *qualia*”, en el sentido de que el conocimiento de los hechos físicos no implica conocimiento de los *qualia*. (Conee, 2004, pág. 212)

Más adelante puede leerse una afirmación similar respecto al aspecto ontológico de la discusión:

El enfoque de la familiaridad no tiene compromisos metafísicos. La relación de la experiencia no tiene que ser una relación simple con una cualidad fenoménica, ya que todo lo que el enfoque implica es que experimentar atentamente una cualidad podría ser un estado cerebral de una naturaleza neurofisiológica compleja, e igualmente podría ser una relación simple e inanalizable entre un alma y una cualidad no física. El materialismo ni está implicado ni está excluido (Conee, 2004, pág. 211).

Y otro tanto afirman Bigelow y Pagetter, que, como vimos, defienden un enfoque afín al de Conee:

En este artículo, mostramos que el argumento del conocimiento no proporciona un caso conclusivo contra el fisicalismo. No probamos, ni presuponemos, que el fisicalismo sea verdadero, solo que puede resistir el argumento del conocimiento (Bigelow & Pargetter, 2004, pág. 179).

Esta sucesión de citas, quizás excesiva, no tiene otra pretensión que demostrar que lo que Coleman plantea respecto a la hipótesis de la habilidad puede adjudicarse perfectamente a la hipótesis de la familiaridad. Incluso aceptando la teoría que sus defensores proponen, dicha aceptación no logra, ni intenta, proveer una defensa acabada del fisicalismo ni del materialismo. Ambas variantes dejan abierta la cuestión de si las propiedades fenoménicas que permiten la adquisición de una habilidad, o a las que accedemos por familiaridad, supervienen nomológicamente a lo físico o no. Al provenir de posturas Tipo-B (con el estatuto dudoso de la posición de Paul Churchland, que parece ser limítrofe), aceptan la no superveniencia lógica de la conciencia fenoménica a lo físico, pero no logran avances respecto a la cuestión nomológica, y nos dejan dando vueltas en círculos.

Como conclusión parcial podemos afirmar que la objeción del conocimiento no proposicional, que incluye a la hipótesis de la habilidad y a la de la familiaridad, aún siendo cualquiera de ellas verdadera, no logra desactivar el argumento del conocimiento ni en su versión débil ni en su versión fuerte. Cada una de ellas podría ser verdadera sin por ello implicar veracidad o falsedad del fisicalismo ni del materialismo. Pasemos ahora a considerar otra línea crítica al argumento del conocimiento.

3.3.3 María aprende nuevo conocimiento proposicional sobre hechos ya conocidos

Una tercera estrategia por la que se ha criticado al AC ha consistido en aceptar que María aprende algo nuevo al salir de su cautiverio, y aceptar que ese aprendizaje consiste en la adquisición de un conocimiento proposicional, con la salvedad no menor de que dicho conocimiento proposicional sería de hechos que

María ya conocía. Por tanto, no se seguiría del AC que existan hechos no físicos no amparados en el conocimiento físico completo de la premisa (1). Se puede afirmar la premisa (1), se puede conceder que María se sorprende al liberarse por premisa (2.a) y que la sorpresa es porque aprende un nuevo conocimiento proposicional por premisa (2.b), y de ello se sigue la conclusión (3) por la que existen conocimientos no físicos de la percepción del color, pero no la conclusión (4) por la que esos conocimientos involucran entidades o propiedades no materiales.

Como se admite la ganancia de nuevo conocimiento proposicional, esta estrategia es notoriamente afín a la intuición de que hay una brecha epistémica por la que la explicación fisicalista no es suficiente para dar cuenta de la experiencia consciente, pero tiene como fin obturar el salto inferencial desde la dimensión epistemológica a la ontológica, defendiendo que no se sigue de la brecha epistémica una brecha ontológica que refute el materialismo. Se trata de una línea afín al materialismo Tipo-B, y puede entenderse que asume alguna variante de la teoría de la identidad psico-física, o de la teoría de la identidad de los estados funcionales, ya que asume que el hecho que María conoce al ser liberada desde una nueva perspectiva experiencial, es el mismo hecho que María ya conocía gracias a su vasto saber académico.

Esta línea ha sido presentada bajo numerosas variantes, y todas ellas suelen englobarse bajo el lema “Nuevo conocimiento / Viejo Hecho” (NC/VH a partir de ahora), en el que se asume que la novedad post-liberación radica en un nuevo modo de presentación de un hecho ya conocido. Resulta imposible revisar las múltiples variantes a esta estrategia, presentaré tan solo algunos planteos clásicos en la literatura con la intención de ilustrar al lector sobre la estrategia y las posibles respuestas que le caben. Puede resultar llamativo el hecho de que esta línea crítica al AC es muy cercana a la de la hipótesis de la familiaridad, ya que en ambos casos se asume que podemos conocer los hechos vía diferentes modos de presentación, con la salvedad de que ésta última constituye una respuesta no conceptualista -ya que se asume que el conocimiento por familiaridad es no conceptual- mientras que

aquí se trata de una alternativa conceptualista que asume la existencia de conceptos fenoménicos con características particulares.

Para defender que el enfoque apropiado es conceptualista, Nida-Rümelin (2004) ha presentado una variante a nuestro experimento que logra separar el progreso epistémico de María en su liberación en dos pasos. Marianna, que puede o no tener un conocimiento físico completo de la fisiología de la visión, lo cual no es relevante en este caso, antes de salir de su cautiverio para poder al fin experimentar como se siente ver un tomate, es confrontada con tarjetas de figuras geométricas cuya forma no es típicamente asociable a colores. En ese contexto, Marianna realiza un primer progreso epistémico: se abre ante sus ojos una nueva categoría fenoménica, la del color, que antes no tenía. Ante figuras de colores primarios, Marianna podría especular respecto a qué color se parece el color del cielo, y a qué color se parece el del tomate maduro. Hasta aquí todo se parece al planteo de Paul Churchland desarrollado en la página 97. Lo diferente radica en que es razonable suponer que Marianna podría equivocarse en su elección. Por ejemplo, podría pensar que, como ha oído a la gente decir que el color del cielo es bello, y como ella siente un placer estético especial al contemplar la tarjeta roja, entonces el cielo ha de ser del color de esa tarjeta. No se trata de un caso de inversión del espectro, simplemente los colores entraron en su mundo generando un nuevo ámbito sobre el qué Marianna puede juzgar, y al juzgar puede acertar o errar. Al salir de su cautiverio y al fin poder mirar el cielo despejado, Marianna está en condiciones de decir “me equivoqué, pensé que el color del cielo era como la tarjeta 1 y en realidad se veía como la tarjeta 2”. Ahí el segundo progreso epistémico de Marianna.

La separación del progreso epistémico en dos partes abre la posibilidad de que Marianna juzgue con error antes de su liberación. Si hay posibilidad de error, entonces hay condiciones de verdad para sus enunciados. Y si hay condiciones de verdad, y posibilidad de juzgar, entonces estamos en el ámbito de lo conceptual y no de lo pre-conceptual. Una vez aceptamos este matiz, que nos permite diferenciar

claramente este enfoque de la hipótesis de la familiaridad, pasemos a considerar la estrategia NC/VH en sí misma.

En un breve artículo publicado en 1984, muy poco después de la publicación del “Epiphenomenal Qualia” de Jackson, Horgan (2004 [1984]) defendió que hay un equívoco en la manera como Jackson entiende el concepto de “información física”, y que de la aceptación de dicho equívoco se sigue que el argumento permite establecer la intuición de la brecha epistémica, pero no permite de ahí saltar a establecer una brecha ontológica del tipo que el autor quería defender. Para ilustrar dicha afirmación nos propone pensar en dos enunciados:

- (i) Luisa Lane sabe que Superman puede volar.
- (ii) Luisa Lane no sabe que Clark Kent puede volar.

Es fácil entender que “Superman puede volar” refiere al mismo hecho que “Clark Kent puede volar”. Sin embargo, Luisa Lane conoce la veracidad de uno de los enunciados y no conoce la veracidad del otro. Esto es así porque, como es de público conocimiento, Luisa no sabe que Clark es Superman. Mientras Luisa ignore ese dato, para ella ambos enunciados expresan información diferente.

La misma estrategia puede aplicarse al caso de María. Al ser liberada, María aprende que “ver tomates tiene *esa* propiedad”, en donde “esa propiedad” refiere ostensivamente al “rojo fenoménico” que ahora puede experimentar. El enunciado expresa nueva información porque María tiene una nueva perspectiva que es la de la primera persona. Esta interpretación del caso no es de ninguna manera incompatible con el fisicalismo, ya que perfectamente el *quale* del “rojo fenoménico” puede ser una propiedad física. La información es nueva por la novedad de la perspectiva, pero el hecho y la propiedad no, por lo que el AC no implica la falsedad del materialismo Tipo-B. El AC no logra eliminar la posibilidad de que la propiedad del “rojo fenoménico” sea idéntica a la propiedad física *p* que se instancia en el sistema nervioso de un observador humano normal, tal que la identidad se postula a partir de una correlación establecida *a posteriori*.

Unos años más tarde, en un influyente artículo, Loar (1990) (2004) ha formulado con mayor precisión una de las versiones más influyentes de esta estrategia, al punto que desde tiendas antifisicalistas (P. ej. (Chalmers, 1999 [1996]) (Nida-Rümelin M. , 2004) (Goff, 2017)) se suele considerar que su postura es posiblemente una de las más desafiantes para con el AC en particular, y con la defensa del dualismo en general. Según Loar, es posible mantener la veracidad de dos tesis aparentemente contradictorias: primero, que los conceptos fenoménicos son conceptualmente irreducibles a conceptos físico-funcionales; y segundo, que las cualidades fenoménicas son idénticas a propiedades físico-funcionales. Esto implica que la intuición de la que parte el AC sería correcta: ninguna cantidad de información física podría permitir a María predecir cómo se sentiría ver algo rojo. Mas también implica que de dicha intuición no se puede deducir que el materialismo sea falso ya que no hay por qué asumir fatalmente que han de existir hechos no físicos. Cuando María es liberada accede al mismo hecho que ya conocía vía un modo de presentación a través de un nuevo, otro, modo de presentación. El AC no resiste, según Loar, la idea de que podemos saber de las cosas mediante diferentes modos de presentación.

Los conceptos fenoménicos refieren a las mismas entidades del mundo que los conceptos físico-funcionales, pero relacionándose de una manera diferente con la entidad referida. Los segundos lo hacen gracias a la mediación de los términos propios de las ciencias físicas (en sentido amplio), mientras que los primeros lo hacen de manera directa, sin mediaciones lingüísticas⁵⁹. De este modo, es perfectamente factible que aquellas propiedades a las que refieren los conceptos fenoménicos sean las mismas a las que pueden referir los conceptos físico-funcionales.

⁵⁹ Posiblemente habría que admitir la presencia de alguna expresión indicativa, del tipo “ese color que veo ahora” o “aquél matiz que vimos hace unos días”, pero ese color, o aquél color, no estarían identificados como “rojo” o “verde”.

El hecho de que en los conceptos fenoménicos la referencia sea directa podría hacernos pensar que la propuesta de Loar se acerca a la hipótesis de la habilidad, ya que podría formularse como la habilidad para referir de manera directa a tales o cuáles propiedades. No obstante, Loar destaca que la línea de Lewis-Nemirow, al negarle estatuto conceptual a dichos estados mentales, imposibilita la posibilidad del error y de la conformación de juicios perceptivos que puedan funcionar como premisa en una inferencia. Lo mismo había surgido del experimento mental protagonizado por Marianna. Por tanto, puede agregarse que ambos tipos de conceptos no solo refieren -de forma diferente- a las mismas propiedades, sino que tienen las mismas condiciones de verdad, a diferencia de lo que ocurre con las habilidades en las que no puede hablarse de condiciones de verdad.

Los conceptos fenoménicos son conceptos reconocionales que permiten emitir juicios gracias a cierta disposición -no habilidad- a discriminar y clasificar objetos (Loar, 1990). No es menor que se trate de un concepto reconocional porque asume una disposición previa, fruto de contar con las redes neuronales oportunas adecuadamente calibradas, tal como vimos en la sección anterior. La propiedad fenoménica relevante -que es físico-funcional en la mirada de Loar- dispara al manifestarse, pone en acto, la disposición a discriminar y clasificar estímulos y por tanto se constituye de forma inmediata en la referencia de los conceptos fenoménicos sin necesidad de una fijación de la referencia de orden superior. De forma análoga a los ejemplos de Luisa Lane, podemos considerar los siguientes juicios:

- (i) “El estado *a* se siente de *esa* manera”.
- (ii) “El estado *a* tiene la propiedad físico-funcional *P*”.

Ambos juicios se diferencian en que uno predica del estado mental *a* un concepto fenoménico, que es reconocional e indicativo, y el otro predica un concepto físico-funcional expresable en el lenguaje de las ciencias cognitivas. Es solo *a posteriori*, gracias a la investigación científica, que podemos constatar que

ambos enunciados son verdaderos gracias a la satisfacción de las mismas condiciones de verdad, por lo que puede establecerse que “sentirse de esa manera” es lo mismo que “P”, tan solo diferenciándose ambas formulaciones en el modo de presentación que tiene la propiedad. “Cuando en nuestras rumiaciones filosóficas juntamos conceptos de los dos tipos, ellos mismos están en estados psicológicos tan diferentes que se crea la ilusión de que sus referencias deben ser diferentes” (Loar, 1990, pág. 90), es decir, la ilusión de que ambos conceptos refieren a propiedades distintas se debe a que nos colocamos en enfoques muy diferentes al juzgar de una u otra manera, y esto genera una colisión de modos de presentación heterogéneos durante la actividad analítica del filósofo. Eso no ocurre en nuestra vida cotidiana⁶⁰

⁶¹ .

En síntesis, según NC/VH el AC logra escenificar la intuición básica por la que se establece la brecha explicativa, por lo que puede concluirse que la conciencia fenoménica no superviene lógicamente a la estructura físico-funcional del organismo. De ahí se sigue que es posible imaginar mundos en los que haya homólogos físico-funcionales a nosotros con estados mentales fenoménicamente diferentes a nosotros, o directamente sin estados fenoménicos de ningún tipo. Sin

⁶⁰ La posibilidad de que el mismo hecho pueda ser conocido mediante diferentes modos de presentación puede ser una explicación de cierta tendencia innata a ser dualistas. Más sobre este punto con las referencias correspondientes en la nota al pie n°35 (pág. 62).

⁶¹ No es objeto central de nuestra investigación, pero vale la pena señalar que los conceptos fenoménicos, reconocionales, cumplen también la importante función de permitirnos adscribir estados de conciencia fenoménica en las otras mentes. Dado que Loar parte de la base de que el AC no logra refutar la teoría de la identidad psico-física, ni un funcionalismo asociado, entiende que es suficiente con atender a las conductas -verbales y no verbales- de otros individuos para suponer, vía proyección de los conceptos fenoménicos propios, cómo se siente para ese individuo estar en un estado. No hay ninguna razón para dudar de que otro individuo, razonablemente similar a mí en términos físico-funcionales, siente un estado físico-funcional de la misma manera que yo. De ahí que, aunque sea lógicamente posible que existan homólogos físico-funcionales con inversión o ausencia de *qualia*, eso no es físicamente factible dada la identidad entre las propiedades a las que refieren los conceptos fenoménicos y los físico-funcionales. Por lo dicho, “...los conceptos pueden en ese sentido ser desde un punto de vista y, no obstante, introducir propiedades que pueden ser capturadas de manera exhaustiva por la ciencia objetiva...” (Loar, 1990, pág. 93). De esa forma, el llamado “problema de las otras mentes” queda disuelto, nuevamente, más que por el aporte de una prueba sustantiva, porque no habría razones para dudar de lo que ocurre en las otras mentes.

embargo, cuando nos referimos a nuestro mundo actual, el AC no permite refutar una tesis de superveniencia nomológica de la conciencia fenoménica a la estructura físico-funcional de los individuos. No puede, en nuestro mundo actual, que sería el más importante, haber cambios en las propiedades físico-funcionales sin que esos cambios estén acompañados de manera inequívoca por cambios en los estados fenoménicos de dichos individuos. Al menos, el AC no estaría proporcionando razones suficientes para dudar de dicha relación de superveniencia nomológica. Los dos tipos de conceptos que utilizamos para referirnos de maneras diferentes a las mismas propiedades son irreducibles entre sí, pero eso no implica que no podamos hacer ciencia acerca de nuestros estados fenoménicos apelando a los conceptos teóricos de las ciencias cognitivas.

Hay una brecha explicativa porque las ciencias cognitivas no podrán teorizar sobre cómo se sienten los estados fenoménicos de forma directa, ya que los conceptos físico-funcionales no pueden capturar las propiedades fenoménicas tal como son experimentadas en primera persona. Aun así, las ciencias cognitivas pueden proceder considerando una dimensión estructural del ámbito de lo fenoménico, que se mantendría estable entre diferentes individuos similares físico-funcionalmente⁶². Esto permite, aun respetando la brecha explicativa, teorizar de manera objetiva acerca de los estados fenoménicos propios y de los demás, siempre y cuando respetemos el límite de no pretender capturar en nuestro discurso lo que el mismo no puede capturar, la experiencia consciente directa de nuestros estados fenoménico-físico-funcionales.

En el siguiente apartado consideraré algunas críticas a la estrategia NC/VH, y posibles maneras de defender el AC de la misma.

⁶² Cfr. Shoemaker (1982) para un argumento similar en el contexto del argumento del espectro invertido.

3.3.3.1 Críticas a la estrategia “nuevo conocimiento / viejo hecho”

Rabin (2016) propone el siguiente argumento (al que le realizo ajustes menores de estilo), con el fin de demostrar que la estrategia NC/VH no logra refutar el AC:

1. Si NC/VH es verdadera, entonces el hecho de que alguien experimenta rojo es idéntico al hecho de que alguien está en el estado P.
 2. Para todo hecho A y B: Si A y B son idénticos, entonces han de ser idénticos en todo mundo posible.
 3. (1, 2) Si NC/VH es verdadera, el hecho de experimentar rojo y el hecho de estar en P son verdaderos en los mismos mundos posibles.
 4. Es posible pensar ejemplos de mundos posibles en los que el hecho de experimentar rojo y el hecho de estar en P no son verdaderos al mismo tiempo.
-
5. (3, 4) NC/VH es falsa.

(1) implementa el núcleo de la tesis NC/VH, más allá de que se refiera a hechos o propiedades, es una razonable expresión de una postura Tipo-B como la de Loar o Horgan. (3) es una instancia de (2) que se enlaza con el antecedente de (1). Por otro lado, Rabin entiende que (2) es muy plausible, y (4), que permite la conformación de una estructura *modus tollens*, es demostrable a partir de un ejemplo. (2) y (4) son las premisas que merecen un comentario.

El ejemplo para demostrar (4) es fácil de concebir, sigo el proporcionado por el propio Rabin. Podemos pensar un mundo posible llamado “Cielo” constituido por ectoplasma, fantasmas que flotan en éter y una fuerza que gobierna todo ello llamada Kismet. Los fantasmas pueden experimentar lo rojo, pero por hipótesis no

pueden estar en un estado físico-funcional P ⁶³. De ahí que Cielo es un mundo posible en el que ambos hechos no son verdaderos al mismo tiempo.

Sin embargo, hago notar que emerge una contradicción al considerar conjuntamente (4) y (2). No una contradicción en el sentido de aquella que propicia el *modus tollens*. Rabin acepta que Cielo es un mundo posible compatible con las tesis básicas del fisicalismo Tipo-B por una parte, asumiendo entonces lo que los propios materialistas Tipo-B entienden, que el fisicalismo es una tesis contingente que refiere a lo que ocurre en este mundo actual y no a lo que ocurre en todo mundo posible (como sí defendería una tesis Tipo-A). Por otro lado, no es tan claro que (2) sea verdadera, pero si no lo fuera, entiende Rabin que se está abriendo la puerta a la posibilidad lógica de homólogos funcionales con ausencia o inversión de *qualia*, cuestión que no concierne al AC, pero que sería inaceptable para el materialista⁶⁴.

La discusión de la veracidad de (2) depende de la valoración de una discusión vinculada a la identidad y la referencia, centrada en el influyente *Naming and Necessity* de Kripke, que escapa a los alcances de la presente tesis. La idea de que los enunciados de identidad han de ser verdaderos en todo mundo posible es una tesis fuerte sobre la cual, al fin de cuentas, termina descansando no solo la factibilidad o no de la estrategia NC/VH, sino también la posibilidad de las estrategias Tipo-B en general. De forma breve, si la identidad entre dos hechos no puede ser contingente, entonces la identidad entre los hechos físico-funcionales y los hechos fenoménicos ha de ser necesaria en todo mundo posible. En ese escenario, el establecimiento de la superveniencia nomológica de las propiedades

⁶³ Estrictamente, los fantasmas podrían estar en el estado funcional P , pero no en el estado físico P , ya que recordemos que el principio de realizabilidad múltiple habilitaría -*prima facie*- un funcionalismo dualista. Esa posibilidad, poco frecuente en la literatura, admitiría concebir explicaciones reductivas aceptando una relación de superveniencia lógica de lo fenoménico a lo funcional, pero negaría que ello implique la instanciación de los estados funcionales en un sustrato material. Por tanto, sostendría la defensa del fisicalismo pero no del materialismo.

⁶⁴ Una presentación simple de dichos argumentos puede encontrarse en Chalmers (1995), y un desarrollo más profundo en Block (1980), Chalmers (1999 [1996]) y Goff (2017).

fenoménicas sobre las físicas, necesariamente habría de traer consigo el establecimiento de la superveniencia lógica. Y por la vía opuesta, la refutación de la superveniencia lógica habría de arrastrar consigo a la superveniencia nomológica. Esto nos dejaría un escenario por el que habríamos de aceptar o un materialismo Tipo-A, o alguna forma de dualismo, o alguna forma de monismo Tipo-F, pero de ninguna manera un materialismo Tipo-B.

Más específicamente, la situación pasa a ser dilemática: o bien mundos como Cielo no son posibles, y de ahí se sigue la superveniencia lógica de lo fenoménico a lo físico-funcional; o bien mundos como Cielo son posibles, y de ahí se sigue o la negación de la premisa (2) de este argumento o la negación del enfoque NC/VH. La primera vía está obturada para los materialistas Tipo-B porque los debería llevar a adoptar una postura Tipo-A en la que los conceptos fenoménicos serían no reductibles a los conceptos físico-funcionales. El panorama para el AC pasaría a ser completamente diferente porque entraríamos en el escenario de negar que María aprenda algo al salir de su cautiverio y aceptar que podría deducir cómo se siente ver algo rojo desde su encierro. Esa vía es transitable, pero ya constituye una modificación relevante de la posición y constituiría la negación de la fórmula NC/VH ya que, como hemos visto, no podría aceptarse ni que hay un nuevo conocimiento ni que hay un logro epistémico de ningún tipo.

El otro cuerno del dilema acepta que mundos como Cielo son posibles, implica que o negamos la tesis de Kripke por la que los enunciados de identidad son necesarios para todo mundo posible, o terminamos por aceptar que la estrategia NC/VH es falsa, tal como lo propone el argumento de Rabin. La cuestión, dentro de este cuerno, pasa a girar en torno a un problema teórico ubicable en el área de la filosofía del lenguaje y la teoría de la referencia. Demos paso a un acercamiento a esta cuestión en el siguiente apartado.

3.3.3.2 Intensión primaria y secundaria

David Chalmers ha desarrollado en varios artículos, y particularmente en (Chalmers, 1999 [1996]), una teoría semántica bidimensional que pretende dar una solución al problema mencionado anteriormente. El núcleo de la teoría consiste en que la referencia de los conceptos funciona diferentemente de acuerdo a si estamos en el mundo real o en mundos posibles contrafácticos. La *intensión primaria* de un concepto fija la referencia en el mundo real. Mientras que la *intensión secundaria* del concepto fija la referencia para todos los mundos contrafácticos, una vez que fue fijada la intención primaria.

Para explicar qué es el agua, primero decimos que es un líquido transparente, inodoro, incoloro, que precisamos para vivir. Esa es la intención primaria del concepto, es lo que saben los niños y es lo que supo la humanidad durante milenios. En tiempos relativamente recientes pudimos establecer que “agua” es “H₂O” gracias al desarrollo de la química, estableciendo la intención secundaria. Si obviamos esta bidimensionalidad de los conceptos, hemos de admitir que en un mundo en el que el líquido transparente imprescindible para la vida está compuesto por XYZ, los habitantes de ese mundo homólogos a nosotros (en sentido primario) no toman agua para hidratarse, ya que agua siempre sería H₂O. Bajo el marco de la semántica bidimensional, los habitantes de ese mundo contrafáctico consumen agua, aun cuando su composición sea XYZ⁶⁵, porque la intención primaria fija la referencia en el mundo actual, y por tanto debemos considerar - según el autor- que fije la referencia en mundos contrafácticos para sujetos centrados en el mundo actual. La intención secundaria se fija *a posteriori*, considerando lo que arroja la ciencia para este mundo, y rigidificándolo para todo

⁶⁵ Para seguir esta discusión cfr. particularmente Block (2003). Se trata de discutir en términos de experimentos mentales la posibilidad de la reducción conceptual, así como de establecer la posibilidad de juicios necesarios *a posteriori* que pueden establecer distintos tipos de necesidad.

mundo posible. Por ello es que la intensión secundaria de “agua” es H_2O aunque sería XYZ si la fijáramos centrados en ese mundo contrafáctico.

Aplicado el marco a la discusión en torno a la superveniencia de lo fenoménico sobre lo físico, puede razonablemente pensarse la cuestión de la siguiente manera. Si las propiedades fenoménicas son lógicamente supervenientes a las propiedades físicas de acuerdo a las intensiones primarias, entonces la relación de implicación entre los conceptos referentes a las propiedades físicas y los conceptos referentes a las fenoménicas ha de establecerse *a priori*. Mientras tanto, si la superveniencia se establece tomando en consideración las intensiones secundarias, entonces la relación de implicación se establece *a posteriori*. La conexión entre “agua” y “ H_2O ” solo puede establecerse *a posteriori*, y lo mismo ocurre con la conexión entre “rojo fenoménico” y “estado físico-funcional *P*”. Por lo tanto, habrá un mundo posible en el que “rojo fenoménico” no es “estado físico-funcional *P*”.

...si María adquiere algún conocimiento fáctico del que anteriormente carecía - aunque solo sea el conocimiento de un hecho viejo bajo un modo diferente de presentación- entonces debe haber algún hecho verdaderamente novedoso del que obtiene conocimiento (Chalmers, 1999 [1996], pág. 189)

Ese hecho verdaderamente novedoso es el que establece que para nuestro mundo “rojo fenoménico” es “estado físico-funcional *P*”, tal como el hecho por el que “agua” es “ H_2O ”, o “dolor” es “excitación de fibras-C”. Es un hecho que conecta ambos modos de presentación, el fenoménico que involucra la intensión primaria y el teórico que involucra la intensión secundaria⁶⁶.

⁶⁶ En respuesta a la crítica de Isabel al dualismo interaccionista de Descartes, este último propuso que además de la idea innata de sustancia extensa, y la idea innata de sustancia pensante, hay que admitir la existencia de la idea innata de la unión entre ambas sustancias (Descartes & Isabel de Bohemia, 2011 [1642]). Si bien la respuesta de Descartes no es aceptable, porque se expone a tener que dar cuenta de la conexión entre las dos primeras ideas innatas y la de la unión, y así *ad infinitum*, resulta interesante notar que se trata del mismo punto. Mientras que Descartes necesita concebir una idea innata que conecte lo mental y lo corporal, Chalmers entiende que se requiere concebir un hecho que conecte ambos modos de presentación.

Este planteo es compatible con el argumento de Rabin. Su premisa (2), según la cual todo enunciado de identidad verdadero ha de ser verdadero en todo mundo posible, es afín a la idea de Chalmers de que para dilucidar este problema debemos centrarnos en las intensiones primarias de los conceptos relevantes que son los que luego servirán para rigidificar las intensiones secundarias.

La brecha explicativa surge de la distancia entre las intensiones primaria y secundaria. Podemos saber cosas sobre el agua, v.g. que se ve transparente, que se siente inodora e insípida, que puede volverse sólida o gaseosa en contextos cotidianos, sin saber que es H₂O. Finalmente, la única manera de seguir sosteniendo la estrategia NC/VH sería excluir del abanico de mundos posibles cualquier mundo del estilo de Cielo. Una restricción de ese tipo sería arbitraria, salvo que estuviera respaldada por algún argumento. Una vez más, la única manera por la que quien escribe puede declarar falsa esa premisa (4) es invirtiendo la carga de la prueba y asumiendo como punto de partida que el dualismo no es posible, y ubicando la negación de la posibilidad de Cielo como conclusión del argumento. No parece ser una movida argumental legítima o razonable partir de una postura metafísica fuerte para concluir acerca de la posibilidad lógica de un caso ejemplar, pero más allá de eso, lo que más deseo transmitir es que nuevamente nos encontramos en un punto en el que la discusión está bloqueada.

Como por la premisa (1) del AC María ya conocía todos los hechos físicos relacionados con la visión, Chalmers entiende que ese hecho por el que ambos modos de presentación están conectados ha de ser un hecho no físico. Ahora bien, ¿hay realmente involucrado como referencia un hecho -con el peso ontológico que ello implica- en la afirmación *a posteriori* de que la propiedad fenoménica es idéntica a la propiedad físico-funcional? No deseo dar respuesta a esta pregunta, pero sí señalar que si atribuimos facticidad a esa identidad, se volvería algo vidrioso el criterio por el cuál atribuimos facticidad en general. Entiendo que la defensa del AC de Chalmers logra, efectivamente, demostrar que el AC es exitoso en su pretensión de exponer la brecha explicativa, pero que el paso por el cual pretende demostrar el dualismo a partir de dicha brecha requiere de un argumento más. Se

trataría de un argumento modal que permita ir desde la posibilidad lógica hasta la realidad ontológica de la existencia de dos sustancias, como el de la Sexta Meditación de Descartes, o el de la concebibilidad de los zombis que él mismo proporciona. Discutir dicho argumento en profundidad podría ser objeto de otra investigación análoga a la presente. El propio Chalmers entiende que ambos argumentos: el AC y el modal, han de funcionar en tándem para lograr defender el dualismo (Chalmers, 1999 [1996]).

A modo de síntesis, de la exploración de la objeción NC/VH, y de la respuesta de Chalmers basada en su semántica bidimensional, surge que la versión fuerte del AC, que atañe a la discusión metafísica, no se sostiene por sí sola y requiere de un argumento adicional; mientras que la versión débil -que refiere a la discusión epistémica- sí puede ser de recibo, ya que los propios materialistas Tipo-B conceden la existencia de la brecha epistémica contenida en la intuición básica del experimento mental.

3.3.4 La paridad de razones contra el dualismo

Hay un argumento simple e ingenioso, que funciona en paralelo a los esbozados anteriormente, no necesariamente invalidando el AC, sino demostrando que el mismo no demuestra lo que Jackson aspiraba a demostrar, y que por el contrario podría volverse en contra de sus propios objetivos. Sigamos a Paul Churchland:

...si el argumento tal como lo formula Jackson fuera correcto, probaría demasiado. Supongamos que Jackson estuviera argumentando, no contra el materialismo, sino contra el dualismo: contra la visión de que existe una sustancia no material -llamémosle “ectoplasma”- cuya constitución oculta y legalidad intrincada fundamenta todo fenómeno mental. Permitamos que nuestra María enclaustrada sea una “ectoplasmóloga” esta vez, y que conozca todo lo que hay por conocer₁ sobre los procesos ectoplasmáticos subyacentes a la visión. Aún habría algo que ella no podría conocer₂: cómo se siente ver algo rojo. ¡El dualismo es, por lo tanto,

inadecuado para dar cuenta de todo fenómeno mental! (Churchland P. M., 1985, págs. 24-25)⁶⁷

El núcleo del argumento citado consiste en mostrar que el mismo perfectamente podría blandirse contra el dualismo, y en particular contra el dualismo de sustancias. Ninguna cantidad de conocimiento acerca del “ectoplasma” podría permitir a la María dualista deducir de antemano cómo podría sentirse ver algo rojo. Nagasawa (2002) reformula el argumento de la siguiente manera, proponiendo a “Marcos” como *alter-ego* dualista de María y X como las propiedades fenoménicas:

1'. Marcos (antes de su liberación) sabe todo lo físico y todo lo X que hay que saber sobre otras personas.

2'. Marcos (antes de su liberación) no sabe todo lo que hay que saber sobre otras personas (porque aprende algo sobre ellos en su liberación).

3'. Hay verdades sobre otras personas (y sobre sí mismo) que escapan a la explicación fisicalista y a la explicación tipo X.

Aquí X refiere a la constitución y legalidad de todo lo que fundamenta los fenómenos mentales, es decir, todo lo relacionado al ectoplasma de Churchland, o a la sustancia pensante cartesiana. La discusión se ve algo contaminada por cierta ambigüedad respecto a si estamos poniendo en juego un dualismo de sustancias o un dualismo de propiedades. Sin duda el uso de “ectoplasma” por Churchland, más allá del sarcasmo, nos remite de forma directa a que se está refiriendo a un dualismo de sustancias, más allá de que se trate de uno interaccionista como el cartesiano, o uno epifenomenista, como el de Jackson. Curiosamente, Jackson plantea en su

⁶⁷ Los subíndices “1” y “2” incluidos en la cita corresponden a la terminología elegida por Churchland para poner de manifiesto el equívoco en el significado de “conocer” que acusa en el AC (recordemos que hemos ubicado otra de sus críticas bajo el paraguas de la hipótesis de la familiaridad). No cumplen funciones relevantes aquí, pero quedan incluidos para conservar la integridad del texto citado.

respuesta que “el caso del ectoplasma es solo un juego” (2003b, pág. 117), descartándolo como una opción seria, y se propone defender un “dualismo de atributos” del argumento de paridad de razones. Resulta extraño porque el epifenomenismo desde el que Jackson propuso el AC, es una postura dualista de sustancia y no de atributos o propiedades. Recordemos que el epifenomenismo asume la existencia de hechos no físicos sin eficacia causal⁶⁸.

Concedamos que Jackson se está refiriendo al dualismo de propiedades, es decir, a la idea de que existen hechos que tienen propiedades no físicas, sin comprometerse con la existencia de hechos o sustancias no físicas. Eso estaría encuadrado o en el materialismo Tipo-B, o en el monismo Tipo-F según la clasificación establecida en el capítulo anterior. Según Jackson (2003b) [1986] el dualismo de propiedades no cae ante su propio AC porque Marcos nunca podría tener un conocimiento completo de la visión del color desde su cuarto monocromático. Es decir, la premisa (1') no podría cumplirse en el encierro. Esta defensa parece caer por su propio peso, ya que algo análogo podría decirse del argumento contra el materialismo: la premisa (1) es tan implausible como la (1'). Esta, de hecho, ha de considerarse como otra línea crítica posible, y se procederá a ello en la siguiente sección.

Robinson, citado por Nagasawa (2002), entiende que la distinción entre dualismo de sustancias y de propiedades es irrelevante en el AC. Afirma que la razón por la que el argumento de la paridad de razones no es legítimo es porque caricaturiza lo que ha de entenderse por sustancia pensante. Por el contrario, la forma como refiere al ectoplasma sugiere que lo que tiene en mente es una versión alterna de otra sustancia extensa. La sustancia mental no ha de entenderse como algo que está hecho de tal o cual cosa, porque así se está reproduciendo una estructura que es característica de lo que es materia. Un dualismo estilo cartesiano

⁶⁸ Cfr. sección 2.1.5 Epifenomenismo o Dualismo Tipo-E a partir de pág. 63.

podrá enfrentar numerosos problemas, pero el AC no le afecta, ya que el argumento de paridad de razones ataca una versión empobrecida o mal interpretada de esa postura.

De lo analizado hasta aquí en esta sección se sigue una afirmación relevante a nuestros fines, la distinción en juego en el AC no es entre dualismo y monismo, sino entre dualismo reductivo o no reductivo.

Lo que hemos incluido bajo la categoría de monismo Tipo-F⁶⁹ es -según Nagasawa (2002)- en donde realmente hace mella el argumento de la paridad de razones. Esto es así porque para ambas variantes se trata de teorías que suponen la existencia de propiedades fenoménicas o protofenoménicas fundamentales de cuya combinación surgirían las propiedades macrofenoménicas de nuestra vida consciente ordinaria. Nagasawa entiende que el panprotopsiquismo es un dualismo reduccionista, mientras que el panpsiquismo sería un dualismo no reduccionista. El panprotopsiquismo es reduccionista porque asume la existencia de propiedades protofenoménicas, que no son fenoménicas por sí solas, sino que el carácter fenoménico es fruto de su combinación en la conformación de entidades de nuestro nivel de organización. Así, las propiedades fenoménicas han de supervenir lógicamente y nomológicamente a las protofenoménicas. Por el contrario, el panpsiquismo sería no reduccionista porque entiende que las propiedades microfenoménicas y las propiedades microfísicas ya son ellas mismas la estructura fundamental de la realidad, y no supervienen a nada.

No obstante lo dicho, cabe pensar que hay una serie de distinciones conceptuales que Nagasawa no está teniendo en cuenta, ya que ambos monismos Tipo-F pueden adoptar una forma constitutiva o emergentista, y ahí está la clave para aceptar el reduccionismo o no. Según Chalmers (2013), a diferencia de la

⁶⁹ Cfr. apartado “2.1.6 Monismo Tipo-F” a partir de pág. 65, en donde se presentan panpsiquismo y panprotopsiquismo y sugieren fuentes para su abordaje.

versión de Nagasawa (2002) en el panpsiquismo hay una relación de enraizamiento (o superduperveniencia⁷⁰) entre lo microfenoménico y lo macrofenoménico, y lo que diferencia la versión constitutiva de la emergentista es si esa relación puede establecerse *a priori* o solo *a posteriori*. El AC, considerando el argumento de la paridad de razones, estaría negando la relación de superveniencia lógica, establecida *a priori*, que es lo necesario para proveer explicaciones reductivas (Chalmers, 1999 [1996]). No hay suficiente conocimiento sobre las propiedades microfenoménicas o protomicrofenoménicas que permita deducir cómo se habría de sentir una experiencia consciente de nivel macro, y como ésta podría surgir de la combinación de las propiedades micro. ¿No es al fin de cuentas una forma alternativa de llegar al problema de la combinación, explicado en el capítulo anterior⁷¹? Eso es lo que representa la gran dificultad del “dualismo ectoplasmático” de Churchland.

En síntesis, lo que el AC está afectando es la posibilidad de generar teorías reduccionistas que expliquen el carácter fenoménico de nuestras experiencias conscientes, sean estas materialistas o dualistas. El argumento de la paridad de razones nos muestra -nuevamente- que el problema central en el AC no es el ontológico sino el epistemológico, al establecer la no superveniencia lógica entre lo fenoménico y la materia por una parte; y, gracias a la versión de Marcos, la no superveniencia lógica entre lo fenoménico y lo microfenoménico. Esto nos deja varias alternativas, pero todas ellas no reduccionistas. O somos materialistas Tipo-A, o dualistas Tipo-D o Tipo-E, o alguna variante no reduccionista (emergentista) de monismo Tipo-F. Sabemos que todas esas posturas presentan sus propios

⁷⁰ Este concepto de superveniencia fuerte fue presentado en el apartado “1.3 La superveniencia y sus formas”, como la forma más exigente de superveniencia, por encima de la de carácter nomológico y la de carácter lógico. Una propiedad es superduperveniente a otra si, además de estar establecida la superveniencia lógica, es posible establecer en términos propios de las propiedades de bajo nivel cuál es la conexión en virtud de la cuál las propiedades de alto nivel les son supervenientes (Horgan T. , 1993).

⁷¹ Cfr. nota 38 en la pág. 66.

problemas, pero al menos podemos afirmar que, si tomamos por válido el argumento, y tomamos por plausibles sus premisas, lo que el AC logra es limpiar el panorama de todo intento de una teoría reduccionista ya que esta -en la forma que sea- siempre deja algo de lo mental sin explicar⁷².

3.3.5 Fisicalismo, materialismo, y la premisa (1)

Para avanzar en este tópico es necesario volver un poco sobre nuestros pasos. Cuando presenté el argumento de la banana azul, página 81, quedó establecido que la posición de un eliminativista Tipo-A como Dennett consiste en afirmar que los defensores del AC subestiman el poder de la premisa (1), ya que, si la tomamos por cierta, María debería poder prever cómo se siente ver algo rojo gracias a un razonamiento deductivo complejo pero posible. Vimos en ese mismo apartado que Robinson opina que Dennett está inflando el alcance de esa premisa (1). Se trata de un nudo cuya resolución puede arrojar luz sobre los temas de fondo que nos ocupan.

Más adelante, en la página 99, en ocasión de discutir la hipótesis de la familiaridad, vinculé las afirmaciones de Dennett con otras similares de Paul Churchland, cuya postura podría ubicarse entre los eliminativistas Tipo-A o entre los materialismos reduccionistas Tipo-B, de acuerdo como se interpreten algunos de sus planteos. En esta sección mostraré cómo diferentes maneras de entender qué entendemos por “información física” determina dos variables relevantes: primera, si es más o menos plausible la premisa (1); segunda, si el AC logra arribar a la conclusión de carácter ontológico (4) tal como era la intención original de Jackson,

⁷² Refiero al lector a la pág. 44 en donde comento la diferencia de exigencia que presentan Chalmers, Churchland y Horgan respecto a las tesis de superveniencia como fundamentación de métodos reduccionistas. Aquí opté por la versión intermedia, por la que la superveniencia lógica es necesaria para el reduccionismo. En la conclusión habrá oportunidad de volver sobre esta cuestión.

o si tan solo arribamos a la conclusión de carácter epistémico (3) manteniéndonos en la versión débil del argumento.

3.3.5.1 ¿Qué es una información física completa?

Establecer qué tipo de informaciones se incluyen en la fórmula “información física completa” debe ir en consonancia con alguna noción acerca del fisicalismo que el argumento pretende refutar. Jackson justamente rehúye a dar una definición precisa de “fisicalismo”, aun cuando la meta del AC es demostrar que el fisicalismo es falso. Lo más aproximado a una definición operacional dice que el fisicalismo implica que “toda información (correcta) es información física” (2003, pág. 96). Al parecer, si nos limitamos a lo que el autor nos propone, vamos en círculos. María tiene información física completa, “fisicalismo” es la tesis de que toda la información es información física. Si del argumento surge que hay información relevante respecto a la percepción del color, y no es información física, entonces el fisicalismo así definido resultará falso y el argumento cumplirá su cometido. Cumplirá su cometido, sí, pero de forma irrelevante, porque si bien podemos concordar con Jackson (2003, pág. 96) en que la discusión respecto al argumento en sí no depende de cómo caractericemos al fisicalismo, sí tenemos que advertir que el argumento se vuelve trivial si no se enmarca bajo una definición más precisa de la postura que busca atacar. Será necesario avanzar en ese sentido.

Para caracterizar el fisicalismo, Philip Goff (2017) parte de una definición del mismo, también trivial, en el que lo presenta como la visión de que la realidad fundamental es enteramente física, es decir, que los únicos hechos y/o propiedades

que existen son físicos. A partir de ahí reconoce la necesidad de determinar con mucha mayor precisión a qué nos referimos con “hechos físicos” y “fiscalismo”⁷³.

Si definimos “hecho físico” como “aquellos hechos sobre los que versa la física”, no queda claro si nos referimos a la física actual o a una física ideal futura. Si se tratara de la física actual, sabemos que la ciencia se encuentra en constante evolución, por lo que seguramente irrumpirá su propia refutación, al menos parcial, en virtud de avances futuros. Esta dificultad se ve reforzada por el hecho de que la física actual incluye una contradicción interna entre lo que surge de nuestras teorías sobre lo muy grande y nuestras teorías sobre lo muy pequeño, por lo que salvo que vayamos en contra de principios fundamentales de la lógica, una caracterización de este tipo del fiscalismo y de la noción de “hecho o propiedad física” no parece ser un proyecto promisorio.

Por otra parte, y más allá de las dificultades que surgen de definir “hecho físico” como aquellos hechos que son objeto de estudio de la física actual, el hecho de que el experimento mental supone que María tiene una información física completa ya nos posiciona en el terreno de una física futura ideal. No es el caso que María conozca todo lo que hoy se conoce sobre la fisiología de la visión, sino que conoce todo lo que hay por conocer sobre el tema. Un nuevo obstáculo se nos antepone a la definición: no tenemos la menor idea acerca de en qué podría consistir una teoría de ese tipo. No sabemos qué tipo de teorías acerca de qué tipo de entidades, procesos o propiedades puede tratar esa física futura, a tal punto que incluso podría incluir hechos de un tipo que los fiscalistas actuales rechazarían como físicos.

⁷³ Otra caracterización, con terminología diferente, aunque similar en su significado, puede hallarse en (Stoljar, 2004), que distingue una concepción de lo físico basada en las teorías de una concepción de lo físico basada en los objetos. Ambas concepciones le permiten distinguir entre propiedades físicas-t y propiedades físicas-o, y desarrollar un argumento análogo al que presento en el cuerpo del texto con la terminología que fue de mi preferencia para el caso.

El caso problemático más notorio para caracterizar lo físico como aquello objeto de una física ideal futura podría surgir del descubrimiento -inverosímil quizás, pero concebible- de que cierta mentalidad fundamental forma parte de la estructura básica del universo. Esto obligaría a los fisicalistas a incluir lo mental dentro del paraguas de lo físico, aceptando la existencia de lo fenoménico como entidad fundamental. Ante esa posibilidad, una caracterización futurista de lo físico habría de tomar una de tres posibilidades. La primera es admitir que, si una física futura estableciera que la conciencia fenoménica es un hecho o propiedad fundamental de la realidad (como ocurre en el monismo Tipo-F), entonces el fisicalista debe aceptar sin medias tintas que lo mental (fenoménico) es un aspecto irrenunciable de la realidad fundamental. Se trataría de un fisicalismo no materialista.

Si seguimos ese camino, quedamos ante un paisaje extraño. Toda la discusión en torno al AC, tanto en su dimensión epistemológica como en su dimensión ontológica se vacía si asumimos que podríamos incluir la conciencia fenoménica como una parte no cuantificable ni material del mundo físico. Obsérvese que de ello se seguiría que “cómo se siente algo rojo” ya debería estar incluido en la información física completa, y, por tanto, termina abonando la interpretación de Dennett. De esa forma se convierte en implausible y carente de sentido el planteo mismo del experimento mental. Por otra parte, parece una posibilidad que pocos fisicalistas estarían dispuestos a aceptar. Sería posible incorporar una restricción *ad hoc* a la caracterización futurista de “hecho físico” que excluya a toda “mentalidad fundamental” del ámbito de lo físico (Goff, 2017, pág. 29). Llegado a este punto, el lector podrá entrever que el problema se está volviendo más terminológico que de fondo, porque la aplicación de una restricción de ese tipo consistiría simplemente en pasar a las siguientes alternativas.

La segunda y la tercera de nuestras posibilidades consisten, directamente, en desterrar a los presuntos *qualia* de toda posible teoría física futura. O han de recibir una explicación reductiva (Tipo-B), que permita comprenderlos en términos de estados neurales, o estados funcionales instanciados en el sistema nervioso; o

han de descartarse del ámbito de la realidad fundamental para pasar a entenderlos como ilusiones inexistentes (Tipo-A), tal como lo propondría un materialista eliminativista. Cualquiera de estas posibilidades implica la adopción de una postura acerca de la estructura fundamental de la realidad, y no acerca de aquellos aspectos de la realidad sobre los que podemos establecer conocimientos científicos. Esa diferencia, relevante, convierte el problema epistémico en ontológico, desdibujando las diferencias entre “fiscalismo” y “materialismo”.

He señalado en varias ocasiones que hay cierta ambigüedad en el uso de los términos “materialismo” y “fiscalismo”. Mientras que el primero se ha utilizado para establecer una postura ontológica, posiblemente “todo lo real es materia” o “la realidad fundamental es material”, el segundo ha sido utilizado especialmente a partir de la obra de los filósofos del Círculo de Viena para establecer una tesis epistémica que tiene un fundamento lingüístico: “todo enunciado significativo puede expresarse en lenguaje fiscalista”. Más adelante, con el correr del siglo XX y el debilitamiento del peso del empirismo lógico, ambas tesis se han ido fundiendo una con otra, y es usual encontrar que se utilizan como términos intercambiables (Stoljar, 2017). ¿Será el caso que la raíz más profunda de las dificultades que he ido presentado respecto al AC y al problema difícil de la conciencia puede atribuirse a dicha tendencia a identificar materialismo y fiscalismo como si fueran términos intercambiables?

3.3.5.2 Fiscalismo en sentido estricto y en sentido amplio

Si aplicamos con mayor rigor la distinción entre materialismo y fiscalismo, demarcando lo epistémico de lo ontológico, podremos avanzar sobre terreno más firme. La materia, siguiendo una caracterización asimilable a la que Descartes hace

de su sustancia extensa⁷⁴, puede pensarse como aquello que incluye entidades ubicables en el espacio y el tiempo, de bajo nivel de complejidad y que no involucran ningún tipo de mentalidad, ni ninguna tendencia espontánea hacia un fin que se asocie a una naturaleza intrínseca. Para un materialista, todo hecho físico es un hecho que concierne nada más que entidades materiales que se relacionan entre sí.

Para mayor precisión, podemos clasificar los hechos físicos considerándolos en sentido estricto o en sentido amplio. En sentido estricto, un hecho es físico si incluye entidades ubicables en el espacio y el tiempo, de baja complejidad, y exentas de mentalidad o finalidad predeterminada. Mientras tanto, un hecho es físico en sentido amplio si superviene de alguna manera sobre algún hecho físico en sentido estricto. Así planteada la distinción, para un materialista Tipo-A todos los hechos físicos son hechos físicos en sentido estricto, y por tanto involucran entidades materiales; al mismo tiempo, para un materialista Tipo-B vale la distinción, ya que defienden la posibilidad de explicar los hechos físicos en sentido amplio -por ejemplo, la experiencia consciente- en términos de hechos físicos en sentido estricto sin pérdida de información en el proceso de reducción teórica.

Llegamos al punto en que podemos aplicar la distinción, como una nueva “vuelta de perilla” al análisis del experimento mental de María. Cuando la premisa (1) establece que María tiene un conocimiento físico completo, ¿se refiere a los hechos físicos considerados en sentido estricto o amplio? La decisión respecto a esta variable tiene consecuencias en la plausibilidad de la premisa (1) así como en

⁷⁴ Discutir qué es la materia podría ser objeto de otra investigación análoga a la presente. Considero suficientemente apropiado caracterizar la materia asociándola a la sustancia extensa cartesiana dado que, como he propuesto en el capítulo 1, el problema difícil de la conciencia parece ser fruto de la manera cómo Descartes la caracteriza en el marco de un mundo fabulado totalmente cognoscible que haya un límite epistémico en el carácter fenoménico de los estados mentales.

la fuerza probatoria del argumento en general. La distinción también nos permite desatar un nudo conceptual que parece más terminológico que de fondo.

Especulemos por un momento qué ocurre si María conoce, antes de su liberación, todos los hechos físicos considerados en sentido estricto sobre la fisiología de la visión. Si asumimos que la conciencia fenoménica superviene lógicamente a los estados físico-funcionales del organismo, entonces el AC permitiría demostrar que hay algunos hechos -que podrían ser físicos en sentido amplio- que María no conocía y que solo puede llegar a conocer gracias a familiarizarse con ellos. Esta línea, como ya vimos al comentar la hipótesis de la familiaridad o la objeción NC/VH, logra establecer la conclusión epistémica (3) del AC, pero no logra por sí misma respaldar el salto hacia la conclusión ontológica (4), ya que no hay razones suficientes para afirmar que ese nuevo conocimiento que María adquiere al ser liberada sea acerca de hechos o propiedades no físicas. Bajo esta línea interpretativa tenemos una premisa (1) plausible pero un argumento con menor fuerza probatoria respecto a lo que era la intención original de Jackson.

Pasemos a considerar qué ocurre si lo que María conoce antes de su liberación son todos los hechos físicos considerados en sentido amplio (ésta es posiblemente la interpretación de Dennett). Si concedemos al AC que María aun así se sorprende al ser liberada, y por tanto que la premisa (2) es verdadera, a pesar de que conocía todos los hechos físicos en sentido amplio, entonces tenemos que admitir la conclusión ontológica (4) que establece que hay hechos o propiedades no físicas y por tanto el fisicalismo es falso. Habría, en ese caso, hechos que ni son ellos mismos físicos, ni supervienen lógicamente a los hechos físicos.

¿Podría María desde su hábitat monocromático cumplir las condiciones de la premisa (1) así interpretada? Es decir, ¿podría estar en conocimiento de todos los hechos físicos en sentido amplio? En el apartado siguiente defenderé la posición de que no sería posible, ya que hay al menos algunos hechos físicos en sentido amplio que María sólo podría conocer si es liberada y está expuesta a la experiencia directa de los colores, v.g. cómo se siente ver algo rojo. El argumento, entonces, funciona

bien y alcanza sus cometidos partiendo de una premisa que dudosamente sea verdadera bajo las condiciones del experimento mental. El escenario sería un escenario imposible y el experimento no tendría lugar.

En resumen: si María conoce solo los hechos estrictamente físicos, el argumento no demuestra lo que pretende porque encierra un equívoco; si María conoce también los hechos ampliamente físicos, el argumento demuestra lo que pretende, pero parte de un escenario imposible. Sea cual sea la manera como configuremos nuestro panel de perillas, el AC a lo sumo logra establecer la brecha explicativa, pero no logra fundamentar el salto hacia una brecha ontológica.

3.3.5.3 Plausibilidad de la interpretación de Dennett

Vimos en el apartado anterior que la manera como interpretemos el conocimiento físico completo establecido en la premisa (1) determina qué lectura podemos hacer del AC. Eso es, si entendemos que María conoce todo lo estrictamente físico, o todo lo ampliamente físico. Cuando Dennett (2004) (2006) señala que subestimamos el alcance de la premisa (1), está bajo la segunda interpretación. O, si se quiere, está negando la distinción desde su marco Tipo-A: si un hecho existe en el mundo físico, solo involucrará entidades materiales; y a la inversa, si involucra entidades no materiales (como sería la conciencia fenoménica), no formaría parte del mundo físico.

La objeción de la paridad de razones⁷⁵ echó por tierra cualquier intento de dar explicaciones reduccionistas a la conciencia fenoménica, por lo que podemos obviar el cuerno del dilema representado por la posición de Paul Churchland que lleva a una interpretación Tipo-B. Lo que queda en juego puede resumirse en la

⁷⁵ Cfr. “3.3.4 La paridad de razones contra el dualismo” en págs. 118 y ss.

siguiente pregunta: ¿por qué sería imposible que María conociera todo lo físico en sentido amplio antes de salir de su encierro? En el apartado inmediatamente anterior di por sentado que una premisa (1) interpretada de esa manera ubicaría el experimento en un escenario imposible. Un argumento empirista clásico podría respaldar esa afirmación. Podría, por ejemplo, apelar a la prioridad gnoseológica de impresiones respecto a ideas, tal como fundamenta Hume (2001 [1739]) (2003 [1748]). No obstante, opino que esa estrategia es problemática, nos hace girar en los mismos círculos que venimos girando. Dado que la posición de Dennett podría verse como objeción a los argumentos de Hume de la misma manera como ocurre en nuestro caso, seguiríamos en un estancamiento dialéctico.

Propongo pensarlo de la siguiente manera: ambas posiciones, la que afirma, y la que niega, que María pueda conocer todo lo físico en sentido amplio sin haber salido de su encierro monocromático, se apoyan en determinada postura establecida de antemano respecto al mobiliario fundamental de la realidad. Cuando Dennett afirma que María podría deducir cómo se siente ver algo rojo a partir de sus conocimientos previos, está asumiendo que todo lo que existe es materia distribuida en el espacio y el tiempo, sin un *telos* ligado a una naturaleza esencial. Cuando lo negamos, lo hacemos parados desde una ontología opuesta.

Por lo dicho es que tiene sentido especular con que se trata de dos posiciones equivalentes desde el punto de vista epistémico. Ambas niegan cualquier posibilidad de brindar explicaciones de los aspectos fenoménicos de nuestra experiencia consciente mediante estrategias reduccionistas. Para el materialismo Tipo-A no hay posibilidad de explicaciones reductivas porque no hay nada que no sea material. Para las alternativas no reduccionistas dualistas o para el monismo Tipo-F, como hay algo no explicable en términos fisicalistas, ha de ser algo no material. Lo que en un caso es premisa en el otro es punto de llegada, pero en ambos casos se recorre la misma calle.

Este tipo de equívocos podría solucionarse si nos apegamos a una distinción más cuidadosa entre “fisicalismo” y “materialismo”. Es notorio que, cuando

Jackson afirma que hay al menos algo sobre la visión del color que María no conocía en su encierro, está haciendo una afirmación de alcance metafísico. No es tan notorio, pero es igual de verdadero, que, cuando Dennett afirma que María podría deducir cómo se siente ver algo rojo desde su encierro, también está apoyando su afirmación en consideraciones implícitas de carácter metafísico para las que no contamos con el tipo de sustento empírico que podría exigirse desde su perspectiva. Pasemos a la fundamentación de esta idea.

En ocasión de caracterizar el lenguaje fisicalista, y particularmente demostrar que el mismo puede ser el fundamento de un conocimiento intersubjetivo, Carnap establece que

Las determinaciones cualitativas pueden ser incluidas en el lenguaje físico dada la condición de que se establezcan reglas para traducir todos los enunciados de ese tipo en determinaciones cuantitativas tal que, por ejemplo, el enunciado “está frío aquí” pueda ser traducido al enunciado “La temperatura aquí está entre 5 y 10 grados Celsius” (Carnap, 1995, pág. 53).

De esta manera, todo hecho de la percepción puede expresarse en lenguaje fisicalista, convirtiendo lo que es privado en intersubjetivo. Pero, esto es importante, las reglas que gobiernan esa traducción son de tal tipo que

no hay palabras del lenguaje físico que se correlacionen en el lenguaje protocolar con palabras que refieran solo a un único campo sensorial (p. ej. nunca hay correlación con determinaciones que son únicamente de color o de sonido). Se sigue que una determinación física permite la inferencia de enunciados protocolares en todo campo sensorial. Las determinaciones físicas son “intersensoriales” (...). Más aún, son también “intersubjetivas”, en acuerdo con las experiencias de varios sujetos... (Carnap, 1995, pág. 57).

En otras palabras, no es posible expresar un juicio perceptivo unimodal en lenguaje fisicalista, por el contrario, la multimodalidad es imprescindible para constituir enunciados comunicables, factibles de integrar el cuerpo de las ciencias cognitivas. Esto es así porque dichos enunciados recogen la estructura del estímulo, que es representable en términos matemáticos, y no su carácter fenoménico que es cualitativo y privado. Así, una postura como la de Dennett, que piensa concebible un razonamiento deductivo que parte de enunciados expresados en términos físicos a enunciados con significados privados no podría ser sostenible debido al carácter

unimodal de lo que María aprende. María podría deducir cómo podría ser, estructuralmente, el ámbito del color. Sabría que puede caracterizarse un color por la determinación de su valor en tales o cuales dimensiones cuantificables (p. ej. matiz, saturación y brillo, o como se las quiera definir). Incluso así, esto no dice nada de cómo se siente para una persona percibir ese color. Esto último dependerá de la ocurrencia de una experiencia que es privada, irrepetible, incomunicable, y que posiblemente esté fuertemente condicionada por el estado peculiar del aparato visual y el sistema nervioso, o los conocimientos previos del individuo (Macpherson, 2015).

De lo dicho se desprende que María puede salir airosa del engaño de la banana azul⁷⁶ porque justamente se trata de una banana, que tiene una forma identificable, y no de una figura no asociable típicamente a un color. Y eso solo sería posible si hubiera otros elementos no pasibles de engaño, como el cielo, para comparar. María se percataría del parecido del color del cielo y del color de eso que a todas luces tiene forma de banana, y recién en esa circunstancia podría -gracias a su conocimiento de la estructura de ese ámbito sensorial- identificar el engaño. Es decir, es necesario que haya transcurrido el primer progreso epistémico que vivió Marianna⁷⁷. Si la figura no fuera típicamente identificable con un color, y si no hubiera otros objetos coloreados para comparar -p. ej., si fuera un círculo que los científicos introducen en el recinto monocromático, y estuviera pintado de azul- María no lograría determinar que la están engañando cuando afirman que es rojo.

De la consideración de las secciones anteriores de este capítulo surge que el AC permite demostrar la conclusión (3) por la que los aspectos cualitativos de la experiencia consciente no son pasibles de una explicación reduccionista. También

⁷⁶ Cfr. pág. 81.

⁷⁷ Véase págs. 104 y ss.

vimos, en especial en la sección anterior, que el AC no permite demostrar la más ambiciosa conclusión (4), de carácter metafísico, que afirma la existencia de hechos o entidades no materiales. La discusión en este apartado, y en especial la reconsideración de la objeción de Dennett, nos permite defender que tampoco es viable -al menos por estas vías- considerar como demostrada la negativa de (4), esto es, la afirmación de que todo estado o evento mental involucra únicamente entidades materiales asimilables a la sustancia extensa cartesiana. En síntesis, ni puede surgir del AC una negación del materialismo, ni de su crítica una afirmación del materialismo. Con todo lo trabajado hasta el momento, pasemos a una revisión general del AC.

3.4 Valoración global del AC para el dualismo

He intentado presentar el AC de la manera más justa posible con su autor, más allá de su ulterior cambio de posición, teniendo en cuenta la formulación esquemática con la que concuerda, y los fines probatorios que persigue al postularlo, en el contexto de un artículo centrado en defender una postura epifenomenista. Vimos que la historia de María pretende ubicarnos en un escenario contrafáctico pero pensable, con la esperanza de que esté tan bien formulado que por sí mismo sea suficiente para hacer irrumpir una intuición fuerte de la manera más prístina posible. En este caso, la intuición que habría de irrumpir es la de la brecha epistémica: ninguna cantidad de información físico-funcional es suficiente para poder dar cuenta del aspecto fenoménico de nuestra experiencia consciente. En ese sentido, el argumento habilita el debate sobre el problema difícil de la conciencia, que justamente parte de la evidencia epistemológica que provee la intuición para discutir si es posible solucionarla, y si es necesario postular la existencia de una sustancia no material para solucionarla.

El AC ha sido muy discutido, especialmente en la formulación del experimento mental de María, al punto que Dennett (2006) -haciendo uso de su

habitual estilo sarcástico- se ha quejado de que ha recibido una atención excesiva para el interés real que tiene. Como es usual, los problemas filosóficos no son resueltos de manera definitiva, y siempre hay razones para progresar en profundidad; sin embargo, las diversas estrategias críticas reseñadas, conjuntamente con los comentarios en defensa, parecen tender hacia un mismo lugar. Señalaré dos aspectos que considero sintomáticos de que la geografía del problema está adoptando esa configuración.

En primer lugar, han sido numerosas las ocasiones en que la disputa argumentativa ha culminado con bloqueo. La discusión de si es posible que María deduzca cómo se siente ver algo rojo a partir de su conocimiento físico completo por parte de Dennett y los defensores del AC quedó bloqueada. El primero exige al dualista que demuestre que la deducción no puede llevarse adelante. Los dualistas piden explicaciones a Dennett acerca de cómo sería una deducción de ese tipo. Finalmente, en la sección “3.3.5.3 Plausibilidad de la interpretación de Dennett” propuse un argumento para demostrar que dichas explicaciones de Dennett no son sostenibles, al menos en opinión de quien escribe.

Más adelante, en ocasión de exponer la hipótesis del conocimiento no proposicional⁷⁸, arribamos a que no es una estrategia conclusiva en contra del AC, porque la discusión en torno a la naturaleza de dicho conocimiento no proposicional no responde ni la cuestión epistémica ni la cuestión ontológica, ya que parecen ser discusiones ortogonales (Coleman, 2009). También vimos que el rechazo de la hipótesis de la habilidad no implica la aceptación de la hipótesis de la información fenoménica, a lo sumo la mantiene provisionalmente vigente y con estatus conjetural. Eso se pone de manifiesto en el siguiente pasaje en el que Churchland comenta el equívoco en el uso de “saber acerca”:

⁷⁸ Cfr. sección “3.3.2.5 ¿Se trata de una estrategia efectiva?” desde pág. 101.

La carga de articular algún sentido específico y unitario de “saber acerca”, y de argumentar que ambas premisas son verdaderas bajo esa interpretación de ese operador epistémico, es una carga no resuelta que sigue siendo responsabilidad de Jackson (Churchland P. M., 2004, pág. 164).

Nuevamente, al presentarse la estrategia NC/VH⁷⁹, he mencionado que el argumento de Loar para disolver el problema de las otras mentes -que no es más que el de la concebibilidad de homólogos funcionales con *qualia* ausentes o invertidos- ha sido el de negar que haya razones para dudar que el otro siente igual que como yo siento. No se ofreció una razón sustantiva sino un argumento negativo que otra vez desplaza la carga de la prueba hacia el escéptico.

Justo después, al considerar el argumento que Rabin esboza para criticar la estrategia NC/VH, en particular su premisa (4) que propone un ejemplo de un mundo posible en donde no hay identidad entre lo fenoménico y lo físico-funcional, la cuestión derivó en determinar qué debe ir como premisa y qué debe ir como conclusión. Si partimos de que el ejemplo es plausible podemos demostrar que la estrategia NC/VH es incorrecta, y que el dualismo defendido por el AC es una posibilidad de recibo. Por la otra senda, si partimos de que el dualismo no puede ser una opción posible, podemos demostrar que el ejemplo de la premisa (4) no es pensable. La diferencia entre la senda de ida y la de vuelta está en qué proposición ubicamos en las premisas y qué proposición ubicamos como conclusión. Esto no es más que una discusión en torno a quién debe tener la carga de la prueba.

Asimismo, la defensa del argumento ha requerido que el lector se pronuncie respecto a marcos teóricos de áreas adyacentes, como ser la semántica bidimensional, o la teoría de la referencia de Kripke. No perdamos de vista que el objeto de la presente investigación ha sido delimitado en estudiar el papel del AC en la dilucidación del problema difícil de la conciencia, y en particular en la

⁷⁹ Cfr. sección “3.3.3 María aprende nuevo conocimiento proposicional sobre hechos ya conocidos” a partir de pág. 104.

determinación de la posibilidad de que las ciencias cognitivas puedan, a través de estrategias reduccionistas, dar cuenta del carácter fenoménico de nuestra experiencia sin pérdida significativa de significado. El hecho de que cada defensa del argumento haya requerido el apoyo de otras líneas argumentales también se me ocurre sintomático de sus limitaciones dialécticas.

Hemos llegado a puntos de bloqueo en los que la cuestión se ha vuelto más terminológica que significativa. El caso más claro es la discusión respecto a cómo hemos de entender “hecho físico” o “información física” en la premisa (1)⁸⁰. Una interpretación quitaba fuerza al argumento, mientras que la otra le daba fuerza, pero quitaba plausibilidad a la premisa. Respecto a la segunda situación, ninguna conclusión significativa puede seguirse de una premisa falsa. Respecto a la primera, tan solo fundamenta la versión débil del argumento, aquella que tan solo logra establecer la brecha epistémica sin afectar la discusión ontológica.

Podemos comparar las diferentes líneas de discusión del AC con una cuenca fluvial. Eso se debe a que la investigación ha devenido en ese resultado, que se ha repetido una y otra vez: el AC logra defender que hay una brecha explicativa, epistémica, entre lo fenoménico y lo físico-funcional, pero no logra por sí mismo fundamentar el salto de la dimensión epistémica a la dimensión ontológica, y por tanto no logra por sí mismo defender el epifenomenismo -como era la intención de Jackson en 1982-. Tampoco logra defender cualquier otra forma de dualismo o de panpsiquismo o panprotopsiquismo.

Sobre este particular, repasemos lo que cada discusión arrojó. Vimos que las hipótesis del conocimiento no proposicional no resultan relevantes para determinar el carácter físico o no físico del aspecto fenoménico de la experiencia. Dicha estrategia, entonces, ha de resultar en la admisión de la brecha explicativa, y

⁸⁰ Cfr. apartado “3.3.5 Fisicalismo, materialismo, y la premisa (1)” (pág. 123-ss.).

en que ni el argumento ni su crítica son concluyentes sobre la cuestión ontológica. Presenté, a partir de la página 103, citas textuales de los defensores de la hipótesis de la familiaridad en las que se muestran enfáticos a la hora de negar que sus consideraciones impliquen pronunciamiento alguno respecto al fondo ontológico de la cuestión.

Otro tanto puede afirmarse sobre la estrategia NC/VH, ya que las propuestas Tipo-B que la impulsan no ponen en duda la intuición básica de la que parte el AC, ni ponen en duda que el argumento permita demostrar su conclusión (3) por la que hay verdades que escapan a la explicación fisicalista. Lo que sí niegan es que (3) permita pronunciarse sobre la conclusión ulterior (4) por la que el fisicalismo -entendido como tesis ontológica, materialismo según la terminología adoptada- sea falso. Y lo mismo surgió del argumento de paridad de razones⁸¹, en el que seguimos la postura de Robinson y Nagasawa, que entienden que la cuestión central del AC es más la de la negación del reduccionismo -por atacar la tesis de superveniencia lógica de lo mental sobre lo físico- que la cuestión ontológica, ya que un dualismo reduccionista como el “ectoplasmático” o algunas formas de monismo Tipo-F sufren de igual manera los efectos del argumento que el materialismo Tipo-B.

En síntesis, una consideración de cierre -aunque siempre provisional- para el presente capítulo, puede pensarse de la siguiente manera. En numerosas ocasiones la disputa en torno al AC termina con apelaciones a la inversión de la carga de la prueba, apelaciones a la necesidad de marcos teóricos o argumentos subsidiarios o cuestiones puramente terminológicas. Valoro esos resultados como sintomáticos de un bloqueo dialéctico. En otras ocasiones, los ataques y defensas del argumento confluyen en el sostenimiento de su versión débil y en la declaración de insuficiencia para sostener la versión fuerte. El experimento mental procuró brindar un escenario óptimo para la irrupción de una intuición fuerte, de carácter

⁸¹ Cfr. sección “3.3.4 La paridad de razones contra el dualismo” (pág. 118-ss.).

epistémico, con la esperanza de que fuera el punto de partida para la defensa de una posición metafísica acerca de lo mental. Al final de la historia, la versión débil del AC no hace más que sostener la intuición epistémica de la que se partía, permite su irrupción con claridad y brinda imaginaria para hacer soportable la discusión. Pero es infértil en el terreno ontológico.

Es posible que sea por razones similares a las expuestas que Carnap se dedica, con detenimiento, a distinguir entre el materialismo tradicional y lo que él llama materialismo metódico, o simplemente fisicalismo. El apelativo “metódico” tiene la intención de expresar que se está refiriendo a

una tesis que versa simplemente sobre la posibilidad lógica de ciertas traducciones y derivaciones lingüísticas, y en ningún sentido sobre la “realidad” o “apariencia” (la “existencia” o la “no existencia”) de lo “dado”, lo “mental” o lo “físico”. Pseudoenunciados de ese tipo ocasionalmente ocurren en las formulaciones clásicas del Positivismo y el Materialismo (Carnap, 1995, pág. 94).

Si bien en el artículo citado Carnap prefiere que el científico utilice, por simplicidad o cercanía al lenguaje cotidiano, el lenguaje en su modo material desde una concepción materialista y no una dualista -por entender que esta última tiende a acarrear más contradicciones que la primera- siempre se debe tener en cuenta que el materialismo tradicional, metafísico, implica un posicionamiento metafísico tanto como el dualismo o el idealismo. De ninguna manera han de entenderse estas líneas como una adhesión global a las tesis del positivismo lógico. Deben entenderse como un intento de defender que la adopción de una posición materialista, que usualmente surge como resultado de perseguir principios de economía ontológica, es ella misma una toma de postura de carácter metafísico. Esto no hace más que explicar por qué las posturas Tipo-A y las posturas Tipo-D, E y F pueden ser empíricamente equivalentes (Goff, 2020); y por qué un intento por brindar una defensa argumental de una por sobre las otras puede arrastrarnos hacia contradicciones. Cualquiera de las dos defensas se extralimitaría respecto a lo que el lenguaje permite expresar.

Dejo para las consideraciones finales de esta tesis, en el siguiente capítulo, el retomar los problemas paralelos del reduccionismo de la conciencia fenoménica y del problema difícil de la conciencia. Volveré brevemente al mundo cartesiano, y evaluaré si el AC constituye algún tipo de novedad realmente significativa en relación a los avatares de la obra cartesiana, o si por el contrario hemos avanzado como Aquiles, sin alcanzar la tortuga.

Conclusiones

Al comienzo de esta tesis resumí en cuatro los objetivos centrales de la investigación. Primero, identificar y formular de manera actualizada el problema de la relación entre la mente y el cuerpo, distinguiendo claramente la dimensión epistemológica y la dimensión ontológica del problema, para lo que se apeló a la distinción entre problemas fáciles y problema difícil de la conciencia, y a una formulación de los mismos en términos de formas de superveniencia. Segundo, proponer una lectura alternativa para el sistema cartesiano que entienda a su posición dualista no tanto como el centro o punto de partida del sistema, sino como una solución de compromiso ante un escollo insalvable para el mecanicismo. Tercero, hacer una presentación de las principales corrientes de los siglos XX y XXI respecto al problema. Y cuarto, analizar el experimento mental de María, propuesto por Frank Jackson como una instancia del argumento del conocimiento para defender el dualismo y atacar el fisicalismo en el ámbito epistémico y el materialismo como postura metafísica, así como los enfoques críticos que he considerado más representativos e influyentes.

Una revisión en sentido inverso facilitará la recapitulación de lo logrado, y de lo no logrado, en la presente investigación. Comenzaré por el final, y arribaré al comienzo, y para las últimas líneas dejaré algunos comentarios relacionados al papel de la especulación metafísica y su relación con la ciencia empírica.

Al final del capítulo anterior establecí que, una vez repasadas y sometidas a crítica las diversas líneas de objeción, el resultado del análisis del AC puede sintetizarse en la aceptación de su tesis epistémica (la conclusión (3)), y el rechazo de su tesis ulterior, ontológica (la conclusión (4)). Esto es, tomo por buena la versión débil del argumento, y no así la versión fuerte por la que se establece la falsedad del materialismo. El rechazo de ésta última no implica afirmar que se trate de una postura falsa, ni adherir a su opuesta, sino a la negación de que el argumento por sí mismo presente elementos contundentes sobre el particular.

Respecto a la conclusión epistemológica, el conocimiento completo de las propiedades físicas involucradas en la percepción visual del color no es suficiente para predecir cómo se siente ver algo rojo (en un caso individual). De aquí se sigue que no podemos postular que lo mental supervenga lógicamente a lo físico. Es posible que haya variaciones en las propiedades fenoménicas sin que haya variaciones en las propiedades físicas. Por ello, podemos imaginar mundos posibles en los que la misma configuración física -del estímulo y del sistema nervioso del receptor- puede arrojar percepciones cualitativamente diferentes. Eso es lo que lleva a María a sorprenderse cuando, al fin, ve algo rojo.

Al presentar la relación entre la tesis de superveniencia y la del reduccionismo (pág. 44) señalé cierta movilidad de criterios en tres niveles. Mientras que Chalmers entiende que la superveniencia lógica es necesaria para el reduccionismo, Churchland entiende que es una exigencia demasiado estricta y Horgan, en el extremo opuesto, piensa que es insuficiente. Pongamos a consideración esta movilidad de criterios.

Como resulta evidente una vez que nos hemos familiarizado con los términos, la negación de la tesis de superveniencia lógica trae consigo la imposibilidad de postular una relación de superduperveniencia (Horgan T. , 1993). Esta última implicaría, además de la identificación *a priori* entre los conceptos fenoménicos y los físicos (neurobiológicos), una explicación de dicha identidad en términos de los segundos, sin ningún concepto puente del que se pueda sospechar un estatuto no físico.

Con este panorama, si aceptamos el criterio de Chalmers, y más aún el de Horgan, debemos aceptar que el programa de las ciencias cognitivas -consideradas en sentido amplio- puede avanzar con estrategias reduccionistas de forma legítima para buscar explicaciones de cualquier tipo de función cognitiva, hasta que lleguemos al límite de la experiencia consciente. Llegados a ese punto, las mismas podrán establecer correlaciones útiles entre estados observables mediante las diversas técnicas disponibles y -en el mejor de los casos- reportes fidedignos por

parte de los sujetos experimentales. Fidedignos, pero susceptibles de escenarios de inversión del espectro o ausencia de *qualia*. La naturaleza de los estados conscientes, como aquella aldea de irreductibles galos, parece ser un último bastión que se resiste a una explicación mecanicista. Para hablar acerca de nuestra experiencia consciente hay que suprimir el enfoque de tercera persona, objetivo y mecanicista. El lenguaje proposicional con pretensiones epistémicas, bien coinciden varios de los autores estudiados, no parece ser el apropiado para ese cometido. Por ello el eliminativismo y el dualismo se encuentran en la insuficiencia del fisicalismo reduccionista: o no hay nada, o hay una dimensión de nuestra mentalidad sobre la que no podemos hacer ciencia. O eliminamos la terminología mentalista de nuestras vidas, o nos tomamos en serio la conciencia (Chalmers, 1999 [1996]), pero no ambas al mismo tiempo. Como dice el filósofo brasileiro, materialista, Gustavo Leal Toledo:

El fisicalismo estricto dice simplemente que la conciencia fenoménica no puede ser construida con los tipos naturales de la física actual y de lo que entendemos por el término “física”. Eso es compatible tanto con el ilusionismo de la conciencia [materialismo Tipo-A] como con el dualismo. La única incompatibilidad está en los “meros términos ontológicos” (Leal Toledo, 2018)

O todos los problemas de la conciencia son fáciles, en principio solubles, o hay un problema difícil que no podemos resolver. No podemos quedar a medio camino. No podemos proponer que las ciencias cognitivas podrían resolver el problema difícil de cómo y por qué algunos estados mentales tienen un componente cualitativo.

Si bien Leal Toledo opta por el eliminativismo, en virtud de otras razones que no vienen al caso, coincide con quien suscribe que de lo anterior no se sigue que el materialismo sea falso. Ni que sea verdadero. De hecho, de la brecha epistémica no se sigue ninguna tesis ontológica. Es perfectamente posible que el materialismo (ontológico) sea verdadero, y que aun así el fisicalismo (epistémico) sea falso. Es decir, puede perfectamente ser verdad que en este mundo solo exista por sí misma una sustancia, material, extensa, medible, y al mismo tiempo que una explicación completa en términos estrictamente fisicalistas del carácter cualitativo

de nuestra experiencia consciente no sea posible. Y también es plausible postular dos sustancias, y que, por ejemplo, por el problema de la interacción entre las sustancias, no tener la clave para saltar la brecha explicativa.

El hecho de que, ni el AC, ni las múltiples objeciones analizadas, hayan sido concluyentes en aportar un sustento sostenible a alguna de las posiciones ontológicas presentadas en el capítulo 2, resulta significativo para poder considerar una nueva cuestión vinculada al sistema cartesiano. He insistido en leer a Descartes descentrando al dualismo y entendiendo su obra como un intento por proporcionar una estrategia reduccionista, mecanicista, para fundamentar una nueva forma de hacer ciencia.

Quien imagina a Descartes meditando acerca de la dualidad irreconciliable cuerpo-alma, sin entender el proceso por el que se arriba a ese punto, se imagina un metafísico, un místico, cerrado a lo que aprehende por sus sentidos, y que desde el sillón frente al fuego pretende dar un cuadro completo de la realidad. En contraste, quien parte de la lectura de sus obras de filosofía natural, y en particular *Le monde* en sus dos volúmenes, puede ver a un científico buscando un nuevo modelo metodológico que con pocos primeros principios pueda explicar la multiplicidad del mundo. Recordemos el pasaje citado en el capítulo 1 (pág. 11), en el que establece que “...hay un medio para explicar la causa de todos los cambios que ocurren en el mundo y de todas las transformaciones que se dan en la Tierra” (Descartes, 1989 [1633], pág. 63). Cuando el autor se topó con la segunda etapa de la sensación, lo que sería la irrupción del carácter cualitativo de algunos de nuestros estados mentales, halló que el mecanicismo es insuficiente para dar cuenta de ello⁸².

El argumento de la Meditación Segunda, posterior a la demostración del *cogito*, no es suficiente para demostrar el dualismo, pero prepara el terreno

⁸² Lo mismo puede afirmarse del experimento del molino de Leibniz citado en la nota al pie 41 (pág. 71).

estableciendo la asimetría epistémica entre los cuerpos y el alma, tal como sugiere el subtítulo de la Meditación. Será recién en la Sexta, a través del argumento modal, que Descartes pretende demostrar el dualismo de sustancias (Rorty, 1989) (Wilson, 1976) . De forma análoga, Chalmers (1999 [1996]) ha defendido que el AC por sí solo no logra demostrar el dualismo, y que precisa colocarse en tándem con otro argumento al que podemos referirnos como argumento modal, y que en su caso adopta la forma de otro experimento mental, el de la concebibilidad homólogos físico-funcionales a los seres humanos (zombis) pero sin experiencia consciente. Un análisis del argumento modal -en versión cartesiana, o en versión de Chalmers- podría constituir otra investigación similar a la presente, y queda como posible línea abierta para un futuro. Lo que en este momento deseo destacar es esa necesaria codependencia de ambos argumentos que Chalmers entiende imprescindible para demostrar el dualismo. El AC establece la no explicabilidad de lo mental mediante enfoques reduccionistas, mientras que el argumento modal parte de dicho resultado para establecer la concebibilidad de mente y cuerpo por separado y así postular el dualismo.

Es notorio que el argumento modal es similar, tanto en su procedimiento como en sus fines, al argumento de la Meditación Sexta para el dualismo. Lo que no he hallado en la bibliografía consultada, y que ahora sí estoy en condiciones de afirmar, es que el AC tal como lo propone Frank Jackson no es nada más ni nada menos que un argumento análogo, *aggiornado*, al argumento de la Meditación Segunda. Cada uno con sus peculiaridades propias de la época y del estilo de cada filósofo, ambos concuerdan en el establecimiento de la asimetría epistémica que permite refutar la superveniencia lógica de lo fenoménico sobre lo físico. En ambos casos la conclusión se halla en el ámbito de lo epistémico, y no es suficiente para concluir algo en el ámbito de lo ontológico. Preparan el terreno para que otro argumento opere en el plano ontológico.

La neurociencia ha avanzado desde entonces, resolviendo problemas fáciles de la conciencia. En contraste, el problema difícil parece resistir cualquier novedad. La actividad filosófica, con su talante inevitablemente metafísico, se ve interpelada

nuevamente. Parece que no hubiese habido ningún progreso genuino, ni en las preguntas ni en las respuestas, desde la inauguración de la Modernidad por el giro cartesiano.

El hecho de que, como vimos, lo ontológico parezca reducirse a una decisión semántica respecto a lo que efectivamente significan términos como “materia” o “físico”, y que numerosas líneas de objeción hayan culminado con un intercambio de la carga de la prueba, podría abogar a favor de la posición de los positivistas lógicos, para quienes tal discusión es pseudoproblemática, y carece de significado. De hecho, he adherido en el cierre de la sección anterior a la distinción de Carnap entre materialismo tradicional y materialismo metodológico para intentar clarificar la cuestión. El propio Goff, panpsiquista entusiasta en la defensa de su postura, ha planteado que el pansiquismo y el fisicalismo son teorías empíricamente equivalentes. Esto significa que no puede haber hallazgos observacionales o experimentales posibles que puedan servir de evidencia para apoyar una por sobre la otra (Goff, 2020).

¿Significa esto que hemos de enviar la bibliografía de esta tesis, y ella misma, a la hoguera? ¿Arrojarla al vacío como la escalera de Wittgenstein? No necesariamente. En todo caso, si deseamos encontrar lugar y sentido a la especulación metafísica, hemos de delimitar claramente su territorio y sus pretensiones, y por sobre todas las cosas, hemos de evitar confundirla con la actividad científica. No es un tópico novedoso, pero no por ello omito abordarlo.

Un experimento mental no es análogo a un experimento científico. El segundo aportará apoyo empírico para el sostenimiento o refutación de alguna hipótesis, por lo que es razonable entenderlo como un argumento. Por el contrario, los experimentos mentales no han de entenderse como argumentos concluyentes. Sus escenarios, y las intuiciones que buscan aislar, son discutibles. Exponentes de posturas opuestas suelen interpretarlos de manera que atraigan agua a sus molinos. Posiblemente sea mejor entenderlos, más que como argumentos filosóficos, como vehículos para la especulación filosófica, y como escenarios para la disputa

argumentativa. Más que un argumento, un experimento mental es el escenario en el que pueden transcurrir los argumentos. No tendrán peso suficiente para hacernos volcar por una u otra posición metafísica, porque después de todo, ¿en virtud de qué es que adoptamos posiciones metafísicas?

En este sentido, reconociendo la oscuridad de la dimensión ontológica del problema mente-cuerpo, deseo cerrar esta investigación reivindicando el pleno sentido de la distinción de Chalmers entre problemas fáciles y difíciles de la conciencia. La distinción se ve apoyada por el AC, ya que la intuición fuerte que el experimento mental recoge resiste el análisis de objeciones: un conocimiento físico completo aún deja algo sin saber, algo acerca del mundo que, particularmente, es algo de nuestra naturaleza humana. Dicha distinción, en opinión exclusiva de quien escribe, resguarda algo de lo humano de la máquina voraz del saber científico-técnico. No se entienda mal, no estoy defendiendo una posición anti-científica. Solo señalo que es en estos resquicios de misterio en donde el pensamiento puede volcarse a la especulación metafísica, con la libertad de aquello que no ha de ser resuelto por un dato, y procurar el sentido, no como algo dado -para eso está la ciencia- sino como algo a determinar o prescribir.

Según la filósofa argentina Diana Sperling, las neuronas “atrasan” en relación al psicoanálisis, o al idealismo trascendental kantiano. Las neurociencias objetivas, reductivas,

intentan conocer esa alteridad como se conoce el glóbulo o la molécula. Se verifica entonces una paradoja: al tiempo que las ciencias buscan ubicar y manipular ese elemento y reducir lo desconocido a lo conocido, lo inasible a lo mensurable y calculable, por otro lado está en desenfrenado auge una multitud variopinta de diversas formas de “espiritualidad”. (...) Pero esta proliferación de promesas da cuenta de un conflicto siempre presente y permanentemente renovado: la dificultad de hacer lugar a eso insabido que nos habita (Sperling, 2018, págs. 323-324).

¿Cómo y por qué algunos de nuestros estados mentales se sienten de cierta manera? Quizás sea “...preciso suprimir *el saber* para dar lugar a la creencia” (Kant, 1970 [1787], pág. 140).

Bibliografía

Armstrong, D. M. (1981). *The Nature of Mind*. Brighton: The Harvester Press.

Balmaceda, T. (2016). El origen de las intuiciones dualistas en la vida cotidiana. *Verba Volant. Revista de Filosofía y Psicoanálisis*.(1), 55-70. Retrieved 08 05, 2016, from http://dspace.uces.edu.ar:8180/xmlui/bitstream/handle/123456789/3577/Origen_Balmaceda.pdf?sequence=1

Barberis, S. (2011). ¿Conocimiento Fenoménico sin Conceptos Fenoménicos? Sobre la Teoría de los Punteros Mentales de Jesse Prinz. *Revista Argentina de Ciencias del Comportamiento*, 3(1), 64-74.

Barberis, S. (2012). Un análisis crítico de la concepción mecanicista de la explicación. *Revista Latinoamericana de Filosofía*, 38(2), 233-265.

Bayne, T., & Montague, M. (2011). *Cognitive Phenomenology*. Oxford: Oxford University Press.

Bayne, T., Cleeremans, A., & Wilken, P. (Eds.). (2009). *The Oxford companion to consciousness*. Oxford: Oxford University Press.

Beaujeu, J. (1985). Galeno. In R. Taton (Ed.), *Historia general de las ciencias* (pp. 441-446). Barcelona: Destino.

Berkeley, G. (1992 [1709]). *Tratado sobre los principios del conocimiento humano*. Madrid: Alianza Editorial.

- Bigelow, J., & Pargetter, R. (2004). Acquaintance with Qualia. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 179-196). Cambridge: MIT Press.
- Block, N. (1980). What is functionalism? In N. Block, *Readings in Philosophy of Psychology Vol. I* (pp. 171-184). Massachusetts: Harvard University Press.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and brain sciences*(18), 227-287.
- Block, N. (2003). La Tierra Invertida. In M. Ezcurdia, & O. Hansberg (Eds.), *La naturaleza de la experiencia. Vol I. Las Sensaciones*. (pp. 289-323). México D.F.: Universidad Nacional Autónoma de México.
- Block, N., & Fodor, J. (2003). Lo que no son los estados psicológicos. In M. Ezcurdia, & O. (. Hansberg, *La naturaleza de la experiencia Vol I: Sensaciones* (A. I. Stellino, Trans., pp. 155-181). México D.F.: UNAM.
- Bloom, P. (2004). *Descartes' Baby*. Londres: Random House Ltd.
- Briscoe, R. (2015). Cognitive Penetration and the Reach of Phenomenal Content. In A. Raftopoulos, & J. Zeimbekis, *The Cognitive Penetrability of Perception: New Philosophical Perspectives* (pp. 192-219). Oxford: Oxford Scholarship Online.
- Brown, J. R., & Fehige, Y. (2017, Junio 21). Thought Experiments. (E. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved Abril 14, 2019, from <https://plato.stanford.edu/archives/sum2017/entries/thought-experiment/>
- Buzsáki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press.
- Carnap, R. (1965). La superación de la metafísica mediante el análisis lógico del lenguaje. In A. J. Ayer (Ed.), *El positivismo lógico* (pp. 66-87). México D.F.: Fondo de Cultura Económica.

- Carnap, R. (1965). Psicología en Lenguaje Fisicalista. In A. Ayer (Ed.), *El positivismo lógico* (pp. 171-204). México D.F.: Fondo de Cultura Económica.
- Carnap, R. (1995). *The Unity of Science*. Bristol: Thoemmes Press.
- Carnap, R., Hahn, H., & Neurath, O. (1929). *La Concepción Científica del Mundo*. (A. Zela, Trans.) Viena: Centro de Estudios de Filosofía Analítica. Retrieved Agosto 18, 2019, from <http://www.vienayyo.com/wp-content/uploads/2011/10/manifiesto.pdf>
- Chalmers, D. (1995, Diciembre). The puzzle of conscious experience. *Scientific American*, 62-68.
- Chalmers, D. (1995b). Facing Up to the Problem of Consciousness. Retrieved 01 10, 2017, from <http://cogprints.org/316/1/consciousness.html>
- Chalmers, D. (1997). Moving Forward on The Problem of Consciousness. Retrieved 01 10, 2017, from <http://cogprints.org/317/1/moving.html>
- Chalmers, D. (1999 [1996]). *La mente consciente: en busca de una teoría fundamental*. Madrid: Gedisa.
- Chalmers, D. (2002). Consciousness and its Place in Nature. In D. Chalmers, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Chalmers, D. (2013). Panpsychism and Panprotopsychism. Retrieved 08 28, 2019, from consc.net/papers/panpsychism.pdf
- Chalmers, D. (2016). The Combination Problem for Panpsychism. Retrieved 09 5, 2019, from consc.net/papers/combination.pdf

- Churchland, P. M. (1985, Enero). Reduction, Qualia, and the Direct Introspection of Brain States. *The Journal of Philosophy*, 82(1), 8-28. Retrieved 04 29, 2018, from <http://www.jstor.org/stable/2026509>
- Churchland, P. M. (1999). *Materia y conciencia*. Barcelona: Gedisa.
- Churchland, P. M. (2004). Knowing Qualia: A Reply to Jackson (with Postscript: 1997). In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 163-178). Cambridge: MIT Press.
- Churchland, P. M., & Smith Churchland, P. (1990, Enero). Could a Machine Think? *Scientific American*, 32-36.
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Londres: The MIT Press.
- Clarke, D. (1986). *La filosofía de la ciencia de Descartes*. Madrid: Alianza Editorial.
- Clarke, D. (2005). *Descartes's Theory of Mind*. New York: Oxford University Press.
- Coleman, S. (2009). Why the Ability Hypothesis is best forgotten. *Journal of Consciousness Studies*, 16(2-3), 74-97.
- Conee, E. (2004). Phenomenal Knowledge. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 197-215). Cambridge: MIT Press.
- Crombie, A. C. (1987). *Historia de la ciencia. De San Agustín a Galileo*. (Vol. 2; La ciencia en la Baja Edad Media y comienzos de la Edad Moderna: siglos XIII al XVII.). Madrid: Alianza Editorial.
- Davidson, D. (1969). True to the Facts. *Journal of Philosophy*, 66(21), 748-764.

- Davidson, D. (1981). *Sucesos Mentales*. México D.F.: UNAM.
- Dennett, D. (1995). Facing Backwards on the Problem of Consciousness. In J. Shear (Ed.), *Explaining Consciousness. The 'Hard Problem'*. (pp. 33-36). Cambridge: The MIT Press.
- Dennett, D. (1995b). *La Conciencia Explicada*. Buenos Aires: Paidós.
- Dennett, D. (2004). "Epiphenomenal" qualia? In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary*. (pp. 59-68). Cambridge: MIT Press.
- Dennett, D. (2006). Lo que sabe RoboMary. In D. C. Dennett, *Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia* (pp. 123-152). Buenos Aires: Katz.
- Dennett, D. (2015). *Bombas de intuición*. México D.F.: Fondo de Cultura Económica.
- Descartes, R. (1968 [1637]). *Discurso del método*. Madrid: Espasa-Calpe.
- Descartes, R. (1984 [1628]). *Reglas para la dirección del espíritu*. Madrid: Alianza Editorial.
- Descartes, R. (1989 [1633]). *El mundo. Tratado de la luz*. Barcelona: Anthropos.
- Descartes, R. (2011 [1633]). Tratado del hombre. In R. Descartes, *Descartes* (pp. 673-736). Madrid: Gredos.
- Descartes, R. (2011 [1641]). Meditaciones metafísicas seguidas de las objeciones y respuestas. In R. Descartes, *Obras* (pp. 153-414). Madrid: Gredos.
- Descartes, R. (2011 [1649]). Las pasiones del alma. In R. Descartes, *Descartes* (pp. 461-548). Madrid: Gredos.

- Descartes, R., & Isabel de Bohemia. (2011 [1642]). Correspondencia con Isabel de Bohemia. In R. Descartes, *Obras* (pp. 549-672). Madrid: Gredos.
- Duncan, S. (2012, Abril). Leibniz's Mill Arguments Against Materialism. *The Philosophical Quarterly*, 62(247), 250-272.
- Fiala, B., Arico, A., & Nichols, S. B. (2012). On the Psychological Origins of Dualism: Dual-Process Cognition and the Explanatory Gap. In E. Slingerland, & M. Collard (Eds.), *Creating Consilience: Integrating the Sciences and the Humanities* (pp. 88-110). New York: Oxford University Press.
- Fodor, J. A. (1986). *La modularidad de la mente*. Madrid: Ediciones Morata.
- Foster, J. (2002). *The immaterial self*. Londres: Taylor & Francis e-library.
- Francescotti, R. (2020). Supervenience and mind. (J. Fieser, & B. Dowden, Eds.) *The Internet Encyclopedia of Philosophy*. Retrieved 02 23, 2020, from <https://www.iep.utm.edu/supermin/>
- Garrison, F. (1966). *Historia de la medicina*. México: Interamericana.
- Goff, P. (2017). *Consciousness and fundamental reality*. New York: Oxford University Press.
- Goff, P. (2020, 02 12). *Special relativity is empirically equivalent to the Lorentzian theory that preceded it, just as panpsychism is empirically equivalent to materialism. Like panpsychists, Einstein sought a simpler & more intelligible interpretation of the data.* Retrieved from @Philip_Goff: https://twitter.com/Philip_Goff/status/1227714460544049152
- Goff, P., Seager, W., & Allen-Hermanson, S. (2017, 12 21). Panpsychism. (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved 8 29, 2019, from <https://plato.stanford.edu/archives/win2017/entries/panpsychism/>

- Gómez-Robledo, L., Valero, E. M., Huertas, R., Martínez-Domingo, M. A., & Hernández-Andrés, J. (2018, Octubre 28). Do EnChroma glasses improve color vision for colorblind subjects? *Optics Express*, 26(22).
- Graham, G. (2019, Marzo 21). Behaviorism. (E. N. Zalta, Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved mayo 2, 2019, from <https://plato.stanford.edu/archives/spr2019/entries/behaviorism/>
- Hamelin, O. (1949). *El sistema de Descartes*. Buenos Aires: Losada.
- Harvey, W. (1994 [1628]). *Del movimiento del corazón y de la sangre en los animales*. México: UNAM.
- Hempel, C. G. (1997). The Logical Analysis of Psychology. In P. Morton, A *Historical Introduction to the Philosophy of Mind: readings with commentary* (pp. 164-174). Ontario: Broadview Press.
- Hierro-Pescador, J. (2005). *Filosofía de la mente y de la ciencia cognitiva*. Madrid: Akal.
- Horgan, T. (1993). From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind*, 102(408), 555-586. Retrieved 03 18, 2019, from <https://www.jstor.org/stable/2254046>
- Horgan, T. (2004 [1984]). Jackson on Physical Information and Qualia. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 301-308). Cambridge: The MIT Press.
- Hume, D. (2001 [1739]). *Tratado de la naturaleza humana*. Albacete: Diputación de Albacete.
- Hume, D. (2003 [1748]). *Investigación sobre el conocimiento humano*. Madrid: Mestas.

- Jackson, F. (1982, Abril). Epiphenomenal Qualia. *The Philosophical Quarterly*, 32(127), 127-136. Retrieved Abril 27, 2018, from <http://www.jstor.org/stable/2960077>
- Jackson, F. (1986, Mayo). What Mary Didn't Know. *The Journal of Philosophy*, 83(5), 291-295. Retrieved 04 26, 2018, from <http://www.jstor.org/stable/2026143>
- Jackson, F. (2003). Qualia epifenoménicos. In M. Ezcurdia, & O. Hansberg (Eds.), *La naturaleza de la experiencia. Volumen I. Sensaciones*. (pp. 95-110). México D.F.: Universidad Nacional Autónoma de México.
- Jackson, F. (2003b). Lo que María no sabía. In M. Ezcurdia, & O. Hansberg (Eds.), *La naturaleza de la experiencia. Volumen I. Sensaciones*. (L. E. Marníquez, Trans., pp. 111-118). Mexico D.F.: Universidad Autónoma de México.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2013). *Principles of Neural Sciences* (5a ed.). McGraw-Hill.
- Kant, I. (1970 [1787]). *Crítica de la Razón Pura* (Vol. 1). Buenos Aires: Losada.
- Kim, J. (2006). *Philosophy of Mind*. Boulder: Westview Press.
- Kulstad, M., & Carlin, L. (2013, Diciembre 21). *Leibniz's Philosophy of Mind*. (E. N. Zalta, Editor) Retrieved Abril 30, 2019, from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/win2013/entries/leibniz-mind/>
- Lain Entralgo, P. (1948). *Vida y Obra de Guillermo Harvey*. Buenos Aires: Espasa-Calpe.
- Latarjet, M., & Ruiz Liard, A. (1998). *Anatomía humana*. Madrid: Panamericana.
- Leal Toledo, G. (2018). Fisicalismo Estrito e Dualismo Enrustido. *Congreso Internacional de la Sociedad Filosófica del Uruguay*. Montevideo: Inédito.

- Leibniz, G. G. (1977 [1714]). *Monadología*. México D.F.: Porrúa.
- Leibniz, G. G. (1977 [1765]). *Nuevo tratado sobre el entendimiento humano*. México D.F.: Porrúa.
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*(64), 354-361.
- Lewis, D. (2003). Lo que la experiencia enseña. In M. Ezcurdia, & O. Hansberg (Eds.), *La naturaleza de la experiencia. Volúmen I. Sensaciones*. (pp. 119-154). México D.F.: Universidad Nacional Autónoma de México.
- Liu, X., Ramirez, S., & Tonegawa, S. (2014). Inception of a false memory by optogenetic manipulation of a hippocampal memory engram. *Philosophical Transactions of The Royal Society*, 369, 1-7.
- Loar, B. (1990). Phenomenal States. *Philosophical Perspectives*, 4, 81-108. Retrieved from <https://www.jstor.org/stable/2214188>
- Loar, B. (2004). Phenomenal States (Revised Version). In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 219-239). Cambridge: The MIT Press.
- Locke, J. (1956 [1689]). L. II, Cap. IX, 8. In J. Locke, *Ensayo sobre el entendimiento humano (compendio)* (pp. 65-66). Buenos Aires: Aguilar.
- Lycan, W. G. (2009). Giving Dualism its Due. *Australasian Journal of Philosophy*, 87(4), 551-563.
- Macpherson, F. (2015). Cognitive Penetration and Nonconceptual Content. In A. Raftopoulos, & J. Zeimbekis, *The Cognitive Penetrability of Perception: New Philosophical Perspectives* (pp. 350-378). Oxford: Oxford Scholarship Online.

- Martínez-Domingo, M. A., Gómez-Robledo, L., Valero, E. M., Huertas, R., Hernández-Andrés, J., Ezpeleta, S., & Hita, E. (2019, Junio 24). Assessment of VINO filters for correcting red-green Color Vision Deficiency. *Optics Express*, 27(13).
- Mason, S. (1985). La revolución científica de los siglos XVI y XVII. In S. Mason, *Historia de las ciencias* (Vol. 2). Madrid: Alianza Editorial.
- McGinn, C. (2003). ¿Podemos resolver el problema mente-cuerpo? In M. Ezcurdia, & O. Hansberg (Eds.), *La naturaleza de la experiencia. Volumen I. Sensaciones*. (pp. 65-94). México D.F.: Universidad Nacional Autónoma de México.
- McLaughlin, B., & Bennett, K. (2018, 03 21). *Stanford Encyclopedia of Philosophy*. (E. Zalta, Ed.) Retrieved 07 18, 2018, from <https://plato.stanford.edu/archives/spr2018/entries/supervenience/>
- Molière. (2000 [1673]). *El enfermo imaginario*. El Aleph.
- Mondolfo, R. (1952). *El pensamiento antiguo. Historia de la filosofía greco-romana*. (Vol. I). Buenos Aires: Losada.
- Nagasawa, Y. (2002). The Knowledge Argument Against Dualism. *Theoria*(68), 205-223.
- Nagel, T. (2000 [1974]). ¿Qué se siente ser un murciélago? In T. Nagel, *Ensayos sobre la vida humana* (pp. 274-296). México D.F.: Fondo de Cultura Económica.
- Nemirow, L. (1980, Julio). Mortal Questions by Thomas Nagel. *The Philosophical Review*, 89(3), 473-474. Retrieved 10 13, 2019
- Nida-Rümelin, M. (2004). What Mary Couldn't Know. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 241-267). Cambridge: The MIT Press.

- Nida-Rümelin, M., & O Conaill, D. (2019). Qualia: The Knowledge Argument. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved 12 18, 2019, from <https://plato.stanford.edu/archives/win2019/entries/qualia-knowledge/>
- Nietzsche, F. (1996 [1896]). *Sobre verdad y mentira en sentido extramoral*. Madrid: Tecnos.
- Papp, D. (1970). Introducción. In W. Harvey, *De motu cordis. Estudio anatómico del movimiento del corazón y de la sangre en los animales*. Buenos Aires: EUDEBA.
- Pellegrino, E. (1963). La medicina y el concepto del hombre. *The annals of The American Academy of Political and Social Science*.
- Place, U. T. (1956, Febrero). Is Consciousness a Brain Process? *British Journal of Psychology*, 47(1), 44-50.
- Platón. (1938). Fedón. In Platón, *Diálogos*. (pp. 16-108). Buenos Aires: Espasa-Calpe.
- Platón. (1999). *La República o El Estado*. Barcelona: Edicomunicación.
- Popper, K. R. (1962). *La Lógica de la Investigación Científica*. Madrid: Tecnos.
- Popper, K. R. (1980). Crítica del materialismo. In K. R. Popper, & J. C. Eccles, *El yo y su cerebro* (pp. 58-112). Barcelona: Ediciones Roche.
- Popper, K. R., & Eccles, J. C. (1980). *El yo y su cerebro*. Barcelona: Ediciones Roche.
- Putnam, H. (1975). The nature of mental states. In H. Putnam, *Mind, language and reality* (pp. 429-439). Cambridge: Cambridge University Press.

- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*(22), 341-432.
- Rabin, G. (2016). Against Old Facts Under New Modes. Retrieved 12 15, 2019, from <http://gabriel-rabin.squarespace.com/s/Against-old-facts-under-new-modes-feb2016.pdf>
- Raftopoulos, A. (2009). *Cognition and perception: how do psychology and neural science inform philosophy?* Massachusetts: Massachusetts Institute of Technology.
- Raftopoulos, A. (2014). The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology*, 27(5), 601-620.
- Ramirez, S., Liu, X., Pei-Ann, L., Junghyup, S., Pignatelli, M., Redondo, R. L., . . . Tonegawa, S. (2013). Creating a False Memory in the Hippocampus. *Science*, 341, 387-391.
- Raymont, P. (1999). The Know-How Response to Jackson's Knowledge Argument. *Journal of Philosophical Research*, XXIV, 113-126.
- Richert, R. A., & Harris, P. L. (2008). Dualism Revisited: Body vs. Mind vs. Soul. *Journal of Cognition and Culture*(8), 99-115.
- Robinson, H. (2004). Dennett on the Knowledge Argument. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 69-73). Cambridge: MIT Press.
- Rorty, R. (1989). *La filosofía y el espejo de la naturaleza*. Madrid: Cátedra.
- Ryle, G. (1949). *The concept of mind*. Londres: Hutchinson's University Library.

- Searle, J. (1990). Is the brain's mind a computer program? *Scientific American*, 23-31.
- Shoemaker, S. (1982). The Inverted Spectrum. *The Journal of Philosophy*, 79(7), 357-381.
- Smart, J. J. (1959). Sensations and Brain Processes. *The Philosophical Review*, 68(2), 141-156.
- Smith Churchland, P. (1994, Enero). Can Neurobiology Teach Us Anything about Consciousness? *Proceedings and Addresses of the American Philosophical Association*, 67(4), 23-40.
- Smith Churchland, P. (1997). The Hornswoggle Problem. In J. Shear (Ed.), *Explaining Consciousness. The 'Hard Problem'*. (pp. 37-44). Cambridge: The MIT Press.
- Sperling, D. (2018). *La diferencia. Sobre filiación y avatares en Occidente*. Buenos Aires: Miño y Dávila.
- Stoljar, D. (2004). Two Conceptions of the Physical. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 309-331). Cambridge: The MIT Press.
- Stoljar, D. (2017). Physicalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved 12 29, 2019, from <https://plato.stanford.edu/archives/win2017/entries/physicalism/>
- Stoljar, D., & Nagasawa, Y. (2004). Introduction. In P. Ludlow, D. Stoljar, & Y. Nagasawa (Eds.), *There's Something About Mary* (pp. 1-36). Cambridge: MIT Press.
- Trappenberg, T. P. (2010). *Fundamentals of computational neuroscience*. New York: Oxford University Press.

- Turing, A. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433-460.
- Turró, S. (1989). Estudio Introductorio. In R. Descartes, *El mundo. Tratado de la luz*. (pp. 7-42). Barcelona: Anthropos.
- Tye, M. (2004). Knowing What It Is Like: The Ability Hypothesis and the Knowledge Argument. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's Something About Mary* (pp. 143-160). Cambridge: MIT Press.
- Van Gulick, R. (2004). So many ways of saying no to Mary. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.), *There's something about Mary*. (pp. 365-406). Cambridge: MIT Press.
- Vanderbroucke, A. R., Sligte, I. G., Fahrenfort, J. J., Ambroziak, K. B., & Lamme, V. A. (2012). Non-Attended Representations are Perceptual Rather than Unconscious in Nature. *PLoS ONE*, 7(11).
- Watson, J. B. (1945 [1924]). *El Conductismo*. Buenos Aires: Paidós.
- Williams, B. (1996). *Descartes. El proyecto de la investigación pura*. Madrid: Cátedra.
- Wilson, M. (1976). Descartes: The Epistemological Argument for Mind-Body Distinctness. *Nous*, 10(1), 3-15.
- Wittgenstein, L. (1973 [1921]). *Tractatus Logico-Philosophicus*. Madrid: Alianza Editorial.
- Zakeri, M., & Ghasemi, M. (2016). The Ability Hypothesis: An eMpirically Based Defense. *Kriterion - Journal of Philosophy*, 30(1), 23-38.
- Zhao, Y. (2012). The Knowledge Argument against Physicalism: Its Proponents and Its Opponents. *Frontiers of Philosophy in China*, 7(2), 304-316.