



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN FILOSOFÍA
FACULTAD DE FILOSOFÍA Y LETRAS
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

**REPRESENTACIONISMO ILUSIONISTA: DESARROLLO Y DEFENSA DE UNA
RESPUESTA AL PROBLEMA DIFÍCIL DE LA CONSCIENCIA**

TESIS QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN FILOSOFÍA

PRESENTA:
DIEGO REYNOSO ORTIZ MONASTERIO

TUTOR PRINCIPAL:
DR. MIGUEL ÁNGEL SEBASTIÁN GONZÁLEZ
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS, UNAM

COMITÉ TUTOR:
DR. JAVIER CUMPA ARTESEROS
UNIVERSIDAD COMPLUTENSE DE MADRID
DR. MARIO GÓMEZ TORRENTE
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS, UNAM

Ciudad de México, Méx., junio 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

En primer lugar, debo agradecer al Doctor Miguel Angel Sebastian Gonzalez, mi director de tesis, a cuya guía y conocimiento sobre el tema debe muchísimo esta investigación. Agradezco infinitamente su tiempo, apoyo, paciencia y dedicación, así como sus observaciones y recomendaciones.

También agradezco enormemente a los otros dos miembros de mi comité tutor, los Doctores Javier Cumpa Arteseros y Mario Gómez Torrente, quienes desde el principio mostraron gran interés en el proyecto y aportaron valiosas observaciones.

Doy gracias también a los Doctores Pedro Stepanenko Gutiérrez y Ricardo Mena Gallardo, quienes fungieron como lectores en la última fase. También estoy muy agradecido con los Doctores Alessandro Torza y Eduardo García Ramírez, quienes me apoyaron en etapas tempranas y con cuyas contribuciones me hubiera gustado contar hasta el final.

Por otra parte, debo agradecer a todas las instituciones que prestaron su apoyo, comenzando por la UNAM, por abrirme sus puertas desde hace no pocos años y por la excelencia que ofrece en estudios de filosofía. Mención especial merecen los programas (dentro y fuera de la UNAM) que me prestaron ayuda económica:

Esta investigación fue realizada gracias al apoyo de una beca de posgrado otorgada por el CONACyT, desde agosto de 2017 hasta julio de 2021. Agradezco al CONACyT por la beca recibida.

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM <<IN400520>> e <<IG400219>>. Agradezco a la DGAPA-UNAM la beca recibida.

También merecen mención todos los investigadores y compañeros de viaje con quienes

tuve la oportunidad de intercambiar ideas en estos últimos años. En especial agradezco a los otros alumnos de Miguel Ángel Sebastián, con quienes llevamos un enriquecedor seminario. También doy las gracias a todos los autores a los que sólo conozco por sus escritos. En general agradezco a todos aquellos que comparten mi pasión por la filosofía y por el conocimiento en general.

Finalmente, doy las gracias a todos mis amigos y familiares, que de alguno u otro modo me apoyaron en esta travesía y mostraron interés en lo que estaba haciendo. En especial agradezco a mi padre, Ricardo Reynoso Monges, quien siguió de cerca todos mis avances.

Índice

Introducción.....	6
1. El problema a resolver: la consciencia fenoménica y el problema difícil.....	10
1.1. Consciencia fenoménica.....	12
1.2. Fisicismo.....	18
1.2.1. Fisicismo respecto a la consciencia fenoménica.....	23
1.3. Internismo fenoménico.....	26
1.3.1. Swampman.....	29
1.3.2. Cerebros en cubeta.....	31
1.4. Contra el fisicismo: la brecha epistémica y el problema difícil de la consciencia.....	33
1.4.1. Formalizando algunos argumentos en contra del fisicismo.....	39
1.5. Sumario 1.....	55
2. Las posibles respuestas I: teorías sobre la presentación tipo E.....	57
2.1. En busca de la distinción clave.....	59
2.2. <i>Acquaintance</i>	61
2.2.1. Realismo directo.....	63
2.2.2. Teorías de los <i>sense-data</i>	66
2.2.3. Contra el <i>acquaintance</i>	67
2.3. Representación y representacionismo.....	73
2.3.1. Representación.....	73
2.3.2. Representacionismo.....	81
2.3.3. Representación naturalizada.....	87
2.3.3.1. Algunas nociones clave.....	87
2.3.3.2. Teleología naturalizada.....	90
2.3.3.2.1. Teleología etiológica.....	92
2.3.3.2.2. Teleología disposicional.....	96
2.3.3.2.3. Posturas híbridas, unitarias y pluralistas.....	103
2.3.3.3. Funciones teleológicas representacionales.....	106

2.3.3.3.1. Representación etiológica.....	106
2.3.3.3.2. Representación disposicional.....	117
2.4. La postura (representacionista) más plausible (hasta ahora).....	122
2.5. Sumario 2.....	124
3. Las posibles respuestas II: teorías sobre los qualia.....	126
3.1. Algunas notas preliminares.....	128
3.2. Fisicismo ingenuo respecto a los qualia: propiedades físicas externas ordinarias.....	129
3.2.1. Contra el fisicismo ingenuo respecto a los qualia.....	131
3.3. Fisicismo no-ingenuo respecto a los qualia: propiedades físicas internas o no-ordinarias.....	136
3.3.1. Contra el fisicismo no-ingenuo respecto a los qualia.....	139
3.4. Anti-fisicismo respecto a los qualia.....	142
3.4.1. Contra el anti-fisicismo respecto a los qualia.....	142
3.5. La taxonomía resultante y el veredicto final.....	145
3.6. Sumario 3.....	150
4. La solución: representacionismo ilusionista.....	152
4.1. La tesis principal y las razones que la respaldan.....	155
4.2. Ilusionismo.....	156
4.3. Representación tipo E' naturalizada.....	161
4.4. El representacionismo ilusionista en el panorama evolutivo.....	165
4.5. Respuesta a los argumentos anti-fisicistas.....	168
4.6. Sumario 4.....	173
5. Conclusiones.....	175
Bibliografía.....	178

Introducción

El objetivo primario de esta investigación es contribuir a explicar algo que para cualquier persona es sumamente familiar, pero que a la vez es tremendamente misterioso: la *consciencia fenoménica*, que abarca todas las experiencias sensibles (o *fenoménicas*), es decir, todos los estados tales que para quien los tiene *es (o se siente) de algún modo* estar en ellos, desde su perspectiva en primera persona. En particular, me interesa dar cuenta de ella en términos compatibles con el *fisicismo* y el *internismo fenoménico*. El primero es la tesis de que el universo es enteramente físico y el segundo es la de que tener consciencia fenoménica depende exclusivamente de factores internos del ente que la tiene. En otras palabras, lo que busco es explicar cómo es que la consciencia fenoménica surge a partir de mera maquinaria física interna del ente que la tiene. Ello es un gran reto porque a la fecha no hay una teoría que vincule la consciencia fenoménica y lo físico de manera transparente e incluso hay quienes sostienen que no puede haberla. Y la tarea se dificulta aun más al limitar los recursos explicativos a sólo maquinaria (física) interna.

Con esa tarea en mente, propondré entender la consciencia fenoménica como un especial tipo de percatación de un especial tipo de objetos, comúnmente llamados ‘qualia’. La idea es que toda experiencia fenoménica consiste en la presentación ante el sujeto de la apariencia que para él tiene alguna porción del mundo (interno o externo) y que los qualia son las propiedades que componen a dicha apariencia. Sobre esa línea, llamaré ‘presentación tipo E’ al especial tipo de percatación de apariencias (compuestas por qualia) que está en juego durante las experiencias fenoménicas y diré que la consciencia fenoménica es idéntica a *presentación tipo E de qualia*. Ello permitirá abordar el problema poniendo atención primero en las diversas teorías sobre la naturaleza de la presentación tipo E y luego en las distintas teorías sobre la naturaleza de los qualia, obteniendo al final una taxonomía general que será producto de meramente listar todas las combinaciones posibles de todas las maneras de entender cada uno de esos dos elementos.

Entonces aparecerá un segundo objetivo de esta tesis: defender y detallar (hasta cierto

punto) la combinación que considero más viable, que es una que hasta ahora ha recibido poca atención en la literatura filosófica: el *representacionismo ilusionista*. A grandes rasgos, esa propuesta consiste en sostener que los qualia no existen y, por ende, que la presentación tipo E de qualia es un tipo de ilusión (representacional) que hace parecer que sí existen (y con mucho poder de convencimiento). La principal razón para negar la existencia de los qualia es sencillamente que hay muy buenas razones para pensar que no son físicos, lo cual orilla a todo físico a negar su existencia. También haré énfasis en que ello no involucra negar la existencia de la consciencia fenoménica (como algunos interpretan), porque sostendré que perfectamente puede haber presentación tipo E de qualia en ausencia de qualia. Cabe señalar que, si bien argumentaré con bastante fuerza a favor de la postura en cuestión, me interesa más mostrar que es perfectamente viable que destruir a sus rivales.

Dicho todo lo anterior, podemos proceder a revisar la estructura general del trabajo, que estará dividido en cuatro capítulos. El primero de ellos estará enfocado en presentar el problema a resolver, es decir, el de explicar la consciencia fenoménica de manera compatible con el fisicismo y el internismo fenoménico. Con ello en mente, comenzaré dedicando una sección a delimitar con mayor claridad a qué me refiero con ‘consciencia fenoménica’, lo cual incluirá motivar la idea de entenderla como presentación tipo E de qualia. Después de eso, explicaré qué son el fisicismo y el internismo fenoménico y qué los respalda. Por último, hablaré de por qué la consciencia fenoménica es problemática para el fisicismo. A grandes rasgos, la idea es que parece haber una desconexión tal entre nuestro entendimiento de lo físico y de la consciencia fenoménica que no sólo no es claro en qué sentido esta última es física, sino que incluso se justifica sostener que no lo es.

Tras haber planteado el problema a resolver, dedicaré los capítulos 2 y 3 a explorar las posibles soluciones. Más específicamente, el capítulo 2 estará enfocado en las dos únicas maneras de entender la presentación tipo E: como percatación directa (*acquaintance*) o como percatación indirecta (*representacional*). Eventualmente terminaré sosteniendo que lo más razonable es sostener que se trata de algún tipo de percatación indirecta, es decir,

algún tipo (E') de representación, porque parece no ser posible hacer sentido de la percatación directa en términos fisicistas. Ello me obligará, entonces, a dedicar una sección a mostrar en qué sentido la representación sí puede ser explicada en esos términos. Ahora, el identificar la presentación tipo E con un tipo de representación convertirá mi postura en un tipo de *representacionismo fuerte e irrestricto*, es decir, en una postura según la cual tener consciencia fenoménica es idéntico a tener ciertas propiedades representacionales: ciertos estados que tienen la peculiaridad de ser *acerca de* algo. Dentro de ese mismo capítulo revisaré qué tipos de representacionismos hay y concluiré algunas cosas sobre cuál es el más plausible.

Después de eso, dedicaré el capítulo 3 a estudiar las distintas teorías sobre la naturaleza de los qualia. En particular me centraré en tres opciones que son exhaustivas y mutuamente excluyentes: identificar a los qualia con *propiedades físicas externas ordinarias* (es decir, propiedades físicas externas, *categoricas, intrínsecas y fundamentales*), identificarlos con *propiedades físicas internas o no-ordinarias* o entenderlos como *propiedades no-físicas*. Eventualmente me inclinaré por la tercera opción, porque parece ser que ninguna propiedad física cumple con lo adecuado para ser identificada con qualia. Por un lado, todos los candidatos parecen ser disociables de los qualia bajo ciertas circunstancias y, por otra parte, las descripciones científicas de lo físico parecen ser insuficientes para capturar aspectos importantes de los qualia. Como ya mencioné, ello me llevará a negar la existencia de los qualia, porque deseo salvar al fisicismo. Hacia el final del capítulo recuperaré las ideas expuestas en el anterior y presentaré una taxonomía completa de posibles maneras de entender la presentación tipo E de qualia, es decir, de posibles teorías sobre la naturaleza de la consciencia fenoménica. Como ya debe ser evidente, me decantaré a favor de entenderla como *representación tipo E' de (ciertas) propiedades no-físicas*.

Por último, en el capítulo 4 ahondaré sobre esa conclusión, detallando mi propia postura *representacionista ilusionista*. Además de plantear y defender sus tesis principales, deberé hacer cuatro cosas. En primer lugar, tendré que aclarar qué son las posturas *ilusionistas* y

relacionarlas con la mía. En segundo lugar, deberé hablar sobre cómo naturalizar la representación tipo E' en términos físicos. En particular, deberé hacerlo respetando la tesis del internismo fenoménico y haciendo énfasis en cómo es posible representar objetos inexistentes con tanto poder de convencimiento de que de hecho sí existen. En tercer lugar, me veré obligado a ofrecer alguna explicación de por qué la naturaleza nos dotó de consciencia fenoménica, dado que básicamente se trata de una ilusión que nos hace creer que el mundo es como de hecho no es, lo cual a primera vista parece ser detrimental. Y finalmente, deberé mostrar cómo es que mi postura es capaz de refutar los principales argumentos en contra del fisicismo.

Capítulo 1

**El problema a resolver: la consciencia
fenoménica y el problema difícil**

1. El problema a resolver: la consciencia fenoménica y el problema difícil

Este primer capítulo está enfocado en presentar el problema a resolver: el de explicar la *consciencia fenoménica* en términos compatibles con el *fisicismo* y el *internismo fenoménico*. Con ello en mente, comienzo especificando a qué me refiero con ‘consciencia fenoménica’ (y otras nociones en la vecindad) (sección 1.1) y luego expongo qué son el *fisicismo* (sección 1.2) y el *internismo fenoménico* (sección 1.3), por qué son importantes, qué los respalda y qué constreñimientos añaden. Finalmente, señalo por qué la *consciencia fenoménica* es especialmente problemática para el *fisicismo* (sección 1.4).

A grandes rasgos, la *consciencia fenoménica* abarca todas las experiencias sensibles, es decir, todos los estados (o procesos¹) tales que para quien los tiene *es (o se siente) de algún modo* estar en ellos, desde su perspectiva en primera persona. Por otra parte, el *fisicismo* es la tesis de que el mundo real es enteramente físico, en el sentido de que todo lo que hay en él es enteramente explicable en términos de algo físico. La *consciencia fenoménica* es problemática para el *fisicismo* porque a la fecha nadie ha sido capaz de ofrecer una teoría transparente sobre cómo es que es basta con mecanismos físicos para generarla: si bien parece estar perfectamente correlacionada con determinados estados físicos, no es claro a qué se debe eso. Debido a ello, hay quienes sostienen que la *consciencia fenoménica* no es física y, por ende, que el *fisicismo* es falso. Lo que trataré de hacer en capítulos posteriores será luchar contra esa idea, ofreciendo una manera de acomodar la *consciencia fenoménica* dentro del panorama *fisicista*. Además, dicha tarea estará limitada por el *internismo fenoménico*, que es la tesis de que la *consciencia fenoménica* depende exclusivamente de factores internos del ente que la tiene, lo cual reducirá considerablemente el campo de explicaciones viables.

Comencemos, entonces, delimitando la cosa a explicar: la *consciencia fenoménica*.

1 En lo sucesivo deberá entenderse que siempre que hable de ‘estados’ estaré incluyendo procesos.

1.1. Consciencia fenoménica

Etimológicamente, la expresión ‘consciencia fenoménica’ refiere al mundo de las *apariencias* o de las *sensaciones*²: a cómo es *para nosotros* el mundo desde nuestro punto de vista *subjetivo* o *en primera persona*³. Teniendo eso en cuenta, es natural entender que la consciencia fenoménica es la conjunción de todas nuestras *experiencias sensibles* (o *fenoménicas*), cada una de las cuales nos presenta una porción de la apariencia que para nosotros tiene el mundo. Por ejemplo, mi experiencia al ver una manzana me presenta la apariencia visual que esa manzana tiene para mí, mi experiencia al oler una taza de café me presenta la apariencia olfativa que ese café tiene para mí y mi experiencia al escuchar una melodía me presenta la apariencia auditiva que esa melodía tiene para mí. Además de las experiencias senso-perceptuales (incluyendo a los sueños), ejemplos paradigmáticos de experiencias fenoménicas son las emociones y cualquier tipo de sensación corporal, como dolores, orgasmos y náuseas. Recurriendo a la formulación popularizada por Nagel (1974) (con antecedente en Farrell (1950)), la consciencia fenoménica abarca cualquier estado tal que *para el sujeto* haya algo tal como estar en él⁴. Otra manera de expresar la misma idea es diciendo que las experiencias fenoménicas son todos los estados tales que *se siente* de alguna manera estar en ellos⁵.

2 ‘Fenoménico’ se deriva de ‘fenómeno’, que etimológicamente proviene del latín ‘phaenomenon’, que a su vez viene del griego ‘φαινόμενον’ (phainómenon), que refiere a lo que se muestra o a la apariencia. Es decir, lo fenoménico es todo lo relativo a las propiedades aparentes de las cosas: a cómo son para quien las percibe, en contraste con cómo son en realidad. Por otra parte, ‘consciencia’ viene del latín ‘conscientia’, que a su vez proviene de ‘con’, que significa unión, junto o con, y ‘scientia’, que se deriva de ‘scire’, que significa saber o discernir, y el sufijo ‘ia’, que indica que se trata de una cualidad. Es decir, la palabra completa significa algo así como tener una unión de saberes o conocimientos o discernir un conjunto de cosas, que básicamente es lo mismo que percatarse de ellas. Si además se toma en cuenta que lo fenoménico refiere originalmente a las apariencias, tiene sentido entender a la consciencia fenoménica como un tipo de percatación de apariencias o como una percatación de algo más pero mediada por apariencias. Ver Van Gulick (2004/14) para un tratamiento más amplio del tema de la consciencia.

3 Para distintas maneras de entender el factor subjetivo, ver Guillot (2017).

4 En ese sentido, incluso es posible contar como experiencias a procesos cognitivos como pensar, imaginar o desear. Para discusión al respecto, ver Siewert (1998, 2011), Horgan & Tienson (2002), Pitt (2004, 2011), Strawson (2004, 2011), Horgan (2011), Kriegel (2011, 2015a, 2015b), Bayne & Montague (2011) y Chudnoff (2015)).

5 No es claro cuál es la mejor forma de traducir la expresión ‘what it is like’ de Nagel (1974). Farrell (2016) señala que es una expresión algo obscura y sugiere no tomarla como término técnico.

Sobre esa línea, cuando se habla del *carácter fenoménico* de una experiencia, lo que se tiene en mente es cómo es la apariencia presentada por ella. Alternativamente, es posible entender la apariencia en cuestión no como la apariencia de alguna otra porción del mundo, sino como la apariencia de la experiencia misma: en ese sentido, el carácter fenoménico es la apariencia que para el sujeto tiene cada experiencia. A su vez, es natural entender cada apariencia como una determinada composición de un especial tipo de propiedades, exclusivas del tipo de experiencia en cuestión: visuales, auditivas, de dolor, etcétera. Por ejemplo, es natural entender que el carácter fenoménico de una experiencia visual es cierta distribución espacial (o espacio-temporal) de ciertas propiedades visuales que podemos estar tentados a llamar ‘colores’. En lo sucesivo llamaré ‘qualia’ (o ‘quale’, en singular)⁶ a las propiedades que componen al carácter fenoménico, independientemente de cuál sea su naturaleza última y de qué tipo de experiencia se trate: sencillamente son las propiedades que experimentamos de primera mano durante nuestras experiencias fenoménicas (qualia visuales, incluyendo los de color, qualia auditivos, qualia de dolor, etcétera).

Cabe insistir en que todo lo anterior pretende ser neutral respecto a qué son en sí los qualia. En particular, es neutral respecto a si los qualia son propiedades de la experiencia o de alguna porción del mundo presentada por ella, así como sobre si son algo interno o externo o si son físicas o no. Sin embargo, por más neutral que me quiera mantener al respecto, una intuición común es que los qualia *parecen* ser un tipo *fundamental* de objetos⁷, en el sentido de que en principio no parecen ser explicables en términos de alguna otra cosa. Es decir, parece imposible describirlos exhaustivamente sin aludir a más qualia. Por ejemplo, parece imposible explicar exhaustivamente cómo es el quale que caracteriza a un determinado tono de verde sin aludir a más qualia de color. Además, también es

6 No hay consenso sobre cómo llamar a las propiedades en cuestión, pero ‘qualia’ es uno de los términos más usados. Otros nombres recurrentes son ‘propiedades fenoménicas’ o ‘propiedades cualitativas’. Tye (1994-2017) ofrece un estudio detallado de distintas acepciones del término ‘qualia’ y posibles naturalezas de los qualia, sobre las cuales ahondaré en el capítulo 3. Por otra parte, autores como Shoemaker (1994b) y Egan (2006) llaman ‘propiedades aparentes’ a cosas como la propiedad de causar qualia o tener la disposición de causarme qualia.

7 Siempre emplearé el término ‘objeto’ en su acepción más amplia, que abarca todo lo que puede ser designado por un sustantivo: desde lo más concreto hasta lo más abstracto y desde lo más particular hasta lo más general, incluyendo objetos concretos, propiedades, particulares, conjuntos, estructuras, sistemas, eventos, procesos, hechos, situaciones, etcétera. Además, emplearé ‘cosa’ y ‘entidad’ como sinónimos de ‘objeto’.

comúnmente aceptado que los qualia parecen ser propiedades *categoricas*, en el sentido que no son *disposicionales*, bajo el entendido de que las propiedades disposicionales son meras capacidades o propensiones⁸. Por ejemplo, el agua tiene la propiedad disposicional de hervir cuando la temperatura rebasa cierto límite (punto de ebullición) y el vidrio tiene la propiedad disposicional de romperse con relativa facilidad (fragilidad). En contraste, las propiedades categoricas son algo que simplemente está “ahí” presente, constituyendo al objeto que las posee y, si acaso, sirviendo como la base concreta que habilita sus capacidades o propensiones. Por ejemplo, es razonable pensar que tanto el punto de ebullición del agua como la fragilidad del vidrio se deben a cómo son sus respectivos componentes y estructura interna, que simplemente están presentes en el agua y el vidrio, constituyéndolos. Entonces la idea es que los qualia parecen no ser disposiciones, sino algo que simplemente está presente de la manera más concreta posible. Por último, ello se conecta con que los qualia parecen ser propiedades *intrínsecas*, en el sentido de que su posesión parece no depender de algo externo al objeto que las posee⁹. Todas estas ideas son cuestionables (como exploraré en el capítulo 3), pero tenerlas en cuenta desde ahora servirá para entender mejor los problemas que la consciencia fenoménica genera.

Ahora, muchos autores (tal vez la mayoría) identifican la consciencia fenoménica con los qualia mismos (o con estados que poseen qualia), pasando por alto u obviando el otro elemento central: el tipo de acceso que tenemos hacia ellos durante nuestras experiencias. Sin embargo, lo cierto es que ese elemento es tan importante que incluso se justifica decir que no puede haber consciencia fenoménica sin él¹⁰. Es decir, debería ser bastante claro que

8 Para un tratamiento más profundo sobre propiedades categoricas y disposicionales, ver Choi & Fara (2006/18) y Orilia & Paolini Paoletti (1999/2020).

9 En la sección 1.3 ahondo en la noción de *propiedad intrínseca*. Ver Marshall & Weatherson (2002/12) para un tratamiento general sobre el tema.

10 Cabe distinguir el tipo de acceso al que me refiero aquí de lo que Block (1995) denomina ‘consciencia de acceso’, que consiste en que cierta información esté disponible para el reporte, la formación de creencias y el control racional de la acción. La razón es que sí es posible ofrecer razones para pensar que la consciencia de acceso es dissociable de la fenoménica (ver también Block (2007) y Sebastián (2013)), pero nadie niega que tener consciencia fenoménica involucra percatarse de qualia. También hay versiones menos exigentes y más cercanas a lo que quiero señalar, que se limitan a hablar de que la información en cuestión esté debidamente posicionada para afectar la toma de decisiones de un organismo (ver Kirk (1994), Dretske (1995) y Tye (1995, 2000), por ejemplo). El punto es que hay un sentido mínimo en el que si un quale es completamente inaccesible para el sujeto, entonces no cuenta como consciencia en ningún sentido.

tener una experiencia no consiste meramente en que haya qualia (o tal vez ni siquiera lo requiere), sino que lo realmente importante es que estos se presenten de algún modo ante el sujeto, habilitando (o constituyendo) algún tipo de percatación de ellos. En otras palabras, no tendría sentido llamar ‘experiencia’ a cierta composición de qualia si nadie la está *experimentando*. Teniendo eso en cuenta, podemos reflexionar sobre el acceso que tenemos hacia los qualia durante nuestras experiencias y notar que es bastante peculiar: los qualia sencillamente parecen presentarse directa, transparente y vívidamente ante nosotros, sin ninguna clase de mediador y revelando su esencia¹¹. Ello, a su vez, parece situarnos en una posición privilegiada para adquirir conocimientos sobre los qualia presentados en nuestras propias experiencias, en el sentido de que parece no dejar espacio para que podamos dudar de su presencia, de que es a nosotros a quienes se le presentan, ni de que son de cierto modo. Por ejemplo, cuando tengo una experiencia visual, tengo la impresión de tener un acceso directo y privilegiado hacia los qualia que esta me presenta (incluyendo los de color): sencillamente se revelan ante mí tal cual son, sin filtros. Por eso a primera vista parece incuestionable que están ante mí, porque mi acceso hacia ellos parece tan directo (o inmediato) que no deja espacio para que haya error.

En lo sucesivo llamaré ‘presentación tipo E’ (por ‘experiencial’) a la manera en que los qualia nos son presentados durante nuestras experiencias y resumiré las características mencionadas en el párrafo anterior meramente diciendo que la presentación tipo E de *x* parece ser una percatación *directa* y, por ende, *infallible* de *x*. Sin embargo, es crucial notar que ello es compatible tanto con que de hecho sea directa e infalible como con que meramente lo parezca, sin serlo, por lo que podemos decir que hay dos posibles versiones de presentación tipo E en juego: una directa e infalible y otra que, aunque también parece ser directa e infalible, de hecho es indirecta y tal vez hasta falible¹², en el sentido de que bien podría resultar que los qualia presentados por ella de hecho no estén presentes o no

11 Lo que presento en este párrafo está basado en los recuentos sobre intuiciones comunes al respecto que hacen Balog (2009) y Chalmers (2003, 2018). Una buena explicación de la consciencia fenoménica debería dar cuenta de esas intuiciones, sean o no acertadas.

12 Por esa razón, tal vez sería mejor meramente hablar de que *s* tiene la impresión de que *x* o de que *s* experimenta *x*, en lugar de decir que *s* se percata de *x*, aunque eso podría invitar a otras confusiones. Para desambiguación entre distintas acepciones de percatación, ver Sebastián (2018, 2022).

sean como parecen ser. Siguiendo a autores como Russell (1905, 1910/11), Fumerton (1995, 2001, 2016), Bonjour (2001, 2003) y Chalmers (2003, 2010), llamaré ‘*acquaintance*’¹³ a la versión directa e infalible. Por otra parte, entenderé que la percatación indirecta es un tipo de *representación*, en el sentido de que es habilitada por un objeto intermediario (z) entre el ente que se percata (s) y aquello de lo que se percata (x): en ese sentido decimos que z *representa* a (o es *acerca de*) x, porque es en virtud de la presencia de z que s se percata de x. La falibilidad de la percatación representacional se deriva de que usualmente z puede existir en ausencia de x, en cuyo caso la percatación de x puede ser ilusoria¹⁴. En contraste, el *acquaintance* con x necesariamente requiere de la presencia de x y, por ende, es infalible. Por último, cabe señalar que el que la presentación tipo E de qualia sea *acquaintance* con qualia es compatible con que también sea representación de alguna otra cosa. Por ejemplo, Russell (1910/11) y Chalmers (2004, 2006c) la entienden como *acquaintance* con qualia, pero sostienen que también es representación de alguna otra porción del mundo, porque los qualia son la apariencia que esta tiene para nosotros.

Otra supuesta característica de la presentación tipo E es que es privada, en el sentido de que cada individuo parece tener ese tipo de percatación sólo de los qualia que se presentan (de modo E) en sus propias experiencias fenoménicas: nadie puede entrar a la cabeza de alguien más y averiguar directamente qué experiencias tiene (y ese es otro sentido en el que la presentación tipo E es privilegiada). Una consecuencia de dicha privacidad y de la aparente fundamentalidad de los qualia es que para saber cómo es una determinada experiencia (y los qualia presentados en ella) parece no bastar con que alguien la platique, sino que es preciso “vivirla en carne propia”. Por ejemplo, imaginemos que tuviéramos que explicarle cómo es ver verde a alguien ciego de nacimiento: parece que nada que podamos decir bastaría para realmente hacerle entender cómo es, sino que la única ruta realmente efectiva sería que dicho alguien viera algo verde (o al menos algunos colores en la cercanía). Nos encontramos en una situación parecida si tratamos de imaginar cómo se sentiría percibir mediante un sentido que no tenemos, como la ecolocalización de los murciélagos: no sólo no sabemos cómo sería, sino que tal vez ni siquiera se parece a nada

13 En español no hay una expresión que capture adecuadamente la noción de *acquaintance*.

14 Para más sobre representación, ver las secciones 2.3.1 y 2.3.3.

que conozcamos (ver Nagel (1974)). Ahora, yo no cuestionaré que la presentación tipo E de hecho es privada, porque sencillamente se la tiene o no se la tiene, pero sí pondré en duda la idea de que no es posible dar cuenta de ella o de los qualia en términos de otra cosa que sí es accesible desde fuera (capítulo 4).

Por último, es razonable pensar que la presentación tipo E de qualia y los qualia mismos (o los acomodos de ellos) agotan todo lo que hay que decir sobre la consciencia fenoménica: por más que busquemos, parece no haber en ella algo además de los qualia que la caracterizan y la manera en que se presentan ante el sujeto. Así que es razonable pensar que no es necesario aludir a algo más para definirla. Entonces no sólo se justifica decir que la consciencia fenoménica involucra presentación tipo E de qualia, sino incluso sostener que eso es todo lo que es. Es decir, es razonable adoptar la siguiente definición:¹⁵

CF) La consciencia fenoménica es idéntica a presentación tipo E de qualia.

Por ejemplo, que un sujeto tenga una determinada experiencia visual es idéntico a que cierta composición de qualia visuales le sea presentada de modo E, que tenga una determinada experiencia olfativa es idéntico a que cierta composición de qualia olfativos le sea presentada de modo E y que tenga un determinado dolor es idéntico a que cierta composición de qualia de dolor le sea presentada de modo E.

Ahora, es crucial dejar en claro que CF es neutral respecto a los detalles realmente importantes sobre la naturaleza de la consciencia fenoménica. Por ejemplo, es neutral respecto a qué son en sí los qualia y la presentación tipo E. En el caso de los qualia, CF meramente los trata como lo que sea que se presenta de primera mano ante el sujeto durante sus experiencias, sin compromisos más profundos sobre su naturaleza (aunque ya mencioné que en principio parecen ser propiedades categóricas, intrínsecas y

15 En el fondo, este razonamiento tiene básicamente la misma estructura que el *argumento del parecer*, presentado por Byrne (2001) y Thau (2002) (ver sección 2.3.2), con la diferencia de que ellos presuponen que la presentación tipo E es representacional, mientras que yo lo estoy dejando abierto (al menos por ahora).

fundamentales). En cuanto a la presentación tipo E, sólo la trata como la manera en que los qualia se presentan ante el sujeto durante sus experiencias, misma que *parece* tener ciertas características, como ser directa e infalible, pero dejando abierta la posibilidad de que de hecho no las tenga. Es decir, CF es neutral respecto a si la presentación tipo E es *acquaintance* o algo que meramente lo parece.

También es muy importante señalar que CF es neutral sobre la relación entre los qualia y la presentación tipo E. En particular, es neutral sobre si los qualia pueden o no existir sin presentarse de modo E y sobre si es posible que algo distinto de qualia se presente de modo E. Ello, aunado a la neutralidad sobre si la presentación tipo E es infalible o no, conduce a concluir que CF es neutral sobre si la presentación tipo E de qualia es o no condición necesaria y/o suficiente para la presencia de los qualia en cuestión. Siendo ese el caso, incluso es posible que CF se mezcle o hasta colapse con la idea de que la consciencia fenoménica son los qualia mismos, dependiendo de qué se sostenga sobre la relación entre qualia y presentación tipo E. Chalmers (1996, 2002a, 2003, 2005, 2006c, 2018), por ejemplo, defiende una postura en la que ambas ideas colapsan, porque para él tener consciencia fenoménica es poseer qualia, pero sostiene que eso es suficiente (y necesario) para tener *acquaintance* con ellos. Para conservarme neutral al respecto, cuando hable de ‘lo fenoménico’, me referiré a la disyunción entre qualia y presentación tipo E de qualia.

Una vez habiendo dejado suficientemente en claro a qué me referiré con ‘consciencia fenoménica’ y ‘fenoménico’, llega el turno de presentar dos tesis fundamentales que deseo respetar en este trabajo: el *fisicismo* y el *internismo fenoménico*. Tener como guía esas dos posturas servirá para descartar algunas posibles explicaciones de la consciencia fenoménica.

1.2. Fisicismo

El *fisicismo* (o *materialismo* o *fisicalismo*¹⁶) (respecto al mundo real) es la tesis de que el

16 Aquí ignoraré las sutiles distinciones que conducen a cada término.

mundo real es enteramente físico, en el sentido de que todo lo que hay en él puede ser explicado sin aludir a algo que no sea físico: todo lo que hay en el mundo real tiene una base física que por sí sola es *suficiente* para constituirlo o generarlo, de modo que tomar en cuenta meramente la parte física del mundo real es tomar en cuenta su totalidad. En otras palabras, el fisicismo es la tesis *monista* según la cual los componentes fundamentales del mundo real son físicos y el resto se deriva de ellos. La principal razón a favor de esa postura es que no se ha encontrado nada en el mundo real cuya presencia no se encuentre perfectamente correlacionada con la de eventos físicos específicos. De hecho, el fisicismo es la postura estándar dentro de la ciencia, aunque no siempre explícitamente, y goza de gran aceptación en el medio filosófico. Por eso es que es bien visto *naturalizar* las teorías filosóficas, bajo el entendido de que ello consiste en explicarlas en términos compatibles con el fisicismo (y con la ciencia)¹⁷. De manera muy general, el fisicismo puede ser definido del siguiente modo (ver Stoljar (2010)):

FIS) El fisicismo respecto a x es verdad si y sólo si todas las entidades que conforman a x son entidades físicas fundamentales o se encuentran en una relación de dependencia D respecto a entidades físicas fundamentales.

Ello bajo el entendido de que la relación de dependencia D refiere al sentido en el que todo lo que hay en $@$ se deriva de su parte física. Sin embargo, definir puntualmente esa relación es todo un reto y en la literatura abundan candidatos que de un modo u otro han probado ser defectuosos¹⁸. Por eso es que yo no intentaré hacer esa tarea en este escrito,

17 Naturalizar algo es explicarlo exclusivamente en términos de lo que existe en el mundo natural. Dado que la ciencia trata justamente de estudiar cómo es el mundo natural, naturalizar algo es cercano a explicarlo en términos científicos. Además, debido a que la ciencia actual entiende todo en términos físicos (y matemáticos), es común que se entienda que naturalizar algo equivale a explicarlo en términos meramente físicos (y matemáticos). Sin embargo, es importante tener en cuenta que podría haber escenarios en los que las tres nociones (lo natural, lo científico y lo físico) se separen. Por ejemplo, bien podría resultar que en la naturaleza hubiese algo más allá de lo físico, en cuyo caso naturalizar algo no necesariamente involucraría explicarlo exclusivamente en términos físicos. Además, el elemento no-físico podría ser inaccesible para la ciencia, de modo que esta se separaría del naturalismo. Pero si el fisicismo es verdadero, las tres cosas sí terminan siendo lo mismo, por lo que no remarcaré la posibilidad de que diverjan.

18 Para un repaso de las principales opciones, ver Stoljar (2001/15, 2010). En general, él ofrece un estudio mucho más extenso y detallado del fisicismo y sus tipos.

sino que meramente me enfocaré en un rasgo mínimo que debe de cumplirse: que todo lo que hay en el mundo real (al cual llamaré '@') sea consecuencia (*metafísicamente necesaria*) de que la parte física de @ sea como es. Entonces la correlación entre lo físico y cualquier aspecto de @ no es un mero accidente propio de @, sino que debe cumplirse en todo *escenario* (y todo *mundo*) *metafísicamente posible*¹⁹. Más específicamente, todo físico debe comprometerse con que duplicar exclusivamente la parte física de @ *necesariamente* implica duplicar la totalidad de @, lo cual suele expresarse diciendo que todo duplicado físico *mínimo* de @ (es decir, exclusivamente de su parte física) es un duplicado *simpliciter* de @ (es decir, un duplicado a secas) (Jackson (1993)). Si P^x es un conjunto de proposiciones tal que de él se sigue necesariamente toda proposición ejemplificada por la parte física de x ²⁰, * indica que no hay adiciones de ningún tipo (de modo que todo lo que cumple con $P^{@*}$ es un duplicado físico *mínimo* de @) y T^x es un conjunto de proposiciones tal que de él se sigue necesariamente toda proposición ejemplificada por x , la idea es que todo físico debe (como mínimo) comprometerse con que $P^{@*}$ necesariamente implica $T^@$ (es decir, $\Box(P^{@*} \rightarrow T^@)$)²¹.

La razón por la que se hace énfasis en que el duplicado físico de @ es mínimo, en contraste con simplemente hablar de que todo duplicado físico de @ es un duplicado simpliciter de @ (es decir, que $\Box(P^@ \rightarrow T^@)$)²², es que esa es una tesis más fuerte (porque implica a la de que $\Box(P^{@*} \rightarrow T^@)$) con la que el físico no necesita comprometerse. Esto último porque bien podría ser falso que $\Box(P^@ \rightarrow T^@)$ y verdadero que $\Box(P^{@*} \rightarrow T^@)$, sin que

19 En este escrito entenderé que algo es *metafísicamente* posible si y sólo si es congruente con todas las leyes lógicas y matemáticas clásicas. Por otra parte, entenderé que los *escenarios metafísicamente posibles* son porciones de *mundos metafísicamente posibles* y estos, a su vez, son totalidades coherentes: para cualquier proposición (o hecho o estado de cosas), cuadran con ella o con su negación y nunca con ambas. De ahí en fuera, no me comprometo con ninguna teoría en particular sobre qué son en sí los mundos posibles. Bien podría tomárselos como objetos concretos, abstracciones o meros contrafácticos. Para un estudio general sobre mundos posibles, ver Menzel (2013/16).

20 Excluyendo referencias a que lo que las ejemplifica es x y cosas por el estilo, porque de otro modo sólo x podría ejemplificar las proposiciones en cuestión. Y lo mismo para T^x .

21 Todo esto suele ser explicado en términos de *superveniencia*: $T^@$ superviene sobre (o en) $P^{@*}$ (ver Stojar (2001/15)). Pero esa noción puede entenderse de varias maneras y ello invita a confusión. Además, la acepción de superveniencia que me interesa es idéntica a la implicación necesaria, así que no hay necesidad de introducir nuevos términos. Para un estudio general sobre superveniencia, ver McLaughlin & Bennett (2005/18).

22 Originalmente Lewis (1986) propuso hablar de que $\Box(P^@ \rightarrow T^@)$ y luego fue corregido por Jackson (1993), quien habló de que $\Box(P^{@*} \rightarrow T^@)$.

ello ponga en peligro al fisicismo. Por ejemplo, podríamos pensar en un duplicado físico de @ en el que además existe algún objeto b (no-físico) tal que pudiese interferir con las consecuencias de $P^@$, de modo que $P^@ \& T^b$ condujese a $\neg T^@$ y, por ende, a la negación de $\Box(P^@ \rightarrow T^@)$ (ver Horgan (1983) y Lewis (1983b)). Pero b sería irrelevante cuando se habla de $P^{@*}$, porque, por definición, $P^{@*}$ es incompatible con T^b , porque b no es físico. Es decir, si bien $P^@ \& T^b \& \neg T^@$ podría ser consistente, $P^{@*} \& T^b \& \neg T^@$ no lo es, por lo que la posibilidad metafísica de T^b (y b) sólo haría peligrar a la tesis de que $\Box(P^@ \rightarrow T^@)$, dejando intacta a la de que $\Box(P^{@*} \rightarrow T^@)$. Finalmente, la razón por la que al físico (sobre @) sólo le interesa sostener que $\Box(P^{@*} \rightarrow T^@)$ y no que $\Box(P^@ \rightarrow T^@)$ es que no tiene por qué negar la posibilidad metafísica de que existan cosas como b, sino sólo negar que las hay en @.

Por otra parte, sostener que $\Box(P^{@*} \rightarrow T^@)$ es distinto de decir que todo lo que hay en @ se *reduce* a algo físico. Ello porque el que $T^@$ se *reduzca* a $P^{@*}$ quiere decir que $T^@$ no sólo es realizado (es decir, constituido, instanciado o generado) por la parte física de @, sino que sólo puede ser realizado de esa forma, pero el físico no requiere negar que algunas cosas que hay en @ podrían ser realizadas de manera no-física (en otros mundos), sino que sólo necesita comprometerse con que todo lo que hay en @ es realizado por algo enteramente físico en @. Por ejemplo, el que sea posible generar una mente como una en @ mediante mecanismos no-físicos es compatible con que baste con mecanismos físicos para generarla y con que todas las mentes en @ sean generadas mediante mecanismos físicos. En otras palabras, el físico no necesita comprometerse con que $T^@$ (ni $T^{@*}$) necesariamente implique $P^{@*}$ (ni $P^@$) (es decir, con que $\Box(T^@ \rightarrow P^{@*})$ o alguna tesis en la cercanía).

Ahora, si todo físico debe comprometerse con que $\Box(P^{@*} \rightarrow T^@)$, se sigue que el fisicismo sería falso si fuese metafísicamente posible un duplicado físico mínimo de @ que no fuese un duplicado de @ en todo sentido (es decir, si $\Diamond(P^{@*} \& \neg T^@)$). Para ello basta con que haya al menos un hecho en @ que no se siga necesariamente de los hechos sobre su parte física, porque entonces no bastaría con duplicar lo físico para duplicar dicho hecho, lo cual llevaría a que no fuese suficiente duplicar lo físico para duplicar la totalidad. Es decir, dado que para toda porción k de @ se cumple que $\neg T^k$ necesariamente implica $\neg T^@$ (es decir,

$\Box(\neg T^k \rightarrow \neg T^@)$), el físico está obligado a sostener que necesariamente $P^{@*}$ implica T^k (es decir, $\Box(P^{@*} \rightarrow T^k)$) para toda k , lo cual hace que para refutar al fisicismo baste con mostrar que es metafísicamente posible $P^{@*} \& \neg T^k$ (es decir, que $\Diamond(P^{@*} \& \neg T^k)$) para alguna k . Por eso, una manera tradicional de atacar al fisicismo es señalar que hay alguna porción de @ cuya presencia no se sigue necesariamente de la de la parte física: entonces esa parte de @ no es explicada por la física y, en ese sentido, no es física.

Sin embargo, es importante señalar que el fisicismo no es propiamente una tesis modal. Siguiendo a Stoljar (2001/15), para mostrarlo basta con señalar que alguien podría sostener que la presencia de ciertos objetos paradigmáticamente no-físicos (d), como Dios, es una verdad necesaria (T^d). Entonces sería verdad que su presencia sería necesariamente implicada por $P^{@*}$ y también por $T^@$ (es decir, $\Box(P^{@*} \rightarrow T^d)$ y $\Box(T^@ \rightarrow T^d)$), con lo cual la verdad de T^d (en @) sería compatible con la tesis de que $\Box(P^{@*} \rightarrow T^@)$ a pesar de ser incompatible con el fisicismo, porque éste querría decir algo más: que D no es el caso en @. Por eso es que el fisicismo no queda agotado por la tesis modal de que $\Box(P^{@*} \rightarrow T^@)$. Para realmente redondear la tesis física, habría que especificar qué tipo de cosas no pueden contar como físicas en ningún sentido²³ y sostener que ninguna de ellas existe en @. Alternativamente, podría especificarse con más detalle la relación de dependencia D, de manera que sea claro qué tipo de objetos pueden depender de ese modo de lo físico.

Suponiendo que ya se ha definido qué tipo de cosas son claramente no-físicas (incluyendo las que no están en relación de dependencia D con lo físico), otra manera de atacar al fisicismo consistiría en simplemente señalar que en @ hay algo claramente no-físico, en el sentido de que estamos tan seguros de que no es físico que no lo contaríamos como tal aun si descubriésemos que de hecho su presencia se sigue necesariamente de la de algo exclusivamente físico. Por ejemplo, si resultase que existe la magia, el *élan vital* o algún dios (en algún sentido tradicional), probablemente insistiríamos en que para dar cuenta de ello habría que hacer algo más que hablar meramente de lo físico, incluso si su presencia fuese una verdad necesaria o estuviese perfecta y necesariamente correlacionada con la de

23 Por ejemplo, Montero & Papineau (2005) y Wilson (2006) han sugerido definir lo físico como todo lo que no es mental o no fundamentalmente mental.

eventos físicos específicos. Nótese que esa estrategia de hecho está íntimamente ligada con la de encontrar una k tal que es falso que $\Box(P^{@*} \rightarrow T^k)$, porque es razonable pensar que muchas de las cosas que definitivamente no son físicas (si no es que todas) son tales que su presencia no se sigue necesariamente de $P^{@*}$ (ni de nada físico) y que el que la presencia de algo en $@$ no se siga necesariamente de $P^{@*}$ es un muy buen indicio de que no es físico, de modo que el que algo cumpla con una de las dos condiciones es un muy buen indicio de que cumple con la otra. Por eso es que, a pesar de que presentaré los argumentos en contra del fisicismo en términos de la primera estrategia, bien pueden ser traducidos a la segunda.

1.2.1. Ficismo respecto a la consciencia fenoménica

La parte de $@$ más aludida por los enemigos del fisicismo, tanto para rechazar $\Box(P^{@*} \rightarrow T^{@})$ como para mostrar directamente que hay algo no-físico, es la mente (que incluye a la consciencia fenoménica), ya que es difícil explicar en qué sentido es física o por qué es que su presencia es necesariamente implicada por la de meros eventos físicos, como debe sostener el fisicista. En particular, no es claro cómo la presencia de consciencia fenoménica se sigue de la de meros acomodos físicos. Así que se suele atacar al fisicismo sosteniendo que es metafísicamente posible un duplicado físico mínimo de $@$ que difiera de $@$ respecto a su parte fenoménica. Es decir, dado que el fisicista está comprometido con que $P^{@*}$ necesariamente implica $Q^{@}$ (es decir, con que $\Box(P^{@*} \rightarrow Q^{@})$), donde $Q^{@}$ es un conjunto de proposiciones tal que de él se sigue necesariamente toda proposición ejemplificada por la parte fenoménica de $@$, es decir, de qué experiencias fenoménicas tiene cada ente y qué qualia hay a cada momento en $@$, se suele tratar de refutar al fisicismo mostrando que es metafísicamente posible $P^{@*}$ sin $Q^{@}$ (es decir, que $\Diamond(P^{@*} \& \neg Q^{@})$).

Sin embargo, el hecho es que toda la evidencia indica que, en $@$, la consciencia fenoménica (y la aparente presencia de qualia) en efecto se encuentra perfectamente correlacionada con eventos físicos: más específicamente, con estados cerebrales (ver Pautz (2014)). Y, dado que es razonable pensar que una correlación tan compleja y estrecha no

puede ser un mero accidente propio de @, sino que debe desprenderse de la naturaleza íntima de los elementos involucrados (en este caso, lo físico y lo fenoménico), es natural concluir que $Q^@$ se sigue necesariamente de $P^{@*}$. En otras palabras, podemos pasar de la observación de que esas regularidades se cumplen en @ a que deben cumplirse en todo mundo metafísicamente posible, porque, si en realidad fuesen contingentes, sería demasiada casualidad estar en un mundo en el que se cumplen a la perfección. Es decir, si en el fondo no hubiese nada que impidiese irregularidades en la correlación entre lo físico y lo fenoménico, sería demasiada casualidad estar justo en un mundo en el que no las hay²⁴. Así que, dado que lo más razonable es sostener que nuestro mundo no es especial en ese sentido, lo más razonable es concluir que de hecho hay algo que impide que se violen las regularidades en cuestión. Por ende, lo más razonable es concluir que esas regularidades deben cumplirse en todo mundo posible, es decir, que $\Box(P^{@*} \rightarrow Q^@)$.

Por otra parte, usualmente reconocemos cierto poder causal de la consciencia fenoménica hacia lo físico, dado que aceptamos que nuestras experiencias subjetivas (y la supuesta presencia de qualia) influyen en nuestro comportamiento: en general, nos apoyamos en nuestras experiencias fenoménicas para navegar el mundo. Pero a la vez sabemos lo suficiente sobre cómo funciona el cuerpo a nivel físico como para concluir que todo nuestro comportamiento puede ser explicado sin recurrir a algo más. Así que, en la medida en que la consciencia fenoménica afecta nuestro comportamiento, es razonable creer que equivale a alguno de los mecanismos físicos involucrados, porque de lo contrario no todo

24 Como se verá en la sección 1.4, esta es una razón relativamente débil, porque el que en @ haya una correlación sistemática entre dos cosas, por más compleja que sea, estrictamente no basta para asegurar que en otros mundos también deba haberla. De hecho, el tipo de razonamiento en cuestión es rechazado por quienes sostienen posturas humeanas respecto a las leyes naturales, según las cuales estas no son otra cosa que descripciones generales de hechos contingentes y por ende son también contingentes y podrían ser distintas en mundos posibles distintos (ver Lewis (1986b)). La idea detrás de ello (Hume (1748)) es que, como no vemos directamente vínculos tales como la causalidad, sino sólo hechos sueltos, uno tras otro, no podemos asegurar que no podrían sucederse de otro modo, así que no podemos asegurar que hay leyes en un sentido más fuerte que meras regularidades contingentes, a pesar de estar habituados a pensar y actuar como si fuese de otro modo. Ahora, el humeano va mucho más allá de eso y afirma que no sólo no podemos asegurar que hay leyes naturales necesarias, sino que podemos asegurar que no las hay. Si bien parece una gran casualidad que nuestro mundo sea tal que en él se cumplen ciertas regularidades a la perfección, lo cierto es que en algún mundo debían cumplirse y era de esperarse que sus habitantes se sintiesen sorprendidos por ello. Para discusión general en torno a leyes naturales, ver Carroll (2016). Para discusión general sobre causalidad, ver Schaffer (2003/16).

nuestro comportamiento sería determinado por lo físico o estaríamos hablando de que dos cosas independientes causan entera y simultáneamente una tercera, lo cual no parece tener mucho sentido (ver Kim (1993), Yablo (1992) y Bennett (2003)). Y si quisieramos apegarnos a que los eventos en cuestión tienen una causa enteramente física y que la consciencia fenoménica no es física, nos veríamos obligados a decir que las experiencias fenoménicas son una causa superflua o ni siquiera tienen poderes causales, lo cual es poco deseable. Ello se relaciona también con el principio de *clausura causal de lo físico*, que dicta que toda causa de un evento físico es a su vez física (ver Davidson (1970)): si la consciencia fenoménica tiene efecto sobre lo físico, debe ser físico en algún sentido. Podemos llegar a la misma conclusión si extendemos el principio de la clausura diciendo que nada físico puede tener consecuencias no físicas y señalamos que en efecto los cambios físicos conducen a cambios fenoménicos. En síntesis, si la consciencia fenoménica no fuese física, no afectaría ni sería afectada por lo físico. Pero el hecho es que existe una correlación tan perfecta entre ambas cosas que parece inviable negar que se afectan mutuamente. Se sigue, entonces, que la consciencia fenoménica sí es física en algún sentido.²⁵

Por último, cabe señalar que la correlación parece ser entre los estados físicos de cada sujeto y sus experiencias fenoménicas, por lo que es razonable sostener que para duplicar las experiencias de un sujeto en particular no se requiere duplicar todo $P^{@*}$, sino que basta con duplicar los estados físicos de dicho sujeto. Es decir, para cada sujeto s es razonable sostener que $\Box(P^{s*} \rightarrow Q^s)$. Incluso es posible acotar la tesis a experiencias particulares y estados físicos al momento de tenerlas. Además, puede bastar con duplicar sólo cierta porción de la parte física del sujeto y/o sólo cierto tipo de propiedades físicas suyas. En particular, a mí me interesa defender que basta con duplicar sus propiedades físicas *intrínsecas*, lo cual equivale a sostener el fisicismo aunado a una tesis conocida como ‘internismo fenoménico’. Así que dedicaré la siguiente sección a estudiar esa idea.

25 Para más sobre estos argumentos, ver Stoljar (2001/15).

1.3. Internismo fenoménico

En general, entenderé que ser *internista* respecto a una propiedad *x* en relación con un objeto *z* (en un tiempo *t*) es sostener que *x* es una propiedad *intrínseca* de *z* (en *t*), es decir, una propiedad cuya posesión por *z* (en *t*) no depende de factores externos a *z*, de manera que duplicar las propiedades internas de *z* (en su estado en *t*) garantiza duplicar *x*. En contraste, ser *externista* respecto a *x* en relación con *z* (en *t*) es sostener que *x* es una propiedad *extrínseca* de *z* (en *t*), es decir, una propiedad cuya posesión por *z* (en *t*) depende de factores externos a *z*, de manera que duplicar las propiedades internas de *z* (en su estado en *t*) no garantiza duplicar *x*. Por ejemplo, (en este momento) mi propiedad de estar frente a un escritorio depende de la presencia de un escritorio, mientras que mi propiedad de tener cuatro extremidades no depende de factores externos a mí: (en este momento²⁶) la primera es una propiedad extrínseca mía, mientras que la segunda me es intrínseca. Nótese que algunos factores externos pueden ser la razón de que se tenga cierta propiedad sin que ello la haga extrínseca: por ejemplo, el que mi cuerpo tenga cuatro extremidades es consecuencia de una historia evolutiva que le es extrínseca, pero lo que importa para tener cuatro extremidades es su estado actual, no cómo llegó a él. Sería distinto, por ejemplo, si considerásemos la propiedad de ser humano, bajo el entendido de que serlo implica estar en cierto sitio en el árbol filogenético, para lo cual no basta con ser un duplicado interno de un humano, sino que también se requiere tener cierta historia evolutiva: por ende, la propiedad de ser humano es extrínseca. También es importante señalar que tener una propiedad intrínseca nunca depende necesariamente de tener una propiedad extrínseca, porque ello la haría depender del factor externo del que la otra depende y, por ende, ser extrínseca. En contraste, sí puede ocurrir que una propiedad extrínseca dependa de una intrínseca, básicamente porque puede depender al mismo tiempo de algo externo. Lo que no puede es depender exclusivamente de ella.²⁷

26 La situación cambiaría, por ejemplo, si tuviese un escritorio integrado a mi cuerpo (en cuyo caso estar junto a un escritorio sería una propiedad intrínseca) o si hubiese perdido un brazo (en cuyo caso ni siquiera tendría la propiedad de tener cuatro extremidades).

27 Ver Marshall & Weatherson (2014) para un tratamiento más completo las propiedades intrínsecas y extrínsecas.

Ahora, dado que estoy llamando ‘fenoménico’ tanto a la consciencia fenoménica como a los qualia, la expresión ‘internismo fenoménico’ podría emplearse para referir a distintas cosas, dependiendo de en qué parte de lo fenoménico se haga énfasis y cómo se lo haga. Pero la versión que (más) me interesará en este escrito es la tesis de que toda propiedad de la forma *tener la experiencia fenoménica q* o *no tener la experiencia fenoménica q* es intrínseca, es decir:

- (IF) Para todo objeto x , toda experiencia fenoménica q y todo tiempo t , el que x tenga o no q en t está enteramente determinado por factores internos de x en t ; de modo que, si x tiene q en t , todo posible duplicado interno de x , como es en t , tiene q ; y si x no tiene q en t , ningún posible duplicado interno de x , como es en t , tiene q .

Es muy importante remarcar que esa tesis es distinta del internismo respecto al hecho de que haya (o se posea) qualia. Es decir, estoy hablando de la presentación tipo E de qualia (para x), no de la presencia de qualia (dentro o fuera de x). La relación entre los correspondientes internismos depende de qué se entienda que son los qualia y de si su presentación tipo E es o no condición necesaria y/o suficiente para que estén presentes. Es decir, depende de factores respecto a los que por ahora me estoy manteniendo neutral. Por eso es que de momento me abstendré de decir más al respecto.

Ahora, en el caso de un ser humano (o de un duplicado interno suyo), por ejemplo, sostener el internismo fenoménico equivale a decir que para duplicar sus experiencias fenoménicas (y la ausencia de las que no tiene) en un momento t , basta con duplicar sus características “de piel para adentro” en t . Es decir, cualquier factor externo es irrelevante, incluyendo la historia de cómo llegó a estar en ese estado (porque pudo haber llegado al mismo estado por distintas vías). La intuición general es que nada podría influir en nuestras experiencias sin hacerlo exclusivamente por medio de modificaciones de nuestros estados internos, en cuyo caso bastaría con esas modificaciones (tal vez causadas por otra cosa) para generar las mismas experiencias. Ello está respaldado porque toda la evidencia

científica apunta a que para manipular las experiencias fenoménicas de un sujeto basta con manipular directamente sus estados cerebrales. Es decir, la evidencia apunta a que aquello con lo que la consciencia fenoménica (y la aparente presencia de qualia) se correlaciona más estrechamente se encuentra al interior del cerebro (ver Pautz (2014)). Aun así, dejaré abierta la posibilidad de que también sea relevante algún otro factor interno del sujeto.

Además, si tener consciencia fenoménica (en general) o tener cierta experiencia fenoménica en particular nunca requiere tener alguna propiedad extrínseca, tampoco lo requiere tener cualquier otra propiedad cuya posesión dependa exclusivamente de tener ciertas experiencias fenoménicas. Esto tendrá importantes consecuencias más adelante, porque permitirá descartar todas las posibles maneras de entender la consciencia fenoménica que involucren elementos externos al sujeto. Más específicamente, permitirá descartar todas las posibles maneras de entender a los qualia o la presentación tipo E que conduzcan a la conclusión de que la consciencia fenoménica es una propiedad extrínseca. Ejemplos de teorías que la entienden así son el *realismo directo* y el *representacionismo de rastreo*: según las primeras, para tener ciertas experiencias fenoménicas es indispensable estar en contacto (o interactuar) con alguna parte del mundo externo al momento de tenerlas, mientras que para las segundas es indispensable haberlo tenido en algún momento previo²⁸. En el capítulo 2 ahondaré en esa discusión, pero por ahora me conformaré con dejar en claro que el internismo fenoménico es uno de mis compromisos y que no es uno demasiado controversial, porque es ampliamente aceptado que rechazarlo es un punto débil de las teorías que lo hacen (ver Block (1990, 1998)). Es decir, la negación del internismo fenoménico suele verse más como un resultado incómodo (aunque interesante) de ciertas propuestas que como una postura que pueda sostenerse por cuenta propia. En otras palabras, respetarlo siempre es una virtud y rechazarlo siempre es un costo. Por lo mismo, yo no lo pondré en duda, sino que me concentraré en estudiar algunas

28 También el *enactivismo* se opone al internismo fenoménico, porque sostiene que la consciencia fenoménica es inseparable de la interacción con el mundo externo. Sin embargo, ello puede entenderse de varias maneras, por lo que no es claro cómo clasificar la objeción. En este escrito me conformaré con señalar que los posibles sentidos en que los enactivistas se oponen al internismo fenoménico quedan capturados en las dos opciones que recién he mencionado (requerir contacto al momento o contacto previo). Algunos ejemplos de enactivistas son Brooks (1987), Varela, Thompson & Rosch (1991), Noë (2004, 2009), Froese & DiPaolo (2011), Hutto & Myin (2013) y Steiner (2014).

de sus consecuencias. Si el lector no se encuentra convencido de aceptarlo, bien puede tomar lo que diré como un condicional: si es verdad el internismo fenoménico, entonces se siguen ciertas cosas que señalaré más adelante.

De especial interés son dos resultados del internismo fenoménico, relativos a dos famosos experimentos mentales: los *cerebros en cubeta* (Harman (1973), con antecedente en Descartes (1641)) y *Swampman* (Davidson (1987)). La observación general que me interesa destacar de momento es que, de acuerdo con el internismo fenoménico, los entes de ambos ejemplos tendrían consciencia fenoménica. Pero veamos con calma cada caso.

1.3.1. Swampman

En el caso de *Swampman*, Davidson (1986) nos pide imaginar que un día él va de excursión a un pantano y que, estando ahí y sin que nadie lo note, un rayo lo golpea y desintegra. Además, por mera coincidencia, otro relámpago cae en el mismo lugar justo después y, por mero azar, reacomoda las moléculas circundantes de manera que se genera un ente (*Swampman*) microfísicamente idéntico a Davidson justo antes de ser impactado.

Entonces podemos preguntarnos algunas cosas respecto *Swampman* en comparación con Davidson. Por ejemplo, podemos preguntarnos qué tanto se parecería su comportamiento al que habría tenido Davidson de haber seguido con vida, ante lo cuál resulta natural pensar que se comportaría exactamente como él lo habría hecho: regresaría a su casa y continuaría la vida de Davidson como si nada hubiese ocurrido. De hecho, sería tentador describir su comportamiento como guiado por los mismos intereses y metas que el de Davidson: a fin de cuentas, procuraría o evitaría las mismas cosas e incluso *parecería* perseguir los mismos planes a largo plazo. En general, parece razonable suponer que habría cierto parecido entre los estados mentales de ambos seres: tendría algo *parecido* a los mismos conceptos, ideas, recuerdos, creencias, deseos, emociones, sensaciones, etcétera; o al menos eso sería lo congruente con su conducta.

Sin embargo, podemos señalar que Swampman diferiría de Davidson en muchos aspectos importantes. Por ejemplo, no sería hijo de los padres de Davidson, ni padre de su hija, ni habría asistido a las mismas escuelas, participado en los mismos eventos o sido amigo de las mismas personas. A grandes rasgos, podemos decir que Swampman carecería de todas las propiedades que dependan de haber tenido la misma historia previa que Davidson, porque carecería de dicha historia. En otras palabras, dado que Swampman sólo compartiría las propiedades intrínsecas de Davidson y su historia le es extrínseca, Swampman no compartiría la historia de Davidson ni nada que requiera compartirla. Incluso podemos decir que Swampman no sería un humano, ni un mamífero, ni un animal y tal vez ni siquiera un ser vivo, si nuestros criterios para clasificarlo de esa manera dependen de que tenga cierta historia evolutiva. En lugar de eso, Swampman meramente *parecería* un humano, un mamífero, un animal y un ser vivo. Siguiendo la misma lógica, incluso podemos concluir que ni siquiera tendría pulmones, cerebro, estómago, corazón, etcétera, porque es razonable sostener que ser cualquiera de esos órganos también requiere tener cierta historia evolutiva. En lugar de todo eso, Swampman meramente tendría algo que *parece* pulmones, cerebro, estómago, corazón, etcétera.

Y un resultado especialmente interesante es que ese razonamiento puede extenderse hacia los estados mentales de Swampman: varios de ellos (si no es que todos) no podrían ser acerca de las mismas cosas que los de Davidson, porque carecerían de una conexión histórica (y causal) apropiada con ellas. Por ejemplo, Swampman jamás habría conocido a los amigos de primaria de Davidson y, por ende, no podría tener recuerdos de ellos, a pesar de que tendría algo muy parecido: duplicados intrínsecos de esos recuerdos. El mismo razonamiento sirve para descartar la posibilidad de que Swampman estuviese continuando los planes que Davison ideó y no concluyó, porque no tendría ninguna conexión con ellos, sino que sólo parecería tenerla. En general, los estados de Swampman no tendrían una historia de uso previa, de modo que carecerían de todo lo que dependa de ella. Desde luego que se comportaría exactamente como Davidson lo habría hecho, pero no sería adecuado describirlo en términos de exactamente los mismos estados.

Pero lo que me interesa señalar no es eso, sino que, dado que Davidson habría tenido consciencia fenoménica, el fisicismo y el internismo fenoménico obligan a concluir que Swampman también la tendría, e idéntica a la que Davidson habría tenido²⁹: ser cualquiera de los dos sería subjetivamente indistinguible. Así que lo que sea que explique a la consciencia fenoménica, debe ser algo con lo que Swampman contaría. Y lo mismo para cualquier propiedad que dependa exclusivamente de tener consciencia fenoménica. Entonces será muy importante descartar todas las explicaciones que aludan, por ejemplo, a la historia previa de Davidson.

1.3.2. Cerebros en cubeta

Cuando en filosofía se habla de *cerebros en cubeta*, lo que se tiene en mente es un caso hipotético propuesto originalmente por Harman (1973), pero con antecedentes en Descartes (1641): el de un cerebro real o un duplicado de un cerebro real, desprovisto del resto del cuerpo, mantenido con vida artificialmente y estimulado eléctricamente por una supercomputadora que emite impulsos eléctricos idénticos a los que un cerebro normal recibiría al estar dentro de un cuerpo e interactuando normalmente con el resto del mundo. En particular me interesan la versión en que el cerebro es un duplicado accidental y la computadora no tiene el propósito de simular un mundo, sino que emite impulsos completamente aleatorios, que sólo por casualidad concuerdan con los que recibiría un cerebro en condiciones normales.

Al igual que en el caso de Swampman, podemos preguntarnos diversas cosas sobre los cerebros en cubeta, en especial sobre cómo sería subjetivamente ser uno de ellos y sobre qué estados compartirían o no con sus contrapartes provistas de cuerpo y expuestas al mundo de manera normal. Y para asegurarnos de que nuestras conclusiones sean compatibles con la extraídas en el caso de Swampman, podemos imaginar que uno de esos

29 Aquí también se asoma un supuesto determinista, es decir, la idea de que duplicar el mundo en un momento t involucra duplicarlo en todo momento posterior a t . Pero no es indispensable comprometerse con eso para estar de acuerdo en que Swampman tendría consciencia fenoménica y que esta estaría enteramente determinada por sus factores internos, aun si pudiese ser distinta de la de Davidson.

cerebros es como el de Davidson o Swampman (ya sea original o duplicado) y que los estímulos que recibe corresponden a los que los suyos habrían recibido después de la excursión al pantano (ya sea intencionalmente o por mero accidente).

Lo primero que podemos notar es que, a diferencia del caso de Swampman, el de los cerebros en cubetas tiene más matices, ya que es muy importante distinguir si el cerebro en cuestión es un original o un duplicado, así como si los estímulos que recibe son al azar o intencionados. Además, si se trata de un duplicado, es crucial saber si fue creado por mero azar o intencionalmente. La importancia de determinar cada opción radica en que de ello depende que el cerebro tenga o no ciertas propiedades extrínsecas. En particular, de ellos depende que sus estados tengan o no una conexión histórica (y causal) con ciertas cosas y de qué tipo, de lo cual depende qué tipo de estados serían. Por ejemplo, si el cerebro fuese un duplicado del de Davidson creado intencionalmente, sus estados estarían inspirados en los de Davidson y, por ende, tendrían una conexión causal con los estados de Davidson y con todo lo que hubiese causado que Davidson los tuviese. Entonces el cerebro en cubeta podría tener, por ejemplo, estados sobre los primos de Davidson, porque tendría estados que serían duplicados intencionados de estados de Davidson que son sobre sus primos. Por otra parte, si los estímulos fuesen intencionados, los estados generados en el cerebro podrían heredar esas intenciones y estarían conectados causalmente con el origen de ellas. Entonces el cerebro en cubeta podría tener estados sobre los primos de Davidson porque tendría estados causados por estímulos que tendrían la intención de ser sobre los primos de Davidson. Así que hay al menos tres maneras en que el cerebro en cubeta podría no ser un duplicado meramente intrínseco del original: si fuese el original, si fuese un duplicado no-accidental o si los estímulos que recibe fuesen intencionados. Por ende, un requisito mínimo para eliminar por completo las propiedades extrínsecas en común entre un cerebro en cubeta y el original es especificar que se trata de un duplicado accidental y que los estímulos que recibe no son intencionados, sino aleatorios. Por eso me interesa especialmente esa variante de la situación.

Una vez acotando el caso de esa manera, sí podemos concluir que el cerebro en cuestión

diferiría del original en todo lo que respecta a sus propiedades extrínsecas. Sin embargo, al igual que en el caso de Swampman, el internismo fenoménico obliga a concluir que compartiría todas las experiencias fenoménicas del original (al igual que cualquier otro cerebro en cubeta). Por ende, tener experiencias fenoménicas no puede depender de ninguna de las propiedades que son eliminadas al especificar que el cerebro es un duplicado accidental y que los estímulos que recibe son aleatorios.

A continuación revisaré los problemas propios del fisicismo para explicar la consciencia fenoménica, los cuales son independientes de si se acepta o no al internismo fenoménico. Sin embargo, dado que algunas propuestas para resolver esos problemas dependen de rechazar el internismo fenoménico, comprometerme con él reduce el campo de soluciones viables (ver capítulo 2).

1.4. Contra el fisicismo: la brecha epistémica y el problema difícil de la consciencia

A grandes rasgos, el principal problema al que se enfrenta el fisicismo es que, a pesar de que parece haber una correlación sistemática perfecta entre estados físicos y experiencias fenoménicas, no es claro a qué se debe esta, sino que parece un mero hecho bruto carente de explicación. Ello debido a que parece haber una desconexión entre nuestros *conceptos* de lo físico y lo fenoménico, que impide explicar una cosa exclusiva y transparentemente en términos de la otra. A su vez, eso ha llevado a algunos autores a pensar que de hecho lo físico y lo fenoménico son cosas fundamentalmente distintas y que, por ende, el fisicismo es falso. Más adelante (en la sección 1.4.1. y el capítulo 2) ahondaré en qué son los conceptos y en qué consiste tenerlos, pero de momento bastará con pensar que un concepto es la idea que una comunidad tiene sobre cómo es determinado objeto, usualmente designado por un término del lenguaje que usa. En principio, poseer el concepto de x implica ser capaz de hacer ciertas cosas, como pensar en x , inferir cosas sobre x , lidiar de cierta manera con x y distinguir entre x y lo que no es x ³⁰. En síntesis, una desconexión conceptual entre x y z es una desconexión en la manera en que una comunidad entiende a x

30 Para un estudio general sobre el concepto de *concepto*, ver Laurence & Margolis (2005/11).

y la manera en que entiende a z, debido a la cual dicha comunidad no puede terminar de entender a x en términos de z ni a la inversa. Entonces el problema es que, dados nuestros conceptos actuales, no podemos terminar de entender lo fenoménico en términos de lo físico, lo cual puede conducir a pensar que de hecho son cosas fundamentalmente distintas y que, por ende, el fisicismo es falso. Pero estudiemos con detenimiento el asunto.

Como punto de partida, se puede señalar que una manera muy efectiva de dar a entender qué es lo fenoménico es mediante un experimento mental que justamente pretende mostrar que no es físico. Jackson (1982, 1986) plantea el caso hipotético de Mary, quien toda su vida ha estado encerrada en una habitación en blanco y negro (o podemos pensar que siempre ha vivido con unos pupilentes que filtran toda la luz y la traducen a blanco y negro). Sus cuidadores se han asegurado de que nunca vea otro color y al mismo tiempo la han educado de tal manera que se ha vuelto una experta en física, especialmente en óptica y en el funcionamiento físico del ojo y el cerebro humanos: sabe todo lo que se puede saber sobre la visión humana desde una perspectiva puramente física. En particular, Mary sabe con detalle qué pasaría físicamente si un rayo de luz roja impactase su retina. Sin embargo, Mary nunca ha visto algo rojo (sin pupilentes) y, cuando por fin sale de su habitación (o se le retiran los lentes) y le es presentada una flor roja, aprende algo nuevo sobre la visión humana que no pudo haber conocido de otro modo: *cómo es ver rojo*. Esa cosa nueva con la que Mary se encuentra es una experiencia fenoménica relativa al rojo, con sus correspondientes qualia (al menos para ella). Pero si el conocimiento exhaustivo sobre física fue insuficiente para transmitirle cómo sería ver rojo, entonces debe haber sido un conocimiento incompleto del mundo. Así que comienza a parecer que se requiere de algo más allá de lo físico para explicar el mundo y, por ende, que el fisicismo es falso. Y esa no es más que una de muchas maneras de darse cuenta de que lo fenoménico es problemático para el fisicismo.

Otra opción parte de notar que, si bien estamos acostumbrados a extraer información sobre el mundo físico a partir de nuestras experiencias fenoménicas (o de los acomodos de qualia que se presentan durante ellas), es razonable pensar que todas ellas podrían ocurrir de

manera independiente de aquello sobre lo cual nos informan, porque bien podríamos estar alucinando o soñando. Es decir, dada cualquier experiencia fenoménica, en principio parece posible replicarla sin conexión alguna con el mundo externo. Ello, a su vez, sirve para entender a lo fenoménico como algo meramente mental, perfectamente dissociable de cómo es el mundo físico del cual estamos acostumbrados a pensar que nos habla. De hecho es posible llevar el cuestionamiento aun más lejos: ¿cómo podemos saber que ciertas experiencias (y qualia) sí se corresponden con el mundo externo? ¿Por qué no dudar de todas ellas? A fin de cuentas, no es claro cómo podemos saber qué pasa en el mundo físico si no presuponemos cierta correlación entre lo físico y lo fenoménico³¹. Si llevamos la duda al extremo, incluso podemos llegar a ser escépticos sobre la existencia del mundo externo en su totalidad (ver Descartes (1641)): si de lo único que podemos tener certeza son nuestros propios estados mentales y estos podrían ser los mismos independientemente de cómo sea el mundo externo o incluso si no hay tal, entonces nuestros estados mentales son evidencia insuficiente para saber cómo es el mundo externo o si siquiera existe³². Estrictamente, lo único que sabemos es que ciertas experiencias suelen seguir a ciertas otras: podrían no tener nada que ver con el mundo físico y ni cuenta nos daríamos. Si bien no quiero poner en duda la existencia de lo físico ni la fiabilidad de nuestros mecanismos perceptivos, señalar esta vía sirve para motivar la idea de que lo fenoménico no implica conceptualmente a lo físico: tal vez hasta podría existir lo primero sin lo segundo.

Por otro lado (y dejando de lado que conocemos el mundo a partir de lo fenoménico), tampoco está más allá de toda duda que los eventos físicos necesariamente conducen a la presencia de experiencias o qualia particulares, a pesar de que estamos acostumbrados a hacer generalizaciones al respecto. Si bien creo que si me lastimo de cierto modo tendré una experiencia de dolor (con su correspondiente carácter fenoménico), porque en el pasado siempre la he experimentado cuando me he lastimado así, me es perfectamente posible dudar de que se repetirá cada vez que me lastime del mismo modo: creemos en la correlación por mera costumbre, no porque entendamos a fondo el vínculo entre lo físico y

31 Nótese que todos nuestros modelos científicos de lo físico se derivan de presuposiciones de ese tipo, ya que descansan en evidencia empírica.

32 Para más sobre el *problema del mundo externo*, ver Lyons (2016).

lo fenoménico. Aunque sistemática, la correlación entre eventos físicos y experiencias (y qualia) ha sido tan falta de una explicación que los conecte de forma transparente, que no es claro que corresponda a una ley inquebrantable³³. Así que no tenemos certeza absoluta de que nuestros estados físicos conducirán siempre a experiencias fenoménicas (o presencia de qualia) en @, específicas o no. Y si ni siquiera tenemos esa certeza sobre @, ¿por qué suponer que en otros mundos metafísicamente posibles debería cumplirse la supuesta correlación? De nuevo, el objetivo no es dudar sobre nuestro conocimiento de la correlación entre lo físico y la consciencia fenoménica en @, sino remarcar que no hemos llegado a él de manera *deductiva*, sino *inductiva*, y que pasar de ahí a una necesidad metafísica implica otro salto igual o más injustificado. De lo que nos terminamos dando cuenta es de que lo físico tampoco implica conceptualmente a lo fenoménico: tal vez hasta podría existir sin ello.

Podemos reforzar el mismo punto si notamos que, aunque solemos suponer que otros individuos también tienen experiencias fenoménicas (o qualia, mente y consciencia en general³⁴), no podemos acceder directamente a ellas y nada de lo que podemos apreciar de ellos desde fuera indica más allá de toda duda que las tengan: hasta donde sabemos, ser ellos podría ser radicalmente distinto de cómo es ser nosotros mismos, o incluso podría no ser de manera alguna (desde su perspectiva en primera persona). En el mejor de los casos, podemos razonar *abductivamente* y sostener que, dado que todos somos objetos del mismo tipo (es decir, humanos), la conclusión más plausible es que tenemos vida mental semejante, pero ello dista mucho de ser un argumento definitivo y no es claro que no se apoye en presuposiciones fisicistas. Estrictamente no sabemos qué pasa en las mentes ajenas, porque es un hecho que no podemos acceder de primera mano a ellas, así que la única evidencia de la que disponemos proviene de lo que sus poseedores reportan, pero

33 Esto no es más que un caso particular del *problema de la inducción* señalado por Hume (1739): la mera repetición sistemática de algo no basta para garantizar que no podría ser de otro modo (para un tratamiento más completo, ver Henderson (2018)).

34 Si bien la consciencia fenoménica es la parte de la mente que más claramente entra en conflicto con el fisicismo, dudas similares pueden ser planteadas básicamente con relación a cualquier otro aspecto de ella: a fin de cuentas, “desde fuera” tampoco podemos tener certeza de lo que otras personas piensan o creen, por ejemplo. Es decir, lo que planteo en este y el siguiente párrafo es un caso particular de algo más amplio: el *problema de las otras mentes* (ver Hyslop (2005/14)).

perfectamente podrían afirmar cosas falsas. Tal vez las experiencias (y qualia) ajenas ni siquiera están correlacionadas sistemáticamente con lo físico o incluso podría ser que los demás carezcan de experiencias (y qualia) y meramente se comporten como si las tuviesen. En general, por más que conozcamos a otra persona (desde fuera) y por más que reporte tener una u otra experiencia fenoménica (o qualia, y tal vez cualquier estado mental), siempre podemos poner en duda todo lo que diga, porque las experiencias fenoménicas (entre otros estados) son privadas y la única manera de tener certezas sobre la vida mental de alguien es ser ese alguien: la única forma de hacer constar que hay cierta experiencia (o qualia) es “sentirla en carne propia”. Ni siquiera basta con tener conocimiento microfísico exhaustivo de la persona en cuestión, porque tampoco a esa escala aparece un puente claro entre lo físico y lo fenoménico. De hecho, todo esto se sostiene para cualquier objeto: no podemos tener certeza sobre si subjetivamente es de algún modo ser una roca o (de nuevo) cómo sería tener algún mecanismo perceptivo distinto de los nuestros, como la ecolocalización de los murciélagos (ver Nagel (1974)). De hecho, si nos alejamos demasiado de ejemplos humanos, hasta nos quedamos sin reportes y argumentos comparativos.

Además, aun si confiamos en el juicio ajeno y concedemos que otros tienen experiencias (o qualia) correlacionadas sistemáticamente con lo físico, no podemos estar seguros de que sea del mismo modo que en nosotros: tal vez alguien tiene la experiencia (o qualia) que yo asocio con el verde cuando ve algo rojo, o ni siquiera tiene una experiencia (o qualia) que yo llamaría ‘visual’. De ser ese el caso, no nos daríamos cuenta, porque ambos habríamos crecido llamando del mismo modo a las experiencias (o qualia) que tenemos ante las mismas cosas, a pesar de que podrían ser distintas: tal vez alguien llama ‘experiencia de rojo’ (o ‘qualia rojo’) a lo que yo llamo ‘experiencia de verde’ (o ‘qualia verde’), porque siempre ha experimentado lo que yo llamo ‘experiencia de verde’ (o ‘qualia verde’) ante lo que a mí me genera lo que llamo ‘experiencia de rojo’ (o ‘qualia rojo’) y cada quién ha aprendido a hablar acorde a ello (ver Shoemaker (1982) y Block (1990)). En síntesis, no sabemos cómo es ser alguien más o si siquiera es de alguna manera, porque lo único que podemos conocer de los demás son sus estados físicos y no sabemos más allá de toda duda

si los estados físicos bastan para generar experiencias fenoménicas o qualia (particulares o no).

En un final, nos damos cuenta de que la evidencia que tenemos a la mano es insuficiente para responder las dudas antes mencionadas, porque, hasta donde podemos asegurar, más de un escenario es compatible con ella. Así que no podemos descartar *a priori* la posibilidad de que eventos físicos del mismo tipo o incluso idénticos podrían ir acompañados de experiencias o qualia distintos o acontecer en completa ausencia de ellos (y viceversa). En otras palabras, nuestro entendimiento actual de lo físico se encuentra separado del de lo fenoménico de manera tal que, aun si reconocemos que en @ hay una correlación estrecha entre ambas cosas, esta parece contingente: simplemente no es claro a qué se debe ni qué tan generalizable es. Estos problemas desaparecerían si pudiésemos encontrar un punto en el que la conexión entre lo físico y la consciencia fenoménica fuese clara, pero al parecer persisten aun si pensamos en un duplicado microfísico nuestro, porque hemos aprendido a abstraer lo físico y lo fenoménico de tal manera que no ha quedado un puente que conecte ambas cosas *a priori*: sencillamente parece que ninguna descripción física capturará la naturaleza de lo fenoménico de una manera transparente, porque nuestros conceptos de ambas cosas son completamente independientes. Siendo ese el caso, nuestros mejores argumentos a favor del fisicismo son inductivos o abductivos, es decir, no son deductivos, lo cual los mantiene débiles³⁵.

Más allá de si de hecho existe o no correlación entre lo físico y lo fenoménico, lo importante es que no es evidente que deba haberla, porque ningún conocimiento sobre lo físico o lo fenoménico basta por sí mismo para inferirlo *a priori*, ya que nuestros conceptos de ambas cosas no incluyen la conexión apropiada para hacerlo³⁶: hay una *brecha (o desconexión) epistémica* en medio (ver Levine (1983)³⁷). Más específicamente, hay una

35 La inducción y la abducción comparten un problema de *subdeterminación*: a diferencia de la deducción, no conducen a la única conclusión compatible con las premisas, sino a una de varias. Para más sobre ese tipo de problemas, ver Stanford (2009/17).

36 A menos que estipulemos la verdad del fisicismo y con ello leyes que vinculen lo físico con la consciencia fenoménica, integrándolas en nuestra concepción de ambas cosas, pero ello sería presuponer la conclusión que está en debate.

37 Levine habla de una 'brecha explicativa', pero yo optaré por el término 'brecha epistémica', porque

brecha epistémica entre las verdades sobre lo físico y las verdades sobre lo fenoménico, de manera que no es *a priori* que las de un tipo se sigan de las del otro. Como resultado, es cuestionable si un duplicado exclusivamente físico del mundo tendría que ser también un duplicado fenoménico y, por ende, es cuestionable la verdad del fisicismo. Lo que hace falta para cerrar la brecha es una explicación transparente de cómo se conecta lo físico con lo fenoménico, es decir, una que vaya más allá de meramente señalar correlaciones. Al reto de encontrar algo así se le conoce como el *problema difícil de la consciencia* (ver Chalmers (1995a)) y es lo que me interesa ayudar a resolver³⁸.

1.4.1. Formalizando algunos argumentos en contra del fisicismo

Además de Jackson (1982, 1986), otros autores, tales como Descartes (1641), Kripke (1972) y Chalmers (2002a, 2005), han inferido la existencia de una brecha ontológica a partir de la brecha epistémica que acabo de exponer, es decir, han concluido que lo físico y lo fenoménico podrían existir por separado a partir de que nuestro entendimiento actual de ambas cosas no se encuentra debidamente conectado. Según Chalmers (2002a)³⁹, la forma general de todos esos argumentos es la siguiente (Arg1):

P1_{A1}) Hay una brecha epistémica entre las verdades sobre lo físico y las verdades sobre lo fenoménico.

P2_{A1}) Si hay una brecha epistémica entre las verdades sobre lo físico y las verdades sobre lo fenoménico, entonces hay una brecha ontológica entre lo físico y lo fenoménico.

P3_{A1}) Si hay una brecha ontológica entre lo físico y lo fenoménico, el fisicismo es

‘explicativa’ puede entenderse como presuponiendo una brecha metafísica, que es justo lo que está en disputa. Es decir, lo que estoy reconociendo es meramente una brecha psicológica, no una ontológica.

38 En contraste, los “problemas fáciles” son aquellos que consisten en explicar comportamientos y procesos cognitivos específicos, como el aprendizaje, la memoria, la integración perceptual y el reporte verbal. Dichos problemas pueden abordarse de manera empírica y la ciencia ya ha avanzado bastante en ellos. También se puede considerar un problema fácil al de simplemente descubrir cómo se correlaciona lo físico con la consciencia fenoménica en @, sin buscar razones más profundas. Nótese que esos otros problemas son fáciles en comparación con el de explicar cómo surge la consciencia fenoménica, porque sólo en el caso de esta no hay consenso acerca de cómo debería ser abordado el problema o qué tipo de respuesta es la buena, pero ello no quiere decir que no impliquen un reto considerable.

39 Chalmers (2002a) fusiona las premisas P2_{A1} y P3_{A3}.

falso.

∴ C1_{A1}) El fisicismo es falso.

El argumento es válido y ya motivé P1_{A1}. Por otra parte, P3_{A1} se sigue de que la consciencia fenoménica existe y de que la ontología fisicista no incluye a nada cuya presencia no se siga necesariamente de la de algo exclusivamente físico: si lo fenoménico no es capturado por esa ontología, entonces el fisicismo es falso. Así que sólo resta justificar P2_{A1}. Sin embargo, hay un obstáculo obvio para lograrlo: que no parece que un hueco en nuestra manera de entender dos cosas necesariamente implique que no haya algo que las vincule de forma relevante. Tal vez sólo quiere decir que no notamos que lo haya, porque nuestras capacidades cognitivas son incapaces de detectarlo, lo cual es a todas luces insuficiente para tener certeza de que no lo hay. Chalmers (2005) toma en cuenta eso y especifica bajo qué condiciones sí se justifica inferir una brecha ontológica a partir de una epistémica. En un final, es razonable entender que delimitar esas condiciones equivale a definir cierto tipo de brecha epistémica (G) y que el argumento resultante es uno que queda capturado en el siguiente (Arg2):

P1_{A2}) Hay una brecha epistémica *tipo G* entre las verdades sobre lo físico y las verdades sobre lo fenoménico.

P2_{A2}) Si hay una brecha epistémica *tipo G* entre las verdades sobre x y las verdades sobre z, entonces hay una brecha ontológica entre x y z.

P3_{A2}) Si hay una brecha ontológica entre lo físico y lo fenoménico, el fisicismo es falso.

∴ C1_{A2}) El fisicismo es falso.

Para evaluar el argumento hay que especificar qué es una brecha epistémica tipo G. Con ese propósito, a continuación explicaré la propuesta de Chalmers (2002b, 2005) y después regresaré sobre Arg2. Él plantea todo en términos de concebibilidad, entendida como la capacidad de pensar o imaginar coherentemente algo que cumpla cierta descripción (*concebibilidad positiva*), o sencillamente la incapacidad de descartar *a priori* la

coherencia de la misma (*concebibilidad negativa*)⁴⁰. La idea general es que, debido a la brecha epistémica (aún por terminar de definir), es posible concebir (en alguno de los dos sentidos) un duplicado físico mínimo de @ sin su correspondiente parte fenoménica ($P^{@*} \& \neg Q^{@}$), a partir de lo cual se concluye que es metafísicamente posible un duplicado físico de @ sin esa parte fenoménica ($\diamond(P^{@*} \& \neg Q^{@})$)⁴¹, lo que es a su vez suficiente para refutar al fisicismo (que está comprometido con que $\Box(P^{@*} \rightarrow Q^{@})$). Sin embargo, hace falta decir más, porque prácticamente todo es concebible bajo una noción de concebibilidad lo suficientemente laxa (incluyendo cosas claramente imposibles), por lo que es sospechoso extraer conclusiones sobre el mundo meramente a partir de ella, así que es indispensable depurar las condiciones que deben cumplirse para validar el salto de la concebibilidad de algo a su posibilidad metafísica.

Con eso en mente, lo primero que se debe notar es que para concebir algo hay que caracterizarlo de alguna manera, es decir, atribuirle ciertas características⁴² que lo definan y distingan de otras cosas. La riqueza de dicha caracterización puede variar, yendo desde prácticamente nada hasta descripciones muy minuciosas, pero hay que tener en cuenta que una caracterización demasiado pobre de un objeto x podría no permitir distinguir entre concebir x y concebir algún otro objeto z. Por ejemplo, si la única idea que se tiene de los tigres y leones es que son felinos, se les confundirá entre sí. Así que, si se quiere usar a A para probar algo, al menos hay que asegurarse de apoyarse en una caracterización que lo

40 La concebibilidad positiva implica la negativa, porque al contemplar algo en toda su extensión se hace palpable su coherencia o incoherencia. Pero es cuestionable la implicación a la inversa, ya que (en principio) no detectar inconsistencias no equivale a contemplar la totalidad. Dado que basta con la concebibilidad negativa para que funcionen los argumentos que mencionaré, no será necesario ahondar más en esto.

41 Para lo cual no es necesario que el duplicado carezca de parte fenoménica, sino que basta con que esta, de haberla, difiera al menos en algo de la de @. De hecho Chalmers (2002a, 2005) habla meramente de $\Box(P^{@*} \rightarrow Q1)$, donde Q1 es alguna parte de $Q^{@}$, como puede ser cualquier proposición que hable sobre alguna experiencia particular en @ (por ejemplo, que cierto qualia de color ocupa ahora determinado lugar de mi campo visual). Dado que, por definición, $\Box(\neg Q1 \rightarrow \neg Q^{@})$, la falsedad de $\Box(P^{@*} \rightarrow Q^{@})$ se sigue de la de $\Box(P^{@*} \rightarrow Q1)$; así que, para refutar al fisicismo, basta con mostrar la posibilidad metafísica de $P^{@*} \& \neg Q1$. En otras palabras, basta con mostrar que alguna experiencia en @ puede variar dejando fija a la parte física (por ejemplo, que otro qualia de color podría ocupar momentáneamente el lugar mencionado, dejando todo lo demás igual). A fin de cuentas pensar en $\neg Q1$ es tan sólo una manera de pensar en $\neg Q^{@}$, por lo que ambas vías son equivalentes y no será necesario insistir en esto.

42 Lo cual requiere la caracterización de dichas características, pero dejemos eso de lado por ahora.

delimite correctamente. Por ejemplo, si lo que se quiere es encontrar tigres, lo ideal será buscarlos mediante una caracterización que no permita confundirlos con leones. Y si lo que se quiere es concebir un mundo metafísicamente posible que cumpla con $P^{@*} \& \neg Q^@$, hay que emplear una caracterización tal que cumplir con ella baste para *ser* $P^{@*} \& \neg Q^@$ y no meramente para parecerlo.

Esas ideas son capturadas por Chalmers (2002b, 2005) dentro del marco del *bidimensionalismo semántico* (ver también Chalmers (2002c, 2002d, 2004, 2005, 2006a, 2006b))⁴³. Según esa teoría, todo término del lenguaje hablado/escrito y todo concepto correspondiente tiene (al menos) dos *intensiones*, es decir, dos funciones (matemáticas⁴⁴) que lo relacionan (o pretenden relacionarlo) con su *referente*, es decir, con el objeto (o tipo de objeto) al que apunta (o *refiere*) en un final⁴⁵ (dentro de todo el espacio de posibilidad): una *subjuntiva* (o *secundaria*) y otra *epistémica* (o *primaria*)⁴⁶. Si ‘x’ es un término o concepto que refiere a un objeto (o tipo de objeto) existente x, su *intensión subjuntiva* es lo que de hecho hace que ‘x’ refiera a x: cuando un objeto z es seleccionado por la intención

43 El *bidimensionalismo semántico* es un intento de conciliación entre dos importantes posturas respecto al significado de los nombres propios y términos de clase natural: el *descriptivismo* y las *teorías de la referencia directa*. Según el descriptivismo, asociado comúnmente a Frege (1892) y Russell (1905, 1910/11, 1912), los términos en cuestión son sinónimos de descripciones definidas de su referente, es decir, de descripciones que sólo cumple éste. En contraste, las teorías de la referencia directa, comúnmente asociadas a Kripke (1972), sostienen que son simples etiquetas sin ninguna descripción asociada. Las teorías de la referencia directa suelen vincularse las teorías causales de la referencia, como la de Putnam (1975), que sostienen que lo que fija el referente es la historia de uso de cada término, que debe incluir un vínculo causal con su referente, que explique cómo es que se le asoció un nombre. Los ataques de los teóricos de la referencia directa y de las teorías causales en contra del descriptivismo han sido tan devastadores que difícilmente alguien sostendría un descriptivismo puro en la actualidad, pero tampoco es fácil descartar del todo la intuición de que comúnmente asociamos descripciones a los referentes de los términos que empleamos, por lo que han surgido teorías híbridas como el bidimensionalismo.

44 Una *función matemática* es una relación entre conjuntos, en la que cada elemento de uno se asocia con uno y sólo uno de los elementos de otro. En la sección 2.3.3 emplearé otras acepciones del término ‘función’.

45 Más adelante, en la sección 2.3, ahondaré en el tema de la referencia. También cabe señalar que estoy entendiendo que los enunciados refieren a hechos o estados de cosas (o tal vez proposiciones). Ello en contraste con posturas como la de Frege (1892), según la cual refieren a valores de verdad. Es decir, entiendo que un enunciado refiere a aquello que lo haría verdadero, no a la verdad o falsedad mismas (ver nota 85).

46 ‘Prefiero usar los términos ‘epistémica’ y ‘subjuntiva’ por encima de ‘primaria’ y ‘secundaria’, porque estos últimos sugieren un orden de importancia con el que no me comprometo y que de hecho parece tergiversar la concepción tradicional, según la cual la intención subjuntiva es la que de hecho fija la referencia.

subjuntiva de 'x', decimos que z *satisface* 'x', o llanamente, que z *es* x. La idea más aceptada es que lo que determina esa intensión es alguna historia causal que vincula a x con el uso de 'x': alguna suerte de bautizo de x como 'x' (ver Kripke (1972)). Sin embargo, lo cierto es que en la práctica rara vez recurrimos conscientemente a esa historia para identificar a los objetos de los que queremos hablar, sino que más bien nos apoyamos en alguna descripción que creemos que cumplen. Por ejemplo, si nos preguntan a qué refiere la palabra 'tigre' difícilmente contaremos la historia de cómo los tigres fueron bautizados de ese modo, sino que más bien ofreceremos alguna descripción que creemos que los tigres cumplen, tal como *felino grande con franjas naranjas y negras*. En términos de Chalmers, lo que ofrecemos es una *intensión epistémica* de 'x': una caracterización (o concepción, descripción o modo de presentación) de x asociada al uso de 'x', de manera que vincula a 'x' con cualquier cosa que cumpla el criterio en cuestión. Y cuando algo es seleccionado por la intensión epistémica de 'x', decimos que *verifica* 'x'. En pocas palabras, la intensión subjuntiva de 'x' selecciona sólo a x, mientras que la epistémica selecciona a todo lo que tiene determinadas características asociadas a x: todo lo que se parece a x en ese sentido verifica 'x', mientras que sólo x satisface 'x'. Por último, cabe señalar que también los enunciados tienen ambos tipos de intensiones, mismas que están determinadas por las de las palabras que los componen. De hecho, las nociones de satisfacción y verificación están pensadas especialmente para enunciados: z satisface un enunciado 'A' si y sólo si z es el tipo de situación a la que refiere 'A' y z verifica 'A' si y sólo si es una situación que cumple con ciertas características asociadas a las situaciones que satisfacen 'A'. En lo sucesivo hablaré indiferentemente de enunciados y palabras, salvo cuando lo especifique de otro modo⁴⁷.

Ahora, lo ideal es que la caracterización de x en la que se apoya la intensión epistémica de 'x' sea tan buena que x y sólo x la cumpla, de tal modo que ambas intensiones de 'x' (epistémica y subjuntiva) *coincidan*, es decir, que seleccionen exactamente a los mismos objetos. Sin embargo, como ya sugerí, ello puede no ocurrir. Por ejemplo, bien podría haber leones mutantes con franjas naranjas y negras o tigres mutantes sin ellas, de modo

47 Recuérdese que las situaciones son objetos (ver nota 7).

que si la intensión epistémica de ‘tigre’ está basada en la descripción *felino grande con franjas naranjas y negras*, seleccionará a los leones mutantes con franjas naranjas y negras y dejará fuera a los tigres mutantes sin ellas, de modo que no seleccionará a todos los tigres ni sólo a los tigres. En general, si la caracterización de x en la que se basa la intensión epistémica de ‘ x ’ puede ser cumplida por algo distinto de x , verificar ‘ x ’ no bastará para satisfacer ‘ x ’, porque ambas intensiones seleccionarán distintos objetos. Ello es especialmente importante cuando se habla de que x es “concebible” o “posible”, porque x bien podría estar siendo confundido con algo que no es x , porque puede ser que se le esté dando prioridad a la intensión epistémica de ‘ x ’ y algo distinto de x cuadre con ella (o x ni siquiera lo haga). La única manera de estar a salvo de eso es si la intensión epistémica de ‘ x ’ captura propiedades esenciales de x , en el sentido de que x siempre las tenga y sólo x las pueda tener: cuando eso pasa, verificar ‘ x ’ sí implica satisfacer ‘ x ’ y viceversa. Entonces decimos que las intensiones de ‘ x ’ *coinciden*, porque ambas seleccionan exactamente los mismos objetos.

Sin embargo, cabe señalar que en la práctica no es claro que los términos y conceptos que empleamos estén siempre asociados a una única caracterización de su referente, sino que es razonable pensar que cada uno puede tener múltiples intensiones epistémicas (dependiendo de sus usos y usuarios) y, por ende, múltiples maneras de confundir a su referente con otra cosa, dependiendo de en qué características se haga énfasis en cada ocasión. En otras palabras, no siempre es claro en qué consiste el concepto de x , sino que lo que se tiene son distintas maneras de entender a x . Adoptando la terminología de Millikan (2013, 2017), lo que se tiene son *uniceptos* de x , es decir, ideas individuales sobre cómo es x ⁴⁸. Si bien lo ideal es que el unicepto que alguien tiene sobre x coincida con el concepto que su comunidad tiene sobre x , usualmente no podemos asegurar que ello suceda. Además, muchas veces ni siquiera el concepto en cuestión está bien definido. Así

48 En general, Millikan (2013, 2017) propone enfocarse en uniceptos en lugar de en conceptos, porque los conceptos son algo que es supuestamente compartido entre individuos y, por lo mismo, es difícil determinar exactamente en qué consiste cada ejemplar de ellos. En contraste, los uniceptos son individuales y, por eso, es más fácil determinar en qué consisten. Por otra parte, Millikan (2017) también habla de *uni-rastreadores*, que son mecanismos que rastrean objetos con la ayuda de sus correspondientes uniceptos.

que, en lugar de presuponer que hay tal cosa como el concepto de x y que de él depende la intensión epistémica de ' x ', es mucho más seguro limitarse a decir que ' x ' tiene una intensión epistémica por cada unicepto de x que se le asocie. Incluso puede suceder que algunas de esas intensiones epistémicas coincidan con la subjuntiva y otras no. Por esa razón, considero que es un error simplemente hablar de la 'intensión epistémica de ' x '' o de la 'verificación de ' x '' a secas (así como de otras nociones asociadas), en lugar de especificar cuál de todas las posibles intensiones epistémicas se está empleando cada vez. Si bien Chalmers (2002b, 2005) habla como si la intensión epistémica fuese única, yo no me comprometeré con que lo sea, por lo que introduciré algunas sutiles diferencias en la manera de hablar al respecto. En particular, siempre especificaré bajo qué intensión epistémica de ' x ' es que algo verifica ' x ': si la caracterización en cuestión es C , hablaré de *C-verificar* ' x '.

Regresando al tema, cuando se habla de la "concebibilidad" o "posibilidad" de x , hay que tener mucho cuidado con qué pasa con las intensiones de ' x ', para evitar confundir a x con algo más. Por eso es que Chalmers (2002b, 2005) distingue entre la *concebibilidad epistémica* (o *primaria*) y la *concebibilidad subjuntiva* (o *secundaria*), así como entre la *posibilidad epistémica* (o *primaria*) y la *posibilidad subjuntiva* (o *secundaria*). Concebir subjuntivamente (o *s-concebir*) x es tomar en consideración algo que es seleccionado por la intensión subjuntiva de ' x ' y, por ende, satisface ' x ': es decir, algo que de hecho es x . En contraste, concebir epistémicamente (o *e-concebir*) x es tomar en consideración algo que es seleccionado por una intensión epistémica de ' x ', es decir, que cumple con cierta caracterización de x asociada a ' x ' y, en ese sentido, verifica ' x ': dicho objeto cumple con algo con lo que se cree que x cumple, pero bien podría no ser x . Si la caracterización en cuestión es C , diré que x se *e-concibe bajo* C o que se *s-concibe* algo que C -verifica ' x '. Por otra parte, la *s-posibilidad* de x no es otra cosa que la posibilidad metafísica de x : que x sea posible subjuntivamente (o *s-posible*) quiere decir que es metafísicamente posible algo que de hecho es x (algo que satisface ' x '). En contraste, que x sea posible epistémicamente (o *e-posible*) quiere decir que es metafísicamente posible (es decir, *s-posible*) algo que es seleccionado por una intensión epistémica de ' x ': un objeto B que cumple con una

caracterización con la que se creó que *x* cumple. Si la caracterización es *C*, diré que *x* es e-posible *bajo C* o que es metafísicamente posible algo que *C*-verifica ‘*x*’. Por último, cabe señalar que si ambas intensiones de ‘*x*’ coinciden (es decir, si cumplir con *C* basta para ser *x*), todas las nociones derivadas de ella también coincidirán: la e-concebibilidad de *x* bajo *C* será lo mismo que su s-concebibilidad y su e-posibilidad bajo *C* será lo mismo que su s-posibilidad.

Por otro lado, es bastante claro que es posible concebir cosas contradictorias si nuestro análisis es superficial: por ejemplo, es posible concebir una esfera cúbica si no se profundiza lo suficiente en qué es ser esférico y cúbico. Para evitar eso, no sólo se requiere una caracterización con el más alto nivel de detalle, sino también que el escrutinio de quien concibe sea impecable, para que no se le escape ninguna posible inconsistencia. En términos de Chalmers (2005), lo que se requiere es *concebibilidad ideal*, en el sentido de que esté libre de errores y no se encuentre restringida por nuestras limitantes naturales, como tener recursos finitos y ser efímeros. Si bien es claro que no somos seres ideales en ese sentido, Chalmers sostiene que hay ocasiones en las que podemos saber cuál sería el veredicto de un ente que sí lo fuese. Así que por lo menos hay que asegurarse de estar en una situación así. De este modo también se garantiza que *x* de hecho cumpla con la caracterización que se le asocia.

Por último, Chalmers (2002b, 2005) se compromete con la tesis del *racionalismo modal*, según la cual parecer coherente bajo razonamiento ideal es lo mismo que ser coherente a secas. De ser así, basta con ser e-concebible (negativa⁴⁹ e) idealmente bajo *z* para ser e-posible bajo *z* y basta con ser s-concebible (negativa e) idealmente para ser s-posible. El racionalismo modal es una tesis atractiva porque en principio no hay razón para negar que sea metafísicamente posible todo aquello que parece coherente, además de que ofrece una guía accesible para saber *a priori* qué es y qué no es metafísicamente posible. Sin embargo, estrictamente no podemos tener certeza de que sea verdad y eso la convierte en uno de los puntos débiles del argumento de Chalmers, porque él la trata como si fuese un

49 En sí no es necesario especificar si la concebibilidad es positiva o negativa, porque la primera implica a la segunda (ver nota 40).

presupuesto inocente cuando en realidad no lo es. De hecho, el racionalismo modal es justamente lo que en el fondo resulta sospechoso de inferir la posibilidad metafísica de algo a partir de su concebibilidad, del tipo que sea. Sin embargo, yo no lo cuestionaré más y simplemente trabajaré como si fuese aceptable, porque lo que me interesa es destacar otros problemas del argumento⁵⁰.

Así que, en síntesis, las condiciones lógicamente suficientes para deducir la s-posibilidad de x a partir de su e-concebibilidad son las siguientes: 1) x debe ser e-concebible bajo z; 2) debe ser verdad el racionalismo modal; y 3) la intensión epistémica de 'x' que se emplee (es decir, z) debe coincidir con la secundaria (es decir, cumplir con z debe ser suficiente para ser x). La razón para incorporar el primer requisito es evidente, ya que cualquier cosa es concebible (del modo que sea) bajo una noción lo suficientemente laxa de concebibilidad, pero es claro que no cualquier cosa es posible (del modo que sea), así que es razonable ser exigentes al respecto. El segundo requerimiento es el puente entre z-concebibilidad y z-posibilidad, que permite inferir la s-posibilidad de un objeto w que se parece a x en cierto aspecto relevante z a partir de la concebibilidad ideal de x bajo ese aspecto. Finalmente, el último punto sirve para asegurar que w de hecho sea x (y no meramente lo parezca). Primero garantizamos la s-posibilidad de un objeto que se parece a x, a partir de que no haya ninguna inconsistencia oculta en z, y luego mostramos que dicho objeto es x: entonces x es s-posible.

Para aclarar la idea, podemos recurrir a un ejemplo inspirado en las ideas de Putnam (1973, 1975): si caracterizamos al agua como *el líquido que cae del cielo, llena los mares, ríos y lagos y en estado puro es incoloro, inodoro e insípido* (como presumiblemente era caracterizada antes de descubrirse su composición química) y llamamos 'C1' a dicha caracterización, entonces es e-concebible bajo C1 que el agua no sea H₂O, porque no parece contradictorio que algo distinto de H₂O cumpla con C1, ni siquiera bajo razonamiento ideal. Por otro lado, el racionalismo modal permite dar el salto de la e-

50 En Reynoso (2017) desarrollo mi rechazo del racionalismo modal. Concluyo que es falso o verdad por mera coincidencia, lo cual lo hace un mal apoyo para concluir lo que sea. Sin embargo, sí es deseable que las teorías no entren en conflicto con él, porque tampoco es claro que sea falso.

concebibilidad bajo C1 de agua que no es H2O a su e-posibilidad bajo C1, es decir, garantiza la s-posibilidad un objeto (XYZ) que cumple C1 pero sin ser H2O, a partir de que C1 no es *a priori* incoherente. Pero aun siendo ese el caso, dado que para ser agua es indispensable ser H2O, entonces no se estaría s-concibiendo al agua y, por ende, lo que se estaría mostrando como s-posible no sería agua que no es H2O, sino XYZ, que cumple C1 pero no es H2O y, por ende, no es agua. Así que en ese caso no se podría dar el salto deseado de e-concebibilidad a s-posibilidad de agua que no es H2O. En general, no bastaría con e-concebir bajo C1 agua que tiene alguna propiedad F (como estar teñido de verde, por ejemplo) para s-concebir agua con la propiedad F, ni bastaría con la s-posibilidad de dicho líquido para garantizar que el agua pueda tener la propiedad F: tal vez en ambos casos se trataría sólo de otra substancia (XYZ) con la propiedad F. Por otro lado, basta con ser H2O para ser agua, así que, si se llama ‘C2’ a la caracterización *ser H2O*, basta con e-concebir bajo C2 agua con la propiedad F para s-concebir agua con la propiedad F, y basta con la e-posibilidad bajo C2 de que el agua tenga la propiedad F para garantizar la s-posibilidad de que la tenga. Todo ello porque C2 coincide con la intensión subjuntiva de ‘agua’.

Entonces, una vez aceptado el racionalismo modal, para refutar al fisicismo basta con que $P^{@*} \& \neg Q^{@}$ sea e-concebible idealmente bajo una caracterización que coincida con la intensión subjuntiva de ‘ $P^{@*} \& \neg Q^{@}$ ’: en tal caso, $P^{@*} \& \neg Q^{@}$ será s-posible, lo cual es suficiente para negar que $\Box(P^{@*} \rightarrow Q^{@})$, refutando así la tesis fisicista. Así que hay que estudiar cuidadosamente la manera en que se puede e-concibe idealmente $P^{@*} \& \neg Q^{@}$, para saber si en algún caso es razonable dar el salto hacia la s-posibilidad de algo que sirva contra el fisicismo. Es decir, hay que averiguar qué intensiones epistémicas tiene ‘ $P^{@*} \& \neg Q^{@}$ ’ y si alguna de ellas es relevante para el fisicista. Dependiendo de eso, podremos determinar si al e-concebir $P^{@*} \& \neg Q^{@}$ se está s-concibiendo $P^{@*} \& \neg Q^{@}$ u otra cosa. Por último, hay que notar que las intensiones de ‘ $P^{@*} \& \neg Q^{@}$ ’ dependen de las de ‘ $P^{@}$ ’ y ‘ $\neg Q^{@}$ ’, que a su vez dependen de las de ‘físico’ y ‘fenoménico’, por lo que son esas en las que hay que concentrarse.

Tomando eso en cuenta, Chalmers (2005) señala que una peculiaridad de lo fenoménico (aunque él se centra en los qualia) es que tenemos acceso privilegiado a ello: se nos presenta de tal manera que no podemos confundirlo con otra cosa. Es decir, nada puede parecer fenoménico sin serlo, ni serlo sin parecerlo: ‘fenoménico’ tiene una intensidad epistémica que coincide con su intensidad subjuntiva. Sin embargo, sí podría ocurrir que confundiésemos dos experiencias fenoménicas o dos qualia porque se parecen mucho. Es decir, puede ser que no tengamos al alcance una intensidad epistémica de ‘Q[@]’ que coincida con su intensidad subjuntiva. Pero lo importante no es Q[@], sino ¬Q[@], de modo que sólo se requiere mostrar que contamos con una intensidad epistémica de ‘¬Q[@]’ que coincide con su intensidad subjuntiva. Siendo ese el caso, basta con que haya ocasiones en que podamos tener certeza de que dos experiencias particulares (o dos qualia) son distintas o que podamos distinguir a una experiencia de su ausencia, lo cual es fácil de conceder. Por ejemplo, si bien podríamos confundir dos experiencias de rojo ligeramente distintas o incluso una de rojo y una de naranja, sería extremadamente anómalo (por no decir que imposible, dado nuestro tipo de acceso) confundir ciertas experiencias de rojo con ciertas experiencias de verde o con la ausencia de experiencia⁵¹. Entonces, aunque tal vez podríamos confundir a Q[@] con alguna otra cosa que sólo se le pareciese, podemos idear una que claramente no sea Q[@], es decir, que sea ¬Q[@]: por ejemplo, podemos imaginar que hubiese una mancha verde o un hueco en nuestro campo visual, donde claramente no hay tales, o que en cierto momento en que de hecho tenemos una experiencia no tuviésemos ninguna. Así que podemos tener la certeza de que es posible ejecutar el argumento sin confundir a ¬Q[@] con algo que no lo sea: es decir, sí podemos lograr que las intensiones epistémica y subjuntiva de ‘¬Q[@]’ coincidan.⁵²

51 Nótese que con ‘experiencia de rojo’ me refiero a la sensación subjetiva que alguien asocia a rojo, no al hecho de estar percibiendo algo rojo. De lo contrario, claramente podríamos confundir una experiencia de rojo con una que no lo es: por ejemplo, podríamos confundir el rojo visto directamente con el blanco (no-rojo) visto a través de unas gafas rojas. Pero debe ser claro que casos como ese no son relevantes, porque en ellos no se está confundiendo lo que importa. Algo análogo deberá pensarse para cualquier otra experiencia.

52 Alguien podría objetar que, dado que la consciencia fenoménica afecta nuestro comportamiento, no cualquier ¬Q[@] es compatible con P[@]. Por ejemplo, si simplemente imagino un duplicado del momento actual salvo porque hay una mancha verde cubriendo mi campo visual, será razonable pensar que en ese escenario no podría escribir esto, porque no sabría qué teclas estaría presionando: es decir, la parte fenoménica de ese escenario desviaría mi conducta, lo cual haría que no se cumpliera P[@]. En respuesta, podría proponerse una ¬Q[@] basada en fenomenología invertida o recorrida, dado que es razonable

El caso de lo físico es más complicado. Chalmers (2005) sostiene, apoyándose en Stoljar (2001), que hay al menos dos maneras de entender lo físico: una apegada a la ciencia actual y otra que alude a nuestra interacción cotidiana e histórica con objetos físicos paradigmáticos⁵³. En cualquiera de los dos casos, P@ es una descripción exhaustiva de la parte de @ conformada por cierto tipo de objetos que hemos conocido independientemente de la ciencia y hemos bautizado como ‘objetos físicos’. De hecho, ambas posturas reconocen que la ciencia actual es nuestro mejor intento de describir ese tipo de objetos. La diferencia radica en que la primera sostiene que la ciencia provee herramientas suficientes para describirlos exhaustivamente, mientras que la segunda lo deja abierto. Siendo ese el caso, sólo la primera alternativa ofrece una manera clara de caracterizar a los objetos físicos: una que emplea exclusivamente las nociones científicas⁵⁴.

Pero Chalmers sostiene que es justamente esa manera de caracterizar a lo físico la que parece inadecuada para capturar la naturaleza de la consciencia fenoménica, debido a que la ciencia física actual habla exclusivamente en términos de relaciones (y estructuras⁵⁵) matemáticas y espacio-temporales⁵⁶ y lo fenoménico parece rebasar cualquier descripción de ese tipo (aun en el caso ideal)⁵⁷: por ejemplo, no parece posible explicar

pensar que los entes resultantes podrían ser funcionalmente idénticos a los originales (ver Shoemaker (1982), Block (1990) y Byrne (2004/15)).

- 53 Según Stoljar (2001), hay dos maneras de entender lo físico, cada una de las cuales da pie a un tipo de fisicismo. Por un lado está el *fisicismo basado en teorías*, según el cual en @ sólo existen los objetos postulados por la ciencia, caracterizados como ella los describe. Por otro lado, el *fisicismo basado en objetos* sostiene que en @ sólo existen objetos del mismo tipo que aquellos que, de manera ostensiva, hemos bautizado como ‘objetos físicos’, sea lo que sea que los caracterice.
- 54 Es decir, sólo la ciencia ofrece una manera clara de definir las intensiones epistémicas de ‘físico’ y ‘P@’, que pueden o no coincidir con sus respectivas intensiones subjuntivas, que, en contraste, más bien dependen de la relación cotidiana e histórica entre el ser humano y los objetos físicos. Siguiendo la nota 52, podemos decir que según el fisicismo basado en teorías, las intensiones epistémicas resultantes coinciden con las subjuntivas, mientras que el fisicismo basado en objetos está abierto a que no lo hagan.
- 55 En lo sucesivo deberá entenderse que cuando hablo de relaciones incluyo estructuras, que son conjuntos de relaciones.
- 56 Se entiende que las entidades fundamentales que reconoce la física no son sino objetos que tienen tales y cuales propiedades matemáticas y espacio-temporales, que incluyen su comportamiento y posibles efectos sobre su entorno.
- 57 Feigl (1958), ha criticado diversas maneras de caracterizar a lo físico en contraste con la consciencia fenoménica; a saber: lo objetivo contra lo subjetivo, lo espacial contra lo no espacial, lo cuantitativo contra lo cualitativo, lo teleológico contra lo no-teleológico, lo no-mnésico contra lo mnésico, lo

exhaustivamente cómo es una experiencia de verde hablando sólo de lo matemático y espacio-temporal. Por más que podamos apreciar relaciones matemáticas y espacio-temporales en la experiencia e incluso situarla dentro de nuestro cuerpo y a éste dentro del mundo físico, parece que hay algo más en ella: los qualia y el cómo se nos presentan. Es decir, describir qué qualia se aprecia dónde y en qué momento no basta para explicar cómo es ni cómo se aparece ante nosotros. Y lo mismo pasa si estudiamos desde fuera los estados físicos internos que supuestamente dan pie a la vida mental: lo único que encontramos son relaciones matemáticas y espacio-temporales, no cómo es que ellas dan pie a experiencias en primera persona como las conocemos introspectivamente. En otras palabras, una manera de e-concebir lo físico que claramente genera una brecha epistémica entre lo físico y lo fenoménico y, por consiguiente, permite e-concebir idealmente $P^{@*} \& \neg Q^{@}$, es aquella según la cual lo físico se reduce a relaciones matemáticas y espacio-temporales, que es justamente la propia de la ciencia actual, según Chalmers. Es decir, si como intensión epistémica de ‘ $P^{@}$ ’ se emplea la descripción de $P^{@}$ que actualmente provee la ciencia y se toman las medidas adecuadas respecto a ‘ $\neg Q^{@}$ ’ (indicadas previamente), resultando en una intensión epistémica de ‘ $P^{@*} \& \neg Q^{@}$ ’ que llamaré ‘CQ’, $P^{@*} \& \neg Q^{@}$ es e-concebible idealmente bajo CQ. Entonces sólo restaría averiguar si CQ coincide con la intensión subjuntiva de ‘ $P^{@*} \& \neg Q^{@}$ ’, para saber si es posible dar el paso de la CQ-concebibilidad de $P^{@*} \& \neg Q^{@}$ a su s-posibilidad y, de ese modo, refutar al fisicismo. Dado que parece ser que las intensiones de ‘ $\neg Q^{@}$ ’ sí coinciden, la clave parece estar en si la caracterización de $P^{@}$ que provee la ciencia actual coincide o no con la intensión subjuntiva de ‘ $P^{@}$ ’, es decir, en si dicha caracterización es o no suficiente para capturar la parte física de @.

atomístico contra lo holístico, lo composicional contra lo emergente y lo no-intencional contra lo intencional. En general señala que no es claro que alguna de las características en cuestión sea exclusiva de lo físico o de la consciencia fenoménica. Pero el punto no es que lo matemático y espacio-temporal sea exclusivo de lo físico, sino que es exhaustivo: es razonable creer que lo físico se reduce a lo matemático y espacio-temporal, mientras la consciencia fenoménica parece requerir de algo más. Sin embargo, para hacer más justicia a Feigl, se podría decir que esa otra cosa es lo cualitativo o lo subjetivo y que no es claro que lo físico no sea también cualitativo o capaz de explicar lo subjetivo. Ello también serviría para atacar la distinción entre físico y fenoménico como relacional (poliádico) contra no-relacional (monádico) o extrínseco contra intrínseco, también empleadas por Chalmers (2005), ya que él entiende que lo espacio-temporal es relacional o extrínseco y la parte de la consciencia fenoménica que queda fuera es intrínseca. Más adelante retomaré algunas de estas ideas, pero de momento lo que me interesa es ilustrar el problema.

Así que, modificando ligeramente la versión de Chalmers (2005)⁵⁸, el argumento resultante puede plantearse del siguiente modo (Arg3):

- P1_{A3}) $P^{@*} \& \neg Q^{@}$ es e-concebible idealmente bajo CQ.
- P2_{A3}) Si algo es e-concebible idealmente bajo x, es e-posible bajo x.
- P3_{A3}) Si $P^{@*} \& \neg Q^{@}$ es e-posible bajo CQ, $P^{@*} \& \neg Q^{@}$ es s-posible o la caracterización científica actual de $P^{@}$ es insuficiente.
- P4_{A3}) Si $P^{@*} \& \neg Q^{@}$ es s-posible, el fisicismo es falso.
- ∴ C1_{A3}) El fisicismo es falso o la caracterización científica actual de $P^{@}$ es insuficiente.

El argumento es válido y ya se ha argumentado a favor de sus premisas. En cuanto a P1_{A3}, no es más que una versión de la brecha epistémica, reforzada por la imagen de la ciencia que ofrece Chalmers⁵⁹, así que ya se ha argumentado a su favor. Por otro lado, P2_{A3} es parte del racionalismo modal, mismo que no pondré en duda en este escrito⁶⁰. Respecto a P4_{A3}, es

58 En el argumento de Chalmers (2005), las premisas son ligeramente distintas, pero en el fondo capturo las mismas ideas. Su primera premisa es que $P^{@*} \& \neg Q^{@}$ es concebible (siempre idealmente). Es decir, lo es en algún sentido, aunque sea en el más débil: e-concebible. Pero la intensión epistémica de ' $P^{@}$ ' con la que trabaja todo el tiempo es la interpretación matemática y espacio-temporal de la ciencia actual y para ' $\neg Q^{@}$ ' se concede que ambas intensiones coinciden, lo cual conduce a P1_{A3}. Su segunda premisa es que si $P^{@*} \& \neg Q^{@}$ es concebible, $P^{@*} \& \neg Q^{@}$ es e-posible, que no es más que un caso particular de P2_{A3}, añadiendo la especificación de que hay una caracterización en juego y de que la concebibilidad en cuestión es e-concebibilidad. Su tercera premisa es que si $P^{@*} \& \neg Q^{@}$ es e-posible, $P^{@*} \& \neg Q^{@}$ es s-posible o es verdad el monismo russelliano (que explico más adelante). Pero, para cualquier x, decir que de la e-posibilidad de x se sigue su s-posibilidad no es otra cosa que decir que las intensiones epistémica y subjuntiva de 'x' coinciden, que a su vez equivale a decir que la caracterización de x que se está tomando en cuenta es suficiente. Por otro lado, como explico más adelante, decir que la caracterización de ' $P^{@}$ ' que toma en cuenta la ciencia no es suficiente es la idea de fondo del monismo russelliano. Así que P3_{A3}, que además es trivialmente verdadera, es la idea de fondo de la tercera premisa de Chalmers. Su cuarta premisa es P4_{A3}, así que no hay nada más que decir ahí. Por último, la conclusión de Chalmers es que *el fisicismo es falso o es verdad el monismo russelliano*. Cuando hablé de P3_{A3} mencioné lo suficiente para ver cómo esa conclusión se vincula con C1_{A3}, pero de todos modos lo retomo a detalle más adelante.

59 Rechazar el argumento negando P1_{A3} (es decir, negando que existe la brecha epistémica), conduce a lo que Chalmers (2002a) llama 'fiscicismo tipo A'. Él señala como algunos representantes de este tipo de fisicismo a Dennett (1991), Dretske (1995), Harman (1990), Lewis (1988), Rey (1995) y Ryle (1949). También Rosenthal (1976) puede ser incluido entre ellos.

60 Para una crítica en esa dirección, ver Reynoso (2017). Ahora, aceptar P1_{A3} pero negar RM lleva a lo que Chalmers (2002a) denomina 'fiscicismo tipo B'. Él considera fisicistas de este tipo a Block & Stalnaker (1999), Hill (1997), Levine (1983), Loar (1997), Lycan (1996), Papineau (1993), Perry (2001) y Tye

verdad por definición del fisicismo. Por último, PSE se sigue de la idea de que CQ está basada en una intensión epistémica de ‘ $\neg Q^@$ ’ que coincide con su intensión subjuntiva: entonces la única manera en que CQ podría ser insuficiente para capturar $P^{@*} \& \neg Q^@$ es que la caracterización científica actual de $P^@$ sea insuficiente para capturar $P^@$. Así que el argumento parece sólido.

Ahora, respecto a la conclusión, Chalmers (2005) señala que una manera de hacer sentido del segundo disyunto es la de Russell (1927): sólo tenemos acceso a las propiedades de lo físico en la medida en que nos afectan, por lo que perdemos de vista su naturaleza intrínseca y categórica, quedándonos sólo con la extrínseca (o relacional) y disposicional (ver también Stoljar (2001)). Siendo ese el caso, una manera de salvar al fisicismo (rechazando el primer disyunto de $C1_{A3}$ pero aceptando el segundo) es sosteniendo que es un *monismo russelliano*, según el cual lo físico es cierto tipo de cosa que científicamente entendemos en términos de propiedades extrínsecas y disposicionales, pero cuya naturaleza puede no quedar enteramente capturada de ese modo, ya que tal vez posee más propiedades (como las categóricas)⁶¹. Por eso es que la caracterización científica actual de $P^@$ es insuficiente (porque es incompleta, a pesar de no ser errada) y, como consecuencia, algo puede parecer físico sin serlo, es decir, verificar ‘ $P^{@*} \& \neg Q^@$ ’ sin satisfacerlo. La idea es que la parte oculta también podría ser física, a pesar de no ser contemplada por la ciencia actual, y podría ser lo que determina a la consciencia fenoménica. En tal caso, $P^{@*}$ sí implicaría necesariamente a $Q^@$, de modo que $P^{@*} \& \neg Q^@$ sería incoherente, a pesar de que la ciencia actual no logra entender el por qué de ello. En ausencia de otra forma en que pueda ser falso el primer disyunto de $C1_{A3}$ sin que lo sea el segundo, el monismo russelliano resulta ser la única manera de salvar al fisicismo. Por ello, la conclusión a la que llega Chalmers (2005) es que ($C2_{A3}$) el fisicismo es falso o es verdad el monismo

(1995).

61 El monismo russelliano es lo que Chalmers (2002a) llama ‘monismo tipo F’ o ‘panprotopsiquismo’, también asociado al *dualismo de propiedades* y al *monismo neutral* (ver Stubenberg (2016)). Chalmers considera monistas-F a Russell (1927), Feigl (1958), Maxwell (1979), Lockwood (1989), Chalmers (1996), Griffin (1998), Rosenberg (1997), Strawson (2000) y Stoljar (2001). Incluso el *misterianismo* de McGinn (1989, 1999) puede ser entendido como una tesis monista-F. Como monistas neutrales podemos incluir (además de Russell) a Mach (1886, 1894, 1905), James (1912), Banks (2014), Heil (2013), Nagel (2012) y Sayre (1976), entre otros. Incluso Spinoza (1677) y Hume (1739) pueden ser entendidos como antecesores del monismo neutral (ver Stubenberg (2016)).

russelliano.

Pero los físicos suelen estar interesados en sostener que los métodos de la ciencia actual son aptos, al menos en principio, para dar cuenta de @ en su totalidad, por lo que el monismo russelliano puede no servirles como respuesta. El problema destacado por el argumento parece no ser simplemente que falte algo por ajustar en las descripciones científicas, sino que las bases mismas con las que se construyen actualmente son defectuosas: parece que ni en el caso ideal son suficientes para cerrar la brecha. Es decir, si no se cambian esas bases, ningún conjunto de ajustes bastará para capturar lo que falta de manera que no dé la impresión de que hay un hueco indeseable. Puesto en otras palabras, los métodos de la ciencia actual parecen no ser aptos, ni siquiera en principio, para dar cuenta de @ en su totalidad sin que quede la sensación de que algo falta, porque bajo ninguna circunstancia podrán capturar lo fenoménico de una manera que no sea indirecta. Así que todo fisicismo que se comprometa exclusivamente con las herramientas de la ciencia actual será defectuoso, porque no será capaz de explicar lo fenoménico en términos físicos que lo capturen de manera transparente.⁶²

Ahora, regresando sobre el argumento Arg2, la brecha epistémica tipo G entre las verdades sobre x y las verdades sobre z puede ser entendida como la situación en que $x \& \neg z$ es e-concebible idealmente haciendo uso de una intensión epistémica de ' $x \& \neg z$ ' que coincide con su intensión subjuntiva. Que eso sea el caso respecto a $P^{@*} \& \neg Q^{@}$ equivale a que sea verdad la premisa $P1_{A3}$ y el primer disyunto del consecuente de PSE, es decir, que $P^{@*} \& \neg Q^{@}$ sea e-concebible bajo CQ y CQ coincida con la intensión subjuntiva de ' $P^{@*} \& \neg Q^{@}$ '. Entonces y sólo entonces, según Chalmers (2005), se justifica inferir que

62 Aquí cabe mencionar el *dilema de Hempel* (Hempel (1969)): dado que las teorías científicas actuales son seguramente falsas (en el sentido de ser erradas o incompletas) y es trivial que las teorías científicas ideales serían verdaderas pero en la actualidad no sabemos qué deberían incluir, definir al fisicismo haciendo referencia a alguna de ellas conduce a que sea falso o trivialmente verdadero pero poco informativo, respectivamente. Sin embargo, siguiendo a Stoljar (2001/15), ello es congruente con que las teorías científicas actuales sean parcialmente acertadas y los métodos científicos actuales sean en principio adecuados para dar cuenta de lo físico (y de @) (ver también Smart (1978), Lewis (1994) y Melnyk (1997, 2003)). Lo que el argumento de Chalmers (2005) pretende evidenciar es la ineptitud de los métodos científicos actuales para dar cuenta de @ en su totalidad, no su aptitud para dar cuenta de lo físico (salvo si se estipula que todo lo que hay en @ es físico, sea lo que sea).

$P^{*} \& \neg Q^{*}$ es s-posible. También cabe señalar que, aunque Arg2 podría no depender de Arg3, porque podría haber otra manera de sostener que hay una brecha tipo G entre P^{*} y Q^{*} , refutar Arg3 de hecho elimina una manera de sostener Arg2, de modo que sí lo debilita un poco. Además, es posible que la forma en que se refute Arg3 evidencie que de hecho no hay brecha tipo G. De hecho, justamente eso es lo que trataré de hacer en este escrito. Para lograrlo, revisaré distintas teorías sobre la naturaleza de la consciencia fenoménica y cómo es que cada una pretende resolver (o no) el problema, evaluando en qué podrían ser acertadas y en qué no, para al final ofrecer mi propia teoría.

1.5. Sumario 1

En este capítulo presenté lo que busco ayudar a explicar y bajo qué marco teórico pienso hacerlo, además de que expliqué por qué esa tarea parece especialmente difícil. Más específicamente, dije que quiero ayudar a explicar la *consciencia fenoménica* bajo el marco *fisicista* y respetando la tesis del *internismo fenoménico*. Además, expliqué por qué la consciencia fenoménica parece no tener lugar dentro ese marco: porque parece haber una desconexión conceptual entre lo físico y lo fenoménico, que impide establecer un puente transparente entre nuestro entendimiento de ambas cosas. Por ende, hay quienes sostienen que no es posible explicar una en términos de la otra.

En el transcurso de la explicación, ofrecí la definición general de consciencia fenoménica con la que trabajaré en este escrito:

CF) La consciencia fenoménica es idéntica a *presentación tipo E de qualia*.

Además, recalqué que esa definición es neutral respecto a los aspectos realmente controversiales sobre la consciencia fenoménica, de modo que es compatible con las principales posturas sobre ella. En los próximos capítulos exploraré las distintas maneras de entender la presentación tipo E (capítulo 2) y los qualia (capítulo 3), lo cual permitirá generar una taxonomía general sobre todas las posibles posturas sobre la naturaleza de la

consciencia fenoménica y apreciar de mejor manera sus similitudes y diferencias, así como sus virtudes y defectos. Ello preparará el terreno para ofrecer mi propia teoría sobre ella (capítulo 4), libre de los problemas fatales que aquejan a la competencia.

Capítulo 2

Las posibles respuestas I: teorías sobre la presentación tipo E

2. Las posibles respuestas I: teorías sobre la presentación tipo E

En el capítulo anterior esboqué qué es lo que quiero contribuir a explicar: la consciencia fenoménica. Además expliqué por qué es problemática para el fisicismo y especificué que la respuesta que busco es una que sea compatible con el internismo fenoménico. A partir de este capítulo me centraré en encontrar tal respuesta. Con ese objetivo, conviene recordar que sugerí entender a la consciencia fenoménica como *presentación tipo E de qualia*, pero manteniéndome mayormente neutral respecto qué son la presentación tipo E y los qualia. Lo que haré a partir de ahora será abandonar esa neutralidad, porque estudiaré las diversas maneras de entender esos dos elementos y decidiré cuáles son las más viables. Con ello en mente, dedicaré este capítulo a revisar los únicos dos candidatos a fungir como presentación tipo E: la percatación directa (o *acquaintance*) (sección 2.2) y algún tipo de percatación indirecta (representacional) (sección 2.3). Pero abriré con una sección (2.1) dedicada a elucidar la distinción fundamental entre ambas opciones. Eventualmente me decantaré por la segunda alternativa, lo cual me comprometerá con un tipo de *representacionismo fuerte*, es decir, con la idea de que tener consciencia fenoménica es idéntico a tener ciertas propiedades representacionales. Hacia el final detallaré con más precisión qué tipo de representacionismo fuerte considero más plausible (sección 2.4), pero será hasta el capítulo 4 que terminaré de definirlo.

Una razón para rechazar la idea de que la presentación tipo E es *acquaintance* se deriva de que algunas teorías que aluden a él (como el *realismo directo*) sostienen que los qualia son objetos externos, lo cual hace que tener *acquaintance* con ellos dependa de que de hecho se encuentren presentes ciertos objetos externos y, por ende, conduce a concluir que la consciencia fenoménica es una propiedad extrínseca, lo cual entra en conflicto con el internismo fenoménico. Sin embargo, también hay posturas internistas que se apoyan en la noción de *acquaintance* (como las *teorías de los sense-data*). Siendo ese el caso, más bien me apoyaré en una razón más general en contra de entender la presentación tipo E como *acquaintance*: que éste no es naturalizable, lo cual entra en tensión con el fisicismo. Sin embargo, esa objeción sólo se podrá mantener en pie en la medida en que la alternativa sí

sea naturalizable, así que dedicaré una buena porción del capítulo (sección 2.3.3) a explicar cómo naturalizar la representación. En particular presentaré dos tipos de teorías al respecto: las *etiológicas* y las *disposicionales*. Las primeras se apoyan en la historia previa del ente que posee los estados en cuestión, mientras que las segunda se fija sólo en las propiedades internas suyas. Mi compromiso con el internismo fenoménico me obligará a sostener que la presentación tipo E debe ser un tipo de representación disposicional, porque la historia previa de algo siempre es una propiedad extrínseca suya.

Comencemos, entonces, presentando la disyuntiva principal sobre cómo entender la presentación tipo E.

2.1. En busca de la distinción clave

Como mencioné en la sección 1.1, una posible diferencia entre ambos tipos de percatación es que el primero no requiere de un intermediario entre el ente que se percata y aquello que de lo que se percata, mientras que el segundo sí lo requiere. Más específicamente, en el caso de la percatación indirecta suele hablarse de que *s* se percata de *x* *mediante* *z*, mientras que en la directa no hay un *z* que habilite o module la percatación, sino que *s* está directamente en contacto con *x*. Sin embargo, esa idea puede no ser de mucha ayuda si se tiene en cuenta que *z* suele ser parte de *s*, en cuyo caso no es claro si la situación debería describirse como un contacto directo entre *s* y *x* o como uno mediado por *z*. Además, incluso quienes hablan de *acquaintance* suelen entenderla como algo dependiente (al menos en parte) de los procesos internos de *s*, dentro de los cuales siempre se pueden identificar estados como *z*. Por ejemplo, ninguna de las dos posturas niega que la visión humana involucra una compleja red causal, tanto externa como interna: los objetos externos emiten o reflejan luz, esta viaja hasta el ojo, atraviesa la córnea y la pupila, es enfocada por el cristalino e impacta la retina, estimulando las células que hay en ella, que a su vez estimulan el nervio óptico y después diversas zonas del cerebro, por medio de muchas conexiones neuronales, dando pie en algún punto a la experiencia visual del sujeto. Por un lado, podemos entender que el ente que se percata está situado en un extremo de la

cadena causal y que lo que se percibe está en el otro, lo cual conduce a pensar que la visión humana verídica es un ejemplo de percatación indirecta. Pero también podemos entender que el ente que se percata es el humano entero y que aquello de lo que se percata es todo lo que está de sus ojos hacia afuera, en cuyo caso la visión humana verídica podría catalogarse como percatación directa.

Ante situaciones como esa, lo razonable es buscar otra manera de especificar en qué sentido una percatación cuenta como directa o como indirecta. Una opción es notar que el contacto entre s y x no puede ser directo si x no está presente ante s . En contraste, s sí podría tener z (o estar en contacto con z) en ausencia de x . Entonces podríamos fijarnos, más bien, en si el tipo de percatación relevante depende exclusivamente de que s tenga z o si necesariamente requiere de la presencia de x . Por ejemplo, ¿para tener una experiencia visual es o no indispensable la presencia de los qualia presentados por ella? Si su presencia es indispensable, entonces podría tratarse de *acquaintance* con ellos, mientras que si no lo es, entonces no podría más que ser un tipo de representación. Es decir, la pregunta es, entonces, si la presentación tipo E de x puede o no tener lugar en ausencia de x . En otras palabras, la diferencia clave es si la presentación tipo E es falible o infalible: si no puede tener lugar en ausencia de x , entonces es infalible y se trata de *acquaintance*, mientras que si sí puede tener lugar en ausencia de x , entonces es falible y se trata de un tipo de representación.

Esa estrategia parece prometedora y es a la que me apegaré principalmente, pero cabe señalar que también tiene sus propias dificultades. Por ejemplo, es razonable pensar que en ocasiones z y x podrían ser un mismo objeto, en cuyo caso s no podría tener z en ausencia de x . Ello sería el caso, por ejemplo, si se piensa que los qualia son propiedades de las experiencias mismas y que además juegan algún papel en que el sujeto se percate de ellos. Entonces habría que decir más para determinar si esos casos son *acquaintance* o representación. Pero de momento pospondré ese problema y utilizaré la falibilidad como guía.

Otra cosa que hay que tener en cuenta es que bien podría sostenerse que algunos casos de presentación tipo E son *acquaintance* y otros no. Por ejemplo, podría pensarse que algunos tipos de experiencias son *acquaintance* y otros representación. O bien, que dos experiencias subjetivamente indistinguibles pueden ser generadas mediante dos mecanismos fundamentalmente distintos, de los cuales uno es *acquaintance* y el otro representación. Conforme vaya siendo pertinente, llamaré la atención sobre algunas teorías de ese tipo. Por ahora, me centraré por separado en los dos tipos de percatación, las posturas que los defienden y los argumentos en que se apoyan. Pero antes de eso, hay que recordar que las diversas teorías sobre la consciencia fenoménica no meramente se pronuncian sobre qué es la presentación tipo E, sino también sobre qué son los qualia, lo cual es crucial para decidir si son o no plausibles. De momento sólo mencionaré de pasada ese factor y ahondaré en él hasta el capítulo 3.

2.2. *Acquaintance*

Sintetizando las ideas recién mencionadas y retomando lo que se dijo en la sección 1.1, podemos decir que el *acquaintance* con x es una percatación directa y transparente de x, que revela la esencia de x y garantiza no sólo que x se encuentra presente, sino también que x es justo como parece ser. Teniendo eso en cuenta, podemos señalar dos razones a favor de que la presentación tipo E es *acquaintance*: una *fenomenológica* y otra *epistemológica*. La primera se apoya en la sencilla observación de que a primera vista la presentación tipo E parece ser *acquaintance*, lo cual se refleja en el hecho de que cotidianamente nos comportamos justo como si eso fuese, porque confiamos (casi) ciegamente en que el mundo es como nos lo presentan nuestras experiencias. Siendo ese el caso, la carga de la prueba está en quienes opinan que la presentación tipo E no es *acquaintance*. Pero incluso puede decirse que esa es una manera demasiado mesurada de exponer la situación real, porque la presentación tipo E no meramente parece ser *acquaintance*, sino que parece serlo con tanta fuerza que hasta parece irracional sugerir que de hecho lo es. En otras palabras, la presentación tipo E tiene tal fuerza de convencimiento que parece absurdo dudar de la presencia de aquello que nos presenta

(cuando nos lo presenta). En el caso particular de la consciencia fenoménica, por ejemplo, parece inadmisibile negar que los qualia presentados por ella se encuentran ante nosotros: tal vez podemos cuestionar la naturaleza última de los qualia, pero de que están presentes, lo están. La fuerza y ubicuidad de esa intuición es tal que quienes hablan de *acquaintance* suelen no argumentar a favor de su existencia y fiabilidad, sino meramente tratarlo como una verdad evidente (ver Grote (1865), Helmholtz (1868), James (1890), Russell (1905, 1910/11), Fumerton (1995, 2001, 2016), BonJour (2001, 2003) y Chalmers (2003, 2010), por ejemplo).

La segunda razón, es decir, la epistemológica, parte de la idea de que nuestras experiencias fenoménicas tienen un papel privilegiado en lo que respecta a justificar nuestras creencias sobre el mundo, mismo que no podrían desempeñar si fuesen falibles. En otras palabras, la idea es que nuestras experiencias fenoménicas son fuentes fiables de conocimiento sólo en la medida en que la presentación tipo E sea *acquaintance*; y, dado que de hecho parecen tener un papel clave en la generación de nuestro conocimiento, lo más razonable es conceder que la presentación tipo E de hecho es *acquaintance*. Sobre esa línea, hay quienes recurren a la noción de *acquaintance* para sostener algún tipo de *fundacionismo* sobre la justificación epistémica, en el sentido de que la consideran como la fuente primordial de creencias justificadas (ver Russell (1905, 1910/11), Fumerton (1995, 2001, 2016) y BonJour (2001, 2003), por ejemplo). En ese sentido, cuestionar que la presentación tipo E sea *acquaintance* podría equivaler a minar las bases de todo nuestro conocimiento.

Los dos principales tipos de posturas que entienden la presentación tipo E como *acquaintance* son el *realismo directo* y las *teorías de los sense-data*. Al revisarlas asomarán algunas maneras de entender a los qualia. Si bien ese tema pertenece más bien al capítulo 3, mencionar algunas opciones desde ahora servirá para aterrizar la noción de *acquaintance* en ejemplos concretos.

2.2.1. Realismo directo

Para los realistas directos (también llamados ‘realistas ingenuos’ o ‘realistas del sentido común’⁶³), la percepción verídica es *acquaintance* con alguna porción del mundo real y, por ende, las experiencias que la acompañan también son justo eso. Por ejemplo, la experiencia visual que tengo al ver una manzana verde es una percatación directa de la manzana y la que tengo al oler café es una percatación directa del café. Más específicamente, cada experiencia parece ser una percatación directa de cierta gama de propiedades del mundo: la experiencia visual parece una percatación directa de propiedades visibles y la olfativa de propiedades olfatibles. De hecho, fijarse en los qualia que caracterizan a cada experiencia no parece ser otra cosa que fijarse en las propiedades del mundo que esta nos revela: el quale que experimento al ver el verde de la manzana no parece ser otra cosa que el verde mismo de la manzana y el que experimento al oler café no parece ser sino el olor mismo del café. Algunos realistas directos son Hinton (1967, 1973), Snowdon (1979, 1980/81, 1990, 2005), McDowell (1982, 1987, 2013), Martin (1997, 1998, 2002, 2004, 2006), Putnam (1999), Campbell (2002a, 200b, 2005a, 2009, 2011), Travis (2004, 2013), Brewer (2004, 2006, 2008, 2011, 2017a, 2017b), Sturgeon (2006, 2008), Byrne & Logue (2008), Fish (2008, 2009), Logue (2012a, 2012b), Hellie (2013), Soteriou (2013) y Campbell & Cassam (2014).

Incluso puede extenderse la misma idea hacia algunos otros tipos de experiencias, como las emociones, dolores y orgasmos: sólo son versiones propioceptivas o introspectivas del mismo principio fundamental. Sin embargo, no en todos los casos es igual de claro qué tipo de propiedades del mundo parecen ser los qualia en cuestión ni dónde parecen estar situados: por ejemplo, si bien es bastante claro que los qualia de color parecen ser propiedades físicas situadas en la superficie de los objetos externos, no es claro qué

63 Una posible distinción entre el realismo directo y el realismo ingenuo (o del sentido común) es que el primero hace énfasis en que la presentación tipo E es *acquaintance*, mientras que el segundo se centra en afirmar que los qualia son propiedades reales del mundo externo. Entendiendo así las cosas, las teorías de los *sense-data* (sección 2.2.2) podrían ser entendidas como un tipo no convencional de realismo directo: uno que se disocia del realismo ingenuo y entiende a los qualia como propiedades de la mente. Sin embargo, evitaré ese uso porque no es el más común en la literatura filosófica.

propiedades del mundo corresponden a los qualia que experimentamos al estar tristes o tener un orgasmo, ni dónde parecen estar situados. Además, si se es lo suficientemente laxo, cualquier teoría que entienda a la presentación tipo E como *acquaintance* podría contar como un tipo de realismo directo. Por eso, mejor me limitaré a hablar del realismo directo sobre la percepción verídica de objetos externos, como la visión, la audición, el olfato, el gusto y el tacto.

El realismo directo se apoya principalmente en dos observaciones, una fenomenológica y otra epistemológica, que son versiones particulares de las razones generales que mencioné previamente a favor de que la presentación tipo E sea *acquaintance*. La primera consiste en señalar que lo que la percepción verídica parece ser a simple vista es una “ventana transparente” hacia alguna porción del mundo, es decir, una percatación directa de (o *acquaintance* con) ella. Siendo ese el caso, parece que lo más natural es concluir que eso es justamente lo que es. A fin de cuentas, cotidianamente nos comportamos justo como si así fuese y es sólo tras sofisticada reflexión filosófica que llegamos a cuestionarlo. Es decir, el realismo directo trata de salvar lo que el sentido común sugiere que es la percepción verídica, por lo que la carga de la prueba recae en sus oponentes (ver Hinton (1967, 1973) y Martin (1997, 1998, 2002, 2004, 2006)). El obstáculo más obvio en contra de esa idea es que resulta igual o más natural entender que las experiencias verídicas tienen exactamente la misma explicación que sus contrapartes no-verídicas (alucinaciones o sueños), de las cuales difícilmente diríamos que son *acquaintance* con algo externo⁶⁴. Ante esa situación, los realistas ingenuos suelen simplemente insistir en que lo que explica las experiencias no-verídicas es fundamentalmente distinto de lo que explica las verídicas, lo cual los convierte en *disyuntivistas*⁶⁵. La idea es que dos experiencias subjetivamente

64 También es posible hablar de alucinaciones en otro tipo de experiencias. El caso más claro es el de los dolores, debido a los casos de miembro fantasma, en los que pacientes amputados dicen seguir sintiendo el miembro faltante e incluso reportan dolores en él (ver Ramachandran & Hirstein (1998)). Por otra parte, si se entiende que las emociones son (al menos en parte) percatación de cambios fisiológicos (como sugieren James (1984) y Lange (1885)) o que los orgasmos parecen estar situados (al menos en parte) en las zonas erógenas, es posible imaginar casos en los que las experiencias correspondientes ocurren sin los cambios fisiológicos pertinentes en las zonas relevantes, debido a la estimulación del sistema nervioso en otros sitios.

65 Para un tratamiento general del disyuntivismo, incluyendo diversas maneras de dar cuenta de las alucinaciones y sueños y qué los distingue de la percepción verídica, ver Soteriou (2009/20).

indistinguibles (ya sea porque tienen el mismo carácter fenoménico o meramente porque no notamos que no lo tienen) pueden ser generadas por mecanismos radicalmente distintos, como el *acquaintance* y la representación. Incluso hay quienes ponen en duda que las experiencias no-verídicas realmente tengan carácter fenoménico: entonces, aun si son subjetivamente indistinguibles de su contraparte verídica, realmente no son fenoménicamente idénticas (ver Sturgeon (2006, 2008) y Fish (2008, 2009)).

La segunda observación (la epistemológica) consiste en señalar que las experiencias correspondientes a la percepción verídica tienen un papel privilegiado en lo que respecta a justificar nuestras creencias sobre el mundo, mismo que no podrían desempeñar si fuesen explicadas por lo mismo que explica a su contraparte no-verídica (ver McDowell (1982, 1987, 2013)). Por ende, deben ser explicadas por otra cosa que sí de cuenta de ese papel privilegiado. Y el candidato obvio es que sean *acquaintance* con el mundo real. Como es evidente, esa estrategia por sí sola también da pie al disyuntivismo.

Sin embargo, no todos los realistas directos son disyuntivistas. Johnston (2004), por ejemplo, sostiene que tanto la percepción verídica como la alucinación son *acquaintance* con las mismas propiedades reales, con la diferencia de que en el primer caso están instanciadas en el mundo real y en el segundo no. La idea es difícil de acomodar porque no es claro en qué sentido una propiedad no instanciada se encuentra realmente presente y, por ende, no es claro en qué sentido puede haber *acquaintance* con ella. Incluso podría parecer que sugiere que puede haber *acquaintance* con objetos inexistentes, lo cual es contradictorio. Pero es posible hacer sentido de la situación si se nota que Johnston también cuestiona la distinción entre percepción directa e indirecta, porque entiende que percatare de algo gracias a un intermediario no es percatare menos directamente de dicho algo, porque la propiedad de causar la presencia del intermediario le es tan propia como todas sus otras propiedades. Entonces podría pensarse que los qualia son propiedades del sujeto que en ocasiones también son propiedades de objetos externos, porque son causadas por ellos. Siendo ese el caso, los qualia siempre estarían instanciados, pero sólo en ocasiones por objetos externos. Entonces la distinción entre experiencias verídicas y no-

verídicas no sería más que la distinción entre las situaciones en que los qualia sólo son instanciados por el sujeto y aquellas en que también son instanciados por algo externo.

Una opción cercana a esa es meramente sostener que los qualia en cuestión de hecho no son algo externo, sino algo interno con lo que tenemos *acquaintance* tanto en las experiencias verídicas como durante su contraparte no-verídica, pero que a su vez empleamos para rastrear algo externo (representándolo). Por ejemplo, tal vez detectamos eventos externos en virtud de que tenemos *acquaintance* con los estados neuronales causados por ellos. Entonces la diferencia entre una experiencia verídica y una no-verídica sería meramente que en la primera están presentes ambos objetos (el interno y el externo), mientras que en la segunda sólo está el interno. Un problema tanto con esa opción como con la de Johnston (2004) es que comienzan a alejarse demasiado de lo que el sentido común sugiere que son los qualia, lo cual les retira parte del apoyo del que goza el realismo directo tradicional. Sin embargo, podría decirse que son más fieles al sentido común en otro aspecto, porque respetan la intuición de que tanto la experiencia verídica como su contraparte no-verídica son *acquaintance* con lo mismo. Por último, cabe señalar que basta con complementar esta última estrategia con algunas afirmaciones sobre la naturaleza de los qualia para convertir la postura resultante en una teoría de los *sense-data*. Así que conviene proceder a estudiar en qué consiste una teoría de ese tipo.

2.2.2. Teorías de los *sense-data*

Apoyándose en la idea de que las alucinaciones y sueños son *acquaintance* con los mismos qualia que sus contrapartes verídicas, los teóricos de los *sense-data* concluyen que estos no pueden ser las propiedades externas que parecen ser a simple vista, sino algo meramente mental y tal vez hasta no-físico, a lo cual llaman ‘sense-data’. Es decir, para ellos la presentación tipo E es *acquaintance* y los qualia son *sense-data*. Entre los autores que sostienen eso, se encuentran Russell (1912), Broad (1925), Price (1950), Ayer (1956), Jackson (1977), Casullo (1987), Robinson (1994), Garcia-Carpintero (2002), O’Shaughnessy (2003) e incluso Chalmers (1996, 2002a, 2003, 2005, 2006c, 2018) (con la

diferencia de que no llama ‘sense-data’ a los qualia). Según ese tipo de posturas, el quale que asocio con el verde no es el verde mismo (entendido como propiedad física externa), sino una mera apariencia que mi mente le asocia al verde y que consiste en algo más: en una suerte de substancia mental con la que mi mente le construye apariencias al mundo. La idea es que, a pesar de que los qualia no se encuentran afuera de la mente y tal vez ni siquiera son físicos, su presencia sí es sistemáticamente causada por eventos físicos externos y por eso es que sirve para detectarlos.⁶⁶

Hay más que decir sobre las teorías de los *sense-data* (y también sobre el realismo directo), pero lo reservaré para el capítulo 3, en el que me enfocaré en estudiar qué son los qualia. De momento procederé a presentar algunas razones en contra de entender la presentación tipo E como *acquaintance*.

2.2.3. Contra el *acquaintance*

Lo primero que podemos señalar es que el realismo directo (tradicional) sobre la percepción verídica de objetos externos es incompatible con el internismo fenoménico, porque sostiene que algunas experiencias fenoménicas literalmente tienen como ingrediente alguna porción del mundo externo, lo cual automáticamente las convierte en propiedades extrínsecas del sujeto de experiencia. En otras palabras, tener dichas experiencias necesariamente requiere que tener contacto o interacción con algo externo al momento de tenerlas, lo cual contradice la tesis de que toda experiencia depende exclusivamente de factores internos del sujeto. Por ejemplo, el realista directo sostendría que la experiencia que tengo al ver el verde de una manzana literalmente tiene como constituyente al verde de la manzana, porque involucra *acquaintance* con él. En cambio, el internista fenoménico sostendría que la experiencia en cuestión depende exclusivamente de factores internos del sujeto y que su relación con el verde de la manzana es contingente. Así que, dado que soy internista fenoménico, me veo obligado a rechazar el realismo directo sobre la percepción verídica de objetos externos.

66 Para un tratamiento general de las teorías de los *sense-data*, ver Huemer (2004/11) y Hatfield (2021).

Sin embargo, convendría decir algo más que meramente estipular la verdad del internismo fenoménico. Con ello en mente, podemos recordar que los realistas directos suelen conceder que dos experiencias subjetivamente indistinguibles, como una verídica y su contraparte alucinatoria, pueden tener distinta explicación. En el caso particular de la percepción verídica de algo externo, la versión verídica tendría una explicación externista (el *acquaintance* con algo externo) y la alucinatoria una explicación internista (el *acquaintance* con algo interno o algún tipo de representación). Además, en congruencia con lo que sugiere la evidencia científica, los realistas directos suelen conceder que el proceso interno que acompaña a cada una de las dos experiencias es básicamente el mismo, diferenciándose sólo por qué lo puso en marcha: un objeto externo en el caso verídico y algo interno en el alucinatorio. Es decir, conceden que el mecanismo responsable de la experiencia no-verídica tiene lugar tanto en la percepción verídica como en la no-verídica. Pero entonces surge un gran problema para el realista directo, porque cada uno de los dos mecanismos parece bastar por sí mismo para producir la experiencia relevante, lo cual conduce a una conclusión sumamente implausible: que cada episodio de percepción verídica está acompañado de dos experiencias, una verídica y una no-verídica (ver Johnston (2004)). Ante esa situación, parece razonable rechazar el realismo directo sobre la percepción verídica de objetos externos y sostener que sólo tiene lugar una experiencia: la que depende exclusivamente del mecanismo interno.

Una alternativa podría ser insistir en que el mecanismo interno no basta para generar una experiencia genuinamente fenoménica, sino meramente algo subjetivamente indistinguible de ella (como sugieren Sturgeon (2006, 2008) y Fish (2008, 2009)): si bien el sujeto reportaría tener la misma experiencia fenoménica con o sin el factor externo, sencillamente estaría equivocado. Sin embargo, puede señalarse que postular diferencias fenoménicas subjetivamente indetectables es sospechosamente *ad hoc* y que, aun perdonando eso, la idea de que las haya carece de sentido, porque el carácter fenoménico consiste justamente en el aspecto subjetivo de la experiencia. Es decir, si las diferencias en cuestión son subjetivamente indetectables, entonces no son fenoménicas. Así que es falso que puede

haber diferencias fenoménicas subjetivamente indetectables⁶⁷ y, por ende, la única opción que queda es rechazar el realismo directo sobre la percepción verídica de objetos externos. Además, cabe notar que no es claro en qué sentido tener como causa al objeto percibido permite convertir en *acquaintance* al proceso que de otro modo sería percatación indirecta o en experiencia fenoménica genuina a algo que meramente parece serlo: si bien es claro que bastaría para convertirlo en percepción verídica, parece que ello no iría acompañado de otra diferencia substancial.

Pero, a pesar de que esas son razones suficientes para rechazar el realismo directo (tradicional) sobre la percepción verídica de objetos externos, no bastan para descartar la posibilidad de que la presentación tipo E sea *acquaintance* en esos casos, porque las teorías de los *sense-data* y ciertos realismos directos no-tradicionales (como el de Johnston (2004)) sí son compatibles con el internismo fenoménico. Así que es indispensable recurrir a otras razones. Una opción es, por ejemplo, señalar que la idea de que los qualia en cuestión son internos es incompatible con cómo nos los presentan nuestras experiencias, lo cual a su vez es incompatible con que la presentación tipo E de ellos sea *acquaintance*, porque entonces es falso que los qualia son tal cual nos los presentan nuestras experiencias. Por ejemplo, los qualia de color son claramente presentados como algo externo, situado en la superficie de objetos externos, de modo que no hace sentido que el *acquaintance* con ello nos los presente como algo interno. Así que la presentación tipo E no puede ser *acquaintance* en esos casos. Uniendo esa idea con la expuesta previamente, podemos obtener el siguiente argumento (Arg4), donde ‘qualia de color’ puede ser substituído por cualquier otro tipo de qualia que sea externo según el realismo directo:

- P1_{A4}) Si la presentación tipo E es *acquaintance*, entonces los qualia de color son propiedades externas al sujeto de experiencia.
- P2_{A4}) Si los qualia de color son propiedades externas al sujeto de experiencia, entonces la presentación tipo E de qualia de color es una propiedad extrínseca del sujeto de experiencia.

67 Nótese que *indetectable* no es lo mismo que *no detectada*: bien podría ser que, a pesar de tratarse de diferencias detectables, el sujeto no las detecte.

P3_{A4}) La presentación tipo E de qualia de color no es una propiedad extrínseca del sujeto de experiencia (sino que le es intrínseca).

∴ C1_{A4}) La presentación tipo E no es *acquaintance*.

El argumento es válido, P3_{A4} es un caso particular del internismo fenoménico, P2_{A4} se sigue de la definición de *acquaintance* y P1_{A4} también, siempre y cuando se conceda que nuestras experiencias presentan a los qualia de color como algo externo, lo cual es bastante plausible, así que el argumento parece sólido. Sin embargo, si bien el realismo directo (tradicional) no puede escapar de él, otras teorías basadas en *acquaintance* sí pueden hacerlo. Por ejemplo, un teórico de los *sense-data* puede simplemente sostener que de hecho los qualia de color sí parecen ser internos: tal vez es sólo una mala costumbre entenderlos como algo externo. Si eso fuese verdad, entonces P1_{A4} sería falsa y el argumento quedaría refutado. Otra opción es, como parece hacer Johnston (2004), sostener que los qualia de color son tanto propiedades internas como externas, en el sentido de que son propiedades internas que sugieren una causa externa. Y respuestas análogas pueden ofrecerse para cualesquiera otros tipos de qualia. Además, aun aceptando el argumento para algunos tipos de experiencias, es posible sostener que en otros tipos sí hay *acquaintance*. Por todas esas razones, es conveniente buscar otra manera de argumentar en contra de que la presentación tipo E sea *acquaintance*. A continuación revisaré algunas estrategias más generales.

Con ello en mente, podemos comenzar señalando que en principio deberíamos estar abiertos a la posibilidad de que la presentación tipo E no sea *acquaintance*. Sobre esa línea, Sellars (1956) acuñó la expresión de ‘el mito de lo dado’ para referirse a la idea de que existe alguna fuente incuestionablemente fiable de conocimiento, como el *acquaintance*. Según él, en principio deberíamos estar abiertos a la posibilidad de que lo que sea que nuestras experiencias nos presenten en realidad no esté ahí. Con ello no se refiere meramente al hecho obvio de que existen cosas como las alucinaciones y sueños, que nos sugieren con bastante fuerza que el mundo es como de hecho no es, sino a que es posible que las experiencias en general no sean base suficiente para asegurar conocimiento acerca

de nada, ni siquiera de la presencia de qualia, entendidos como algo meramente mental. Por otra parte, Sellars (1956) también señala que, aun suponiendo que de hecho existiese algo como el *acquaintance*, no es claro cómo la infalibilidad que lo caracteriza podría conservarse al extraer su contenido, traducirlo y trasladarlo a otros estados, como las creencias. La razón es que esos procesos parecen requerir de la intervención de otras herramientas que sí son indirectas y falibles. Es decir, incluso suponiendo que existe el *acquaintance* y que consiste en una percatación infalible de algo, no parece posible formular con certeza acerca de qué es, porque para formularlo hay que sacrificar la infalibilidad. Como señala Frankish (2016), las opciones son quedarnos con una fuente de conocimiento infalible pero inerte o convertirla en falible. Siendo ese el caso, nada de lo que podamos extraer de la presentación tipo E de x basta para saber qué es x y si en realidad se encuentra presente, lo cual conduce a no poder garantizar que la presentación tipo E sea *acquaintance*. En síntesis, en general deberíamos ser escépticos ante la idea de que realmente exista algo como el *acquaintance*, por más que algo parezca serlo.

Además, podemos inclinar la balanza a favor de la inexistencia del *acquaintance* meramente señalando que nadie parece tener idea de cómo ocurre o podría ocurrir, si es que realmente ocurre o podría ocurrir (y menos dentro del marco fisicista) (ver Frankish (2016)). Ello se refleja en el hecho de que quienes hablan de *acquaintance* suelen tratarlo como un mero hecho bruto, primitivo e inanalizable, que simplemente tiene lugar en ciertas situaciones pero no puede ser explicado más allá de eso (ver Russell (1905, 1910/11), Fumerton (1995, 2001, 2016), Bonjour (2001, 2003) y Chalmers (2003, 2010), por ejemplo). En contraste, la percatación indirecta (representacional) es mucho menos misteriosa y hay varias propuestas sobre cómo naturalizarla (sobre las cuales ahondaré en la sección 2.3.1). Incluso, como detallaré más adelante (especialmente en el capítulo 4), es posible plantear un tipo de percatación indirecta que parecería ser *acquaintance* para el ente que la tiene. Siendo ese el caso, parece mucho más razonable sostener que lo que realmente tiene lugar es ese tipo de percatación indirecta, en lugar de algo tan inexplicable como el *acquaintance*. Ese razonamiento puede ser complementado diciendo que, en lugar de conceder que tiene lugar algo tan misterioso como el *acquaintance*, más bien

deberíamos fijarnos en los mecanismos físicos que subyacen a la presentación tipo E y concluir que esta es tan directa o indirecta como ellos. Sobre esa línea, podemos señalar que la ciencia ha revelado que los mecanismos en cuestión son bastante complejos y siempre involucran sofisticadas cadenas causales con múltiples estados intermediarios, de los que no siempre nos percatamos y siempre distancian al estado final de su objeto y permiten errores de diversa índole, lo cual sugiere que la percatación en cuestión es opaca, indirecta y falible.

En un final, parece que lo más razonable es entender a la percatación tipo E como algo que introspectivamente parece ser *acquaintance* pero en realidad no lo es: de alguna manera, los complejos mecanismos que están en juego producen la ilusión de que hay *acquaintance* cuando en realidad no la hay. Ello tiene mucho sentido si tomamos en cuenta que nuestros mecanismos senso-perceptivos fueron seleccionados naturalmente para informarnos sobre el mundo y son lo suficientemente fiables como para que no sólo no haya sido útil, sino hasta detrimental, incorporar un mecanismo que haga tomarlos con cautela. Como consecuencia, los mecanismos en cuestión evolucionaron para hacernos creer que el mundo es de cierto modo y no meramente para sugerirnos que lo es (dejando espacio a la duda). Por eso es que lo natural es esperar que introspectivamente parezcan ser un contacto directo con el mundo y, por ende, arrojar certezas, a pesar de no serlo. Así que, a fin de cuentas, lo más razonable es concluir que la apariencia de *acquaintance* no es más que eso, una mera apariencia, y que lo que está realmente en el fondo es otro tipo de percatación: uno representacional.

Por último, también es posible armar un argumento (Arg5) a partir de la idea de que los qualia podrían no ser físicos (la cual será explorada con detenimiento en el capítulo 3), en el sentido de que no son entidades físicas fundamentales ni se encuentran en una relación de dependencia D respecto a entidades físicas fundamentales:

P1_{A5}) Si los qualia existen y no son físicos, el fisicismo es falso.

P2_{A5}) Si la presentación tipo E es *acquaintance*, los qualia existen.

P3_{A5}) (Por P1_{A5} y P2_{A5}) Si la presentación tipo E es *acquaintance* y los qualia no son físicos, el fisicismo es falso.

∴ C1_{A5}) Si los qualia no son físicos, el fisicismo es falso o la presentación tipo E no es *acquaintance*.

Las premisas P1_{A5} y P2_{A5} son verdad por definiciones y el argumento es válido, así que también parece ser sólido. Así que, si por alguna razón el fisicista llega a la conclusión de que los qualia no son físicos, no le queda otra razón que negar que la presentación tipo E sea *acquaintance*. Más adelante, en el capítulo 3, argumentaré a favor de esa conclusión, así que podrá completarse el argumento de manera que la conclusión sea llanamente que la presentación tipo E no es *acquaintance*.

Así que hay muy buenas razones (aunque una de ellas terminará de concretarse hasta el siguiente capítulo) para pensar que la presentación tipo E no es *acquaintance*, sino algún tipo de percatación indirecta. A continuación estudiaré esa opción y las posturas que se derivan de ella.

2.3. Representación y representacionismo

En esta sección hablaré de las posturas que entienden la presentación como un tipo de percatación indirecta. Como he mencionado recurrentemente, lo natural es, además, entender la percatación indirecta como un tipo de *representación*. Así que las posturas en cuestión entienden la consciencia fenoménica en términos representacionales. Sin embargo, he dicho muy poco sobre qué es la representación, por lo que primero dedicaré un espacio a hablar de ella con relativo detalle.

2.3.1. Representación

El tema de la representación (o *significación*) es muy amplio y ha sido abordado desde diversas disciplinas, dentro y fuera de la filosofía, enfocándose en distintos tipos de ella y

acuñando mucha terminología especializada pero desconexa⁶⁸. Por eso, si bien intentaré rescatar lo que las principales teorías y tipos de representación tienen en común, haciendo un esfuerzo para identificar qué elementos son fundamentales y que términos refieren a lo mismo y unificarlos⁶⁹, más bien me concentraré en aclarar las nociones que emplearé (y he empleado) en este trabajo. Dado que trataré de entender el tema de forma muy amplia, muchos detalles serán pasados por alto y no abordaré las reglas y significados de representaciones o sistemas representacionales específicos.

En una primera aproximación, representar (o significar) es *ser acerca de* algo, en el sentido de estar en su lugar, con el propósito de apuntar o remitir a ello. Ejemplos paradigmáticos de representaciones (o *signos*) son las palabras, enunciados, ilustraciones, descripciones, teorías, modelos, disfraces, actuaciones y los estados de artefactos de medición, comunicación y computación, entre otros. Por ejemplo, la palabra ‘silla’ es acerca de sillas, el enunciado ‘las sillas sirven para sentarse’ es acerca de que las sillas sirven para sentarse, un retrato de Aristóteles es acerca de Aristóteles (y su aspecto) y los estados de un termómetro son acerca de temperaturas: todos esos son objetos que tienen la intención de apuntar hacia algo y, por ende, lo representan. Pero además de esas, que son creaciones culturales, podemos identificar como representaciones a toda una gama de estados y procesos biológicos, que sirven para almacenar, manipular o transmitir información de diversa índole. De particular interés son las representaciones mentales⁷⁰, como las ideas, conceptos (y uniceptos), pensamientos, actitudes proposicionales y (según argumentaré) estados senso-perceptuales. Por ejemplo, mi experiencia visual en este momento es acerca de las palabras que estoy escribiendo y mi creencia de que la gravedad existe es acerca de que la gravedad existe. En general, podemos decir que la representación es algo que puede

68 Para un tratamiento general de la representación lingüística, ver Speaks (2010-14). Para representación pictórica, ver Hyman & Bantinaki (2017). Para representación mental, ver Pitt (2000/12).

69 Ver nota 79.

70 La representación mental suele ser llamada ‘intencionalidad’ (desde Brentano (1874)). De hecho, debido a que es razonable pensar que muchas representaciones no-mentales (si no es que todas) se derivan de las mentales (ver sección 2.3.3), incluso se llega a emplear ‘intencionalidad’ como sinónimo de ‘representación’, pero yo separaré ambas nociones. Cabe señalar que las palabras ‘intencionalidad’, ‘intención’ e ‘intensión’ tienen un origen en común y no es difícil ver la relación entre las nociones correspondientes, pero aun así no hay que confundirlas. Para un tratamiento general del tema de la intencionalidad, ver Jacob (2003/19).

tener lugar a través de muchas *modalidades representacionales*, como todas las que he mencionado en este párrafo (ver Crane (2003)⁷¹). Además, muchas modalidades son o tienen subtipos: por ejemplo, la visión y la audición son sub-modalidades de la senso-percepción, que a su vez es una sub-modalidad de lo mental, así como lo son las creencias y deseos.

Ahora, si bien puede ser tentador entender a la representación como una relación entre dos objetos (o de uno consigo mismo), en la que uno es acerca del otro (o acerca de sí mismo), hay buenas razones para pensar que no necesariamente es así, porque hay algunas representaciones que, si bien pretenden ser acerca de algo, en un final hay un sentido en que no lo logran. Ejemplo de ello son los nombres vacíos, como ‘Sherlock Holmes’, ‘Santa Claus’, ‘magia’ o ‘unicornio’. La razón de fondo por la que no logran apuntar a algo en particular es que no es claro a qué pretenden apuntar. En general, para crear una representación de un objeto (o tipo de objeto) en particular es indispensable determinar de qué objeto (o tipo de objeto) se trata, ya sea ostensivamente (mostrándolo) o por medio de una descripción que sólo ese objeto cumpla⁷², lo cual es imposible en algunos casos. Por ejemplo, es imposible mostrar a Sherlock Holmes, Santa Claus, magia o unicornios y no es claro cuáles son las condiciones necesarias y suficientes para ser cada uno de esos objetos (y no meramente parecerlo): para cualquier descripción que se ofrezca, siempre se puede plantear el caso hipotético de dos objetos distintos que la cumplan, ante lo cual no hay manera de saber cuál es el indicado (además de que no es claro en qué sentido algo inexistente podría cumplir con una descripción). Así que es más seguro entender a la representación meramente como el intento de ser acerca de algo, en un sentido abierto de ‘algo’.⁷³

71 Crane (2003) llama ‘modos intencionales’ a las modalidades (y sub-modalidades). Yo no emplearé ese término, porque no quiero comprometerme con que la representación y la *intencionalidad* sean lo mismo (ver nota 70).

72 Esto está directamente relacionado con los dos tipos de conocimiento que plantea Russell (1910/11): *directo y por descripción*. Es decir, para tener la intención de remitir a algo, hay que conocerlo de alguna manera, ya sea directamente o por medio de una descripción suya. Cabe señalar que Russell considera que el conocimiento directo en su forma más básica tiene lugar en las experiencias fenoménicas e involucra *acquaintance*, pero es posible debilitar la noción lo suficiente como para incluir cualquier tipo de conocimiento que se apoye en tener contacto con ejemplares paradigmáticos de algo.

73 Algunos autores, como Meinong (1904), Mally (1912) y Parsons (1980) sostienen que sí es posible

Aun así, usualmente podemos distinguir distintas partes de una representación. Por un lado, decimos que el objeto que tiene la propiedad de representar es el *vehículo representacional* y aquello a lo que apunta (cuando lo hace) es su *contenido representacional* (o *significado*). Por ejemplo, en el caso de las palabras escritas, el vehículo representacional es el rastro visible que las conforma y el contenido representacional es aquello acerca de lo cual son. A su vez, entenderé que el contenido puede tener hasta dos componentes: *referente* (o *extensión* o *denotación*) y *caracterización* (o *sentido* o *connotación*)⁷⁴. El referente es aquello a lo que la representación apunta primariamente (a lo que *refiere* o *denota*)⁷⁵ y la caracterización es un *modo de presentación* del referente, es decir, lo que se predica de él (o qué propiedades que se le atribuyen) mediante la representación en cuestión. Por ejemplo, la palabra ‘soltero’ refiere a los solteros y los caracteriza como hombres no casados, la descripción ‘el tercer planeta del Sistema Solar’ refiere a la Tierra y la caracteriza como siendo el tercer planeta del Sistema Solar, y un retrato realista de Aristóteles refiere a Aristóteles y lo caracteriza como teniendo cierta apariencia visual⁷⁶. Por otra parte, en la sección 1.4.1 introduje la noción de las *intensiones*, que son los vínculos (o intento de vínculos) entre vehículo y referente: funciones (o intentos de funciones) que seleccionan (o intentan seleccionar) al referente. Como se dijo previamente, Chalmers (2002d) contempla dos tipos de intensiones: subjuntivo y epistémico⁷⁷. La primera es determinada por la historia de uso del vehículo para apuntar al referente y las segundas son determinadas por descripciones asociadas al referente por medio de la representación en cuestión (o usos particulares de la misma)⁷⁸. Como resulta evidente, cada intensión epistémica determina una caracterización y

referir a objetos particulares inexistentes.

74 Mill (1843) propuso los términos ‘denotación’ y ‘connotación’, mientras que Frege (1892) propuso ‘referente’ y ‘sentido’. Ver también nota 79.

75 La referencia es un tema extenso y complicado que no pretendo agotar en este escrito. Para un repaso general sobre ella, ver Eliot & Reimer (2003/19).

76 Dado que para atribuirle una propiedad a algo es indispensable referir a ella, puede entenderse que la caracterización es una referencia secundaria.

77 Chalmers (2002d) reserva el término ‘contenido representacional’ para las intensiones, no para algo determinado por ellas (como hago yo). Otros autores que hacen eso son Kaplan (1989) y Speaks (2010/19). Hasta donde alcanzo a ver, el desacuerdo conmigo es sólo terminológico (ver nota 79).

78 Los dos tipos de intensión están estrechamente relacionados con las dos maneras de delimitar al referente: la intensión subjuntiva corresponde a la manera ostensiva y las epistémicas a la descriptiva.

viceversa. Por último, resta señalar que algunos signos son mezcla de otros, como los enunciados son combinaciones de palabras, en cuyo caso su contenido representacional se deriva del de sus componentes (lo cual puede ocurrir de distintas maneras y depender de reglas arbitrarias en las que no ahondaré).⁷⁹

Dicho lo anterior, es razonable pensar que algunas representaciones carecen de referente, caracterización y/o algún tipo de intensión. Por ejemplo, parece ser que los nombres propios de cosas existentes funcionan como meras etiquetas que refieren al mismo objeto en todo escenario, sin importar qué características tenga (es decir, son *designadores*

79 No todos los términos son empleados del mismo modo por todos los autores, por lo que a continuación presento un breve recuento de algunos usos. Las palabras ‘extensión’ e ‘intensión’ fueron propuestas por Carnap (1947): la extensión es el conjunto de todos los objetos seleccionados por el signo y la intensión es una función matemática que asocia objetos con valores de verdad: *verdadero* si el objeto es un elemento de la extensión del signo y *falso* si no lo es. En otras palabras, la intensión es una función que determina a la extensión, porque para cada objeto indica si se encuentra o no dentro de ella. La idea original es que cada signo tiene una única intensión, pero algunas posturas proponen variantes sobre eso, como cuando Chalmers (2002c, 2002d, 2004, 2005, 2006a, 2006b) distingue entre intensión subjuntiva y epistémica: la primera se apoya en la historia de uso del signo para seleccionar a su extensión (o para asignar valores de verdad), mientras que la segunda se apoya en una descripción del referente. Por otro lado, Chalmers (2004, 2006c) llama ‘contenido russelliano’ al referente y ‘contenido fregeano’ a algo ambiguo entre la intensión epistémica y la caracterización que le corresponde (nótese que él trabaja como si la intensión epistémica fuese única, mientras que yo sostengo que es múltiple). Mill (1843), por su parte, denomina ‘denotación’ al referente y ‘connotación’ a la caracterización. Frege (1892) llama ‘referente’ al referente y ‘sentido’ o ‘modo de presentación’ a la caracterización. Cabe señalar que los descriptivistas (ver nota 43) consideran que las intensiones epistémica y subjuntiva son lo mismo, porque sostienen que lo que determina al referente es una descripción. Kaplan (1989) llama ‘referente’ al referente y ‘contenido’ a algo parecido a la intensión epistémica, pero señala que éste contenido (kaplaniano) se deriva de la interacción entre el contexto en que se emplea el signo y ciertas reglas que lo acompañan, que Kaplan llama ‘carácter’ (y no es claro si están incluidas en una intensión epistémica o sólo ayudan a determinarla). Además el contenido kaplaniano es sensible al contexto al que es aplicado (circunstancias de evaluación) y de ello depende la extensión. La idea es que el carácter es una función que vincula al contexto con un contenido kaplaniano y, a su vez, ese contenido es una función que relaciona al signo con una extensión de acuerdo a las circunstancias en que es evaluado. Por ejemplo, el carácter de la expresión ‘el presidente de este país’ es sensible al lugar en que es proferida y determina un contenido kaplaniano, que a su vez es sensible al momento en que es evaluada, de modo que selecciona a distintos individuos en diferentes situaciones (ver también Speaks (2010-14)). Por otro lado, Peirce (1931/35/58) presenta una complicada clasificación con base en triadas y subtriadas en la que no profundizaré, salvo para decir que llama ‘objeto’ al referente e ‘interpretante’ a la intensión, distinguiendo tres tipos de ella (ver también nota 84). Este repaso de términos no es exhaustivo, pero no me extenderé más en ese asunto.

rígidos (ver Kripke (1972))⁸⁰) y, por ende, sólo tienen referente e intensión subjuntiva⁸¹. Por otra parte, podría pensarse que las descripciones de objetos imposibles, como ‘el mayor número primo’, sólo tienen caracterización e intensión epistémica. Los nombres vacíos, en cambio, parecen carecer de ambas cosas, porque es razonable pensar que no tienen referente ni caracterización ni intensiones (por lo que señalé hace un par de párrafos). Por último, algunos signos son sensibles al contexto en que son empleados o evaluados y, por ende, tienen distinto referente en diferentes situaciones. Ejemplo de ello son los *indéxicos* o *deícticos* (como ‘yo’, ‘ayer’ o ‘aquí’) y descripciones como ‘el presidente de México’⁸². Una manera de hacer sentido de ello es pensar que por sí mismos sólo tienen caracterización e intensión epistémica, y que sólo al ser empleados adquieren un referente (y tal vez hasta una intensión subjuntiva). Yo no me comprometeré con que esa sea la manera correcta de describir cada una de las representaciones mencionadas, sino que sólo me interesa destacar que es razonable pensar que pueden existir las combinaciones señaladas.

Por otra parte, es posible distinguir entre representaciones públicas y privadas, así como entre los contenidos e intensiones que son “tomados prestados” al emplear una representación preexistente y los que son añadidos al momento de emplearla. Por ejemplo, si tomo prestada del español la palabra (pública) ‘olmo’, esta viene con una intensión subjuntiva determinada por su historia de uso, en virtud de la cual refiere a los olmos (ver Putnam (1975) y Burge (1979)), pero bien puedo equivocarme y emplearla para referirme (de manera privada) a un sauce, añadiendo una nueva intensión que lo vincula con él (y que claramente no pertenece al español). De manera similar, puedo confundirme y referirme a alguien como ‘el joven de la copa de vino’, aun si en realidad es un viejo o si

80 De manera más formal, un designador rígido es un signo que (desde el mundo real) selecciona siempre a un mismo objeto (o tipo de objeto), y a ninguno otro, en cada escenario o mundo posible en el que dicho objeto (o tipo de objeto) existe y nada en los mundos o escenarios posibles en los que no existe. Por ejemplo, el nombre ‘Saul Kripke’ selecciona a Saul Kripke y sólo a Saul Kripke, independientemente de qué otras características cumpla éste: bien podría no haber hecho nada de lo que lo hizo famoso sin que por ello dejase de ser Saul Kripke.

81 En contraste, los descriptivistas, como Frege (1892) y Russell (1905), sostienen que los nombres propios sí tienen contenido descriptivo. Y el bidimensionalismo semántico (Chalmers (2002c, 2002d, 2004, 2005, 2006a, 2006b)) puede entenderse como un intento de reivindicación de esa postura (ver nota 43).

82 Ver Kaplan (1989), Speaks (2010-14) y nota 79.

su copa sólo tiene agua de jamaica (ver Donellan (1966)). Es decir, un mismo objeto puede ser una representación en varios sentidos al mismo tiempo, algunos heredados y otros generados al momento de emplearlo. Yo me mantendré neutral respecto a qué contenidos e intenciones son propios de cada representación preexistente y cuáles son añadidos. De nuevo: sólo me interesa señalar que pueden existir las combinaciones mencionadas.

Por último, resta hacer énfasis en el requisito *teleológico* de la representación, es decir, en el hecho de que para representar x no basta con apuntar hacia x, sino que es preciso tener la *intención* (o *propósito*, *objetivo*, *meta*, *pretensión*, *fin*, *finalidad*, etcétera) de hacerlo. Ello permite descartar muchas cosas que no son representaciones pero suelen ser confundidas con ellas, como cuando un objeto remite a otro meramente porque se le parece. Por ejemplo, una mancha generada por un derrame accidental de pintura o la forma de una nube podrían parecerse mucho a otros objetos, impulsándonos a pensar en ellos cuando los vemos, pero no por eso son representaciones de aquello a lo que se parecen, sencillamente porque no tienen el propósito de remitir ello: su parecido es meramente accidental (ver Putnam (1981)). Por otra parte, hay objetos que están relacionados causalmente con otros y por eso nos hacen pensar en ellos o inferir su presencia, pero tampoco son representaciones suyas. Por ejemplo, el sarampión es causa de ronchas y el fuego es causa de humo y es peligroso, lo cual incluso da pie a expresiones como ‘las ronchas son *signo* de sarampión’, ‘el humo *significa* que hay fuego’, ‘que haya fuego *significa* que habrá humo’ o ‘el fuego *representa* peligro’ (ver Grice (1948)⁸³). Pero ninguno de esos objetos tiene el propósito de remitir a sus causas o consecuencias y, por ende, no son representaciones de ellas en el sentido relevante. De hecho, ni siquiera basta con que haya al mismo tiempo parecido y relación causal para que haya representación: las huellas de los animales suelen parecerse a la planta de sus patas y ser causadas por ellas, pero no por ello las representan, por la misma razón de que carecen de ese propósito. En general, estas ideas pueden extenderse hacia cualquier tipo de correlación sistemática: ninguna es suficiente para que haya

83 Grice (1948) llama ‘signos naturales’ a los objetos que remiten a otros porque son evidencia de ellos y ‘signos convencionales’ a las representaciones propiamente dichas, porque estas últimas usualmente dependen de un acuerdo entre usuarios (con un fin) y las otras no.

representación, sino que siempre hace falta el factor teleológico⁸⁴. En otras palabras, siempre que hay representación es porque hay teleología involucrada, de un modo u otro (ver sección 2.3.3.2).

También es muy importante señalar que el sentido teleológico de la representación es inseparable de un aspecto normativo: todas y sólo las cosas que tienen un propósito pueden estar bien o mal, en el sentido de que pueden cumplir o no con su propósito. Ello se refleja en que la referencia y caracterización, así como las intensiones que las determinan, pueden ser entendidas en términos de condiciones de adecuación, es decir, de qué se le pide a un objeto para que el signo en cuestión sea aplicado correctamente a él o, por lo menos, para que alguna de las intensiones lo seleccione (ver Speaks (2010-14) y MacBride (2013-19)). La extensión equivale a de qué tipo de objeto debe tratarse y la caracterización a qué propiedades debe poseer. Por ejemplo, el concepto de *pez* refiere a los peces y los caracteriza como siendo de cierto modo. Las condiciones de adecuación son lo que está detrás de que clasifiquemos las aplicaciones de signos como correctas o incorrectas, las afirmaciones y creencias como verdaderas o falsas, las memorias como precisas o imprecisas, los deseos como satisfechos o insatisfechos, las experiencias sensoriales como verídicas o ilusorias (o eso argumentaré), las descripciones como acertadas o erradas, las ilustraciones como realistas o no, etcétera.⁸⁵

84 No estoy diciendo que la correlación sistemática no pueda ser importante en algunas representaciones, sino sólo que no basta para que haya representación: desde luego que algunos objetos pueden ser elegidos como vehículo representacional porque se parecen al referente o son evidencia de él, pero no lo son sino hasta que se les asigna ese propósito (ver también Goodman (1972), Godfrey-Smith (1996) y Suárez (2003)). Godfrey-Smith (2006) y Shea (2018) reconocen ese hecho y llaman ‘relaciones explotables’ a las características de un objeto que lo hacen ser un buen candidato para ser elegido como representación. Por otro lado, Peirce (1931/35/58) clasifica los tipos de signos de acuerdo a la razón por la que su vehículo fue elegido: los *símbolos* son signos cuyo vehículo fue elegido de manera arbitraria; los *índices* son signos cuyo vehículo fue elegido porque guarda una (supuesta) relación evidencial con el referente; y los *íconos* son signos cuyo vehículo fue elegido debido a que (supuestamente) comparte ciertas características con el referente, es decir, porque se parece a él en algún sentido relevante. Goodman (1968), por su parte, llama ‘símbolo’ a cualquier representación y ‘ejemplificación’ al recurso de apoyarse en el parecido entre vehículo y referente. En la sección 2.3.3 se retomará este tema.

85 El hecho de que las representaciones puedan ser correctas o incorrectas ha servido de punto de partida para que algunos autores entiendan las intensiones como funciones que asocian al signo con valores de verdad (verdadero o falso), según candidatos a referente, lo cual a su vez ha llevado a que se entienda que el referente de algunos signos (como los enunciados afirmativos) es su valor de verdad (esto lo encontramos desde Frege (1892) hasta Speaks (2010-14)). A pesar de su popularidad, esa me parece una interpretación poco intuitiva y hasta tergiversada de la referencia, por lo que no me apegaré a ella

2.3.2. Representacionismo

Las posturas representacionistas (o *intencionalistas*⁸⁶) son todas aquellas que sostienen que las experiencias fenoménicas son *necesariamente* representacionales, ya sea porque ciertas experiencias siempre involucran propiedades representacionales o porque ciertas propiedades representacionales siempre involucran experiencias. Más específicamente, entenderé que las posturas representacionistas respecto a una experiencia fenoménica (en particular) son todas aquellas que buscan explicarla en términos representacionales y, con ello en mente, sostienen que tener esa experiencia es condición necesaria y/o suficiente para instanciar una determinada propiedad representacional⁸⁷. De especial interés son las dos siguientes tesis generales:

QRx) Para cada experiencia fenoménica q del tipo x , hay una propiedad representacional r tal que tener q necesariamente implica instanciar r .

RQx) Para cada experiencia fenoménica q del tipo x , hay una propiedad representacional r tal que instanciar r necesariamente implica tener q .

Consideraré que cada una de esas tesis es un *representacionismo débil* (respecto a x) y llamaré ‘representacionismo fuerte’ (respecto a x) a la siguiente mezcla de ambas⁸⁸:

RFx) Para cada experiencia fenoménica q del tipo x , hay una propiedad

(para mí, el referente de los enunciados afirmativos es la situación que afirman, por ejemplo) (ver nota 45).

86

Ver nota 70.

87 El representacionismo fuerte contemporáneo apareció por primera vez en los escritos de Hall (1961), Anscombe (1965), Armstrong (1968) y Hintikka (1969). Después se unieron autores como Kraut (1982), Lewis (1983c), Rosenthal (1986, 2002), Lycan (1987, 1996), Harman (1990), Shoemaker (1994b), Tye (1994, 1995, 2000), Dretske (1995), Carruthers (2000), Rey (1998), Clark (2000), Byrne (2001), Crane (2001, 2003), Kriegel (2002, 2003, 2006), Thau (2002), Droege (2003), Jackson (2003), Levine (2003), Egan (2006), Prinz (2011), Sebastián (2012, 2017) y McClelland (2013, 2016). Lycan (2000-2019) ofrece un tratamiento general sobre el tema.

88 Nótese que el representacionismo fuerte no es meramente es la conjunción de QRx y RQx, porque la q y la r bien podrían ser distintas en cada tesis débil. En un final, RFx es la conjunción de QRx y RQx aunada a la especificación de que la q y la r son las mismas en ambas.

representacional r tal que tener q necesariamente implica instanciar r e instanciar r necesariamente implica tener q.

Una manera natural de darle sentido a RFX es entendiendo que cada experiencia fenoménica del tipo x es idéntica a alguna propiedad representacional. Por otro lado, hay que aclarar que x puede abarcar tanto como se desee: por ejemplo, podría incluir sólo una experiencia fenoménica en particular, todos los casos de percepción verídica y alucinación, sólo emociones o llanamente toda experiencia fenoménica. Variando sobre la terminología de Byrne (2001), llamaré ‘representacionismo restringido a x’ al que se limita a las experiencias de un cierto tipo x y ‘representacionismo irrestricto’ al que incluye a toda experiencia (o ‘restringido a toda experiencia’)⁸⁹. Entenderé, entonces, que RFI, QRI y RQI son las variantes irrestrictas de RFX, QRx y RQx, respectivamente. Además, dado que la consciencia fenoménica es (idéntica a) presentación tipo E que qualia, es natural entender que el representacionismo fuerte irrestricto (RFI) es la tesis de que la presentación tipo E es un tipo de representación (E’, que bien podría ser disyuntivo) y que los qualia (o composiciones de qualia) son (al menos parte de) su contenido representacional. Por ende, otra manera de formular RFI podría ser la siguiente:

RFI) La consciencia fenoménica es idéntica a representación tipo E’ de qualia.⁹⁰

89 Byrne (2001) llama ‘restringido’ (a secas) al caso en que el tipo x es la percepción verídica (visión, olfato, gusto, etc.) y sus contrapartes alucinatorias, ‘irrestricto’ al caso en que se añaden sensaciones corporales, como dolores, comezón y orgasmos, y ‘completamente irrestricto’ al caso en que además se agregan las emociones, estados de ánimo, imaginación y lo que sea que se considere que involucre consciencia fenoménica, es decir, a lo que yo llamo ‘irrestricto’. No sigo su nomenclatura al pie de la letra porque considero que su división es demasiado arbitraria.

90 Chalmers (2004) llama ‘representacionismo’ (a secas) a RFI y tanto él como Lycan (2000/19) llaman ‘representacionismo débil’ a QRI. Además, Lycan llama ‘representacionismo’ (a secas) y ‘representacionismo fuerte’ a una misma cosa: a la tesis de que cierto tipo de representación es suficiente para que haya consciencia fenoménica y que dicho tipo de representación es expresable en términos exclusivamente físicos. Ello es compatible con que haya experiencias que no son explicables en términos representacionales, pero me interesa excluir esa posibilidad. Además, esa definición incorpora un elemento reduccionista que yo estoy abordando por separado. Por otra parte, Byrne & Tye (2006) llaman ‘intencionalismo débil’ a RQI e ‘intencionalismo fuerte’ una versión de RFI. Por último, Crane (2003) llama ‘representacionismo’ (a secas) o ‘intencionalismo fuerte’ a RFI e ‘intencionalismo débil’ a QRI aunada con la negación de RQI. Como es evidente, no hay acuerdo sobre cómo llamar a qué. Sin embargo, considero que la nomenclatura que elegí es más intuitiva.

Los argumentos a favor de RFI suelen consistir en motivar o presuponer QRI y después señalar que las propiedades representacionales aludidas en QRI parecen agotar todo lo que hay que decir sobre la consciencia fenoménica: si no hay algo más involucrado, es natural concluir que la consciencia fenoménica es idéntica a esas propiedades representacionales⁹¹. La manera usual de defender QRI parte de señalar que tener una experiencia fenoménica (y ponerle atención) necesariamente implica tener la impresión de que el mundo es de cierto modo, porque las experiencias siempre sugieren que el mundo es de algún modo tan sólo en virtud de su carácter fenoménico (ver Siewert (1998)⁹², Byrne (2001), Thau (2002) y Horgan & Tienson (2002)). Estrictamente, ello es compatible tanto con que se trate de *acquaintance* como con que se trate de representación, pero suele darse por sentado que la única opción viable es la representacional, lo cual conduce inmediatamente a QRI. Una estrategia más precavida puede dar ese salto apoyándose en las razones expuestas en la sección 2.2.3. Así que, por lo menos, es razonable sostener QRI.

Además, como señalan Byrne (2001) y Thau (2002), no sólo es el caso que el carácter fenoménico determina un contenido representacional, sino también que el contenido representacional determina un carácter fenoménico: absolutamente todo elemento del carácter fenoménico es inseparable de algún elemento de cómo nos parece que es el mundo. Para motivar esa idea, basta con señalar que nos parece que el mundo cambia si y sólo si notamos que nuestras experiencias cambian y exactamente en la medida en que nos parece que lo hacen: si nos parece que dos experiencias son subjetivamente idénticas, no tenemos razones para pensar que por sí mismas sugieren que el mundo es distinto. En otras palabras, toda coincidencia en el carácter fenoménico de dos experiencias equivale a una coincidencia en cómo sugieren que es el mundo y toda diferencia en su carácter fenoménico equivale a una diferencia en cómo sugieren que es. En otras palabras, toda similitud o diferencia en el carácter fenoménico es una similitud o diferencia en qué qualia

91 En cuanto los representacionismos restringidos, es fácil construir argumentos a su favor a partir de los que apoyan a los fuertes: sólo hay que acotar cada idea para un tipo particular de experiencia. De hecho, los argumentos a favor de las versiones irrestrictas suelen ser generalizaciones de argumentos para las restringidas (ver nota 92, por ejemplo).

92 Siewert (1998) se enfoca en el caso de la visión, de modo que su argumento es más bien a favor de la variante de QRx para experiencias visuales. Pero es fácil extender la idea para otras modalidades sensoriales y, en general, para cualquier experiencia fenoménica.

parecen estar presentes. Así que el que una experiencia tenga un determinado carácter fenoménico no sólo implica que tiene cierto contenido representacional, sino también a la inversa, lo cual permite dar el paso hacia RFI.

Una manera alternativa de llegar a la misma conclusión es la de Harman (1990), Tye (1995, 2000) y Crane (2003), quienes hacen énfasis en que lo más natural es entender que los qualia que se nos presentan en la experiencia no son sino propiedades que esta nos sugiere que el mundo tiene: la consciencia fenoménica parece no ser más que una “ventana transparente” hacia el mundo y aquello que lo puebla, de modo que poner atención en el carácter fenoménico no es otra cosa que poner atención en el mundo que la experiencia nos presenta. Así que es razonable concluir que el carácter fenoménico no es otra cosa que parte del contenido representacional de la experiencia. Además, no parece que la consciencia fenoménica involucre algo más allá de la percatación de dicho contenido: ni siquiera parece requerir que el mundo de hecho sea como sugiere, porque sabemos que hay alucinaciones y sueños. Siendo ese el caso, es razonable sostener que la consciencia fenoménica sólo consiste en representar de cierta manera (tipo E’) que el mundo es de tal o cual modo (compuesto por qualia). Es decir, tener consciencia fenoménica es idéntico a instanciar ciertas propiedades representacionales, de lo cual se sigue RFI.

Ahora, dentro de los representacionismos fuertes (restringidos o no), podemos distinguir entre *reductivos* y *no-reductivos*, dependiendo (respectivamente) de si las propiedades representacionales de las que hablan son o no explicables sin aludir directamente a la consciencia fenoménica o a los qualia (Chalmers (2004)⁹³). En otras palabras, los reductivos entienden a la representación como algo más fundamental que la consciencia fenoménica y que sirve para explicarla de manera exhaustiva, mientras que los no-reductivos sostienen que la consciencia fenoménica es igual o más fundamental que la representación, de modo que es inevitable hablar directamente de ella al plantear las

93 Dado que Chalmers (2004) entiende a la consciencia fenoménica como instanciación de qualia, meramente explica el representacionismo reductivo en términos de no aludir directamente a la consciencia fenoménica. Sin embargo, debido a que yo distingo entre la consciencia fenoménica (presentación tipo E de qualia) y los qualia, me veo obligado a especificar que tampoco se puede aludir directamente a los qualia.

propiedades representacionales relevantes. Ejemplos de representacionistas (fuertes) reductivos son Lycan (1987, 1996), Harman (1990), Tye (1994, 1995, 2000), Dretske (1995), Carruthers (2000), Clark (2000), Droegge (2003) y Jackson (2003). Por otra parte, Chalmers (2004) explícitamente se manifiesta a favor de un representacionismo (fuerte) no-reductivo y señala a McGinn (1988), Searle (1990), Strawson (1994), Siewert (1998), Horgan and Tienson (2002) y Loar (2003) como otros posibles ejemplos.⁹⁴

Otra distinción es entre representacionismos *puros* e *impuros*, dependiendo (respectivamente) de si las propiedades representacionales a las que aluden son *puras* o *impuras* (Chalmers (2004) y Egan (2006)). Las puras tienen la forma *representar x*, mientras que las impuras son de la forma *representar x bajo la modalidad representacional z*. Como ya se mencionó, algunos ejemplos de modalidades (y sub-modalidades) representacionales son la mental, la percepción, la escritura, el dibujo, la visión y el pensamiento. Incluso se pueden incluir especificaciones sobre qué papel desempeña la representación en cuestión dentro de un sistema, así como de qué manera y para qué está disponible su contenido. En general, cualquier especificación sobre una representación que vaya más allá de su contenido es una especificación sobre su modalidad representacional. Por otra parte, hay que notar que toda propiedad representacional impura necesariamente implica una pura: para representar x bajo la modalidad representacional z, es indispensable representar x. Por esa razón, el representacionismo QRx impuro necesariamente implica al QRx puro⁹⁵, pero no a la inversa. Por lo mismo, el representacionismo RQx puro necesariamente implica al RQx impuro, pero no a la inversa.

Dado que la consciencia fenoménica es presentación tipo E de qualia y el representacionista fuerte irrestricto sostiene que la presentación tipo E es un tipo de representación, podría ser tentador concluir que especificar la modalidad representacional

94 Nótese que esta es una noción de *reducción* distinta de la expuesta en la sección 1.2, según la cual reducir x a z es sostener que x puede ser realizado mediante z y sólo mediante z. Sin embargo, ambas nociones están íntimamente relacionadas, porque ambas giran en torno a la idea de que x es algo que no va más allá de z y, por ende, para explicar x no es necesario recurrir a algo que no esté ya en z.

95 Aquí estoy yendo más allá de lo que Chalmers (2004) dice, porque él se concentra en el representacionismo fuerte y no habla de la variante débil de cada tipo de representacionismo.

equivale a describir la presentación tipo E (o representación tipo E'). Sin embargo, hay que tener en cuenta que bien podría resultar que el contenido representacional relevante fuese más allá de los qualia, de manera que describir la presentación tipo E podría requerir especificar partes del contenido y no sólo de la modalidad. Por ejemplo, podría sostener que el contenido no son meramente los qualia, sino también su presencia en cierta ubicación. Así que, si bien es razonable pensar que definir la presentación tipo E involucra determinar la modalidad representacional, bien podría ser que requiera aun más cosas. También cabe señalar que el mero hecho de aludir a un tipo E' de representación no basta para descartar al representacionismo RQx puro, porque bien podría sostenerse que lo que distingue a la representación tipo E' de qualia de la mera representación de qualia es sólo más contenido representacional, es decir, que representar de modo E' qualia sólo consiste en representar z, donde z incluye qualia. Entonces podría salvarse un representacionismo irrestricto y puro según el cual la consciencia fenoménica es idéntica a representación de z.

Por otra parte, podemos separar a los representacionismos *amplios* de los *estrechos*, donde los primeros sostienen que las propiedades representacionales relevantes dependen de factores externos al sujeto de experiencia, mientras que los segundos aluden exclusivamente a factores internos suyos (Chalmers (2004)). Es decir, los amplios aluden a propiedades representacionales extrínsecas (o amplias) del sujeto y los estrechos sólo de intrínsecas (o estrechas). Y ambos tipos de representacionismo son compatibles con ser fuertes o débiles, reductivos o no-reductivos y puros o impuros. Sin embargo, cabe señalar que el representacionismo QRx amplio es incompatible con el internismo fenoménico (IF), porque es incoherente decir que basta con algo que depende exclusivamente de factores internos para que haya algo que depende de factores externos. Por lo mismo, el representacionismo fuerte y amplio es incompatible con IF. Algunos ejemplos de autores que rechazan IF porque son representacionistas amplios son Dretske (1995, 1996), Tye (1995, 1998) y Lycan (1996, 2001). En cuanto al representacionismo estrecho, es defendido por autores como Shoemaker (1994b), Rey (1998), Kriegel (2002, 2003, 2006), Levine (2003) y Egan (2006).

Ahora sólo falta hacer dos cosas para terminar de concretar cada postura representacionista. Una de ellas es especificar el contenido representacional al que se alude en cada caso, lo cual involucra determinar qué son en sí los qualia según cada teoría, porque estos siempre son parte del contenido. Pero ese es un trabajo que dejaré para el siguiente capítulo. La otra tarea es naturalizar la representación, para tender el puente entre lo enteramente físico y lo representacional, lo cual es indispensable para terminar de tender el puente entre lo enteramente físico y lo fenoménico. En eso se enfocará la siguiente sección.

2.3.3. Representación naturalizada

Como he señalado ya varias veces desde la sección 2.2.3, la principal razón en contra de entender la presentación tipo E como *acquaintance* es que el *acquaintance* parece no ser naturalizable. Sin embargo, ello es una razón para entenderla como representación sólo si la representación sí es naturalizable. En esta sección me enfocaré en explicar en qué sentido sí lo es. Con ello en mente, presentaré las dos principales maneras de naturalizar la representación: la etiológica y la disposicional. Pero antes de eso, es importante aclarar algunos detalles, lo cual haré a continuación.

2.3.3.1. Algunas nociones clave

Lo primero que hay que notar es que algunas representaciones son *derivadas*, en el sentido de que para explicarlas hay que aludir a otras representaciones. Por ejemplo, las palabras son representaciones *derivadas*, porque se apoyan en conceptos, que son representaciones mentales. En general, es claro que todas las representaciones culturales se derivan de las mentales, pero también es razonable creer que algunas representaciones mentales son derivadas: por ejemplo, si me fijo en mi unicepto⁹⁶ de *silla*, lo usual es que vengan a mi mente otros uniceptos relacionados con sillas, así como imágenes de sillas y otras

96

Recuérdese que estoy tomando prestada la terminología de Millikan (2013, 2017), quien propone centrarse en uniceptos en lugar de en conceptos (ver nota 48).

sensaciones que asocio con ellas y las representan de diversas maneras. Así que para realmente naturalizar la representación hay que identificar y naturalizar las representaciones que no dependen de ninguna otra, es decir, las *no-derivadas*. Para encontrar representaciones estrictamente de ese tipo, tenemos que fijarnos en algo más básico o primitivo, como pueden ser nuestros estados senso-perceptuales más rudimentarios, antes de ser procesados por el resto de nuestro sistema cognitivo. Otro candidato son los diversos estados que informan al sistema nervioso autónomo y permiten el buen funcionamiento de cosas como los arcos reflejos, la regulación de la frecuencia cardíaca, la digestión, la salivación, la sudoración, etcétera. Así que primero me centraré en representaciones de esos tipos, bajo el supuesto de que el resto se deriva de ellas. De especial interés serán los estados senso-perceptuales, debido a que son los que típicamente están asociados con la consciencia fenoménica, en contraste con los estados del sistema nervioso autónomo, que son inconscientes.

Por otra parte, una manera natural de expresar la idea de que representar es tener el propósito de apuntar hacia algo es diciendo que apuntar hacia algo es una *función* del vehículo representacional. Sin embargo, la noción de *función* requiere desambiguación (y también la requiere la noción de *apuntar hacia algo*, que será atendida en la sección 2.3.3.3). En particular es indispensable aclarar tres acepciones del término ‘función’⁹⁷: *rol funcional*, *función sistémica* y *función teleológica*. Un *rol funcional* (o *rol causal*, si involucra causalidad) es cualquier papel desempeñado o a desempeñar por algo en algún contexto, es decir, *f* es un rol funcional *de* un objeto *z* en un contexto *c* si y sólo si *f* es algo que *z* hace o tiene la disposición (o capacidad o propensión)⁹⁸ de hacer en *c*. Por ejemplo, mi corazón tiene los roles funcionales de bombear sangre, emitir un sonido característico y ejercer atracción gravitacional sobre el Sol. Por otro lado, en ocasiones interesa hacer énfasis en que *z* es parte de un sistema *s* dentro de *c* y que ciertos roles funcionales de *z* son especialmente relevantes para *s*: entonces decimos que esos roles funcionales son

97 Previamente, en la sección 1.4.1, recurrí a otra acepción de ‘función’, para definir la noción de *intención*: la de *función matemática* (ver también nota 44).

98 Tener la disposición de hacer *f* es tener ciertas propiedades (en el presente) tales que conducen a hacer *f* (en un futuro hipotético). Por ejemplo, un vaso de vidrio tiene la disposición de romperse si se le estrella contra el suelo. Ello debido a que su estructura física presente es de cierto modo.

funciones sistémicas de x , relativas a s . Por ejemplo, el que mi corazón lata es especialmente relevante para el sistema que es mi cuerpo, mientras que ejercer atracción gravitatoria sobre el Sol no lo es. Por último, tener una *función teleológica* es tener el *propósito* de desempeñar un determinado rol funcional. Es decir, un rol funcional f es una función teleológica de z si y sólo si z tiene el propósito de hacer f . Por ejemplo, bombear sangre es una función teleológica de mi corazón, mientras que emitir sonido no lo es, porque sólo lo primero es su propósito: mi corazón existe *para* bombear sangre, no para sonar como suena.

Es importante señalar que para tener a f como función teleológica no es necesario tener a f como rol funcional: hay cosas que sencillamente son incapaces de cumplir con su propósito. Por ejemplo, un avión descompuesto tiene la función teleológica de volar, aun si no puede hacerlo. De hecho, podemos decidir emplear cualquier cosa para hacer cualquier tarea, otorgándole de ese modo esa función teleológica, sea o no útil para realizarla. Por ejemplo, puedo emplear mi zapato para tratar de leer mentes. Una consecuencia curiosa de ello es que para naturalizar la función teleológica de hacer f podría no hacer falta naturalizar el rol funcional f , porque es perfectamente posible tener un objetivo imposible: para explicar cómo es que mi zapato tiene el propósito de leer mentes, no necesito naturalizar la lectura de mentes.

Entonces, dado que representar es tener el propósito de apuntar hacia algo, podemos puntualizar su definición del siguiente modo:

(DR) Representar x es tener como función teleológica el rol funcional de apuntar hacia x .

Cabe notar que, a pesar de que parece haber algunas representaciones que no logran el propósito de apuntar hacia algo, como los nombres vacíos⁹⁹, sabemos que otras sí lo hacen, de modo que no es permisible eludir la tarea de naturalizar ese rol funcional. Entonces lo

⁹⁹ Nótese que no estoy sugiriendo que los nombres vacíos son inútiles, sino sólo que tienen una función teleológica que no cumplen, lo cuál es compatible con que tengan y cumplan otras.

que hay que hacer es explicar en qué consiste dicho rol y cómo es que puede ser una función teleológica de algo; todo ello exclusivamente en términos físicos. En los siguientes apartados me enfocaré en la naturalización de las funciones teleológicas en general y después me centraré en el caso particular de la representación, es decir, cuando el rol funcional relevante es el de apuntar hacia algo.

2.3.3.2. Teleología naturalizada

La *teleología* refiere a que algunas cosas tienen un *propósito*, es decir, existen *para* algo. Es una noción difícil de acomodar en el mundo natural bajo el marco fisicista, porque lo usual bajo ese marco es pensar que la naturaleza simplemente es como es, sin un propósito intrínseco. Por lo mismo, la idea suele ser repudiada dentro de la ciencia y la filosofía naturalista (ver Colin & Neal (1996/2020)). Según Mayr (1988), es un concepto espurio porque parece sugerir la acción de alguna fuerza vital o mental donde no hay tales, además de que parece requerir algo imposible: que un evento futuro (el fin) sea capaz de explicar, en un sentido causal, uno pasado (el medio). Por otro lado, no es claro cómo es que las afirmaciones que implican la presencia de teleología pueden ser probadas empíricamente (Allen & Bekoff (1995)). Por inquietudes como esas, Pittendrigh (1958) prefirió introducir la noción de *teleonomía* (retomada por Mayr (1965, 1974) y Williams (1966)), que refiere a la mera apariencia de teleología donde en realidad no la hay. Así que el proyecto de naturalizar la teleología puede ser entendido como el de encontrar algo en el mundo natural que se le parezca suficientemente, independientemente de si hay una noción no-natural de ella. En lo personal prefiero pensar que ambas opciones son teleología genuina, con la diferencia de que una es una idealización irreal y la otra una noción más realista. Pero en un final lo realmente importante es ofrecer una noción que permita distinguir entre meramente tener un rol funcional o sistémico f y tenerlo como algo que merezca ser llamado ‘propósito’, aunque sea en un sentido muy débil.

Otra razón por la que hay que tener cuidado al hablar de teleología es que es una noción *normativa* y la normatividad misma también es misteriosa y requiere ser naturalizada.

Como ya he señalado, es normativa en el sentido de que tener un propósito implica estar bien en ciertas situaciones y mal en otras: bien cuando se lo cumple y mal cuando no. Es decir, si un objeto z tiene al rol funcional f como función teleológica, entonces z *debe* hacer f : si z hace f , entonces está funcionando bien, y si no, está funcionando mal. Pero (bajo el marco fisicista) el *deber* no es algo que suela reconocerse como parte del mundo natural: las cosas simplemente son como son, sin “bien” o “mal”. En general, el naturalismo se centra en describir cómo de hecho son las cosas y no parece válido extraer principios normativos a partir de ello: es decir, no es válido inferir el cómo deberían ser las cosas tan sólo a partir de cómo son, por más que en ocasiones sea tentador hacerlo (ver Hume (1739)). Así que, de nuevo, lo más que una buena teoría naturalista sobre las funciones teleológicas puede hacer es reemplazar la parte normativa por algo que no sea normativo pero que tenga la apariencia de serlo. En especial, debe hacerlo de manera que sea posible distinguir entre funcionar bien y funcionar mal. En el caso particular de la representación naturalizada, de ello dependerá que tenga sentido hablar de condiciones de adecuación, es decir, de que un signo esté bien o mal aplicado.

Ahora, una forma evidente de adquirir propósito es a través de un agente que lo asigne: basta con simplemente decidir emplear algo con cierto fin para que lo tenga como propósito, independientemente de si es o no bueno para ello. De hecho, como ya mencioné, cualquier objeto puede tener cualquier función teleológica, mientras alguien se la asigne. Pero toda asignación de propósito por medio de un agente depende de los propósitos de dicho agente, mismos que requieren explicación. Así que, al igual que con la representación, podemos distinguir entre teleología *derivada* y *no-derivada*: la primera depende de teleología previa y la segunda no. De hecho, la distinción entre representación derivada y no-derivada no es más que un caso particular del caso teleológico. Entonces, al igual que con la representación, me concentraré principalmente en la teleología no-derivada.

Con todo lo anterior en mente, revisaré las dos principales familias de teorías que tratan de naturalizar las funciones teleológicas: las *etiológicas* y las *disposicionales*. Cabe señalar

que no son posturas rivales, sino complementarias: dos maneras distintas de obtener teleología (ver Millikan (1989), Godfrey-Smith (1994) y Allen & Bekoff (1995)). En un final, cada una corresponde a un sentido bastante intuitivo en el que un objeto *z* puede considerarse como poseedor de propósito: las teorías disposicionales se enfocan en lo que *z* de hecho hace, que en muchos casos puede ser entendido como la persecución de una meta, mientras que las etiológicas toman en cuenta el origen de *s*, que determina por qué y para qué está presente. Pero también están quienes intentan unificar ambas opciones, subsumiendo una bajo la otra, como hacen Griffith (1993), Walsh (1996), Walsh & Ariew (1996) y Mossio, Saborido & Moreno (2009). A continuación revisaré cada vertiente y después las aterrizaré en el caso particular de la representación.

2.3.3.2.1. Teleología etiológica

Las teorías *etiológicas* sobre la teleología, popularizadas por Wright (1973, 1976)¹⁰⁰, parten de la idea de que el propósito de algo (su *para qué*) suele formar parte de la explicación de *por qué* se encuentra presente. Si bien ello puede sonar a que un evento futuro puede ser responsable de uno pasado, lo cual no parece tener sentido dentro del marco naturalista, lo que en realidad se tiene en mente es que algunos objetos hacen algo que contribuye a perpetuar su propia existencia. Sobre esa línea, lo que hacen las teorías etiológicas es recurrir a la historia del objeto en cuestión, haciendo énfasis en algún mecanismo selectivo que le ha dado origen y para el cual ha sido relevante lo que el objeto o sus ancestros han hecho en el pasado. La opción más obvia es aludir a la selección natural y a los rasgos biológicos que han sido fijados mediante ella: características cuya presencia se ha perpetuado debido a que han hecho algo que ha contribuido a la supervivencia de la especie que los posee (ver especialmente Millikan (1984) y Neander (1991)). De ese modo, un mero rol funcional o sistémico de los ancestros del rasgo termina siendo la razón por y para la cual se encuentra presente y, por ende, convirtiéndose en su propósito. Rescatando a Neander (1991:74), una manera de exponer la idea es la siguiente:

¹⁰⁰ Una propuesta aun previa es la de Ayala (1970), pero Wright (1973) fue quien popularizó la estrategia. Para un repaso general sobre teorías etiológicas de la teleología, ver Neander & Schulte (2004-20).

Una función (teleológica) propia¹⁰¹ de un rasgo z de un organismo s es hacer aquello que los rasgos del mismo tipo que z hicieron para contribuir a la aptitud inclusiva de los ancestros de s , que fue lo que causó que el genotipo del cual z es la expresión fenotípica fuese seleccionado naturalmente.

Pero la idea puede ser expandida para abarcar otros mecanismos de selección, tales como la selección artificial, la cultural, el aprendizaje y la mera contribución de z a favor de la preservación de s . Por ejemplo, Millikan (1984) habla sobre selección cultural de memes, Papineau (1984) de aprendizaje en general, Dretske (1986) de aprendizaje mediante condicionamiento, Garson (2011) de aprendizaje en conexiones neuronales y Shea (2018) de la contribución de z a favor de la preservación de s . Sin embargo, hay que notar que algunos de esos mecanismos sólo sirven para generar teleología derivada, porque dependen de más teleología. Por ejemplo, la selección artificial necesita agentes que impongan sus propios objetivos, el aprendizaje puede requerir metas previas hacia las cuáles estar encaminado y la selección cultural actúa sobre la cultura, que por sí misma ya involucra agentes con propósitos. Por eso es razonable pensar que el mecanismo fundamental es la selección natural o incluso algo más básico, como algún tipo de selección inorgánica.

Otro mecanismo que merece especial mención es la asignación deliberada de propósito (incluyendo al diseño), porque parece ser válido para generar teleología a pesar de que no requiere que la función en cuestión haya sido realizada previamente por el objeto seleccionado: hasta es posible asignarle un propósito a algo que claramente no sirve para ejecutarlo. Si bien algunos autores (como Shea (2018)) tratan a la asignación deliberada como un caso aparte (haciendo énfasis en el diseño), yo sí la incluiré como mecanismo de selección relevante. Sin embargo, es claro que depende de los objetivos previos del agente que asigna el propósito, mismos que debieron haber surgido de otra manera, de modo que sólo sirve para generar teleología derivada. Así que ello nos vuelve a dejar sólo con la

101 Neander (1991) habla sólo de un tipo de funciones teleológicas: las *funciones propias*, que son el tipo de funciones teleológicas que en biología se atribuye a los rasgos de organismos vivos que son producto de selección natural: aquello que dichos rasgos deberían hacer en las circunstancias para las cuales fueron seleccionados.

selección natural u otro mecanismo más básico.¹⁰²

Por otro lado, los objetos también pueden perder sus funciones teleológicas o adquirir nuevas. Si pensamos en la selección natural, el que z haga f puede dejar de ser útil o z puede dejar de ser útil para hacer f , en cuyo caso z (o su tipo) eventualmente perdería el propósito de hacer f . Entonces será accidental que z perdure, a menos que resulte útil para hacer otra cosa g , en cuyo caso g se convertirá en el nuevo propósito de z . O simplemente podría resultar más útil hacer g que hacer f , en cuyo caso la función de hacer g prevalecería, incluso a costa de la otra. Por ejemplo, las alas de los pingüinos perdieron la función teleológica de ayudar a volar y adquirieron la de ayudar a nadar, seguramente porque a sus ancestros les fue más útil nadar que volar. Por eso es que, si se van a investigar las funciones teleológicas de z en un tiempo t , sólo es relevante la parte más reciente de la historia de selección de z antes de t , porque lo que importa son sus funciones teleológicas en t , no las previas (ver Griffiths (1993) y Godfrey-Smith (1994)). Incluso puede ser que la historia de selección relevante sea extremadamente corta, como cuando el mecanismo selectivo es la asignación deliberada de propósito, porque podemos asignar y retirar propósitos arbitraria y casi instantáneamente. Entonces, tomando en cuenta todo lo anterior, podemos adoptar la siguiente definición:

(FTE) Un rol funcional f es una función teleológica *etiológica* de un objeto z si y sólo si z ha sido seleccionado (recientemente) para hacer f (mediante un proceso selectivo relevante).

Ello independientemente de si z de hecho hace f : por ejemplo, un órgano malformado tiene el mismo propósito que su versión sana, aun si es incapaz de realizarlo. En general, z bien podría haber sido seleccionado porque en el pasado hacía f , porque sus predecesores hacían f , porque otros objetos del mismo tipo hacen f o porque un agente exigió

102 Por otra parte, Kitcher (1993) emplea el término ‘diseño’ de una manera laxa, que abarca todos los mecanismos selectivos relevantes (incluyendo la asignación deliberada). Siguiendo esa idea, podría ser válido entender que *ser seleccionado para hacer f* equivale a *ser diseñado para hacer f* . Sin embargo, yo evitaré el uso que sugiere Kitcher, porque el término ‘diseño’ podría invitar a la confusión, debido a que suele estar fuertemente asociado con la acción de un agente consciente.

arbitrariamente que hiciera f, todo lo cual es compatible con que z de hecho no haga f. Por otra parte, el que f sea una función teleológica de z automáticamente da pie a un aspecto normativo, porque entonces podemos decir que z funciona bien cuando hace f y mal cuando no lo hace¹⁰³. Por último, cualquier otra cosa que z haga o sea capaz de hacer (en un contexto c) es un mero rol funcional (si acaso también sistémico) de z (en c), a menos que también tenga una historia de selección relevante.

El principal problema de las teorías etiológicas es que es difícil determinar qué es lo mínimo necesario para que algo cuente como proceso selectivo en el sentido relevante. Por ejemplo, si bien es claro que es válida la selección natural y cualquier mecanismo selectivo que la presuponga, no es claro qué tanto se puede debilitar esa exigencia: es decir, no es claro si algún tipo de selección inorgánica bastaría. Por otra parte, no es claro qué tipo de sistema debe ser s para que la mera contribución de z a favor de su preservación cuente como mecanismo selectivo relevante: si bien ser un organismo vivo parece bastar, no es claro si otro tipo de sistema podría funcionar. De hecho, si no se delimitan esos aspectos con cuidado, podrían incluirse como funciones teleológicas cosas que claramente no lo son. Por ejemplo, Wright (1973) entiende que f es una función teleológica de z si y sólo si 1) z hace f y 2) z está presente porque hace f; lo cual permite llegar a la conclusión implausible de que una roca que se mantiene en su sitio dentro de un río porque sostiene a otra roca que la protege de la corriente tendría el propósito de sostener a la segunda roca (o la segunda la de cubrir a la primera); o que una fuga de gas que se mantiene abierta porque envenena a todo quien se acerca a intentar cerrarla tendría el propósito de emitir gas o envenenar gente (ver Boorse (1976)). Ahora, la mayoría de las teorías etiológicas, como las de Millikan (1984) y Neander (1991), evaden ese problema porque se enfocan en casos que están lejos de los márgenes, como rasgos que son producto de selección natural sobre organismos vivos, pero ello no desaparece la tarea pendiente de definir la frontera. Además, aun teniendo ya definidos los mecanismos selectivos relevantes, resta determinar qué tanto tiempo debe transcurrir para considerar que una función teleológica ha quedado

103 Nótese que no se exige que z haga f en todo momento, sino sólo en las situaciones para las cuales fue seleccionado para hacer f. O puesto de otro modo: f puede tener la forma de *hacer g en la situación c* o *dotar al sistema s de la capacidad de hacer g*.

establecida o eliminada.

Una última cosa que hay que señalar es que el tener una función teleológica etiológica siempre es una propiedad extrínseca del objeto que la posee, porque tenerla siempre depende de su historia previa, misma que le es extrínseca. Siendo ese el caso, ninguna propiedad intrínseca puede ser explicada en términos teleológicos etiológicos. Por consiguiente, las teorías etiológicas no sirven para explicar la consciencia fenoménica si además se sostiene el internismo fenoménico. Siendo ese el caso, podemos adelantar que la teoría que defenderé no se apoya en ese tipo de funciones teleológicas, sino en el que expongo a continuación.

2.3.3.2.2. Teleología disposicional

La principal razón para pensar que las teorías etiológicas no bastan para dar cuenta de todas las funciones teleológicas es que el mero comportamiento de algunos organismos hace pensar que tienen un propósito, en el sentido de que parecen perseguir metas, pero para duplicar dicho comportamiento no hace falta duplicar su historia previa (que incluye los procesos selectivos que pudieron haberles dado origen). Una manera de reforzar la idea es apelando al internismo fenoménico y a la introspección, porque parecen revelar que nuestros propósitos y la capacidad de asignar nuevas metas en un momento t depende exclusivamente de cómo es nuestra vida mental en t , no de cómo llegamos a ser como somos. Es decir, parece ser que cualquier duplicado interno nuestro (tipo Swampman) compartiría el aspecto subjetivo de nuestra vida mental, lo cual bastaría para que tuviese y pudiese asignar propósitos. Incluso es razonable pensar lo mismo de los cerebros en cubetas. En general, dado que no podemos estar seguros de no ser nosotros mismos como Swampman o los cerebros en cubetas, pero sí es innegable que poseemos propósitos, lo más razonable es concluir que hay un sentido en que estos no dependen de propiedades extrínsecas. Las teorías *disposicionales* (o *sistémicas*) son aquellas que atribuyen funciones teleológicas a los objetos tan sólo con base en lo que de hecho hacen o son capaces de hacer y, por ende, pueden dar cuenta de funciones teleológicas intrínsecas. Dado que se

prescinde de la explicación sobre cómo llegó a existir el objeto en cuestión, el discurso sobre existir *por* algo debe ser reemplazado por el de estar *enfocado hacia* algo.

Un antecedente de las estrategias disposicionales es la teoría sobre análisis funcional que ofrece Cummins (1975), como un refinamiento libre de teleología de las propuestas de Hempel (1959) y Nagel (1961). La idea general es que, si un sistema *s* tiene una disposición o capacidad general *f* en un contexto *c* y una parte suya *z* contribuye a ello gracias a que es capaz de hacer *f* (en *c*), entonces es razonable decir que *f* es una función sistémica de *z* relativa a *s*, *f* y *c*¹⁰⁴. Por ejemplo, mi sangre circula por todo mi cuerpo gracias (en parte) a que mi corazón la bombea (en mi contexto cotidiano), de modo que bombear sangre es una función sistémica de mi corazón relativa a la capacidad de mi cuerpo de mantener mi sangre circulando (en mi contexto cotidiano). Versiones más recientes de la misma idea pueden encontrarse en Craver (2001) y Davies (2001). Para ellos no tiene sentido decir que algo tiene un propósito ni que funciona bien o mal, sino meramente describir lo que de hecho hace.

Sin embargo, muchos autores consideran que un análisis funcional sin teleología es sencillamente incompleto: para ellos lo más importante no es determinar lo que *s* o sus partes de hecho hacen o son capaces de hacer, sino lo que *deberían* hacer. Y es claro que la estrategia recién mencionada no puede (ni pretende) hacer eso, porque trata del mismo modo a cualquier cosa relativamente compleja que *s* sea capaz de hacer (Wouters (2005)). En particular, es incapaz de explicar la parte normativa que parece regir a muchas funciones: sencillamente no deja espacio para que se pueda hablar de que algo funciona bien o mal, porque para ello habría que determinar qué es lo que *s* y sus partes deben hacer, no meramente lo que de hecho hacen o son capaces de hacer (Millikan (1989),

104 Tanto Hempel (1959) como Nagel (1961) hablan de que *s* de hecho haga *f* y que *z* de hecho contribuya a ello (y no mencionan a *f*). Además entienden la situación en términos teleológicos, porque entienden que hacer *f* es un propósito de *s* y que ello explica la presencia de *z*. Cummins (1975), en contraste, no entiende a *f* como propósito de *s* ni como algo que explique la presencia de *s* ni como algo que *s* de hecho haga, sino como una mera capacidad de *s*, que en todo caso es explicada por la presencia de *z* (y sí menciona a *f* y lo entiende como una mera capacidad de *z*, que contribuye a que *s* pueda hacer *f*). La principal diferencia entre las propuestas de Hempel y Nagel es que el primero habla de que la presencia de *z* sea *suficiente* para que *s* haga *f*, mientras que el segundo habla de que la presencia de *s* sea *necesaria* para que *s* haga *f*.

Neander (1991)). Por lo mismo, tampoco permite distinguir entre roles funcionales relevantes o irrelevantes para *s* y sus partes (Bigelow & Pargetter (1987), Millikan (1989)), ni entre contribuciones accidentales y no-accidentales para la consecución de *f*' (Millikan (1993, 2002)). En un final, las teorías de ese tipo parecen pasar por alto muchas distinciones importantes, que en conjunto son un reflejo de que el aspecto teleológico no debe ser dejado fuera.

Entonces se pueden retomar las propuestas originales de Hempel (1959) y Nagel (1961, 1977), que sí se enfocan en roles funcionales especiales que pueden considerarse como metas del sistema en cuestión. Es decir, no se trata de elegir cualquier rol funcional complejo de *s*, sino uno que parezca tener prioridad, porque *s* tiene una *marcada disposición* de comportarse como si lo tuviese como propósito. Los primeros en proponer esa idea fueron Rosenblueth, Wiener & Bigelow (1943) y Sommerhoff (1950, 1959), cuya estrategia fue retomada por Adams (1979) y Boorse (1976, 2002)¹⁰⁵. Su criterio para considerar que *f*' es una función teleológica de *s* es que *s* no sólo tenga la disposición de hacer *f*', sino que tenga la disposición de hacerlo por medio de distintas rutas y bajo una amplia gama de condiciones posibles, lo cual incluye (intentar) adaptarse a cualquier adversidad que amenace con impedirlo. Es decir, si se aprecia cierta tendencia general en el comportamiento de un objeto (dentro de cierto contexto), con varios mecanismos que la refuerzan, tiene sentido pensar que es su propósito, aun si no hay una historia previa de selección. En lo sucesivo, adoptaré la expresión 'marcada disposición' como término técnico para referirme a las tendencias generales de ese tipo. Y una vez identificada una

105 Nagel (1977) también habla sobre las posturas de Woodfield (1976) y Mayr (1965). El primero parte del hecho de que las acciones humanas en pos de una meta suelen partir del deseo de lograr dicha meta y sugiere que algo parecido pasa con otros tipos de sistemas: de algún modo tienen una representación de la meta y actúan para lograrla. Nagel descarta esa propuesta porque hay claros casos en los que se persigue una meta sin tener una representación de ella. Por ejemplo, parece completamente inapropiado decir cosas como que un organismo se desarrolla hasta la adultez porque tiene una representación de ella. Otro problema es que recurre a la representación para explicar la teleología, lo cual es circular si se entiende que la representación es una noción teleológica. Por otra parte, Mayr sostiene que un comportamiento es la persecución de un propósito sólo cuando es resultado de un *programa* interno del objeto que lo presenta, en el sentido de que está regulado por factores internos a él. Nagel objeta que no es claro cuándo ocurre eso, además de que en la práctica nos fijamos en el comportamiento y no en sus causas. También señala que tener propósito dependería de qué programa se tiene, no del mero hecho de tenerlo: en particular se requiere persistencia y plasticidad (ver más adelante).

meta f^* de s (en un contexto c), es posible distribuir funciones teleológicas entre sus partes, de acuerdo a cómo contribuyen a lograr f^* (en c): si f^* es una función teleológica de s (en c), z una parte de s , f un rol funcional de z (en c) y f contribuye a favor de la ejecución de f^* (en c), entonces podemos decir que f es una función teleológica de z (en c). A fin de cuentas es la misma idea de Cummins (1975), pero enfocándose sólo en cierto tipo de roles funcionales que merecen ser considerados persecución de metas.

En particular, los organismos vivos parecen tener como meta sobrevivir y reproducirse¹⁰⁶, porque tienen una marcada disposición de hacerlo, dado que hacen cosas como evitar situaciones peligrosas, expulsar ciertos materiales y consumir otros, regular sus procesos internos y otras tantas cosas que ayudan a mantener sus parámetros internos dentro de ciertos márgenes idóneos (homeostasis), además de procurar aparearse (si son sexuados), reproducirse y (en muchos casos) proteger a sus crías. Algunos autores que han explicado las funciones teleológicas biológicas aludiendo a ese tipo de contribuciones son Wouters (1995, 2007) (supervivencia), Schlosser (1998) (reproducción) y Edin (2008) (auto-mantenimiento). Entonces se puede decir, por ejemplo, que el corazón tiene la función teleológica de bombear sangre, porque gracias a ella el organismo que lo posee puede llevar a cabo uno de sus propósitos, que es sobrevivir.

Sin embargo, si bien esa estrategia es más sensible que la anterior, también es demasiado inclusiva, debido a que no ofrece una manera de distinguir el límite entre los sistemas que genuinamente tienen propósito (como los organismos vivos) y los que carecen de él pero sí tienen una tendencia comportamental clara que cumple con los criterios señalados (como pueden ser ciertos sistemas de equilibrio físico) (Nissen (1980) y Bedau (1992)¹⁰⁷). Además, sigue teniendo el problema de enfocarse sólo en lo que s y sus partes de hecho hacen o tienen la capacidad de hacer, a partir de lo cual no se justifica inferir qué es lo que *deberían* de hacer. Es decir, por más que el comportamiento de s parezca encaminado en

106 Nótese que la reproducción es una contribución a la supervivencia del linaje del organismo en cuestión, de modo que podría subsumirse bajo la noción de *supervivencia*.

107 Bedau (1992) propone fijarse en si f^* es benéfica para s , pero no deja claro en qué sentido se puede determinar eso sin presuponer metas.

una dirección, estrictamente siempre es posible que ello sea un error o mera casualidad, de modo que no basta para asegurar que *s* tiene una meta. Entonces sigue sobreviviendo el problema de distinguir entre funcionar bien y funcionar mal, así como entre funciones teleológicas genuinas y meras contribuciones accidentales en una dirección. Por otra parte, incluso si se sabe que *s* tiene una meta *f**, no es claro cómo distinguir entre funciones teleológicas genuinas y meras contribuciones accidentales a favor de *f**, ni entre disfunciones y meras contribuciones accidentales en contra de *f**. Incluso podría tomarse como función teleológica algo que de hecho es una disfunción, pero que accidentalmente contribuye a favor de *f** (Bedau (1992) y Melander (1997)). Por ejemplo, tener un ala deforme podría ser útil si hay depredadores acechando en el aire, porque entonces es útil no volar, pero no por eso decimos que el propósito del ala deforme es impedir el vuelo.

Una manera de lidiar con esos problemas es presuponer que *s* tiene una meta *f** y enfocarse no en cómo las partes de *s* de hecho contribuyen a hacer *f**, sino en cómo su presencia afecta la probabilidad (o propensión) de hacer *f**. Sobre esa línea, autores como Canfield (1964), Ruse (1971) y Bigelow & Pargetter (1987) se centran en el objetivo de sobrevivir (y reproducirse) y en cómo las partes de *s* afectan su propensión a lograrlo. De ese modo, pueden atribuir a *s* y a sus partes funciones teleológicas que de hecho no realizan (sino que meramente tienen la propensión de realizar) e incluso decir que un rasgo funciona mal cuando disminuye las probabilidades de perdurar o reproducirse de *s*. Además, ello permite distinguir entre rasgos que genuinamente tienen una función teleológica y otros que meramente contribuyen a ella por accidente en una ocasión aislada, así como identificar rasgos que están funcionando mal, a pesar de que accidentalmente contribuyan a la meta en alguna ocasión, porque lo importante es qué probabilidades aumentan, no qué es lo que de hecho ocurre. Por ejemplo, para una paloma, tener alas aumenta la probabilidad de poder volar y, con ello, la de sobrevivir, aun si muere antes de poder volar o si sus alas están deformes y nunca le permiten volar. Por ende, es razonable concluir que un propósito del ala de la paloma es ayudarle a volar. Por otra parte, no negaríamos que ese es un propósito suyo, ni mucho menos diríamos que lo es evitar que la paloma vuele, sólo porque en alguna ocasión aislada sería conveniente que no vuele (por ejemplo, si hubiese un

depredador acechando en el aire).

Un primer problema de esa estrategia es que no se justifica presuponer arbitrariamente que el objetivo de un sistema es uno en específico. En particular, no se justifica presuponer que es sobrevivir o reproducirse. Sin embargo, ese problema ha sido mayormente pasado por alto en la literatura filosófica. Otro problema es que es completamente circunstancial qué rasgos aumentan las probabilidades de qué: para evaluar si un rasgo ayuda o no a algo, hay que determinar primero qué ambiente le corresponde al sistema que lo posee, lo cual no se puede saber sólo con base en cómo es el rasgo o sistema. Bigelow & Pargetter (1987) reconocen eso y señalan que su propuesta debe ser relativizada al hábitat natural de s, el cual sugieren determinar históricamente, es decir, fijándose en dónde han surgido los sistemas como s. Sin embargo, como señala McLaughlin (2001), incluso fijando el ambiente, para todo rasgo es posible encontrar un sentido en el que aumenta o disminuye las probabilidades de sobrevivir en él, lo cual lleva a atribuir demasiadas funciones teleológicas y muchas de ellas intuitivamente irrelevantes. Por ejemplo, si bien no se justificaría concluir que un propósito de un ala deforme es impedir el vuelo (a secas) sólo porque en ocasiones hay depredadores acechando en el aire, sí podrían admitirse propósitos relativos y concluir que uno de ellos es *impedir el vuelo cuando hay depredadores acechando en el aire*, lo cual no parece aceptable.

Para lidiar con el problema de la arbitrariedad del propósito, la estrategia se podría combinar con la anterior: identificar *marcadas disposiciones* de hacer algo, pero contemplando propensiones y no meramente lo que de hecho ocurre. Otra opción sería asignar funciones teleológicas a s y sus partes con base en cómo se comporta o tiene la propensión de comportarse la mayoría de los objetos de su mismo tipo en su hábitat natural y no meramente s y sus partes. Y para saber qué tipo de objetos son s y sus partes, se puede recurrir a su pasado evolutivo para establecer homologías¹⁰⁸ (ver Amundson & Lauder (1994)). Por ejemplo, con base en su origen, se puede juzgar que un objeto x es una paloma adulta viva y que, por ende, cierto órgano suyo z es un ala de paloma adulta viva.

108 En biología evolutiva, dos rasgos son homólogos cuando descienden de un mismo rasgo.

Por otra parte, con base en cómo funcionan (y han funcionado) las alas en la mayoría de las palomas adultas vivas, se puede concluir que *z* tiene la función teleológica de ayudar a volar a *x*, la realice o no. En general, sólo los roles funcionales de *z* compartidos con la mayoría de las alas de paloma adulta viva serán candidatos a ser funciones teleológicas de *z*. Sin embargo, cabe destacar que recurrir a la historia evolutiva de *s* y sus partes es una manera (a veces camuflada) de introducir etiología, lo cual hace que los criterios que lo hacen no sean estrictamente disposicionales, sino más bien etiológicos o híbridos (ver Artiga & Martínez (2016)).

Por esa razón, entenderé como funciones teleológicas *genuinamente* disposicionales sólo a aquellas en las que lo que *z* y *s* deberían hacer se deriva exclusivamente de lo que de hecho hacen o son propensos a hacer, y no de lo que otros objetos hacen, han hecho, son propensos a hacer o han sido propensos a hacer. Así que el tipo de funciones teleológicas disposicionales que me interesarán son las que son capturas del siguiente modo:

(FTD') Un rol funcional *f* es una función teleológica (genuinamente) *disposicional* de un objeto *z* (en un contexto *c*) si y sólo si se cumple alguna de las siguientes condiciones:

- *z* es un sistema que (en *c*) tiene una *marcada disposición* (incluyendo propensiones) de comportarse como si tuviera el propósito de hacer *f*.
- *z* es parte de un sistema *s* que (en *c*) tiene una *marcada disposición* (incluyendo propensiones) de comportarse como si tuviera el propósito de hacer *f* y *z* contribuye a favor de que *s* tenga esa marcada disposición, en virtud de que *z* es propenso a hacer *f*.

También cabe notar que ese criterio es el único que realmente da como resultados funciones teleológicas intrínsecas, porque todas las alternativas recurren a la historia de *z* (que le es extrínseca).

2.3.3.2.3. Posturas híbridas, unitarias y pluralistas

Como comenté previamente, las teorías etiológicas y disposicionales sobre la teleología no son opuestas, sino complementarias. Algunos autores incluso ofrecen teorías en las que ambas se combinan o tratan de explicar un tipo de función teleológica en términos de la otra. Por ejemplo, Griffiths (1993) sugiere que las funciones teleológicas etiológicas son una subclase de las disposicionales: aquellas que se obtienen si se presupone que el objetivo es la supervivencia o reproducción del organismo; y Walsh (1996) y Walsh & Ariew (1996) desarrollan ideas similares.

Por otra parte, hay quienes plantean teorías mayormente disposicionales, pero retoman la idea de que el propósito de un rasgo es lo que explica su presencia, de modo que se enfocan en rasgos de sistemas que contribuyen a su propia preservación. Ejemplos de eso son las posturas de McLaughlin (2001), Christensen & Bickhard (2002) y Mossio, Saborido & Moreno (2009), de las cuales me concentraré en la última, porque es la más reciente y precisa. Mossio, Saborido y Moreno atribuyen funciones teleológicas sólo a objetos que describen como *sistemas que se auto-mantienen y que están organizacionalmente cerrados y diferenciados*, siendo los seres vivos el ejemplo paradigmático de ello. Un sistema que se auto-mantiene es uno que se preserva idéntico a sí mismo mediante procesos de auto-reparación y auto-replicación (McLaughlin (2001)). Que sea organizacionalmente cerrado quiere decir que tenga una dinámica causal circular, en el sentido de que posea partes que contribuyen al mantenimiento de otras partes que a su vez mantienen a las primeras. Por último, los sistemas organizacionalmente diferenciados son aquellos en los que es posible distinguir partes que contribuyen de forma distinta al funcionamiento del sistema. Este último requisito está pensado para evadir ejemplos de sistemas que cumplen con los otros dos pero intuitivamente son demasiado simples como para atribuirles teleología. La idea general, tomando en cuenta algunas consideraciones señaladas por Artiga (2011)¹⁰⁹, es la siguiente:

109 Mossio, Saborido & Moreno (2009) se expresan diciendo meramente que un rasgo z (de s) tiene una función teleológica si y sólo si: 1) s contribuye al mantenimiento de la organización O de s , 2) z es mantenido y producido por O y 3) s es organizacionalmente diferenciado. Así que parte de lo que hace

Un rol funcional f es una función teleológica disposicional de z , dentro de un sistema s , si y sólo si se cumplen tres condiciones:

- El que z haga f contribuye al mantenimiento de la organización o' de s .
- z es producido y mantenido por o' .
- s es organizacionalmente diferenciado respecto a z .

Cabe señalar que hay algunas funciones teleológicas, como la reproducción, que en principio no contribuyen a mantener al sistema que las realiza (y a su organización interna) y, por ende, parecen ser dejadas fuera. Sin embargo, la teoría las acomoda aludiendo a sistemas de orden superior, como puede ser la especie a la que s pertenece o la rama entera en la que se encuentra dentro del árbol filogenético: es claro que la reproducción de los organismos individuales ayuda a persistir a la especie a la que pertenecen (y la organización que la caracteriza, incluyendo la de sus miembros). De hecho, esa estrategia sirve para entender cualquier función teleológica etiológica como disposicional. El problema es que también puede entenderse como una manera camuflada de introducir etiología. De hecho, no es claro cómo es que es posible determinar a o' sin saber qué tipo de objeto es s y no es claro cómo determinar qué tipo de objeto es s sin aludir a su historia, es decir, sin introducir etiología (ver Artiga & Martínez (2016)). Pero tal vez Mossio, Saborido & Moreno (2009) aceptarían reconocer su postura como etiológica, dado que de hecho sostienen que su estrategia es expandir las ideas etiológicas para subsumir lo disposicional dentro de lo etiológico (a pesar de que su propuesta es marcadamente disposicional).

Por otro lado, Shea (2018) ofrece una teoría mayormente etiológica, pero con algunos elementos disposicionales. Él se enfoca en cierto tipo de funciones teleológicas, que denomina 'funciones de tarea':

Artiga (2011) es hacer énfasis en f y en que s sea organizacionalmente diferenciado específicamente respecto a z . También agrega la especificación de que s sea organizacionalmente cerrado, pero ello ya quedaba garantizado por las primeras dos condiciones (y por eso lo omito). Después ofrece más refinamientos de las condiciones, pero no las incluyo porque no son relevantes para expresar la idea general.

Un rol funcional f^* de un sistema s es una función de tarea de s si y sólo si se cumplen dos condiciones:

- f^* es una función de consecuencia robusta de s .
- f^* es una función estabilizada de s o s ha sido diseñado para hacer f^* .

Donde:

Un rol funcional f^* de un sistema s (en un contexto c) es una función de consecuencia robusta de s (en c) si y sólo si (en c) se cumplen dos condiciones:

- s' hace f^* en respuesta a diversos estímulos.
- s' hace f^* ante diversas condiciones externas relevantes.

Un rol funcional f^* de un sistema s es una función estabilizada de s si y sólo si hacer f^* ha sido sistemáticamente estabilizado de alguna de las siguientes maneras:

- Contribuyendo directamente al éxito evolutivo de sistemas del tipo de s haciendo f .
- Contribuyendo por medio de aprendizaje a la disposición de z de hacer f^* .
- Contribuyendo directamente a la persistencia de s , donde s es un organismo vivo.

Así que, en síntesis, una función de tarea es cierto tipo de función teleológica disposicional (una versión debilitada de marcada disposición, porque no exige mecanismos de refuerzo) que al mismo tiempo es una función teleológica etiológica. Por ejemplo, el estómago tiene como función de tarea el desintegrar lo que se introduce en su interior, porque ello es una función de consecuencia robusta suya y también una función estabilizada suya. Es una función de consecuencia robusta suya porque lo hace con la mayoría de lo que llega a su interior, recurriendo a diversas estrategias y bajo prácticamente cualesquiera condiciones externas. Y es una función estabilizada suya porque es gracias a que sus antepasados han hecho eso que se ha preservado como rasgo del tipo de organismos al que pertenece.

Además, una vez identificando una función de tarea f^* de s , es posible distribuir funciones teleológicas a algunas partes de s , dependiendo de cómo contribuyen a la realización de f^* .

2.3.3.3. Funciones teleológicas representacionales

Ya he ofrecido distintas vías naturalistas para entender las funciones teleológicas, así que sólo resta explicar con qué y cómo deben de combinarse para obtener representaciones, es decir, qué puede desempeñar el rol funcional de apuntar hacia algo y cómo puede convertirse en una función teleológica según cada postura. Debido a cómo ha sido abordado el tema en la literatura filosófica, estudiaré por separado las representaciones que se apoyan en teorías etiológicas y las que se apoyan en teorías disposicionales, pero procuraré señalar algunas convergencias.

2.3.3.3.1. Representación etiológica

Siguiendo los pasos de autores como Stampe (1977) y Dretske (1981, 1986, 1988)¹¹⁰, una manera natural de entender el rol funcional de apuntar hacia algo es en términos de *portar información* (sobre dicho algo), en un sentido correlacional o evidencial. Ello parece buena idea porque tiene sentido pensar que las representaciones de más bajo nivel, como las que están en la base de la percepción verídica, no son otra cosa que detectores de aquello que representan, gracias a que son regularmente causados por ello y, por ende, son evidencia suya. Por ejemplo, los estados del sistema visual, desde los estados de la retina hasta las experiencias visuales, son evidencia de que hay cierta distribución de objetos en el mundo externo, porque son regularmente causados por los tipos de luz que estos regularmente emiten o reflejan hacia el ojo. Y lo mismo aplica para los procesos del sistema nervioso autónomo, porque también se apoyan en transmisión y procesamiento de información.

110 Sus trabajos se apoyan en la teoría de la información de Shannon (1948). Sin embargo, él se enfoca en la cantidad de información, en el sentido de cuánto reduce la incertidumbre sobre algo, mientras que ellos hablan también del tipo de información en cuestión, es decir, acerca de qué es y de qué manera reduce la incertidumbre.

En general, todo efecto que un objeto tiene sobre su entorno es evidencia de su presencia y la presencia de todo objeto sirve para predecir sus efectos, siendo así evidencia de ellos. En ese sentido, todo porta *información* sobre todas sus causas y consecuencias. Por ejemplo, si el fuego genera humo, el humo podrá ser tomado como evidencia de fuego y el fuego servirá como evidencia de que habrá humo. Regresando al ejemplo de la percepción visual, podemos decir que cada eslabón en la cadena causal que conduce a una experiencia visual es evidencia de cómo es cada otro eslabón, ya que permite averiguarlo. Además la idea puede ser expandida hacia cualquier correlación sistemática, involucre o no causalidad: para justificar inferir un evento B a partir de un evento A, basta con que A se correlacione con B de forma suficientemente fuerte, sin importar si ello se debe a alguna relación causal¹¹¹. En otras palabras, podemos hablar meramente de que la ocurrencia de algunos eventos aumenta la probabilidad de la ocurrencia de otros, o incluso la garantiza. En términos más formales, cuando en un contexto c la probabilidad de un evento (o tipo de evento) B dado un evento (o tipo de evento) A es mayor que la de B a secas (es decir, cuando $P(B|A) > P(B)$ en c), decimos que, en c, A porta *información correlacional* a favor de B (o que A porta información correlacional *positiva* sobre B)¹¹²; y si la ocurrencia de A garantiza la de B (en c) (es decir, si $P(B|A) > P(B) = 1$ en c), decimos que, en c, A es un

111 Ello a pesar de que sea tentador pensar que la causalidad siempre es indispensable para que haya tal correlación. Para un tratamiento general de las nociones de *correlación, evidencia e indicación*, ver Kelly (2006/14).

112 Y si la probabilidad de B dado A es menor que la de B a secas (es decir, si $P(B|A) < P(B)$), decimos que la información es *en contra* de B (o que es información *negativa* sobre B). En general, lo que importa para hablar de información correlacional es que ambas probabilidades sean distintas (es decir, que $P(A|B) \neq P(A)$) (ver Shea (2018)). Pero $P(B|A) < P(B)$ es equivalente a $P(\neg B|A) > P(\neg B)$, porque aumentar la probabilidad de un evento es equivalente a disminuir la probabilidad de su negación, de modo que se puede pensar todo en términos de información positiva.

Otra característica de la relación de portar información correlacional es que siempre es simétrica: A porta información correlacional a favor de B si y sólo si B porta información correlacional a favor de A. Es decir, $P(B|A) > P(B)$ si y sólo si $P(A|B) > P(A)$. Para demostrarlo, tomemos en cuenta que, en general, la probabilidad de algo siempre es igual a la cantidad de casos en que aparece dividida entre la cantidad total de casos tomados en cuenta. Sin pérdida de generalidad, podemos estipular que la cantidad de casos de A&B es f, la de A& \neg B es g, la de \neg A&B es h y la de \neg A& \neg B es i. Entonces $P(A|B) = f/(f+h)$, $P(A) = (f+g)/(f+g+h+i)$, $P(B|A) = f/(f+g)$ y $P(B) = (f+h)/(f+g+h+i)$. Y resulta ser que $f/(f+h) > (f+g)/(f+g+h+i) \Leftrightarrow f/(f+g) > (f+h)/(f+g+h+i)$, es decir, que $P(A|B) > P(A) \Leftrightarrow P(B|A) > P(B)$.

Sin embargo, puede ocurrir que la relación sea más fuerte en una dirección que en la otra. Por ejemplo, tener sarampión aumenta la probabilidad de tener ronchas y tener ronchas aumenta la probabilidad de tener sarampión, pero, dado que hay muchas fuentes posibles de ronchas además del sarampión (y más comunes), pero es muy poco probable tener sarampión sin ronchas, tener sarampión aumenta la probabilidad de tener ronchas más de lo que tener ronchas aumenta la probabilidad de tener sarampión.

indicador de B (ver Dretske (1986, 1988)). Así que resulta natural pensar que los roles funcionales relevantes para las representaciones de más bajo nivel son portar información correlacional positiva o indicar. Entonces sólo resta explicar cómo esos roles funcionales pueden dar pie a funciones teleológicas.

Para motivar (más) la necesidad de ese factor teleológico, cabe notar que para averiguar el contenido representacional de algo no basta con fijarse en la información correlacional que porta (incluyendo aquello que indica¹¹³), porque todo porta mucha más información correlacional que la que propiamente representa. Por ejemplo, los estados de la retina no sólo portan información correlacional positiva sobre la luz que la impacta y las fuentes de la misma, sino también sobre los futuros estados del nervio óptico y de las zonas visuales del cerebro, así como de las futuras experiencias del sujeto, sus reacciones ante ellas, los cambios en el mundo externos como consecuencia de ello, la propia existencia de la retina, del ojo, del nervio óptico, del sujeto, de sus antepasados y descendientes, de la humanidad, del planeta Tierra, del Sistema Solar, etcétera: básicamente de todo lo que influyó en que tuviesen lugar los estados retinales en cuestión y todas su consecuencias futuras. Así que, a menos que estemos dispuestos a decir que todo eso es contenido representacional de los estados retinales, es indispensable encontrar algo que permita destacar una información sobre otra. Esa dificultad es parte de un problema más amplio, conocido como el *problema de la indeterminación funcional del contenido*, mismo que admite múltiples variantes (ver Dretske (1986), Fodor (1990), Neander (1995) y Neander & Schulte (2004-20)). En síntesis, la idea general es que los roles funcionales (como portar información correlacional) suelen resultar insuficientes para determinar que algo es una representación y cuál es su contenido. En el caso particular recién señalado, el problema de fondo es uno que ya había mencionado antes (en la sección 2.3.1): que la mera correlación sistemática es insuficiente para que haya representación. Pero la solución general es la misma que se señaló entonces: introducir teleología. Es decir, si lo que se quiere averiguar es cuál de

113 Nótese que el único caso en que indicar B no implica portar información correlacional a favor o en contra de B es cuando la probabilidad de B a secas es 1, porque entonces nada puede alterarla. Es razonable sostener que en tal caso no debería hablarse de información correlacional sobre B, porque no es un caso en el que se reduzca la incertidumbre sobre la ocurrencia de B.

entre toda la información portada por algo es su contenido representacional, lo que hay que hacer es determinar cuál de toda ella es la que dicho algo tiene el propósito de portar. En otras palabras, el contenido representacional de las representaciones de más bajo nivel es la información correlacional que tienen el *propósito* de portar.

Sobre esa línea, Dretske (1986, 1988) sostiene que las representaciones senso-perceptuales son estados que tienen la función teleológica *etiológica* de *indicar* algo. En otras palabras, *z* es una representación sensorio-perceptual de *x* si y sólo si *z* tiene el propósito de garantizar la presencia de *x*, porque fue seleccionado para ello. Por ejemplo, las experiencias visuales tienen el propósito de garantizar que el mundo externo es de tal o cual modo, porque esa es la razón por la que fueron seleccionadas naturalmente. De manera similar, los estados de un termómetro son representaciones de temperaturas porque tienen el propósito de garantizar que dichas temperaturas son el caso, porque para eso fueron seleccionados por quien diseñó el termómetro. Así que resulta razonable sostener lo siguiente:

RE1) Un objeto (o tipo de objeto) *z* es una representación (etiológica) de un objeto (o tipo de objeto) *x* si *z* fue seleccionado para indicar *x*.

Cabe agregar que todo lo que tome el lugar de *z* es un vehículo representacional y todo lo que tome el lugar de *x* es su contenido. Por otra parte, el que *z* (o los objetos de su tipo) haya sido seleccionado para indicar *x* no necesariamente se debe a que *z* de hecho sea o haya sido un indicador de *x*, sino que bien puede deberse sólo a que porta o ha portado información correlacional positiva sobre *x*. En otras palabras, *z* puede ser seleccionado para garantizar la presencia de *x* meramente porque su presencia aumenta o ha aumentado la probabilidad de la de *x*, aunque no la garantice ni la haya garantizado. Por ejemplo, es razonable pensar que nuestros estados senso-perceptuales fueron seleccionados para garantizar que el mundo es como sugieren, porque se correlacionan fuertemente con que lo sea, a pesar de que sean falibles. Ejemplos de autores que se apoyan en variantes de RE1 son Millikan (1984, 1989, 2017), Papineau (1984, 1987), Fodor (1987), Lloyd (1987,

1989), Neander (1995, 2017), Rupert (1999), Usher (2001), Ryder (2004) y Shea (2018).

Una razón para modificar RF1 es que no es claro que indicar sea un rol funcional, porque no es claro que sea una acción, así que no es claro qué es exactamente lo que es seleccionado ni por qué (ver Neander & Schulte (2004-20)). Otro problema es que la noción de indicación es demasiado fuerte, en el sentido de que si z indica x siempre que haya z habrá x , lo cual conduce a que las representaciones en cuestión siempre sean exitosas o nunca lo sean, lo cual entra en conflicto con la idea de que deberían poder ser exitosas en algunas ocasiones y fallidas en otras (ver Fodor (1990)). De hecho, en la práctica son raros los eventos que garantizan la ocurrencia de otros y hasta es claro que nuestros estados senso-perceptuales no lo hacen, porque bien podrían ser alucinaciones, de modo que esos estados no son indicadores y, por ende, nunca cumplen su propósito según RE1, lo cual parece inaceptable. Así que es natural concluir que el propósito relevante debe ser (al menos ligeramente) distinto de la indicación. También cabe señalar que no puede ser meramente portar información correlacional positiva, porque eso es satisfecho hasta por las alucinaciones.

Por otro lado, aun ignorando esos conflictos y a pesar de que RE1 evita muchos problemas sobre la determinación del contenido, algunos sobreviven. En particular son problemáticos los casos en los que z porta información correlacional sobre x porque porta información correlacional sobre otro evento x' , que a su vez porta información correlacional sobre x , en cuyo caso no es claro si z tiene el propósito de portar información sobre x , sobre x' o sobre alguna combinación de x y x' . Dretske (1986) se centra en el ejemplo de las bacterias magnetotácticas, que son anaerobias, viven en los océanos y tienen pequeños magnetos (magnetosomas) que son atraídos por el norte magnético y de esa manera direccionan a la célula hacia zonas más profundas y relativamente libres de oxígeno. El problema es que no es claro si lo correcto es decir que los magnetosomas tienen la función teleológica de portar información sobre las zonas relativamente libres de oxígeno, sobre norte magnético, sobre conjunción de ambas cosas o sobre su disyunción. Ante esa situación, si se pusiera un magneto cerca de la célula, conduciéndola en una dirección distinta y probablemente

hacia su muerte, no sería claro si sus magnetosomas estarían funcionando bien o mal: ¿estarían funcionando bien porque seguirían apuntando hacia el norte magnético o mal porque no estarían conduciendo a la bacteria hacia las zonas libres de oxígeno? Otro problema es que en ocasiones más de un contenido puede explicar una misma conducta. Por ejemplo, las ranas intentan comerse cualquier objeto negro pequeño que estimule su sistema visual y no es claro si lo que representan sus estados visuales es *objeto negro pequeño*, *mosca*, *comida* o alguna combinación de esas cosas.

Siguiendo a Neander (2012), una manera de lidiar con esos problemas es simplemente especificar qué es lo que en última instancia ha sido importante detectar en la historia evolutiva del sistema en cuestión. Además, es posible plantear la idea sin hablar de indicación, si en lugar de concentrarnos en estados particulares nos enfocamos en la maquinaria que los produce:

- RE2) Un estado (senso-perceptual) z de un sistema s tiene el contenido representacional x (y no x') si se cumplen dos condiciones:
- s fue seleccionado para producir z en respuesta a x .
 - Si s fue seleccionado para producir z en respuesta a x' (distinto de x), lo fue porque era un medio para producir z en respuesta a x y no viceversa.¹¹⁴

En el ejemplo de la bacteria magnetotáctica, ello da como resultado que los magnetosomas tienen como contenido representacional las zonas relativamente libres de oxígeno y no el norte magnético, porque en un final lo útil para el organismo es que indiquen lo primero, y es por eso que se seleccionó el rasgo en cuestión. En cuanto al caso de las ranas, Neander (2012) sostiene que lo razonable es concluir que el contenido es *objeto negro pequeño*,

¹¹⁴ Neander (2012) habla específicamente de estados que son representaciones y de contenidos descriptivos. Además especifica que x y x' son eslabones de una misma cadena causal que conduce a la producción de z , en la que x' es más cercano a la producción de z que x . Por último, ella plantea la segunda condición del siguiente modo: Si s fue seleccionado para producir z en respuesta a x y a x' , fue seleccionado para producir z en respuesta a x' porque era un medio para producir z en respuesta a x y no viceversa. Yo omito todos esos detalles porque me interesa una noción más general, que incluye a la que se refiere Neander (2012).

porque el sistema visual de la rana no es sensible a propiedades como ser comida o mosca y, por ende, no pudo haber sido seleccionado para detectar eso. Sin embargo, considero que ese razonamiento es errado, porque es un hecho que los estados en cuestión portan información correlacional a favor de la presencia de comida (para rana) o mosca y, de hecho, lo más plausible es que hayan sido seleccionados por eso. Es decir, es accidental que también ocurran cuando hay cualquier otro objeto negro pequeño. Además, detectar moscas es útil porque son comida, por lo que considero que la conclusión adecuada es que el contenido representacional en cuestión es *comida*. Otra manera de capturar la misma idea es la de Millikan (1984, 1989) y Papineau (1984, 1987, 2016), quienes sugieren pensar en *s* como un *consumidor de representaciones* y enfocarse en qué es lo realmente importante para él: detectar *x* o detectar *x'*. De hecho, fijarse en lo que *s* hace ante *x* puede ayudar a delimitar no sólo qué objeto es el representado, sino qué significa su presencia para *s*, permitiendo determinar, por ejemplo, que el contenido es *comida* y no *mosca* ni *objeto negro pequeño*. De otra manera no se explicaría por qué las ranas siempre tratan de comerse dichos objetos.

Una forma más general de entender la situación parte de notar que la única razón por la que es útil tener estados que son producidos en respuesta a algo es porque sirven para reaccionar adecuadamente ante dicho algo. En otras palabras, tener estados que portan información correlacional sobre *x* es útil sólo en la medida en que tener esos estados habilita algún comportamiento apropiado ante *x*. Por consiguiente, los estados que son seleccionados naturalmente como representaciones de *x* dentro de un sistema *s* son sólo aquellos que, además de tener el rol funcional de portar información a favor de la presencia de *x* tienen el de contribuir a que *s* se comporte adecuadamente ante *x*. Siendo ese el caso, para identificar el contenido representacional de *z* es indispensable fijarse en cómo ha sido y hacia qué ha estado dirigido el comportamiento de los antecesores de *s* resultante de tener estados como *z*, no meramente en las causas usuales de los estados como *z*. Entonces es razonable decir que *z* ha sido seleccionado para indicar aquello hacia lo que el comportamiento *útil* desencadenado por los antecesores de *s* ha estado enfocado, que no es más que una manera de decir que eso es lo realmente importante para *s*. Detrás de ese

razonamiento también se puede encontrar la idea de que portar información correlacional o ser causado por algo no son propiamente roles funcionales, sino meramente propiedades: no son algo que *z haga*, sino sólo algo que caracteriza a *z*. Por eso es que Papineau (1987) y Millikan (1989) hacen énfasis en los efectos derivados de tener *z*, no en sus causas, porque *z* es seleccionado sólo en virtud de los primeros. Sin embargo, también es razonable sostener que la utilidad de dichos efectos se debe a que *z* tiene las propiedades correlacionales que tiene, por lo que no es del todo inapropiado decir que *z* ha sido seleccionado debido a ellas (y *para* tenerlas) y seguir tratandolas como roles funcionales.

Por otra parte, diversos estados de *s* pueden ser representaciones de un mismo objeto (o tipo de objeto) o de distintos aspectos de él, con múltiples comportamientos asociados. Además, la selección (del tipo que sea) también actúa sobre los vínculos que un sistema hace entre sus distintos estados. Así que es natural que surjan redes relativamente complejas de representaciones en torno a un mismo objeto (o tipo de objeto), que abarcan diversas maneras posibles de interacción e incluso relaciones con otros objetos (a través de sus respectivas representaciones), lo cual ya podríamos comenzar a considerar *proto-conceptos* o, más bien, *uniceptos* (Millikan (2013, 2017)¹¹⁵). En otras palabras, lo que se tiene son diferentes representaciones de un mismo objeto, que apuntan a él empleando distintas propiedades suyas como puente y, al integrarse, lo *caracterizan* como algo que las tiene a todas ellas. Por ejemplo, tengo diferente formas de identificar a una persona en particular, porque hay varias maneras en las que ella me puede afectar, creando en mí diversos estados que son representaciones suyas y de las características mediante las cuales me afecta: puedo distinguir distintas cualidades de su voz, de su aspecto visual, de su olor, de su forma de ser y de pensar, de las actividades que realiza frecuentemente, de mi trato con ella, de cómo la conozco, etcétera. Y puedo integrar todo eso en una representación compleja de ella, que la caracteriza como teniendo todas esas propiedades.

Cabe señalar que para ello es útil que los vehículos representacionales vinculados guarden

115 Ver sección 1.4.1 y nota 48.

relaciones entre sí que reflejen las de aquello que representan, es decir, que la estructura resultante sea una suerte de modelo a escala de su referente. Ello por la sencilla razón de que facilita el procesamiento de la información: mientras más arbitrarias y enredadas sean las asociaciones, más complicado será procesarlas. Así que lo ideal es que las relaciones entre las partes del referente puedan ser inferidas casi directamente de las de las partes del vehículo. Shea (2018) expresa esa idea diciendo que, además de la información correlacional, un elemento explotable para generar representaciones es la *correspondencia* (o *isomorfismo*) *estructural* entre vehículo y contenido. Un ejemplo útil para aclarar la idea son los mapas cartográficos: son muy útiles para representar espacios y ubicaciones porque las relaciones espaciales entre las partes del mapa corresponden directamente (o casi) a relaciones espaciales (a escala) entre sus referentes. En contraste, sería considerablemente más laborioso extraer esa información a partir de un texto que describiese a detalle el espacio y ubicaciones en cuestión, a pesar de que podría contener exactamente la misma información. Ello explica, por ejemplo, que algunas relaciones espaciales dentro del sistema visual literalmente reflejen (a escala) las de aquello que rastrean (como sucede en la retina y la corteza visual primaria). Sin embargo, si bien todo eso puede ser útil para representar, no es indispensable para que haya representación.

También es importante señalar que incluso puede ocurrir que una representación incluya algunas asociaciones erradas, de modo que se caracterice al referente como teniendo una propiedad que de hecho no tiene. Un ejemplo de ello podría ser la humanización de la naturaleza que suele encontrarse en diversas religiones, en particular cuando se le atribuye agencia consciente a fenómenos naturales. Si esa es una asociación errada, su existencia podría ser resultado accidental o guardar una utilidad especial. Por ejemplo, bien podría ser mero resultado accidental de tener una tendencia a atribuir agencia consciente a lo que sea, producto de haber evolucionado en un entorno social, dentro del cuál es muy común interactuar con entes que tienen agencia consciente. O simplemente podría ser que atribuirle ese tipo de orden al mundo natural pueda ser más llevadero que pensarlo como un mero caos sin propósito. También es útil, por ejemplo, que un organismo le atribuya un valor intrínseco a su propia vida o a la vida en general, aun si no lo tiene, porque ello lo

impulsa a luchar por su propia preservación (y la de su comunidad).

Por otro lado, es útil poder reclutar algunas representaciones y manipularlas aun en ausencia de su referente, para poder prever situaciones y hacer planes, así como para comunicarse con otros entes en ausencia de lo que se comunica, rasgos que también pueden estar sujetos a selección, eventualmente dando paso a la imaginación y el pensamiento abstracto. Además, si el sistema en cuestión requiere comunicarse con otros, es indispensable que comparta contenidos representacionales con ellos y alguna manera de expresarlos, lo cual da pie a la capacidad (colectiva) de adquirir y generar *lenguajes* y *conceptos* propiamente dichos.

Otro detalle importante, es que es poco plausible que nazcamos con representaciones predeterminadas para todo tipo de objeto, sobre todo si se trata de objetos con los que prácticamente no hemos interactuado en nuestro pasado evolutivo. Parece mucho más razonable sostener que nacemos, si acaso, con unas cuantas representaciones preestablecidas, pero con la maquinaria necesaria para adquirir más, es decir, para otorgar a diversos objetos (internos o externos) la función teleológica (derivada) de remitir a otros. Dicha maquinaria también es producto de selección natural y por eso es que tiene un propósito y es capaz de asignar o aprender objetivos derivados (ver Millikan (1984)). Tiene sentido pensar que las representaciones con las que se nace están relacionadas con procesos vitales y senso-percepción muy básica, mientras que las demás, especialmente las relacionadas con aspectos culturales, aparecen después: son aprendidas. Es decir, el proceso de selección comienza siendo selección natural y continúa mediante aprendizaje y condicionamiento cultural (o lo equivalente en cada ser vivo), a través de lo cual el organismo integra dentro de su sistema nuevos signos: algunos empleados previamente por su comunidad, producto de selección cultural en generaciones anteriores, otros que son individuales pero sirven como puente con los de la comunidad (como los uniceptos son un puente con los conceptos) y algunos completamente nuevos. A fin de cuentas, lo importante es que haya algún proceso de selección, no si éste es biológico o no. A grandes rasgos, se trabaja por medio de prueba y error, dentro de una misma generación o a través

de varias, reforzando unas asociaciones y debilitando otras, dependiendo de los resultados obtenidos con ellas. Así vinculamos algunos objetos (internos o externos) con otros y nos acoplamos a nuestro entorno. En un final, cada organismo termina con su propia red de asociaciones, pero congruente con la ya disponible a su alrededor.¹¹⁶

Ahora, siguiendo a Peacocke (1992) y Neander & Schulte (2004-20), un problema de las teorías etiológicas sobre la representación es que no es claro cómo es posible generar ciertos contenidos por la vía que proponen, porque no es claro en qué sentido su representación puede ser adaptativa. Especialmente problemáticos con los contenidos sobre cosas abstractas, futuras o inexistentes, porque no parece posible haber interactuado con ellas en el pasado. Ejemplos de ello son la democracia, la virtud, el mañana y los unicornios. Por otra parte, algunas cosas, como los quarks y las verdades lógicas y matemáticas, pueden tener una influencia tan general que no es claro cómo podrían tenerse detectores específicos para ellas. En respuesta, se puede señalar que de hecho sí es útil tener representaciones de todas esas cosas, aunque no lo haya sido en nuestro pasado evolutivo distante. Además, también existen procesos de selección posteriores a la historia evolutiva de cada organismo, como la selección cultural y la mera capacidad de generar representaciones nuevas a partir de viejas. Por último, ninguno de esos contenidos parece imposible de construir a partir de otros que son más concretos y tienen una utilidad más obvia. Así que el problema no parece imposible de resolver, por sencillamente se puede aludir a esos mecanismos.

De hecho, lo único que se requiere para dar cuenta de todas las representaciones es dar cuenta de las más básicas y de la capacidad de generar nuevas, de lo cual parecen perfectamente capaces las teorías etiológicas. El resto del trabajo consiste meramente en identificar qué representaciones básicas y comportamientos asociados hay que manipular y

116 Skyrms (2010) ofrece un estudio matemático detallado sobre cómo la repetición de una misma situación, inicialmente con respuestas aleatorias ante ella, pero con recompensas asociadas a cada una, desemboca en patrones estables (nótese que la supervivencia es en sí misma una recompensa). Por otra parte, Houghton (2005), Thomas & McClelland (2008), Griffith y Kemp & Tenenbaum (2008) presentan algunas explicaciones computacionales de cómo funciona la cognición en general empleando principios probabilísticos como los que he mencionado, incluyendo lo relativo a aprendizaje.

de qué manera para construir cada otra representación y comportamiento. Por ejemplo, si se tiene una representación de rocas y una de cuadrados, cada una asociada a determinado comportamiento, bastará con unir esas cosas adecuadamente para construir una representación de rocas cuadradas, con su correspondiente comportamiento asociado. Incluso puede emplearse esa estrategia para generar representaciones de combinaciones imposibles de propiedades: caracterizaciones sin referente. Por ejemplo, bastará con vincular apropiadamente representaciones de círculos con representaciones de cuadrados para producir una de círculos cuadrados. Justamente ahí se puede fundar la posibilidad de representar objetos inexistentes y comportarse como si fuesen reales. Sin embargo, cabe señalar que no toda mezcla inexistente de propiedades corresponde a una caracterización sin referente: bien puedo atribuirle a un objeto existente una propiedad que de hecho no tiene, sin que por ello deje de referirme a él. A fin de cuentas hay que estudiar cada caso por separado, pero lo importante es que es posible representar cosas inexistentes, lo cual es central para algunas tesis que expondré más adelante y para la que yo mismo defenderé.

Sin embargo, sí parece haber representaciones de las que no es posible dar cuenta etiológicamente, como las que parece que tendrían Swampman o los cerebros en cubetas. Así que parece indispensable recurrir a las teorías disposicionales para completar el panorama.

2.3.3.3.2. Representación disposicional

Dar cuenta de la representación en términos disposicionales consiste en identificar comportamientos o disposiciones comportamentales que sugieran fuertemente que hay representación. O en otras palabras, consiste en determinar qué quiere decir que algo tenga una *marcada disposición* a comportarse como si tuviese el propósito de apuntar hacia algo. Un primer paso en esa dirección es fijarse en casos paradigmáticos de sistemas representacionales etiológicos y estudiar qué es lo que queda si se hace caso omiso de su historia previa, es decir, si se trabaja bajo la hipótesis de que son como Swampman. Una opción más radical es omitir también su contexto externo y concentrarse sólo en los

factores internos, es decir, trabajar bajo la hipótesis de que son como cerebros en cubetas. La ventaja de esta última opción es que así se garantiza obtener propiedades representacionales intrínsecas, porque no dependen del entorno. Por eso es que me concentraré en esa alternativa.

Sobre esa línea, si se quiere averiguar el contenido representacional de una parte z de un sistema s , una opción es fijarse en cómo es que s usa (o tiene la disposición de usar) a z . Eso es lo que sugieren autores como Field (1977), Woods (1981), Block (1986, 1998b), Harman (1987), Block & Campbell (1987), Chalmers (2002d), Greenberg & Harman (2006) y Cohen (2007), quienes sostienen que algunos significados están enteramente determinados por el rol inferencial, conceptual o funcional del vehículo representacional dentro del sistema al que pertenece, es decir, por cómo se encuentra conectado causalmente con otras partes del sistema, incluyendo otras representaciones, y qué permite hacer al sistema en virtud de ello. Por ejemplo, es razonable pensar que las conectivas lógicas son definibles exclusivamente en términos del tipo de inferencias que permiten hacer, por lo que es razonable concluir que un sistema posee esas nociones tan sólo en virtud de que tenga la disposición de hacer las inferencias relevantes (descritas en términos de generar ciertos estados a partir de ciertos otros), tras lo cual sólo hace falta averiguar qué parte del sistema realiza el rol funcional crucial para ello. Por otra parte, es posible definir algunas nociones meramente a partir de otras, como suele pasar con conceptos técnicos en ciencia o filosofía, de modo que poseerlos puede traducirse a poseer las conexiones apropiadas entre ciertas nociones relevantes. Sin embargo, un problema es que las representaciones descritas de esa manera son claramente derivadas, porque dependen de otras representaciones: hacer inferencias requiere manipular representaciones para extraer nuevas y tener las conexiones apropiadas entre cualesquiera nociones involucra tener representaciones de ellas, conectadas apropiadamente. Así que primero habría que dar cuenta de las representaciones más básicas, para asegurarse de que las que las emplean sean enteramente disposicionales.

Schellenberg (2018), por ejemplo, propone enfocarse en las capacidades de discriminación

e individuación de objetos que cada sistema tiene en cada ambiente, es decir, en sus disposiciones a tener respuestas diferentes ante distintos tipos de estímulo y estables ante un mismo tipo. Pero esa propuesta tiene dos problemas: por un lado, los mismos mecanismos pueden seleccionar objetos distintos en contextos distintos y no es claro cómo determinar qué contexto es el adecuado; por otra parte, aun fijando el contexto, los mismos mecanismos podrían ser entendidos como rastreando distintas cosas si estas se encuentran correlacionadas sistemáticamente. En síntesis, reaparecen los ya familiares problemas de indeterminación funcional del contenido. Si bien Schellenberg (2018) está convencida de que esos problemas no surgen para su teoría e incluso afirma que no es necesario recurrir a las teorías etiológicas para resolverlos, lo cierto es que no termina de quedar claro cuál es la alternativa que propone para lidiar con ellos. Aun así, la idea es rescatable si se agregan especificaciones como las que se mencionaron en la sección sobre teleología disposicional: meramente hay que exigir que las disposiciones en cuestión incluyan los suficientes mecanismos de refuerzo como para contar como *marcadas disposiciones* de comportarse como si se tuviese el propósito de apuntar hacia algo. Por ejemplo, pueden exigirse mecanismos de discriminación tan precisos, variados e interconectados que no dejen lugar para ambigüedad respecto a cuál es el contenido.

Nótese que el objetivo no es sostener que no hay representaciones etiológicas ni que alguna en particular no lo sea, sino sólo que puede haber algunas que no lo son: las que cumplan los criterios señalados. De hecho, es razonable pensar que cualquier estrategia exclusivamente disposicional tiene algunos límites, porque parece ser que algunos contenidos sencillamente no pueden ser representados sin la ayuda de propiedades extrínsecas. Para evidenciarlo, sirve estudiar el caso de la *Tierra gemela*, propuesto por Putnam (1973, 1975). La idea es imaginar un duplicado del mundo real salvo porque en él no hay agua, sino que en su lugar hay otra substancia, XYZ, que es indistinguible del agua a simple vista. Además, se aclara que es un duplicado previo al descubrimiento de maneras alternativas de distinguir al agua. Si entonces nos fijamos en una persona real y pensamos en su contraparte en la Tierra gemela, descubriremos que tendrían exactamente las mismas disposiciones pero dirigidas a distintos objetos: H₂O en un caso y XYZ en el otro.

Además, en principio ninguno de los dos contextos sería privilegiado: cualquiera de los dos sería igual de válido para determinar el contenido de los estados en cuestión. Así que es razonable conceder que dichos estados representarían al agua en un caso y a XYZ en el otro. Pero entonces el contenido representacional de esos estados dependería de cuál de las dos sustancias puebla cada mundo, lo cual es una propiedad extrínseca de ambos sujetos¹¹⁷. Por ende, no bastaría con fijarse en sus propiedades internas para saber qué sustancia es representada por sus estados. Como conclusión, parece ser que la representación de ciertos objetos siempre es una propiedad extrínseca del sujeto. En específico, ese parece ser el caso para objetos externos particulares, porque no parece posible fijar su referencia mediante mecanismos exclusivamente internos.

Sin embargo, ello es compatible con que las disposiciones en juego sean inseparables de cierto contenido representacional. Por ejemplo, puede sostenerse que las disposiciones internas determinan una caracterización que se mantiene estable de un contexto a otro, pero que puede seleccionar referentes distintos en contextos distintos, dependiendo de qué la satisfaga en cada uno. Siguiendo a autores como White (1982) y Fodor (1987), puede pensarse a las disposiciones en cuestión como una suerte de descripción o mapa, cuyas partes seleccionan distintos objetos en diferentes contextos. Sin embargo, según Fodor (1997), se trata de una descripción (o mapa) inefable, en el sentido de que no puede ser expresada en términos del lenguaje ordinario, porque para ello habría que asociarle propiedades ordinarias, lo cual a su vez requeriría presuponer un contexto de evaluación. De hecho, ni siquiera es claro que lo que hay sea digno de ser llamado ‘contenido representacional’ y no sólo meras disposiciones para seleccionar uno dado un contexto (Block (1986, 1998b). Una alternativa es no tratar de describir el (supuesto) contenido, sino meramente concentrarse en que lo que hay es una intensión epistémica, es decir, una función definible por qué selecciona en cada contexto (ver Lewis (1979, 1994), Stalnaker (1999) y Chalmers (1996, 2002d)¹¹⁸).

117 Putnam (1973, 1975) no toma en cuenta el hecho de que el sujeto en cuestión tendría H₂O dentro de su cuerpo en un caso y de XYZ en el otro, lo cual puede ser problemático. Pero la idea general puede replicarse aludiendo a alguna sustancia que no se encuentre dentro del cuerpo del sujeto, como plutonio, por ejemplo.

118 Chalmers (2002c, 2004) denomina ‘contenido fregeano’ a algo ambiguo entre una intensión epistémica

Otra posibilidad es sostener que el contenido intrínseco se reduce a combinaciones de propiedades internas de *s*, que se mantienen fijas de un contexto a otro. Siendo ese el caso, el contenido sobre objetos externos es contingente e indirecto: es aquello que en cada contexto produce los estados internos en cuestión. Para ello se puede recurrir a las mismas ideas que plantea Schellenberg (2018), pero enfocándose en discriminación e individuación de estados internos de *s* y sosteniendo que eso es lo que está en la base de la discriminación e individuación externa: desde el punto de vista de *s*, lo que distingue a los objetos externos es cómo le afectan, es decir, qué estados internos le producen. Por ejemplo, tiene sentido pensar que el sistema visual, al monitorear sus propios estados internos, monitorea lo que sea que los causa desde el exterior. Otra manera de entenderlo es como un rastreo de *propiedades disposicionales* externas de causar ciertos estados internos: en virtud de detectar una determinada propiedad *categorica* interna, *s* es capaz de detectar la disposición externa de causarla.

En síntesis, es bastante razonable entender que algunas de nuestras disposiciones delimitan contenidos representacionales, en el sentido de que conforman marcadas disposiciones de comportarse como si se apuntara hacia ellos. Además, cabe señalar que gran parte de lo que cada contenido es *para* el sujeto está determinado por cómo éste está dispuesto a reaccionar ante él, lo cual depende de cómo los estados que sirven para discriminarlo están conectados causalmente con otros estados, que conducen a acciones específicas o a relacionarlo con otras nociones, que a su vez pueden conducir a acciones específicas. Es decir, el contenido intrínseco no sólo está determinado por disposiciones a discriminar e individuar objetos, sino también por disposiciones a comportarse ante ellos de cierta manera o a relacionarlos con otras nociones, incluyendo la disposición a entenderlos como objetos de cierto tipo o como objetos a secas.

Así que hay al menos dos maneras de naturalizar la representación (la etiológica y la

y la caracterización correspondiente. Dicho contenido se opone al *contenido russelliano*, que es algo ambiguo entre la intensión subjuntiva y el referente. De hecho, Schellenberg (2011) explícitamente se compromete con que el contenido representacional de las experiencias fenoménicas es fregeano.

disposicional), en contraste con el *acquaintance*, que parece no ser naturalizable. Siendo ese el caso, lo más razonable para el fisicista es concluir que la presentación tipo E es un tipo de representación. Además, dado que la consciencia fenoménica siempre involucra presentación tipo E de qualia y nada más, es natural concluir que es verdad el representacionismo fuerte e irrestricto. Sin embargo, como se señaló en la sección 2.3.2, hay varios tipos de representacionismo fuerte e irrestricto. A continuación regresaré sobre las distinciones que los definen y diré qué combinación considero más plausible y por qué. Después de eso, sólo faltará determinar cuál es el contenido representacional relevante, lo cual se hará en los capítulos 3 y 4, y, con base en ello, detallar con mayor precisión qué tipo de representación es la presentación tipo E, lo cual se hará en el capítulo 4.

2.4. La postura (representacionista) más plausible (hasta ahora)

Recapitulando y como recién mencioné, la principal razón para defender el representacionismo fuerte e irrestricto respecto a la consciencia fenoménica es que toda experiencia fenoménica parece reducirse a presentación tipo E de qualia y la presentación tipo E, a su vez, sólo parece ser naturalizable si se la entiende como un tipo de representación. Sin embargo, resta decidir si el representacionismo a defender es reductivo o no-reductivo, puro o impuro y amplio o estrecho.

Respecto a la primera disyuntiva, me decantaré por un representacionismo reductivo, porque me interesa explicar la consciencia fenoménica en términos exclusivamente físicos, de manera que debo abandonar el lenguaje explícitamente fenoménico en alguna parte. Nótese, sin embargo, que la idea es sólo que ciertos mecanismos físicos bastan para generar consciencia fenoménica, no que dichos mecanismos sean necesarios para ello. En otras palabras, la postura que defenderé será compatible con que la consciencia fenoménica sea realizable mediante mecanismos no-físicos (en otros mundos). Con lo que no será compatible es con que sea realizable de manera no-representacional.

En cuanto a la disyuntiva entre representacionismo puro e impuro, lo cierto es que

prácticamente nadie defiende el representacionismo RQx puro para ninguna x, es decir, la idea de que hay un contenido tal que basta con representarlo para generar consciencia fenoménica. La principal razón es que en principio es fácil representar cualquier cosa mediante representaciones que claramente no involucran consciencia fenoménica, como las palabras. De hecho, para que cualquier objeto z represente a cualquier objeto w, basta con convenirlo así, pero no queremos sostener que los objetos pueden adquirir consciencia fenoménica por mera convención nuestra ni mucho menos que cualquier cosa puede tener cualquier experiencia fenoménica¹¹⁹. Por otra parte, si la consciencia fenoménica es algo mental, las representaciones involucradas también deben ser mentales, lo cual ya implica una modalidad representacional. E incluso parece hacer falta especificar más, ya que comúnmente es aceptado que podemos tener estados mentales representacionales de los que no nos percatamos, tales como creencias inconscientes e incluso percepciones. Así que, en síntesis, todo indica que el representacionismo RQx puro es inviable, independientemente de qué x se trate. Por ende, el representacionismo fuerte, irrestricto y puro también es inviable. Chalmers (2004) incluso señala que una opción obvia es simplemente decir que la modalidad indicada es una *fenoménica*, es decir, una que involucre consciencia fenoménica. Sin embargo, ello conduce al representacionismo no-reductivo. Además, la tesis resultante es poco informativa y hasta trivial, lo cual es deseable evitar. Por lo tanto, yo defenderé un representacionismo impuro que especifique la modalidad representacional de otra manera.

Por último, dado que no cuestionaré el internismo fenoménico, me veré obligado a defender un representacionismo estrecho, es decir, uno según el cual las propiedades representacionales relevantes son propiedades intrínsecas del sujeto de experiencia. Es muy importante notar que el representacionismo estrecho no niega que las experiencias fenoménicas tengan propiedades representacionales extrínsecas, sino que simplemente sostiene que no son esas las que realmente explican la consciencia fenoménica. En otras

119 Lloyd (1991) y Thau (2002) sugieren que sí hay ciertos contenidos tales que basta con representarlos para que haya consciencia fenoménica, pero no es claro por qué no podrían ser representados de otra manera. Por ejemplo, no parece que algo impida meramente asignarles un nombre. Además, sus ejemplos sí parecen involucrar medios de representación como el mental o el perceptivo.

palabras, las propiedades representacionales extrínsecas de la consciencia fenoménica le son contingentes, mientras que las intrínsecas le son esenciales. Cabe notar que las únicas propiedades representacionales que son intrínsecas son las (genuinamente) disposicionales, porque las etiológicas dependen de la historia previa del objeto que las tiene, la cual siempre le es extrínseca. Así que deberé aludir sólo a ese tipo de propiedades representacionales.

En síntesis, lo que defenderé será un *representacionismo fuerte, irrestricto, reductivo, impuro y estrecho*. *Fuerte e irrestricto* porque es razonable entender que la consciencia fenoménica siempre involucra presentación tipo E de qualia y que esta, a su vez, basta para que haya consciencia fenoménica. *Representacionismo* porque parece que la única manera naturalista de entender la presentación tipo E es la representacional, porque la única alternativa es el *acquaintance* y no es naturalizable. *Reductivo* porque busco explicar la consciencia en términos exclusivamente físicos. *Impuro* porque el puro conduce a atribuirle consciencia fenoménica a cosas que claramente no la tienen. Y *estrecho* porque el amplio es incompatible con el internismo fenoménico. Una vez dejando eso en claro, podemos pasar a explicar el elemento central del contenido representacional de la experiencia: los qualia. En ello se enfocará el siguiente capítulo.

2.5. Sumario 2

En este capítulo se revisaron las dos únicas maneras de entender la presentación tipo E: como percatación directa (*acquaintance*) o como percatación indirecta (representacional). Además, dado que sólo la segunda es naturalizable, se concluyó que la presentación tipo E debe ser algún tipo de representación. Ello conduce a una postura *representacionista fuerte*, es decir, una según la cual tener consciencia fenoménica es idéntico a tener ciertas propiedades representacionales. Más específicamente, se abogó por una teoría representacionista *fuerte, irrestricta, reductiva, impura y estrecha*. *Irrestricta* en el sentido de que es una teoría sobre toda experiencia fenoménica: dado que la consciencia fenoménica siempre involucra presentación tipo E de qualia y nunca requiere de algo más,

es razonable sostener una postura representacionista fuerte e irrestricta. *Reductiva* porque en un final se busca recurrir sólo a terminología física para describirla, para eliminar la brecha epistémica entre lo físico y lo fenoménico. *Impura* en el sentido de que las propiedades representacionales relevantes involucran una modalidad representacional en particular (la presentación tipo E), lo cual es indispensable, dado que hay modalidades representacionales que claramente no involucran consciencia fenoménica y bajo las cuales es posible representar cualquier contenido. *Estrecho* en el sentido de que las propiedades representacionales relevantes son intrínsecas del sujeto de experiencia, lo cual es indispensable para salvar el internismo fenoménico. Más específicamente, se habló de que hay dos maneras de naturalizar la representación, una *etiológica* y una *disposicional*, y se concluyó que la presentación tipo E debe ser un tipo de representación disposicional. La razón es que las representaciones etiológicas dependen de la historia previa del ente que posee el estado en cuestión, la cual siempre es una propiedad extrínseca suya. La disposicional, en cambio, sólo alude a propiedades internas suyas, lo cual la torna compatible con el representacionismo fuerte y estrecho.

Para completar el estudio de las posturas disponibles y refinar la mía, sólo hace falta determinar los posibles contenidos representacionales relevantes. Ello, a su vez, depende principalmente de las diversas teorías sobre qué son los qualia, porque los qualia son la parte central del contenido representacional de la experiencia. Por eso es que el siguiente capítulo se enfocará en revisar las distintas propuestas sobre la naturaleza íntima de los qualia.

Capítulo 3

Las posibles respuestas II: teorías sobre los qualia

3. Las posibles respuestas II: teorías sobre los qualia

En el capítulo anterior se exploraron las dos maneras de entender la presentación tipo E: como *acquaintance* y como representación. En este capítulo me enfocaré en estudiar el único otro elemento central de la consciencia fenoménica: los qualia. Más específicamente y tras unas notas preliminares (sección 3.1), revisaré tres maneras lógicamente exhaustivas de entenderlos: como *propiedades físicas externas categóricas, intrínsecas y fundamentales* (abreviado como propiedades físicas externas *ordinarias*) (sección 3.2), como *propiedades físicas internas o no-ordinarias* (sección 3.3) o como *propiedades no-físicas* (sección 3.4). Al final (sección 3.5) combinaré los resultados obtenidos con los del capítulo anterior, lo cual arroja seis posibles maneras de entender la consciencia fenoménica:

- *Acquaintance* con propiedades físicas externas ordinarias.
- *Acquaintance* con propiedades físicas internas o no-ordinarias.
- *Acquaintance* con propiedades no-físicas.
- Representación tipo E' de propiedades físicas externas ordinarias.
- Representación tipo E' de propiedades físicas internas o no-ordinarias.
- Representación tipo E' de propiedades no-físicas.

Por un lado, descartaré las tres primeras porque parecen no ser naturalizables, porque el *acquaintance* parece no serlo. Por otra parte, rechazaré la primera y la cuarta porque están en contra del internismo fenoménico. Así que sólo sobrevivirán las dos últimas posturas de la lista: representación tipo E' de propiedades físicas internas o no-ordinarias y representación tipo E' de propiedades no-físicas (o *representacionismo ilusionista*). Entonces señalaré que la primera de ellas tergiversa la naturaleza (no-física y en ocasiones externa) que la experiencia atribuye a los qualia a cambio de sostener que existen, mientras que la segunda niega su existencia (para salvar al fisicismo) a cambio de respetar la naturaleza que la experiencia atribuye a los qualia. A pesar de que consideraré que la

segunda postura sale mejor librada, ofreceré otras dos razones para enfocarse en ella: que es una postura poco trabajada en la literatura filosófica y que ofrece una respuesta original y poderosa (que expondré en el último capítulo) a los argumentos anti-fisicistas que se apoyan en la brecha epistémica entre lo físico y lo fenoménico.

Comencemos, entonces, señalando algunas cosas que servirán para darle forma a la discusión posterior.

3.1. Algunas notas preliminares

Al igual que cuando presenté las posibles naturalezas de la presentación tipo E, podemos comenzar revisando lo que en primera instancia sugiere el sentido común sobre los qualia, es decir, lo que estos parecen ser a simple vista. En otras palabras, qué propiedades parecen tener cuando son presentados de modo E. Ello puede ser un asunto complicado, porque parecen ser objetos bastante diversos, pero lo cierto es que sí hay algunas cosas que parecen tener en común, mismas que mencioné en la sección 1.1: a grandes rasgos, parecen ser *propiedades categóricas, intrínsecas y fundamentales*. Categóricas en el sentido de que meramente parecen estar “ahí” presentes, constituyendo al objeto que las posee, en contraste con ser disposiciones suyas; intrínsecas porque parecen no depender de algo externo al objeto que las posee; y fundamentales (o *primitivas, simples o irreducibles*) en el sentido de que no parecen ser explicables en términos de otra cosa. En particular, parecen no ser explicables meramente en los términos matemáticos y espacio-temporales propios de la ciencia (como se señaló en la sección 1.4.1). Además, se puede agregar que los qualia parecen ser objetos claramente *existentes*, en el sentido de que parecen estar claramente presentes ante nosotros durante nuestras experiencias, en contraste con ser meras sugerencias de nuestra mente. De ahí en fuera, hay que estudiar los tipos de qualia caso por caso para poder decir algunas cosas más. Por ejemplo, sobre los qualia de color se puede agregar que parecen ser *externos e independientes del sujeto de experiencia*, porque parecen estar *situados en objetos externos*¹²⁰; mientras que los qualia propios de emociones

120 Esto es parte de lo que sostienen los *primitivistas* respecto al color, como Hacker (1987), Campbell (1994, 2005), McGinn (1996), Watkins (2005) y Gert (2006, 2008). En el fondo, lo que hacen es

más bien parecen ser *internos* o incluso meramente *mentales*. Dependiendo de qué se sostenga que es en sí cada tipo de qualia, es posible aceptar o rechazar que tiene cada una de esas características o incluso es posible rechazar que parece tenerlas, pero en principio es una virtud rechazar lo menos posible.

Por otra parte, en principio es posible distinguir entre los qualia en sí y los objetos físicos en respuesta a los cuales los experimentamos. Por ejemplo, es posible distinguir entre un quale de color tal cual lo experimento y diversos elementos involucrados en la cadena causal que me lleva a experimentarlo, tales como mis estados retinales, la luz que los causa y las propiedades de luminiscencia o reflectancia lumínica de los objetos que emiten o reflejan esa luz. Pero, si bien es posible sostener que los qualia son independientes de los objetos que llevan a experimentarlos, también es posible identificarlos con algunos de ellos. Por ejemplo, es posible identificar los colores con tipos de luz o con propiedades de luminiscencia o reflectancia lumínica. Parte de lo que debe hacer cualquier teoría sobre la naturaleza íntima de los qualia es pronunciarse sobre esas posibles relaciones. Además, debe hacerlo de manera que se pueda conectar coherentemente con alguna teoría sobre la presentación tipo E de los qualia en cuestión.

A continuación exploraré las principales posturas sobre la naturaleza de los qualia, comenzando por la que posiblemente es más intuitiva. Cabe señalar que es posible que algunas de las teorías se traslapen, dependiendo de sobre qué tipo de qualia se hable, por lo que haré énfasis en un tipo de qualia sobre el cual todas tienen opiniones claramente distintas: los qualia de color.

3.2. Fisicismo ingenuo respecto a los qualia: propiedades físicas externas ordinarias

Como es de esperarse, la opción más natural es entender que ciertos qualia, como los de color, son propiedades categóricas, intrínsecas y fundamentales, porque justo eso parecen ser a primera vista. Pero no sólo eso, sino que también es natural entender que son

sostenerlo sobre los qualia de color y luego identificar a estos con los colores.

propiedades externas e independientes del sujeto de experiencia, porque parecen estar situados en objetos externos. Además, dado que lo más natural es entender que dichos objetos son físicos, también es natural entender que los qualia en cuestión son justamente propiedades físicas suyas. A fin de cuentas, la noción de lo físico del sentido común es más cercana a las apariencias compuestas por qualia presentadas por nuestras experiencias que a las abstractas formulaciones matemáticas de la ciencia, que parecen sólo capturar el aspecto espacio-temporal. Eso es básicamente lo que sugieren los *realistas directos* (ver sección 2.2.1) y algunos representacionistas fuertes, comúnmente denominados ‘representacionistas de rastreo’, como Dretske (1995), Tye (1995, 2000) y Crane (2003). Todos ellos se apoyan en la intuición de que la percepción verídica parece ser una “ventana transparente” hacia el mundo físico y sus propiedades físicas, difiriendo sólo respecto a qué naturaleza la adjudican a la “ventana” en cuestión (*acquaintance* o representación). Como consecuencia, identifican a los qualia con las propiedades físicas que parecen ser a simple vista y entienden que la experiencia simplemente muestra (o sugiere) que hay cierta distribución de ellas en el mundo real. Entonces sólo hace falta averiguar, con ayuda de la ciencia, exactamente de qué propiedades físicas se trata. En lo sucesivo llamaré ‘propiedades físicas ordinarias’ a las propiedades físicas que son categóricas, intrínsecas y fundamentales.

Antes de revisar las objeciones en contra de ese tipo de postura, conviene señalar que es muy obvio en qué sentido sirve para salvar al fisicismo: si los qualia no son más que propiedades físicas, entonces su existencia sencillamente no puede poner en peligro al fisicismo. Siendo ese el caso, la única parte de la consciencia fenoménica que podría ser peligrosa sería la presentación tipo E, y ello sólo si se la entiende como *acquaintance*, porque la representación parece perfectamente naturalizable. Así que para terminar de salvar al fisicismo bastaría con sostener que la presentación tipo E es representacional, que es justamente lo que hacen los representacionistas de rastreo.

3.2.1. Contra el fisicismo ingenuo respecto a los qualia

El principal problema de la postura ingenua es que parece ser que toda propiedad física externa ordinaria falla en cumplir algún requisito para ser adecuadamente identificada con los qualia de color. De hecho, el mismo problema surge para cualquier quale que se pretenda identificar con alguna propiedad física externa ordinaria. Por un lado, hay buenas razones para pensar que la asociación entre qualia y propiedades físicas externas ordinarias podría variar de persona en persona, sin que se siga que una de ellas estaría equivocada. En particular, se han planteado muchos escenarios en los que dos personas tienen distintas experiencias ante un mismo color o iguales ante distintos colores (entendidos ya sea como tipos de luz o como propiedades de luminiscencia o reflectancia lumínica)¹²¹, sin que sea posible decidir cuál de ellas está en lo correcto, si es que sólo una lo está (ver Shoemaker (1982) y Block (1990)). Por otra parte, hay buenas razones para pensar que el mundo físico real de hecho no es como nos lo presentan nuestras experiencias.

Para ejemplificar la primera ramificación del problema, podemos fijarnos en el caso de la *Tierra invertida*, presentado por Block (1990, 1998a). Él pide imaginar un planeta intrínsecamente idéntico a la Tierra salvo por un par de diferencias. Por un lado, en él los colores de todas las cosas están invertidos: lo que aquí es rojo allá es verde (y viceversa), lo que acá es amarillo allá es azul (y viceversa), etcétera. Más específicamente, las longitudes de onda de la luz emitida o reflejada por los objetos que pueblan esa *Tierra invertida* son las correspondientes a invertir las longitudes de onda del espectro lumínico aquí en la Tierra. Por otra parte, esa inversión de colores está acompañada de una correspondiente inversión lingüística: los habitantes de la Tierra invertida llaman ‘verde’ a lo que nosotros llamamos ‘rojo’ (y viceversa), ‘amarillo’ a lo que llamamos ‘azul’ (y viceversa), etcétera. En otras palabras, los habitantes de la Tierra invertida no hablan español, sino *español invertido*. Siendo ese el caso, ellos también llaman ‘verde’ al color

121 Hay debate sobre qué son en sí los colores. Aquí estoy trabajando con la idea de que son propiedades físicas externas ordinarias, pero hay otras opciones. Si el lector favorece una acepción distinta de la que manejo aquí, perfectamente puede leerme con los ajustes terminológicos correspondientes y sin que se pierda la esencia de lo que digo. Para discusión sobre el tema, ver Maund (1997/2018).

de las hojas y ‘azul’ al color del cielo, a pesar de que desde nuestra perspectiva sus hojas serían rojas y su cielo amarillo.

Ahora, una vez planteado ese escenario, Block pide imaginar que alguien de la Tierra es intervenido quirúrgicamente del siguiente modo: en sus ojos son implantados lentes que invierten la longitud de onda de la luz que llega a ellos y cada parte de su cuerpo es teñida del color inverso al original. Después de eso, la persona es trasladada a la Tierra invertida e intercambiada por su contraparte allá. Además, todo lo mencionado ocurre sin su conocimiento, porque todo es realizado mientras se encuentra inconsciente y sin dejar rastro (fácilmente) perceptible del cambio. Entonces podemos preguntarnos qué pasaría allá con esa persona, ante lo cual es razonable pensar que no notaría nada nuevo: para ella todo sería exactamente como en la Tierra y se desenvolvería exactamente igual allá que acá. Sin embargo, lo cierto es que sus experiencias estarían invertidas respecto a los colores que le rodean: ante cierto rojo tendría la experiencia correspondiente a cierto verde (y viceversa), ante cierto amarillo la correspondiente a cierto azul (y viceversa), etcétera. Es decir, asignaría erróneamente la apariencia de la Tierra a la Tierra invertida: la representaría como de hecho no es, debido a que sería engañado por sus implantes. Además, por la misma razón, llamaría ‘verde’ al rojo de las hojas y ‘azul’ al amarillo del cielo. Sin embargo, él no estaría hablando español invertido, sino empleando mal el español de la Tierra: también sus representaciones lingüísticas serían erradas.

Sin embargo, la persona en cuestión sería perfectamente funcional en la Tierra invertida y, por ende, es razonable pensar que perfectamente pudimos haber evolucionado con sus características: con algo análogo a sus lentes inversores y con la pigmentación corporal invertida. Del mismo modo, perfectamente pudimos haber desarrollado un lenguaje igual al español invertido. En otras palabras, el hecho de que consideremos que su asignación sería errada y la nuestra acertada depende de considerar que nuestra historia evolutiva (natural y cultural) es la única adecuada, lo cual es injustificado. De hecho, si la persona en cuestión (o su swampman) se quedara a vivir en la Tierra invertida, tuviese descendencia con otra persona en su misma situación y sus rasgos alterados fuesen heredables, es

razonable pensar que eventualmente los contenidos de sus estados y de su lenguaje cambiarían, acoplándose a las asociaciones que al principio eran erradas, porque tendría lugar una nueva historia de selección que los fijaría de ese modo. Así que, a fin de cuentas, lo más razonable es concluir que nuestra particular asociación entre qualia y colores es tan acertada como muchas otras, lo cual a su vez obliga a concluir que los qualia visuales no son los colores mismos, sino algo distinto que contingentemente ha sido asociado a ellos. Y un razonamiento análogo puede ser desarrollado para la relación entre cualesquiera qualia y cualesquiera propiedades físicas externas ordinarias, conduciendo a la conclusión de que ningún quale es idéntico a una propiedad física externa ordinaria.

La intuición de fondo que motiva ese razonamiento es que el representar propiedades físicas externas ordinarias siempre depende de factores independientes de los que determinan a las experiencias, por lo que es posible modificar unas sin alterar las otras. Más específicamente, representar propiedades físicas externas ordinarias parece siempre depender de relaciones contingentes entre el sujeto y las propiedades en cuestión, mientras que las experiencias parecen depender sólo de sus propiedades intrínsecas. Siendo ese el caso, es fácil idear situaciones en las que las mismas propiedades intrínsecas están emparejadas con distintas relaciones externas o distintas propiedades intrínsecas con las mismas relaciones externas, conduciendo a que las mismas experiencias estén asociadas a distintos elementos externos o distintas experiencias a uno mismo. Entonces tener consciencia fenoménica no puede ser idéntico a representar (de ninguna manera) propiedades físicas categóricas externas, porque la primera es una propiedad intrínseca y la segunda extrínseca. Así que, si la consciencia fenoménica es un tipo de representación de qualia, los qualia no pueden ser propiedades físicas externas ordinarias. Pero no sólo eso, sino que la misma idea puede ser extendida para abarcar la relación de *acquaintance*, porque el *acquaintance* con propiedades físicas externas ordinarias también es una propiedad extrínseca, porque requiere de la presencia de dichas propiedades. Por ende, los qualia no pueden ser propiedades físicas externas ordinarias, independientemente de qué se entienda que es la percatación tipo E.

Una manera alternativa de capturar la misma idea consiste en señalar que la idea de que la consciencia fenoménica es idéntica a algún tipo de percatación de propiedades físicas externas ordinarias implica la negación del internismo fenoménico, porque equivale a hacer depender la consciencia fenoménica de propiedades extrínsecas del sujeto. Así que defender el internismo fenoménico puede ser una razón para sostener que los qualia no son propiedades físicas físicas externas ordinarias. Sin embargo, como ya había mencionado, los representacionistas de rastreo y muchos realistas directos sí están dispuestos a renunciar al internismo fenoménico, porque consideran que así se salvan cosas más importantes, como las intuiciones sobre la transparencia de la experiencia e inquietudes epistemológicas como las que expuse en la sección 2.2.

Otro problema para la postura ingenua es que de hecho hay muy buenas razones para concluir que el mundo físico externo de hecho no es como nuestras experiencias sugieren que es¹²². En primer lugar, podemos identificar algunos fenómenos que se derivan del hecho de que nuestra visión no capta los objetos externos directamente, sino sólo la luz que emiten o reflejan. Por ejemplo, experimentamos los objetos visibles con las deformaciones propias de la proyección cónica (se ven más pequeños mientras más lejos estén y las paralelas parecen coincidir en el horizonte, por ejemplo), lo cual hace parecer que tienen formas y tamaños que sabemos que de hecho no tienen (Russell (1912)). Además, los qualia de color experimentados al ver un objeto cambian dependiendo de la iluminación y del punto de vista desde el cual se vea, a pesar de que las propiedades de su superficie no cambien, lo cual impide asociarle qualia a sus colores, al menos si estos son entendidos como propiedades de su superficie (Russell (1912)). De hecho, sabemos que dos objetos de color claramente distinto pueden parecer del mismo bajo las condiciones adecuadas y que distintas combinaciones de luz nos pueden generar experiencias idénticas (metamerismo en ambos casos (ver Boyton (2003))), porque a fin de cuentas estimulan del mismo modo a nuestra retina y, por ende, les asignamos los mismos qualia. Por otra parte, somos incapaces de detectar una porción considerable del espectro electromagnético, lo cual conduce a que quede sin qualia asignados. Pero no sólo eso, sino que nos parece que los

122 Los teóricos de los *sense-data*, como Russell (1912), Broad (1925), Price (1950) y Ayer (1956)), se han esmerado especialmente en hacer ver ese punto.

extremos del espectro visible se juntan y forman un círculo cromático cerrado, en el que los colores cercanos al infrarrojo se parecen a los que se acercan al ultravioleta, siendo que en realidad las longitudes de onda que les corresponden están más alejadas entre sí que respecto a cualquier otro color visible. De forma más general, podemos decir que las propiedades físicas que son candidatos naturales para ser identificadas con los qualia de color no están relacionadas entre sí por las mismas relaciones de similitud que los qualia de color que les corresponderían. Por otra parte, ahora sabemos que los objetos físicos macroscópicos son mayormente vacío y que las superficies que a simple vista nos parecen continuas en realidad distan mucho de serlo. De hecho, a pesar de que nos parece que podemos tocarlas y sentirlas directamente, lo cierto es que siempre hay un espacio entre nosotros y ellas (lo que nos mantiene separados y evita que las atravesemos es la repulsión entre nuestros campos magnéticos). Además, sabemos que nuestra manera de experimentar el tiempo y el espacio entra en conflicto con lo que dicta la relatividad general y que la mecánica cuántica sugiere que los eventos a nivel microscópico obedecen principios que desafían lo que experimentamos a nivel macroscópico. Y todo eso por mencionar sólo algunos ejemplos.

En un final, si bien todo nuestro conocimiento sobre lo físico parte en primera instancia de cómo nos lo presentan nuestras experiencias, nuestro entendimiento se ha ido alejando cada vez más de cómo estas sugieren que es el mundo, gracias a que hemos encontrado métodos alternativos y más precisos de averiguar cómo es en realidad (ver también Pautz (2018)). Gracias a ello, ahora sabemos que los qualia que la consciencia fenoménica asigna al mundo cuadran sólo muy pobremente con las propiedades físicas externas ordinarias reales. Así que hay muy buenas razones para concluir que los qualia no pueden ser propiedades físicas externas ordinarias. Sin embargo, ello no obliga a abandonar la empresa de identificar los qualia con propiedades físicas, porque también hay otros tipos de ellas, tales como propiedades físicas internas, disposicionales, extrínsecas o incluso no-fundamentales. Si bien recurrir a ellas nos alejaría de lo que el sentido común sugiere que son los qualia, tal vez es preferible pagar ese precio que renunciar al fisicismo (o al internismo fenoménico). Revisemos ahora algunas opciones en esa dirección.

3.3. Físicismo no-ingenuo respecto a los qualia: propiedades físicas internas o no-ordinarias

Ante los problemas recién expuestos, es razonable renunciar a la idea de que los qualia son propiedades físicas externas ordinarias, lo cual deja sólo dos alternativas: sostener que son otro tipo de propiedades físicas o que no son físicos. La segunda opción no es muy popular, porque parece ir en contra del físicismo, así que para muchos sólo queda una alternativa: que los qualia sean propiedades físicas distintas de las externas ordinarias. Ello, a su vez, abre dos opciones: que sean propiedades físicas internas (del sujeto o de ciertos estados suyos) o que sean propiedades físicas no-ordinarias, es decir, que sean disposicionales, extrínsecas o no-fundamentales.

A favor de que los qualia sean propiedades físicas internas, se puede señalar que es mucho más tolerable concluir eso que sostener que son disposicionales, extrínsecas o no-fundamentales, porque es mucho menos claro que parezcan ser internas. De hecho, en la práctica no hay nadie que literalmente sostenga que los qualia son propiedades físicas no-ordinarias, a pesar de que sí hay quienes sostienen que el contenido intrínseco de las experiencias es sobre propiedades físicas no-ordinarias, tales como la de *causar x qualia* (Shoemaker (1994b)) o *la disposición de causar x qualia en mí* (donde ‘mí’ refiere a quien tiene la experiencia) (Egan (2006)). Sobre esa línea, la idea es que nos percatamos de las propiedades físicas externas debido a que nos percatamos de cómo nos afectan, es decir, gracias a que detectamos los cambios físicos internos causados por nuestra interacción con ellas. Por ejemplo, para Shoemaker (1994b) los qualia son estados neuronales y lo que hacemos es detectar por medio de ellos las propiedades externas que los causan.

Cabe señalar que para determinar qué propiedad interna puede ser identificable con un determinado quale es indispensable averiguar qué porción de todo lo que hay dentro del sujeto es la mínima suficiente para generar la experiencia de ese quale, porque es indispensable que la propiedad elegida sea parte de esa porción. De lo contrario, siempre

se podrán idear escenarios hipotéticos en que la experiencia se disocie de la propiedad en cuestión, simplemente siguiendo la misma lógica empleada en escenarios como el de la Tierra invertida. Por ejemplo, dado que al parecer para generar una experiencia visual basta con lo que acontece en las áreas visuales del cerebro, será inviable identificar los qualia relevantes con estados de activación retinal, porque siempre se podrá pensar en un escenario hipotético en el que las conexiones neuronales que van de la retina a las áreas visuales del cerebro estén modificadas de tal modo que generen experiencias distintas pero no por ello erradas.

Ahora, dado que nadie identifica a los qualia con propiedades físicas no-ordinarias, no dedicaré más espacio para esa opción, pero aun así conviene estudiar las principales ventajas de postular propiedades no-ordinarias como parte del contenido. Con ello en mente, me centraré en la propuesta de Egan (2006), que es un refinamiento de la de Shoemaker (1994b) y sostiene que el contenido de las experiencias son cosas como *la disposición de causar x qualia en mí*. Una primera virtud de esa propuesta es que permite salvar la idea de que es posible que dos individuos tengan experiencias subjetivamente distintas pero a la vez correctas ante una misma propiedad categórica externa, porque la misma propiedad externa puede tener distinto efecto en cada uno y, por ende, servir de base para dos propiedades disposicionales distintas: la de causar el quale x en el sujeto s y la de causar el quale w en el sujeto z. Por ejemplo, el mismo tipo de luz puede generar una experiencia de color en un individuo y otra distinta en otro, por lo que ambos estarían en lo cierto al entenderla como una propiedad disposicional distinta. Una segunda virtud es que ello también sirve para salvar la idea de que dos individuos pueden tener experiencias subjetivamente idénticas pero a la vez correctas ante distintas propiedades externas, porque es posible que lo que una le causa a un sujeto sea lo mismo que la otra le causa al otro sujeto. Una tercera virtud es que también se permite salvar la idea de que los qualia son en algún sentido propiedades externas e independientes de la mente, porque ciertamente las disposiciones en cuestión son externas y permanecen constantes aun cuando no se les esté percibiendo.

Otro autor que entiende a los qualia como propiedades internas (aunque tal vez no necesariamente físicas), pero sugiere que el contenido representacional de la experiencia son propiedades disposicionales (de causar qualia) es Kriegel (2009a, 2009b). Lo que distingue a su postura es que entiende que los qualia en cuestión son parte del vehículo representacional y que, por ende, éste se representa a sí mismo (y por eso la postura suele ser llamada ‘representacionismo del mismo orden’). Ello es una virtud si se tiene en cuenta que hay que cuidar que los qualia sean parte de la porción del sujeto mínima suficiente para generar la experiencia. Por otra parte, sostiene que el factor auto-representacional explica por qué para el sujeto hay algo como tener la experiencia (elemento *subjetivo*), mientras que los qualia determinan cómo es tenerla (elemento *cualitativo*). En cuanto a cómo se logra la auto-representación, lo que sugiere Kriegel es que el estado que tiene el sujeto está conformado por dos partes, una de las cuales representa a la otra, representado de esa manera al estado completo. McClelland (2013, 2016) ofrece una teoría parecida, con la diferencia de que él sostiene que no se representan propiedades disposicionales externas, sino sólo categóricas internas.

Cabe señalar que, a pesar de que tanto Shoemaker (1994b) como Egan (2006) y Kriegel (2009a, 2009b) entienden la presentación tipo E en términos representacionales, en principio se puede formular una teoría análoga a la suya pero con *acquaintance*. Por ejemplo, alguien podría sostener que nos percatamos (representacionalmente) de las propiedades disposicionales externas por medio de tener *acquaintance* con sus efectos sobre nosotros¹²³. De hecho eso es básicamente lo que sostiene Chalmers (2006c), porque para él durante las experiencias fenoménicas tenemos *acquaintance* con qualia, que a su vez son modos de presentación de propiedades externas, porque son causados por ellas (aunque hay que notar que para Chalmers los qualia no son propiedades físicas internas, sino algo no-físico).

123 Johnston (2006) sostendría que en todos los casos se trata de *acquaintance*, porque él rechaza la idea de que exista algo tal como la percatación indirecta: para él percatarse a través de un mediador también es percatarse directamente, porque para él la propiedad de haber generado al mediador es tan propia del objeto percibido como las propiedades categóricas en virtud de las cuales fue capaz de generarlo.

Por último, hay que notar que identificar los qualia con propiedades físicas internas (o no-ordinarias) sirve para salvar al fisicismo por la misma razón que sirve identificarlas con propiedades físicas externas ordinarias: si los qualia son físicos, entonces su existencia no es problemática para el fisicismo. Y ello puede ser acompañado con la idea de que la presentación tipo E es naturalizable si se la entiende como un tipo de representación, lo cual permite terminar de erradicar el problema presentado por la consciencia fenoménica en general.

3.3.1. Contra el fisicismo no-ingenuo respecto a los qualia

A pesar de que puede resultar atractiva la idea de identificar los qualia con propiedades físicas internas, lo cierto es que esa estrategia enfrenta varios problemas. El primero de ellos es una extensión de un problema que también aqueja a la propuesta de identificarlos con propiedades físicas externas ordinarias: que hay buenas razones para pensar que el mundo físico *en general* no es como nos lo presentan nuestras experiencias. Es decir, no sólo es el caso que hemos descubierto que el mundo externo es distinto de como sugieren nuestras experiencias, sino que también hemos descubierto que el interno tampoco lo es. Incluso es mucho más fácil sostener esa idea respecto al mundo físico interno, porque en su caso hasta es difícil de acomodar la idea de que los qualia sean propiedades internas, dado que a simple vista parecen no serlo. Y es mucho más difícil acomodar la idea de que sean estados neuronales, por ejemplo. Es decir, la idea de que son idénticos a propiedades físicas internas parece más una medida desesperada que otra cosa.

Sin embargo, ello puede ser enfrentado simplemente negando la fuerza de las intuiciones en cuestión, sosteniendo que son erradas o aludiendo a otras contrarias, lo cual conduce a un mero choque infructuoso de afirmaciones. Debido a eso, es deseable decir algo más. Una opción es, por ejemplo, recordar que la ciencia describe a lo físico en términos meramente matemáticos y espacio-temporales (ver sección 1.4.1), lo cual parece dejar fuera importantes aspectos sobre los qualia. Siendo así, se sigue inmediatamente que el mundo físico no es como nos lo presentan nuestras experiencias, independientemente de

qué tipo de propiedades físicas se esté hablando. Es decir, parece poco útil reemplazar las propiedades físicas externas ordinarias por propiedades físicas internas, porque la ciencia no las trata de manera distinta. En otras palabras, la aparente fundamentalidad de los qualia se conserva y bloquea cualquier intento de identificarlos con algo físico.

Ante esa situación podría volverse más tentador renunciar a la fundamentalidad de los qualia o sostener que son propiedades físicas no-ordinarias en un sentido muy fuerte: tal vez los qualia son algo que necesariamente aparece cuando se dan ciertas condiciones físicas, por más que parezca no ser físico. Sin embargo, ambas opciones son problemáticas: la primera más bien parece un cambio de tema y la segunda conserva la brecha epistémica entre los qualia y lo físico, lo cual nos regresa a los argumentos anti-fisicistas que se apoyan en ella. En un final, estas opciones parecen aun más desesperadas que la de entender a los qualia como propiedades físicas internas, porque ambas parecen consistir en simplemente hacer caso omiso de un problema claramente palpable.¹²⁴

Cabe concluir esta sección haciendo temporalmente a un lado las inclinaciones representacionistas y reconsiderando brevemente la posibilidad de que la presentación tipo E sea *acquaintance*. Si así lo fuese, se podría señalar que nos percatamos de los qualia de forma transparente, directa e infalible, pero no parece posible tener ese tipo de acceso hacia propiedades físicas. De hecho, si tomamos en cuenta que el mundo físico parece no ser como nos lo presentan nuestras experiencias, es natural entender que sólo podemos llegar a conocerlo inferencialmente, a partir de estudiar muchas experiencias distintas, descartando unas cosas y quedándonos con otras (ver Russell (1912) y Ayer (1967)). Otra manera de plantear la idea es diciendo que los objetos físicos no siempre son como parecen ser, en contraste con los qualia, que siempre lo son. Entonces parece ser que nuestro conocimiento sobre los qualia es directo o no-inferencial, mientras que las propiedades físicas sólo pueden ser conocidas inferencialmente, a partir de los qualia, que funcionan como mediadores imperfectos entre nosotros y el mundo físico (Ayer (1967)). De hecho, nuestro

124 En general, los fisicistas tipo B (ver nota 60) son los que optan por la opción que involucra sostener el fisicismo a pesar de reconocer la brecha epistémica. Para exploraciones en esa dirección, ver Reynoso (2017).

conocimiento sobre nuestros sistemas sensoriales sugiere que nuestra manera de percatarnos del mundo físico siempre requiere de cadenas causales, que a su vez requieren de estados intermediarios, de los que no siempre nos percatamos y siempre distancian al estado final de su objetivo y permiten errores de diversa índole, lo cual refuerza la idea de que nuestra percatación de lo físico siempre es opaca, indirecta y falible.

Por otra parte, también es popular la idea de que los qualia son dependientes de la mente (ver Hatfield (2021)), en el sentido de que no pueden existir fuera de las experiencias fenoménicas en las que se manifiestan, es decir, sin presentarse de modo E ante alguien (aunque esa idea tiene algunos detractores, como Russell (1914)). En lugar de argumentar a favor o en contra de ella, simplemente señalaré que es una posible distinción entre los objetos físicos y los qualia, la cual podría impedir identificarlos entre sí. Chalmers (1996, 2002a, 2003, 2005, 2006c, 2018), por ejemplo, estaría de acuerdo con ella, porque considera que la posesión de qualia es condición suficiente para que haya *acquaintance* con ellos.

Sintetizando los últimos párrafos, es razonable concluir que los qualia cumplen con tres condiciones esenciales (aunque la primera sólo convencerá a los que entienden la presentación tipo E como *acquaintance*): 1) nos percatamos de ellos directa, transparente e infaliblemente, 2) dependen de la mente y 3) tienen las propiedades que los objetos físicos *parecen* tener (porque son la apariencia misma). En contraste, los objetos físicos externos (y sus propiedades) tienen las siguientes características: 1') nos percatamos de ellos indirectamente (por medio de estados u objetos intermediarios, como los qualia), 2') son independientes de la mente y 3') pueden no tener las propiedades que aparentan, porque su apariencia, compuesta por qualia, es sólo un intermediario contingente entre nosotros y cómo son en realidad. Así que ahí hay al menos tres sentidos en que los qualia parecen no ser propiedades físicas, del tipo que sea.

3.4. Anti-fisicismo respecto a los qualia

Ante tantas razones para pensar que los qualia no son físicos, tal vez lo más razonable es concluir que de hecho no lo son, como hacen los teóricos de los *sense-data* y los dualistas en general. Incluso algunos fisicistas lo aconsejan, siempre y cuando además se niegue la existencia de los qualia (ver Frankish (2016), por ejemplo). Una variante menos radical consiste en sostener que la experiencia presenta a los qualia como de hecho no son: si bien nos hace creer que no son físicos, en realidad sí lo son (Frankish (2016) también contempla esa opción). Pero, dado que aquí estoy entendiendo a los qualia como aquello presentado de primera mano por la experiencia, tal cual nos lo presenta esta, sería más adecuado expresar esa última opción de otro modo: la experiencia nos presenta qualia en lugar de alguna otra cosa o, más bien dicho, la experiencia nos presenta alguna otra cosa como si se tratase qualia. En cualquier caso, la idea es que la experiencia nos presenta una apariencia que no es física, sea o no una representación de algo más.

Una manera de hacer compatible esa postura con el fisicismo (sin negar la existencia de los qualia) es sostener que los qualia necesariamente aparecen cuando se cumplen ciertas condiciones físicas. Sin embargo, ya he dicho por qué esa opción es problemática: porque no disuelve la brecha epistémica entre los qualia y lo físico y, por ende, dificulta responder a los argumentos que se apoyan en ella.

3.4.1. Contra el anti-fisicismo respecto a los qualia

El principal problema de negar que los qualia son físicos es que la existencia de algo no-físico basta para refutar al fisicismo, por lo que todo fisicista que entienda a los qualia como algo no-físico debe, además, sostener que no existen. Pero ello a su vez es problemático, porque para muchos es inadmisibles negar que los qualia existen. Ello se debe principalmente a que nuestras experiencias nos los presentan tan vívidamente que rechazar su presencia durante ellas parece absurdo y, por ende, negar su existencia también lo parece. En otras palabras, la idea de fondo es que la presentación tipo E es claramente

infalible y, por ende, la mera existencia de una experiencia fenoménica garantiza la existencia de los qualia presentados (de modo E) por ella. Ello no necesariamente quiere decir que se entienda a la presentación como *acquaintance*, porque también es posible sostener que se trata de algún tipo de representación infalible.

Puede decirse, entonces, que una razón para rechazar la idea de que los qualia no son físicos es que sostener que sí lo son obliga a rechazar el fisicismo o a sostener que la presentación tipo E es falible, como ilustra el siguiente argumento (Arg6):

- P1_{A6}) La consciencia fenoménica (existe y) es presentación tipo E de qualia.
- P2_{A6}) (Por P1_{A6}) Si los qualia no son físicos y la presentación tipo E es infalible, existe algo que no es físico.
- P3_{A6}) Si existe algo que no es físico, el fisicismo es falso.
- P4_{A6}) (Por P2_{A6} y P3_{A6}) Si los qualia no son físicos y la presentación tipo E es infalible, el fisicismo es falso.
- ∴ C1_{A6}) Si los qualia no son físicos, la presentación tipo E es falible o el fisicismo es falso.

El argumento es sólido y el problema radica en que cualquiera de los dos disyuntos del consecuente de C1_{A6} parece inaceptable. Sin embargo, lo cierto es que la idea de que la presentación tipo E sea falible no debería tomar a nadie por sorpresa: a fin de cuentas, justo se propuso entenderla como representación porque se tenían razones independientes para considerar que es falible (a saber, las alucinaciones y sueños). De hecho, una de las supuestas características centrales de la representación es que es falible, por lo que la idea de que haya representación infalible debería resultar bastante sospechosa. Así que, en principio, todo representacionista fuerte debería estar al menos abierto a la idea de que los qualia no existan. En otras palabras, debería estar abierto a la idea de que la presentación tipo E no sea más que una ilusión muy poderosa, que nos hace creer que hay qualia cuando en realidad no los hay. De hecho, es posible que la renuencia a aceptar que los qualia bien podrían no existir en el fondo es resultado de no haberse librado por completo de la idea de

que la presentación tipo E es *acquaintance*.

Sin embargo, hay otros dos problemas con negar la existencia de los qualia. El primero es que no es claro cómo es posible representar objetos inexistentes, y más con semejante poder de convencimiento sobre su presencia. El segundo es que conduce a concluir que la consciencia fenoménica es esencialmente ilusoria (por no decir alucinatoria): somos víctimas de un engaño perpetuo. Ello es un problema no sólo por lo contraintuitivo y radical que pueda parecer, sino porque entonces no es clara la utilidad adaptativa de tener consciencia fenoménica. De hecho, a primera vista podría ser tentador pensar que estar sujetos a semejante error sistemático no sólo no es adaptativo, sino que hasta es detrimental. Siendo ese el caso, resulta difícil explicar por qué la naturaleza nos dotó de consciencia fenoménica e incluso cómo es que logramos sobrevivir con ella. En el capítulo 4 responderé detenidamente a ambos problemas, pero esbozaré de una vez ambas respuestas. Respecto a cómo representar objetos inexistentes con tal poder de convencimiento, la idea es que equivale a tener una *marcada disposición* (ver secciones 2.3.3.2.2 y 2.3.3.3.2) de comportarse como si de hecho existiese o incluso estuviese incuestionablemente presente algo como los qualia. Por otra parte, respecto al problema de la utilidad adaptativa, la idea es que para sobrevivir no es indispensable saber a la perfección cómo es el mundo, sino que basta con tener una idea relativamente aproximada de cómo son ciertas porciones relevantes de él, lo cual deja espacio para rellenar el resto libremente. A fin de cuentas, algunas ilusiones pueden ser útiles a pesar de que se alejen de la realidad. Pero más aún: algunas pueden ser útiles justamente porque la manera en que se alejan de la realidad, porque a veces es útil creer cosas falsas.

Sea como sea, considero que las razones expuestas previamente son suficientes para tomarse en serio la idea de que los qualia no son físicos y no existen. Por otra parte, hay que recordar que rechazar la existencia de los qualia inevitablemente conduce a algún tipo de representacionismo fuerte (a menos que se niegue la existencia de la consciencia fenoménica, lo cual es aun más implausible), porque automáticamente queda descartada la posibilidad de que la presentación tipo E sea *acquaintance*, dado que esta última es

infalible por definición (ver Arg5, en la sección 2.2.3). Entonces no queda otra opción que sostener que la presentación tipo E es algún tipo de representación. La postura resultante ha sido defendida de manera muy escueta por Jackson (2003) y es del tipo que yo mismo sostendré y detallaré en el siguiente capítulo. La llamaré ‘representacionismo ilusionista’, bajo el entendido de que las teorías ilusionistas son todas aquellas que niegan la existencia de los qualia o al menos sostienen que no son como parecen ser: sólo hay una ilusión que hace parecer lo contrario (ver Frankish (2016) y siguiente capítulo). En contraste, llamaré ‘representacionismos realistas’ al resto de los representacionismos.

Con esto concluyo mi estudio sobre las posibles naturalezas de los qualia, así que sólo resta juntar los resultados con lo que ya se dijo sobre presentación tipo E y extraer conclusiones, en lo cual se enfocará la siguiente sección.

3.5. La taxonomía resultante y el veredicto final

Dado que la consciencia fenoménica es presentación tipo E de qualia, las diversas posturas sobre su naturaleza pueden clasificarse según qué entiendan que son la presentación tipo E y los qualia. Por un lado, podemos identificar dos maneras generales de entender la presentación tipo E: como percatación directa (*acquaintance*) y como percatación indirecta (representación tipo E’). Por otra parte, podemos fijarnos en la gama de opciones respecto a los qualia, que, si bien es bastante amplia y complicada, se puede capturar en las siguientes opciones exhaustivas y mutuamente excluyentes: propiedades físicas externas ordinarias, propiedades físicas internas o no-ordinarias y propiedades no-físicas. Así que tenemos dos opciones para la presentación tipo E y tres para los qualia, las cuales dan pie a seis combinaciones posibles:

- ***Acquaintance* con propiedades físicas externas ordinarias**, como sostienen los *realistas directos* respecto cualquier percepción verídica de algo externo. La principal virtud de ese tipo de postura es que es la más apegada al sentido común. Sus tres grandes desventajas son que parece no ser naturalizable (porque el

acquaintance parece no serlo), que ninguna propiedad física externa ordinaria parece ser apta para ser identificada con los qualia y que está en contra del internismo fenoménico. Además, tiene problemas para acomodar las experiencias no-verídicas.

- ***Acquaintance* con propiedades físicas internas o no-ordinarias.** Hasta donde sé, no hay una postura claramente identificable que sostenga eso. Tal vez sea porque se aleja demasiado del sentido común: sencillamente entiende a los qualia como algo que claramente no parecen ser. Además de ese problema, tendría el de no ser naturalizable (porque el *acquaintance* parece no serlo) y el de que la ciencia sugiere que el mundo físico interno no incluye qualia. Una ventaja sería que podría ser compatible con el internismo fenoménico.
- ***Acquaintance* con propiedades no-físicas,** como sostienen los *teóricos de los sense-data*. Sus principales virtudes son que respeta la intuición de que la presentación tipo E es *acquaintance*, que trata del mismo modo a las experiencias verídicas que a su contraparte no-verídica y que es compatible con el internismo fenoménico. Su principal desventaja es que no es naturalizable por dos razones: porque el *acquaintance* parece no serlo y porque se compromete con la existencia de objetos no-físicos. Además, tiene el defecto de que no respeta la intuición de que ciertos qualia son externos, como los de color.
- **Representación tipo E' de propiedades físicas externas ordinarias,** como sostienen los *representacionistas de rastreo* respecto a experiencias como las visuales. Sus principales ventajas son que es naturalizable, que trata del mismo modo a las experiencias verídicas que a su contraparte no-verídica y que respeta la intuición de que los qualia son propiedades físicas externas ordinarias. Sus principales defectos son que no respeta la tesis del internismo fenoménico y que ninguna propiedad física externa ordinaria parece ser apta para ser identificada con los qualia. Ello en parte porque parece que la relación entre experiencias y contenidos sobre propiedades físicas externas ordinarias es contingente. Además no respeta la intuición de que la presentación tipo E es *acquaintance*.
- **Representación tipo E' de propiedades físicas internas o no-ordinarias,** como

sostiene Shoemaker (1994b). Sus principales virtudes son que es naturalizable, que respeta la tesis del internismo fenoménico, que trata del mismo modo a las experiencias verídicas que a su contraparte no-verídica y que el contenido que propone no es fácilmente dissociable de las experiencias en cuestión. Sus principales defectos son que ciertos qualia (como los de color) no parecen ser propiedades internas y que la ciencia sugiere que el mundo físico interno no incluye qualia. Además no respeta la intuición de que la presentación tipo E es *acquaintance*.

- **Representación tipo E' de propiedades no-físicas**, como sostiene el representacionismo ilusionista. Sus principales virtudes son que es naturalizable, que es compatible con el internismo fenoménico, que trata del mismo modo a las experiencias verídicas que a su contraparte no-verídica y que respeta todas las razones en contra de que los qualia son físicos. Además respeta la intuición de que algunos son externos (como los de color). Sus principales desventajas son que es la única que está en contra de la intuición de que los qualia existen y que se compromete con que somos víctimas de un error sistemático.

Nótese que es posible sostener que distintas experiencias fenoménicas particulares o distintos tipos de ellas tienen distinta explicación. Por ejemplo, es posible sostener que percepción verídica de objetos externos es *acquaintance* con propiedades físicas externas ordinarias, mientras que las emociones son *acquaintance* con propiedades físicas internas. Además, hay que señalar que hay muchas posturas que no es del todo claro dónde hay que acomodar (aunque es claro que no quedan fuera de la clasificación).

Por ejemplo, si bien Egan (2006), Kriegel (2009a, 2009b) y McClelland (2013, 2016) son representacionistas fuertes y entienden a los qualia como propiedades categóricas internas, no son claros respecto a si son o no físicos, porque sus principales intereses están en resolver otros problemas sobre el contenido. Entonces sólo podemos asegurar que son representacionistas de alguno de los dos últimos tipos.

Por otro lado, los *representacionistas de orden superior*, como Rosenthal (1986, 2002),

Lycan (1987, 1996), Carruthers (2000) y Gennaro (2012), defienden teorías según las cuales tener una experiencia fenoménica consiste en tener un estado (de orden superior) que representa que se tiene otro estado (de orden inferior), que a su vez tiene otro contenido representacional sobre el mundo externo. Por ejemplo, tener una experiencia visual consiste en representar (mediante un estado de orden superior) que se tiene cierto estado (de orden inferior), que a su vez representa que el mundo externo es de cierta manera. Siendo ese el caso, alucinar consiste en tener el estado de orden superior sin que sea el caso lo que el de orden inferior sugiere sobre el mundo externo¹²⁵. Cabe señalar que Rosenthal y Gennaro sostienen que los estados de orden superior relevantes son pensamientos, mientras que Lycan y Carruthers los entienden como percepciones, pero no me adentraré en esa discusión. Por otra parte, la idea es que los qualia son propiedades del estado de orden inferior, pero no es claro si hay que entenderlas como físicas o no (ver Rosenthal (1976, 2010)¹²⁶). Así que entenderé que hay dos posibles versiones de ese tipo de teorías: las que entienden a los qualia como propiedades físicas internas y las que las entienden como no-físicas. Así que puede tener una versión dentro de los dos últimos tipos de representacionismo.

Sebastián (2012, 2017), por su parte, sostiene que lo que se representa en la experiencia son relaciones en perspectiva entre el sujeto y los qualia. Según entiendo, para él representar qualia de esa manera por sí mismo sugiere un punto de vista y, por lo tanto, la existencia y localización del sujeto. Ello es compatible con diversas interpretaciones sobre qué son en sí los qualia y no considero pertinente asociar la teoría con una en particular. En lugar de eso, me conformaré con señalar que la idea de que los qualia tengan puntos de vista integrados es compatible (en principio) con que su tipo de propiedad sea cualquiera de los que he mencionado, así que puede tener una versión dentro de cada tipo de representacionismo.

125 Aunque también hay otros dos sentidos en que se podría decir que hay alucinación: cuando se tiene el estado de orden superior sin el de orden inferior o cuando se tiene el de orden inferior sin que sea el caso lo que sugiere sobre el mundo externo.

126 Rosenthal (1976) señala que no es claro qué criterio hay que seguir para clasificar una propiedad como física, mental o neutral.

También es importante mencionar que algunos autores, como Byrne (2001), son muy cautelosos y no se comprometen con que los qualia sean algo en específico, sino que se limitan a decir que son contenido representacional de las experiencias fenoménicas, sea lo que sea que sean además de eso. Ello sólo permite distinguir que su postura es algún tipo de representacionismo, pero no cuál. Por otra parte, hay quienes más bien señalan lo que no son los qualia. Por ejemplo, Thau (2002) es un representacionista que se limita a argumentar que los qualia no son propiedades físicas externas ordinarias, lo cual sólo permite descartar uno de los tres tipos de representacionismo.

Ahora, en lugar de asignar pesos arbitrarios a cada ventaja y desventaja, me centraré en las dos principales condiciones que establecí desde el inicio: respetar el fisicismo y el internismo fenoménico. Las tres posturas que se apoyan en *acquaintance* son descartadas por la primera condición (porque parecen no ser naturalizables) y las dos posturas que identifican a los qualia con propiedades físicas externas ordinarias son descartadas por la segunda. Ello nos deja sólo con dos opciones: representación tipo E' de propiedades físicas internas o no-ordinarias o representación tipo E' de propiedades no-físicas. La única ventaja de la primera sobre la segunda es que respeta la intuición de que los qualia existen, mientras que la segunda tiene las ventajas de que se apega más fielmente a la intuición de que los qualia no son explicables en términos físicos y a la de que ciertos qualia (como los de color) son claramente externos. En otras palabras, la primera es realista a cambio de tergiversar la naturaleza que la experiencia atribuye a los qualia, mientras que la segunda sacrifica el realismo a cambio de no tergiversarla. En lo personal considero que la segunda postura queda mejor parada y por eso me inclino por ella, pero principalmente he decidido enfocarme en ella porque ha sido menos trabajada en la literatura filosófica. Además, como se verá en el siguiente capítulo, permite ofrecer una respuesta original y muy poderosa ante los argumentos anti-fisicistas que se apoyan en la brecha epistémica entre lo físico y lo fenoménico. Dedicaré, entonces, el último capítulo a detallar esa postura.

3.6. Sumario 3

En este capítulo se expusieron tres distintas maneras lógicamente exhaustivas de entender la naturaleza de los qualia: como *propiedades físicas externas categóricas, intrínsecas y fundamentales* (abreviado como propiedades físicas externas *ordinarias*), como *propiedades físicas internas o no-ordinarias* (haciendo énfasis en las internas) o como *propiedades no-físicas*. Al final se combinaron los resultados obtenidos con los del capítulo anterior (sobre las dos posibles naturalezas de la presentación tipo E: *acquaintance* y representación tipo E'), lo cual dio como resultado seis posibles maneras de entender la consciencia fenoménica:

- *Acquaintance* con propiedades físicas externas ordinarias.
- *Acquaintance* con propiedades físicas internas o no-ordinarias.
- *Acquaintance* con propiedades no-físicas.
- Representación tipo E' de propiedades físicas externas ordinarias.
- Representación tipo E' de propiedades físicas internas o no-ordinarias.
- Representación tipo E' de propiedades no-físicas.

Después se señaló que las tres primeras parecen no ser naturalizables, porque el *acquaintance* parece no serlo, y que la primera y la cuarta están en contra del internismo fenoménico. Como consecuencia, sobrevivieron sólo las dos últimas posturas de la lista: representación tipo E' de propiedades físicas internas o no-ordinarias y representación tipo E' de propiedades no-físicas (o *representacionismo ilusionista*). Entonces se señaló que la primera de ellas tergiversa la naturaleza (no-física y en ocasiones externa) que la experiencia atribuye a los qualia a cambio de sostener que existen, mientras que la segunda niega su existencia (para salvar al fisicismo) a cambio de respetar la naturaleza que la experiencia atribuye a los qualia. A pesar de que se consideró que la segunda postura quedaba mejor posicionada, se ofrecieron razones extra para enfocarse en ella: que es una postura poco trabajada en la literatura filosófica y que ofrece una respuesta original y

poderosa (a desarrollar en el siguiente capítulo) a los argumentos anti-fisicistas que se apoyan en la brecha epistémica entre lo físico y lo fenoménico. Por todas esas razones, dedicaré el siguiente y último capítulo a detallar mi propia postura representacionista ilusionista.

Capítulo 4

**La solución: representacionismo
ilusionista**

4. La solución: representacionismo ilusionista

En este capítulo junto las piezas que fui presentando y defendiendo en los capítulos anteriores, para darle forma a mi propia postura, que es un *representacionismo ilusionista*. Más específicamente, se trata de un *representacionismo fuerte, irrestricto, reductivo, impuro y estrecho* que también es un *ilusionismo fuerte*, aunque estas últimas piezas las proveeré más adelante.

Es un *representacionismo fuerte e irrestricto* porque identifica el tener consciencia fenoménica (de cualquier tipo) con tener ciertas propiedades representacionales. Ello porque entiende a la consciencia fenoménica como presentación tipo E de qualia y a la presentación tipo E como un tipo E' de representación. Es *reductivo* porque pretendo explicar todo en términos exclusivamente físicos, para que de esa manera sea genuinamente una explicación de la consciencia fenoménica dentro del marco fisicista. Es *impuro* porque para la representación tipo E' importa la *modalidad representacional* y no sólo el contenido. Y es *estrecho* porque está comprometido con el *internismo fenoménico*, lo cual obliga a sostener que las propiedades representacionales relevantes son intrínsecas del sujeto de experiencia (más específicamente, son disposicionales). Por otra parte, es un *ilusionismo fuerte* porque niega la existencia de los qualia, debido a que los entiende como propiedades no-físicas pero está comprometido con el *fisicismo*.

En la sección 4.1 presento la propuesta general, conformada por la tesis de que la consciencia fenoménica es representación tipo E' de qualia, aunada a la idea de que los qualia son propiedades no-físicas y, por ende, no existen. Ello conduce a concluir que la consciencia fenoménica sugiere que el mundo es como de hecho no es y, por lo tanto, se trata de una suerte de ilusión, lo cual conecta mi postura con el ilusionismo. Por eso dedico la sección 4.2 a explicar qué es el ilusionismo y cómo se relaciona con mi propuesta y con el representacionismo en general.

Después dedico dos secciones a detallar un par de aspectos cruciales para mi propuesta, en

respuesta a dos problemas que enfrenta. El primero de estos es el de cómo representar objetos inexistentes, además de hacerlo con tanto poder de convencimiento sobre su presencia. En respuesta a ese problema, dedico la sección 4.3 a esbozar cómo podría ser naturalizado el tipo de ilusión al que aludo. A grandes rasgos, la idea es explicarla en términos de tener una *marcada disposición* (ver secciones 2.3.3.2.2 y 2.3.3.3.2) de comportarse como si los qualia relevantes estuviesen incuestionablemente presentes, lo cual no resulta tan misterioso. El segundo problema es el de explicar por qué la naturaleza nos dotó de consciencia fenoménica, a pesar de que esta consiste en un engaño perpetuo y masivo. Por eso dedico la sección 4.4 a reflexionar cuál podría ser su utilidad adaptativa. En pocas palabras, lo que propongo es que para sobrevivir no es indispensable saber con exactitud cómo es el mundo, sino que basta con tener un conocimiento aproximado, lo cual permite rellenar el resto de manera imprecisa. Pero no sólo eso, sino que hasta puede ser útil creer que el mundo es como de hecho no es.

Por último, en la sección 4.4 regreso sobre los argumentos anti-fisicistas y muestro cómo responder a ellos desde mi postura. La idea general es que entender la consciencia fenoménica del modo que propongo permite cerrar la brecha explicativa entre lo físico y la presentación tipo E de qualia, de modo que $P^{@*} \& \neg Q^@$ se torna idealmente inconcebible. Más específicamente, es idealmente inconcebible un duplicado físico mínimo de @ sin las disposiciones a las que aludo en mi caracterización de la representación tipo E', pero con ellas basta para obtener $Q^@$, así que $P^{@*} \& \neg Q^@$ se torna contradictorio y, por ende, idealmente inconcebible. Por ende, fallan los argumentos que se apoyan en la concebibilidad ideal (y posibilidad metafísica) de $P^{@*} \& \neg Q^@$. Lo único que sobrevive es una brecha epistémica entre lo físico y los qualia, por lo que sí resulta ser concebible idealmente un duplicado físico mínimo de @ sin los qualia que las experiencias fenoménicas en @ sugieren que hay. Pero eso no es problemático para el fisicismo, porque ni siquiera en @ mismo están presentes esos qualia.

Comencemos, entonces, exponiendo la postura general.

4.1. La tesis principal y las razones que la respaldan

La tesis que detallaré en este capítulo es el resultado de juntar las piezas que fui presentando en capítulos anteriores. Del capítulo 1 retomo la idea de entender las consciencia fenoménica como *presentación tipo E de qualia*, así como los compromisos con el fisicismo y el internismo fenoménico. Del capítulo 2 recupero la conclusión de que la presentación tipo E es un tipo de representación. Y del capítulo 3, tomo la conclusión de que es razonable entender a los qualia como propiedades no-físicas. Entonces podemos decir que la tesis que defiendo es una elaboración, dentro del marco fisicista e internista fenoménico, de la idea de que la presentación tipo E es algún tipo (E') de representación y los qualia son propiedades no-físicas: es decir, de la idea de que la consciencia fenoménica es idéntica a representación tipo E' de ciertas propiedades no-físicas. Como es evidente, la postura en cuestión es un tipo de *representacionismo fuerte e irrestricto*, porque se pronuncia sobre toda experiencia fenoménica e identifica tener consciencia fenoménica con tener ciertas propiedades representacionales. Además, al estar comprometido con el internismo fenoménico, se trata de un representacionismo *estrecho*, es decir, uno según el cual tener consciencia fenoménica es una propiedad intrínseca del ente que la tiene, por lo que las propiedades representacionales relevantes también deben ser intrínsecas. Por otra parte, dado que la representación tipo E' es un tipo particular de representación, también se trata de un representacionismo *impuro*. Por último, procuraré explicarlo todo en términos exclusivamente físicos, para genuinamente proveer una explicación de la consciencia fenoménica en términos compatibles con el fisicismo, de modo que el resultado será un representacionismo *reductivo*. En síntesis, la postura que defenderé es un *representacionismo fuerte, irrestricto, reductivo, impuro y estrecho* (ver también secciones 2.3.2 y 2.4).

En breve, una fuerte razón para sostener que la presentación tipo E es representacional es que la única alternativa es que sea *acquaintance*, pero el *acquaintance* no es naturalizable (ver sección 2.2.3). Por otra parte, sostengo que los qualia no son físicos porque ninguna propiedad física cumple con lo necesario para ser un quale (ver secciones 3.2.1, 3.3.1 y

3.4), ya sea porque los candidatos relevantes fallan en aspectos específicos o porque, en general, no es posible capturar cómo son los qualia en los términos exclusivamente matemáticos y espacio-temporales propios de las descripciones científicas de lo físico. Por otra parte, como señalé en la sección 3.4, un requisito obvio de toda postura fisicista que entienda a los qualia como propiedades no-físicas es negar la existencia de los qualia. Y de ahí se desprende otra razón para sostener que la presentación tipo E no es *acquaintance*, porque por definición es imposible tener *acquaintance* con algo inexistente (ver Arg5, en la sección 2.2.3).

Además, se sigue que la consciencia fenoménica nos sugiere que el mundo es como de hecho no es, porque nos sugiere erróneamente que en él hay qualia. Eso convierte a mi postura en un tipo de *ilusionismo*, así que conviene dedicar una sección a detallar en qué consiste eso y cómo se relaciona con el representacionismo en general.

4.2. Ilusionismo

Las posturas ilusionistas son todas aquellas que sostienen que los qualia no existen o no son como parecen ser, sino que meramente hay una ilusión que hace parecer lo contrario (ver Frankish (2016)). La principal razón para sostener ese tipo de postura es que el fisicismo parece ser verdad y, sin embargo, hay muy buenas razones para pensar que los qualia parecen no ser físicos. Ante esa situación, antes de decantarnos por teorías que incluyan la existencia de algo tan problemático para el fisicismo, deberíamos tomar en serio la idea de que los qualia, tal como parecen ser, simplemente no existan (ver también Dennett (2016)). Ello se traduce en que lo más razonable es hacer una de dos cosas: conceder que los qualia no son físicos, pero negar su existencia; o bien, conceder que parecen no ser físicos, pero insistir en que de hecho sí lo son. En cualquiera de los dos casos, hay que complementar la afirmación explicando por qué parece ser el caso lo contrario, ante lo cual lo más natural es meramente sostener que hay una ilusión en juego, que nos inclina a pensar que el mundo es como de hecho no es: que hay qualia, a pesar de que de hecho no existen; o que los qualia no son físicos, a pesar de que de hecho sí lo son.

Frankish (2016) llama ‘ilusionismo fuerte’ a la opción que niega la existencia de los qualia e ‘ilusionismo débil’ a la que meramente sostiene que no son como parecen ser. Cualquiera de las dos permite defender al fisicismo argumentando que la parte problemática de la consciencia fenoménica es meramente ilusoria. En general, puede entenderse que el ilusionismo reemplaza el problema difícil de la consciencia por el *meta-problema de la consciencia*, que consiste en explicar no los qualia, sino nuestra impresión de que existen y nuestras creencias en torno a ellos y a la consciencia fenoménica (ver Chalmers (2018)).

Si bien la idea central puede ser sumamente contraintuitiva, porque nuestras experiencias sugieren con muchísima fuerza que los qualia sí existen y están presentes, el punto es que es más razonable dudar de ellas que renunciar al fisicismo. Otra manera de plantearlo es diciendo que la ciencia debe ser nuestra guía primaria para descubrir cómo es realmente el mundo, así tengamos que poner en duda cualquier otra fuente de conocimiento, incluidas nuestras propias experiencias. A fin de cuentas, es razonable pensar que nuestros mecanismos naturales para percibir y entender el mundo son demasiado limitados, rudimentarios y guiados por intereses distintos del conocimiento más puro, en comparación con los que la ciencia ha logrado desarrollar. Como consecuencia, es razonable pensar que nuestras experiencias son defectuosas en ese sentido y que en un final resultan ser alguna especie de engaño o ilusión. Entre los ilusionistas podemos encontrar a Dennett (1988, 1991, 2005, 2016), Rey (1992b, 1995, 2007), Jackson (2003), Hall (2007), Pereboom (2011), Tartaglia (2013), Frankish (2016) y Kammerer (2021).

En particular, Pereboom (2011) propone una tesis ilusionista débil y llama ‘hipótesis de imprecisión cualitativa’ a la idea de que al inspeccionar los qualia somos víctimas de un engaño que nos hace pensar que son como de hecho no son. Por ejemplo, puede ser que los qualia de hecho no sean propiedades intrínsecas, categóricas, fundamentales, inefables, privadas y/o conocidas directa o infaliblemente. Frankish (2016), en contraste, se inclina por un ilusionismo fuerte, porque duda que lo que sea que quede al retirar las propiedades que convierten a los qualia en algo no-físico sea algo que sea digno de ser llamado

‘qualia’. Otra manera de entender la objeción es notando que llamamos ‘qualia’ a la apariencia misma, tal cual nos la presenta la experiencia, no al objeto que pueda estar detrás de ella, de modo que tratar de identificar a un quale de una manera que no sea hacer caso literal a lo que la ilusión nos sugiere es confundirlo con otra cosa. Es decir, si acaso hay algo físico que representamos en forma de qualia no-físicos, a la fuerza es algo distinto de dichos qualia. Yo concuerdo con esas observaciones y es por eso que mi postura es un ilusionismo fuerte. También por eso, en lo sucesivo reservaré los nombres ‘qualia’ y ‘quale’ para la versión no-física de las propiedades en cuestión y meramente diré que la ilusión en juego hace parecer que hay qualia (en ese sentido) cuando en realidad no los hay.

Ahora, Frankish (2016) se centra principalmente en la idea de que la ilusión es introspectiva: atribuimos erróneamente qualia a nuestros propios estados mentales (como las experiencias mismas). Sin embargo, también está la opción de que la ilusión sea extrospectiva: atribuimos erróneamente qualia a otra porción del mundo (lo cual abarca atribuir qualia a estados internos distintos de los mentales)¹²⁷. Una distinción cercana a esa podría ser entre ilusionismos de primer orden (o de orden inferior) y de segundo orden (o superior): en los primeros la ilusión se debe a la experiencia misma, que atribuye erróneamente qualia a alguna otra porción del mundo, mientras que en los segundos la ilusión se debe a un estado de orden superior, que monitorea a la experiencia y le atribuye erróneamente qualia. Digo que es una distinción cercana a la otra porque es natural entender que la ilusión introspectiva es de orden superior y la extrospectiva de orden inferior. Sin embargo, no necesariamente deben ser emparejadas de esa manera: bien podría ser que la experiencia misma se autoatribuya erróneamente qualia, por ejemplo, constituyendo lo que podríamos llamar una ‘ilusión del mismo orden’. Como es evidente, esas opciones muestran un marcado paralelismo con distintos tipos de representacionismos, de modo que es posible idear una versión ilusionista de cada uno.

También es posible que algunas experiencias involucren ilusiones de un tipo y otras de

¹²⁷ Frankish (2016) llama ‘ilusionismos volcados hacia afuera’ a los extrospectivos y sólo los menciona brevemente.

otro. Por ejemplo, es razonable pensar que las experiencias visuales son ilusiones extrospectivas de primer orden, porque sitúan a los qualia de color en la superficie de los objetos externos, mientras que las emociones son ilusiones introspectivas del mismo orden, porque sus correspondientes qualia son situados en ellas mismas. Y también podría ser que haya más de una ilusión en juego en un determinado caso. Por ejemplo, bien podría ser que los qualia propios de las emociones sean primero atribuidos a cambios fisiológicos y luego reasignados introspectivamente a estados mentales del sujeto. Otra opción, como sugiere Humphrey (2011), es que los qualia sean primero asignados a estados del sujeto y luego proyectados sobre el mundo externo. Pero en cualquiera de esos casos, lo que realmente es de interés es la primera ilusión, porque es la que explica a la consciencia fenoménica. Es decir, hay que distinguir entre la ilusión de que hay qualia (o de que son de cierta manera) y la que meramente consiste en confundir su ubicación aparente, que no juega un papel en explicar la primera. En lo personal considero que lo más plausible es que la mayoría de las experiencias sean ilusiones extrospectivas de primer orden, porque es razonable pensar que la mayoría de ellas tiene como propósito primario informarnos sobre cómo es el mundo más allá de nuestros propios estados mentales. Tal vez las únicas excepciones sean las emociones (y tal vez procesos cognitivos como el pensamiento, el deseo y la imaginación, suponiendo que tengan fenomenología propia¹²⁸).

Por último, cabe señalar que algunos autores, como Lycan (2000/19), consideran que el ilusionismo fuerte es incompatible con el representacionismo fuerte, porque consideran que el primero niega la existencia de lo que el segundo busca explicar, a saber, la consciencia fenoménica. De hecho, es cierto que algunos ilusionistas (como Kammerer (2021)) explícitamente dicen que el ilusionismo niega la existencia de la consciencia fenoménica (mientras que otros, como Frankish (2016), son más cautelosos). Sin embargo, ello se debe a que dichos autores entienden que tener consciencia fenoménica consiste en tener estados que de hecho poseen qualia. Así que para mostrar que el ilusionismo fuerte y el representacionismo fuerte sí son compatibles, basta con señalar que la consciencia fenoménica también puede ser entendida como mera presentación tipo E de qualia, como

128 Ver nota 4.

he estado haciendo en este escrito, y que bien podría haber presentación tipo E de qualia en ausencia de qualia. Incluso puede entenderse que una virtud del trabajo que presento es justamente que estoy ofreciendo una manera de liberar al ilusionismo del peso de hacer algo tan implausible como negar la existencia de la consciencia fenoménica: para mí es claro que esta existe; lo único que no existe son los qualia. Por otra parte, hasta es natural entender que la presentación tipo E es justamente la ilusión relevante y que es representacional, de modo que el ilusionismo fuerte naturalmente da paso al representacionismo fuerte. Así que estoy ofreciendo una manera bastante elegante de conjuntar ambas propuestas.

Ahora, lo cierto es que para muchos autores también parece inadmisibile negar la existencia de los qualia (ver por ejemplo Chalmers (1996, 2018) y Tye (1997/2017)) ante lo cual sólo queda enfatizar la fuerza de la ilusión en juego y señalar que todo representacionista fuerte, por el mero hecho de ser representacionista fuerte y no realista directo o teórico de los *sense-data*, debería estar en principio abierto a la posibilidad de que los qualia no existan. A fin de cuentas, justo una de sus motivaciones es que la representación no necesariamente requiere de la existencia o presencia de lo representado, de modo que recurrir a ella permite tener experiencias en ausencia de qualia. Lo que hace el representacionismo fuerte ilusionista fuerte no es otra cosa que explotar al máximo ese hecho, notando que el representacionista fuerte puede conceder cualquier tipo de contenido representacional cuya realidad sea incompatible con cualquier otra teoría que se sostenga (como el fisicismo), siempre y cuando al final se cuide de negar su existencia o su adecuación respecto al mundo real (ver también Jackson (2003)¹²⁹). La principal diferencia entre el representacionismo fuerte ilusionista y el ilusionismo a secas es que el primero afirma que las experiencias mismas son la ilusión, mientras que el segundo es compatible con que la ilusión sea generada por un mecanismo distinto de ellas (aunque también

129 Jackson (2003) también propone una teoría representacionista ilusionista fuerte (aunque no la llama así), pero no la detalla. Tal vez la diferencia más notoria entre su teoría y la mía es que él entiende que los qualia son parte tanto del contenido como del vehículo representacional de las experiencias fenoménicas, mientras que yo considero que sólo son contenido y que el vehículo son estados del sujeto de los que el sujeto no se percata (salvo por inferencia). Me parece que lo que está en el fondo de esa idea de Jackson es que no termina de decidirse entre si la ilusión es introspectiva o extrospectiva.

representacional).

Así que el representacionismo fuerte ilusionista fuerte es una postura coherente y hay muy buenas razones para aceptarlo, por lo que es una teoría que debería ser tomada bastante en serio. Otra razón a su favor es que permite salvar al fisicismo de los argumentos más fuertes en su contra e incluso promete cerrar la brecha epistémica entre lo físico y la consciencia fenoménica. Pero antes de ahondar en eso, conviene detallar un poco más la propuesta. En especial conviene explicar en qué consiste representar de modo E' qualia no-físicos bajo un marco fisicista, sobre todo respetando el internismo fenoménico y teniendo en cuenta la idea de que los qualia en cuestión no existen.

4.3. Representación tipo E' naturalizada

Antes de proceder a intentar naturalizar la representación tipo E' de qualia no-físicos, hay que señalar que es inviable aludir a propiedades representacionales etiológicas. En primer lugar, dichas propiedades siempre son extrínsecas, por lo que recurrir a ellas involucra renunciar al internismo fenoménico, lo cual es indeseable. Pero aun suponiendo que estuviésemos dispuestos a ello, tendríamos que enfrentar otro problema: que los qualia no existen, lo cual impide que se pueda tener una historia de interacción con ellos y, por ende, cumplir con un requisito mínimo para la representación etiológica. Así que lo más viable parece ser recurrir a propiedades representacionales disposicionales (Frankish (2016) también lo considera así).

Por otra parte, debo aclarar que lo que presento no es una teoría acabada sobre la representación tipo E', sino sólo un esbozo general de cómo podría ser desarrollada. La principal razón por la que no voy más allá es que, además de que el espacio del que dispongo es limitado, aún hace falta mucha investigación empírica para saber exactamente de qué disposiciones hay que dar cuenta y cuáles de las estructuras que podrían servir como su base categórica de hecho se encuentran presentes en nuestro cuerpo (especialmente en el cerebro). Otra razón es que para defender al fisicalismo no hace falta

mostrar los mecanismos específicos que están en juego, sino sólo motivar la idea de que, sean cuales sean, no parece que para explicarlos vaya a ser necesario ir más allá de lo físico.

Ahora, si recordamos que la representación tipo E' de x es algo que para quien la instancia parece ser *acquaintance* con x, resulta natural pensar que lo que hay que hacer es describir en qué consiste tener una *marcada disposición* de comportarse como si se tuviese *acquaintance* con x. Con ello en mente, conviene recordar cómo parecen ser nuestras experiencias fenoménicas a primera vista: un contacto directo e incuestionable con la apariencia (compuesta por qualia) que para nosotros tiene alguna porción del mundo, porque esta parece simplemente presentarse vívidamente ante nosotros sin ninguna clase de filtro, revelándonos su esencia. Entre otras cosas, de ahí podemos extraer la idea de que representar x de modo E' involucra o conduce a atribuirle a x ciertas propiedades, tales como ser de cierta manera, estar incuestionablemente presente y ser accesible de forma directa. A su vez, es posible dar cuenta de eso en términos de tener un conjunto de disposiciones que llevan al sujeto a actuar como si x fuese de cierto modo, se encontrase incuestionablemente presente ante él y no hubiese ninguna clase de mediador entre ambos. Y esas disposiciones pueden ser entendidas en términos de disposiciones más finas o incluso en términos de la ausencia de ellas. Por ejemplo, podría aludirse a disposiciones a tratar de interactuar con x de cierto modo, procurar o evadir a x, tomar en cuenta su presencia para algo más, privilegiar las experiencias fenoménicas sobre cualquier otra fuente de información (lo cual podría incluir la disposición a negar la posibilidad de que fuesen erradas), así como a la ausencia de disposiciones tales como la de dudar sobre el sentido de tomar esas acciones o la de comportarse como si hubiese un mediador. Pero incluso esas disposiciones (y ausencias de disposiciones) parecen no ser lo suficientemente finas, así que habría que seguir descomponiéndolas y, sobre todo, en algún punto habría que postular mecanismos físicos que sirvan como su base categórica.

Por otra parte, hay que recordar que lo que nos interesa principalmente es la representación tipo E' de un particular tipo de objetos: los qualia. Así que hay que aterrizar todas las

disposiciones propias de la representación tipo E' en el caso específico en el que lo que toma el lugar de x son qualia. En síntesis, lo que hay que explicar es cómo caracterizamos a los qualia, para saber qué tipo de propiedades los definen y, especialmente, qué disposiciones tenemos ante ellos. Por ejemplo, pueden señalarse las disposiciones a tomarlos como objetos concretos, abstractos, internos, externos, propiedades categóricas, propiedades intrínsecas o entidades fundamentales, simples o primitivas, así como las disposiciones a situarlos en lugares y momentos específicos (posiblemente dentro de una espacio-temporalidad también fenoménica) y las disposiciones a clasificarlos por tipos y subtipos y ordenarlos de acuerdo a relaciones de similaridad. Incluso puede ser que involucren algún componente afectivo, como resultar agradables o desagradables, lo cual se traduce en disposiciones a procurarlos o evitarlos, entre otras. De especial interés son las disposiciones a vincularlos o no con las entidades físicas, porque eso, en combinación con cómo es la representación tipo E', puede explicar disposiciones y comportamientos tales como sentir perplejidad ante los qualia, ser incapaces de vincularlos *a priori* con lo físico e incluso pensar en elaborados argumentos en contra del fisicismo.

También es importante señalar que la mayor parte de nuestras disposiciones son inaccesibles introspectivamente, sencillamente porque no tenemos integradas representaciones suyas del tipo apropiado. Siendo ese el caso, meramente tenemos acceso al resultado de ellas, que son nuestras experiencias y lo que estas nos sugieren sobre el mundo, no a los mecanismos subyacentes. De hecho, es razonable pensar que por eso es que nuestra percepción de los qualia parece directa y es tan difícil dudar de ella, a pesar de que sea indirecta y falible: sencillamente no poseemos los mecanismos apropiados para descubrir el engaño introspectivamente. A fin de cuentas, de lo que somos conscientes en primera instancia es del contenido representacional de nuestros estados, no de los estados mismos. Si acaso, podemos inferir a partir de nuestro acceso a dicho contenido que debemos tener estados que lo representen, pero eso no es lo mismo que tener acceso de primera mano hacia estos últimos. Además, cabe señalar que es de esperarse que nuestras experiencias estén protegidas contra el escrutinio, porque su propósito es hacernos creer que el mundo es de cierto modo, no meramente sugerirnos que lo es. Todo eso contribuye

a explicar que parezcan ser *acquaintance* con algo.

Ahora, la labor de identificar todas las disposiciones relevantes y explicarlas en términos físicos requeriría de mucho más espacio del que dispongo, y más si lo que se quiere no es meramente proponer mecanismos hipotéticos, sino presentar los que de hecho se encuentran funcionando dentro de nosotros. Por eso es que habremos de conformarnos con un tratamiento general y algo escueto, como el que he presentado hasta ahora. Por otra parte, ahondar en el tema no es necesario para notar lo que más me interesa: que en principio no parece haber ningún impedimento para aterrizar todas las disposiciones relevantes en mecanismos exclusivamente físicos, porque no parece que nuestro comportamiento en torno a los qualia sea algo especialmente misterioso para la física. De hecho, es comúnmente aceptado que nuestro Swampman compartiría con nosotros absolutamente todas nuestras disposiciones comportamentales, incluyendo todas las relativas a los qualia, tan sólo en virtud de compartir nuestra composición física.¹³⁰

Entonces, dado que ya tenemos una visión suficientemente completa de cómo la representación tipo E' de qualia puede ser explicada exhaustivamente en términos físicos, básicamente hemos terminado de construir un puente que vincula *a priori* la consciencia fenoménica con procesos físicos, en el sentido de que es *a priori* que ciertos acomodos físicos son representación tipo E' de qualia, porque es *a priori* que bastan para tener las disposiciones relevantes. Entonces es *a priori* que son consciencia fenoménica. En un final, lo que hemos hecho es describir qué se requiere para que un ente tenga la disposición de comportarse como si tuviese consciencia fenoménica, que no es otra cosa que describir nuestro propio comportamiento y disposiciones en relación a los qualia. La idea de fondo es que tener la disposición de comportarse como si se tuviese consciencia fenoménica es lo mismo que tenerla: a fin de cuentas, tanto nosotros como nuestro Swampman nos sentiríamos igual de seguros de que la tenemos e igual de perplejos ante los problemas que

130 Nótese que en principio estas ideas sirven para explicar la representación de cualquier tipo de entidad inexistente o incluso imposible. Además pueden ser complementadas con representaciones etiológicas para explicar representaciones sobre las que no se requiera ser internista. Por ejemplo, es razonable sostener que una interacción adecuada entre representaciones etiológicas de cuadrados y círculos basta para generar representaciones de círculos cuadrados.

presenta. Tener en claro todo eso servirá para entender cómo responder a los argumentos anti-fisicistas desde mi postura.

Pero antes de proceder en esa dirección, conviene atender un desafío que enfrenta la teoría: el de explicar por qué es que la naturaleza nos dotó de consciencia fenoménica, dado que a fin de cuentas es un engaño perpetuo y masivo.

4.4. El representacionismo ilusionista en el panorama evolutivo

Es claro que el representacionismo fuerte ilusionista fuerte es una teoría del error, porque sostiene que estamos bajo una ilusión perpetua y masiva: sistemáticamente representamos que el mundo es como de hecho no es y en una muy amplia gama de aspectos. Pero no sólo eso, sino que la ilusión es especialmente fuerte y podría parecer que es errada respecto a cosas básicas para nuestra supervivencia. Siendo ese el caso, es natural preguntarse cómo es que nuestra especie sobrevivió tanto tiempo así y por qué es que la naturaleza se esfuerza tanto en engañarnos de esa manera. Mi respuesta a ese problema tiene dos partes. Primero argumentaré que en realidad es natural, si no es que inevitable, que nuestros mecanismos naturales representen mal el mundo. Después, argumentaré que la ilusión en cuestión sí es útil para la supervivencia, tanto por su contenido como por su fuerza. Así que no sólo no es extraño que la naturaleza haya permitido el engaño, sino que es razonable pensar que lo seleccionó naturalmente.

En cuanto a la inevitabilidad de la ilusión, lo primero que podemos señalar es que somos capaces de detectar lo que hay a nuestro alrededor sólo en la medida en que nos afecta, lo cual es limitado. Por otra parte, un mismo conjunto de afectaciones puede ser compatible con más de una configuración del mundo externo. Como resultado, nuestra mente no sólo carece de herramientas para averiguar cómo es gran parte del mundo, sino que aquellas con las que sí cuenta son insuficientes para determinar cómo es la porción del mundo a la que apuntan. El resultado natural de eso es que la mente se equivoque en algunos aspectos y termine modelando al mundo como de hecho no es: por un lado es natural que omita por

completo algunas cosas y, por otro lado, es natural que reconstruya mal las que no omite. Por ejemplo, nuestras experiencias visuales y táctiles omiten el hecho de que los objetos físicos macroscópicos son mayormente vacío, debido a que los estímulos que recibimos de ellos, como la luz que emiten o reflejan (en el caso de la visión) y la presión resultante del choque de nuestros respectivos campos magnéticos (en el caso del tacto), son insuficientes para determinar (al menos a simple vista) su estructura interna.

Sin embargo, tiene sentido pensar que lo que a nuestra mente le importa, en términos de supervivencia (incluyendo reproducción), no es obtener un modelo perfecto de la totalidad del mundo, sino sólo un modelo útil de las partes relevantes para sobrevivir en él. Por eso es que no importa si se le escapan algunas cosas, ni si no reconstruye a la perfección las que no se le escapan. Ahora, cuando hablo de ‘un modelo útil’, me refiero a uno tal que tratarlo como si fuese el caso sea útil para lidiar adecuadamente con el mundo real. En otras palabras, se trata de uno que provoque los comportamientos apropiados ante las situaciones apropiadas. Para ello es útil, por ejemplo, ignorar todos los elementos del mundo que son incapaces de afectarnos significativamente, como puede ser el hecho de que los objetos macroscópicos sean mayormente vacío, porque a fin de cuentas el choque entre nuestros respectivos campos magnéticos nos impiden atravesarnos mutuamente. Por otra parte, puede ser más importante identificar que dos objetos son del mismo tipo o que se encuentran en cierto sitio, así como responder adecuadamente ante su presencia, que saber exactamente cómo son. Como consecuencia, si bien el mundo real no contiene qualia, el modelo que nuestra mente construye incluyéndolos puede ser útil porque nuestra mente los emplea para mapear propiedades que de hecho tiene el mundo real, de modo que similitudes, regularidades y cambios en ella corresponden a similitudes, regularidades y cambios en el mundo, a pesar de que los qualia no sean idénticos a las propiedades que mapean. Por ejemplo, tiene sentido pensar que los qualia de color mapean las propiedades de reflejar o emitir cierto tipo de luz, a pesar de no ser idénticos a ellas ni mapearlas a la perfección. A fin de cuentas, es útil que la mente le asigne una apariencia a esas propiedades, a pesar de que sea errada, porque sólo entonces cobran existencia para nosotros. En síntesis, parece razonable concluir que en efecto se trata de una ilusión útil,

porque lidiar con ella equivale a navegar el mundo adecuadamente, debido al paralelismo entre aspectos considerables de ambas cosas. Dennett (1991) expresa una idea similar comparando la consciencia fenoménica con las interfases generadas por computadora: son espacios que realmente no están ahí, pero nos ayudan a manipular adecuadamente la información que debemos manipular.

Y en lo que respecta a la representación tipo E', también tiene sentido que parezca una percatación directa e infalible, porque en la práctica es útil no dar cabida a dudas respecto a nuestro acceso cognitivo hacia el mundo, así como lo es no distinguir entre apariencia y realidad cuando lo que nos conviene es comportarnos como si fuesen la misma cosa. Es decir, para nuestra supervivencia es útil que la experiencia nos presente tan vívidamente el modelo que nuestra mente ha generado, como si fuese una certeza indudable (a pesar de ser errado en algunos aspectos), porque de otro modo no nos comportaríamos como si el mundo real existiese, sino que lo tomaríamos como una mera hipótesis o una mera sugerencia. En una línea cercana, Humphrey (2011) sugiere que la ilusión es altamente adaptativa justamente porque el dotarnos de la impresión de que nuestra vida mental es especial en algún sentido (tal vez incluso mágica o no-física) puede potenciar nuestro interés en ella y en lo que nos sugiere que es el caso: es como una especie de espectáculo diseñado para que nos involucremos más intensamente con el mundo. Así que todo el engaño resulta ser útil para nuestra supervivencia, por lo que tiene sentido que haya sido seleccionado naturalmente.

Por último, es razonable pensar que lo que hace la ciencia es retomar la tarea de modelar el mundo donde nuestros mecanismos perceptivos natos la han dejado, incorporando nuevos datos y estrategias que permiten encontrar un modelo más completo y preciso, que inevitablemente se distancia del que nuestras experiencias fenoménicas ofrecen. Por eso es que surge una brecha epistémica entre cómo nos dice la ciencia que es el mundo y cómo nos lo presenta nuestra consciencia fenoménica: porque básicamente la ciencia termina por evidenciar la ilusión. En tal caso, lo mejor es hacerle caso a la ciencia y tomar todo lo demás como una mala interpretación por parte de nuestra mente, que es justo lo que hace el

ilusionismo.

Con eso concluyo la exposición de mi postura, así que podemos proceder a revisar cómo responde a los principales problemas que aquejan al fisicismo, es decir, a los argumentos apoyados en la brecha epistémica entre lo físico y lo fenoménico.

4.5. Respuesta a los argumentos anti-fisicistas

Como se señaló en la sección 1.4.1, los argumentos anti-fisicistas que se apoyan en la existencia de la consciencia fenoménica suelen tener la siguiente forma general (Arg2):

- P1_{A2}) Hay una brecha epistémica *tipo G* entre las verdades sobre lo físico y las verdades sobre lo fenoménico.
 - P2_{A2}) Si hay una brecha epistémica *tipo G* entre las verdades sobre x y las verdades sobre z, entonces hay una brecha ontológica entre x y z.
 - P3_{A2}) Si hay una brecha ontológica entre lo físico y lo fenoménico, el fisicismo es falso.
- ∴ C1_{A2}) El fisicismo es falso.

P2_{A2} se deriva del racionalismo modal, mismo que no cuestionaré en este escrito (pero véase Reynoso (2017) para una defensa del fisicismo en esa dirección). Por otra parte, P3_{A2} es un compromiso básico del fisicismo, así que tampoco lo pondré en duda. Por ende, sólo resta atacar P1_{A2}, es decir, sostener que de hecho no hay una brecha epistémica tipo G entre las verdades sobre lo físico y las verdades sobre lo fenoménico. Y justamente todo el trabajo que he hecho en los últimos tres capítulos ha consistido en tratar de cerrar la supuesta brecha, motivando la idea de que la consciencia fenoménica es idéntica a alguna cosa x tal que no hay una brecha tipo G entre las verdades sobre x y las verdades sobre lo físico. Más específicamente, lo que he puesto en el lugar de x es la representación tipo E' de qualia: por una parte, he motivado la idea de que la consciencia fenoménica es idéntica a ella y, por otra parte, he explicado en qué sentido basta con ciertos acomodados físicos para

obtener representación tipo E' de qualia. Como consecuencia, he explicado en qué sentido basta con tener esos acomodos físicos (tipo E'') para obtener consciencia fenoménica. La idea puede compactarse en el siguiente contraargumento (Arg7), donde $R^@$ es un conjunto de proposiciones tal que de él se sigue necesariamente toda proposición ejemplificada por la totalidad de los acomodos físicos tipo E'' que tienen lugar en @ y $R'^@$ es un conjunto de proposiciones tal que de él se sigue necesariamente toda proposición ejemplificada por la totalidad de las instancias de representación tipo E' de qualia en @¹³¹:

- $P1_{A7}) \quad Q^@ \leftarrow R'^@$
 $P2_{A7}) \quad \Box(R^@ \rightarrow R'^@)$
 $P3_{A7}) \quad \Box(P^{@*} \rightarrow R^@)$
 $\therefore \quad C1_{A7}) \quad \Box(P^{@*} \rightarrow Q^@)$

El argumento es válido, $P3_{A7}$ es verdad por definición del fisicismo. Por otra parte, entender adecuadamente la consciencia fenoménica conduce a adoptar el representacionismo fuerte e irrestricto, lo cual equivale a entender que es *a priori* que lo que satisface ' $R'^@$ ' es lo mismo que satisface ' $Q^@$ ' y, por ende, a concluir que es *a priori* que $P1_{A7}$. Por último, entender a fondo la representación conduce a entender *a priori* que basta con $R^@$ para obtener $R'^@$, lo cual conduce a aceptar que es *a priori* que $P2_{A7}$. En un final, la idea es que he ofrecido una manera de cerrar la supuesta brecha tipo G entre las verdades sobre lo físico y las verdades sobre la consciencia fenoménica, de tal manera que se torna *a priori* que $\Box(P^{@*} \rightarrow Q^@)$. Siendo ese el caso, la premisa $P1_{A2}$ de Arg2 resulta ser falsa, lo cual estropea todo el argumento.

Pero tal vez conviene aterrizar las ideas en un caso menos general, como puede ser la instancia particular del argumento anti-fisicista propuesta por Chalmers (2005). Recordemos, entonces, mi formulación de dicho argumento (Arg3), donde CQ es la intención epistémica de ' $P^{@*} \& \neg Q^@$ ' que resulta de combinar la intención epistémica de ' $P^@$ ' que actualmente provee la ciencia con una intención epistémica de ' $\neg Q^@$ ' que está

131 Ver nota 20.

basada en nuestro trato cotidiano con la consciencia fenoménica y coincide con su intención subjuntiva:

P1_{A3}) $P^{@*} \& \neg Q^{@}$ es e-concebible idealmente bajo CQ.

P2_{A3}) Si algo es e-concebible idealmente bajo x, es e-posible bajo x.

P3_{A3}) Si $P^{@*} \& \neg Q^{@}$ es e-posible bajo CQ, $P^{@*} \& \neg Q^{@}$ es s-posible o la caracterización científica actual de $P^{@}$ es insuficiente.

P4_{A3}) Si $P^{@*} \& \neg Q^{@}$ es s-posible, el fisicismo es falso.

∴ C1_{A3}) El fisicismo es falso o la caracterización científica actual de $P^{@}$ es insuficiente.

Podemos notar que el argumento es válido, P4_{A3} es verdad por definición del fisicismo, P2_{A3} se sigue del racionalismo modal, mismo que no cuestionaré, y P3_{A3} es verdad por definición de CQ, una vez que se ha aceptado el racionalismo modal. Así que sólo resta atacar P1_{A3}. En su contra, lo que sugiero es meramente recordar Arg7 y concluir que la falsedad de P1_{A3} se sigue de la verdad de C1_{A7}, porque C1_{A7} es verdad *a priori* y de ella se sigue también *a priori* que $\neg \diamond (P^{@*} \& \neg Q^{@})$. En el fondo, lo que pasaría sería que toda mente ideal tendría en cuenta las premisas de Arg7 y ello le impediría considerar que $P^{@*} \& \neg Q^{@}$ es coherente.

Sin embargo, es natural que permanezca la intuición de que P1_{A3} es verdad y conviene ofrecer una explicación de por qué sucede eso. Considero que lo que pasa es que en el fondo sobrevive el impulso de asumir que la consciencia fenoménica involucra la presencia de qualia, lo cual lleva a confundir $P^{@*} \& \neg Q^{@}$ con algo más, como resultado de confundir $\neg Q^{@}$ con otra cosa. Más específicamente, se confunde $\neg Q^{@}$ con la ausencia de algún quale cuya presencia es sugerida por alguna experiencia fenoménica en @, lo cual conduce a confundir $P^{@*} \& \neg Q^{@}$ con un duplicado físico mínimo de @ que no incluye al correspondiente quale. Pero lo cierto es que ningún duplicado físico mínimo de @ sin cualquier quale es un problema para el fisicismo, porque los qualia sencillamente no existen en @. Otra manera de exponer la idea es diciendo que lo que se está teniendo en

mente al pensar en $\Box(P^{@*} \rightarrow Q^{@})$ no es $Q^{@}$ entendido (adecuadamente) como una descripción completa y acertada de todos los episodios de presentación tipo E de qualia en $@$, sino $Q^{@}$ entendido (erróneamente) como una descripción completa y acertada de todos los acomodos de qualia presentados de modo E por todas las experiencias fenoménicas en $@$. Pero lo cierto es que el fisicismo no está comprometido con $\Box(P^{@*} \rightarrow Q^{@})$ entendido del segundo modo, sino sólo con $\Box(P^{@*} \rightarrow Q^{#@})$ entendido del primer modo. Si llamamos ' $Q^{#@}$ ' a la segunda descripción, la idea es que el fisicismo no está comprometido con $\Box(P^{@*} \rightarrow Q^{#@})$, sino sólo con $\Box(P^{@*} \rightarrow Q^{@})$. Por poner un ejemplo burdo, lo que ocurre es algo parecido a confundir los pensamientos sobre la presencia de magia con la presencia de magia, para después concluir que un duplicado físico mínimo de $@$ sin magia es un contraejemplo para el fisicismo, lo cual es a todas luces descaminado.

Ahora, la única manera de evadir mi respuesta a Arg3 sería insistir en que realmente no hemos terminado de explicar la presentación tipo E de qualia en términos físicos: tal vez lo que he descrito sólo alcanza para lograr una semejanza superficial y no para realmente capturar lo que se pretende, porque ser funcionalmente idéntico y referir a las mismas cosas y de la misma manera que algo que tiene consciencia fenoménica no basta para tenerla también. Se podría insistir, por ejemplo, en que también se requiere de la presencia de qualia y que para ello no basta con la representación tipo E' de ellos.

Para anular definitivamente ese tipo de objeciones, conviene reflexionar un poco más sobre cómo sería la situación de alguien real en contraste con la de un duplicado exclusivamente físico suyo: independientemente de si tuviese o no consciencia fenoménica, dicho duplicado tendría exactamente las mismas disposiciones comportamentales que el original y aseguraría sí tener consciencia fenoménica con la misma vehemencia. De hecho, cualquier duplicado exclusivamente físico del mundo real sería tal que sus habitantes (nuestros duplicados) elaborarían sofisticados argumentos acerca de otros seres hipotéticos, físicamente idénticos a ellos pero sin consciencia fenoménica. Y si bien alguien podría insistir en que algunos de ellos estarían equivocados y nosotros no (respecto a tener consciencia fenoménica), el hecho es que no hay manera de saberlo desde nuestra

situación: hasta donde podemos asegurar, tanto nosotros como todos nuestros duplicados microfísicos seríamos subjetivamente indistinguibles. Es decir, no nos encontramos en una posición privilegiada para saber que no somos exactamente como esos seres, porque nuestro mismo nivel de convencimiento sobre tener consciencia fenoménica puede alcanzarse meramente copiando la parte física del mundo. Siendo ese el caso, resulta injustificado afirmar que nosotros tenemos consciencia fenoménica y alguno de nuestros duplicados no; o más bien dicho, estaríamos tan justificados para hacer esa afirmación sobre ellos como ellos para hacerla sobre nosotros. En el fondo hemos descrito tanto nuestra situación como la suya de forma completa y averiguado que en realidad no tenemos razón para asegurar que son distintas. Insistir en lo contrario sería mera necesidad.

Por último, podemos decir algo acerca del argumento anti-fisicista de Jackson (1982, 1986) mencionado en la sección 1.4, en el que se planteaba que parece no bastar con ser un experto en cómo funciona la visión a color a nivel físico para saber cómo es para el sujeto ver algo rojo. Si he hecho bien mi trabajo, la idea es que ahora debería ser relativamente claro que esa intuición es errada, porque el experto en cuestión sabría exactamente qué disposiciones (relevantes) tendría el sujeto que ve algo rojo y sabría cómo sería tener dichas disposiciones, de modo que tendría una idea precisa sobre cómo sería ver algo rojo. En otras palabras, la idea es que si el experto después viese algo rojo diría algo como “vaya, ciertamente ahora tengo nada más y nada menos que esas disposiciones, justo como esperaba”. Claro que eso es suponiendo que el experto fuese una mente ideal. Un experto más realista meramente diría algo como “pues sí, algo así me esperaba que sería”. Si al lector, como a mí, le parece más plausible esta última observación, es porque no somos mentes ideales ni expertos en visión humana al nivel relevante (lo cual incluiría no pasar por alto absolutamente ningún detalle sobre la visión humana y todos los microestados físicos que la habilitan en el caso particular de ver algo rojo). Así que es razonable concluir que la fuerza del argumento descansa en un estudio pobre de la situación. De hecho, el propio Jackson (2003) lo considera así y favorece un representacionismo fuerte ilusionista fuerte¹³².

132 Ver nota 130.

Entonces es bastante claro que el representacionismo fuerte ilusionista fuerte que propongo basta para lidiar con los principales problemas que enfrenta el fisicismo.

4.6. Sumario 4

En este capítulo expuse propiamente mi postura, que es un *representacionismo fuerte, irrestricto, reductivo, impuro y estrecho* que además es un *ilusionismo fuerte*. Del capítulo 1, retome la idea de entender a la consciencia fenoménica como presentación tipo E de qualia, así como los compromisos con el *fisicismo* y el *internismo fenoménico*. Del capítulo 2, rescaté la idea de que la presentación tipo E debe ser algún tipo E' de representación, porque la única alternativa (el *acquaintance*) parece no ser naturalizable. Por ende, me comprometí con un tipo de *representacionismo* en particular: *fuerte* porque identifica tener consciencia fenoménica con tener ciertas propiedades representacionales, *irrestricto* porque se pronuncia sobre todo tipo de experiencia fenoménica, *impuro* porque especifica un tipo E' de representación, *reductivo* porque busca eliminar la brecha epistémica entre lo físico y lo fenoménico y *estrecho* porque se compromete con el internismo fenoménico. Por otra parte, del capítulo 3 extraje la conclusión de que lo más razonable es conceder que los qualia no son físicos, lo cual a su vez me llevó a concluir que no existen, porque soy fisicista. Entonces lo único que hay en su lugar es representación tipo E' de qualia: una muy fuerte ilusión que hace parecer que los qualia están presentes cuando en realidad ni existen. Por eso es que mi postura es un *ilusionismo fuerte*.

Por otra parte, mi compromiso con el internismo fenoménico me llevó a ofrecer una explicación *disposicional* de la representación tipo E', porque en principio es imposible representar *etiológicamente* entidades inexistentes. A grandes rasgos, lo que sugerí es que representar x de modo E' consiste en tener la *marcada disposición* de comportarse como si x estuviese presente ante uno, lo cual es compatible con que x de hecho no exista.

Además, indagué en las razones por las que la naturaleza nos dotó de consciencia

fenoménica y señalé que, a pesar de ser una ilusión, es útil, porque lo que nos sugiere sobre el mundo es análogo a como de hecho es éste en los aspectos realmente importantes para la supervivencia. Por ende, comportarse como si fuese el caso lo que la ilusión sugiere que es el caso conduce a lidiar adecuadamente con el mundo real. Siendo ese el caso, incluso es razonable pensar que es útil que la ilusión sea tan fuerte como es, porque dudar de ella no es conveniente para sobrevivir.

Por último, mostré que la postura resultante también tiene la virtud de permitir evadir los argumentos anti-fisicistas que se apoyan en la concebibilidad ideal de $P^{@*} \& \neg Q^@$ y acomodar las intuiciones que los respaldan. Por un lado, ofrece una caracterización de la consciencia fenoménica cuya presencia claramente se sigue de la de lo físico, de tal modo que termina siendo *a priori* que $\Box(P^{@*} \rightarrow Q^@)$. Entonces $P^{@*} \& \neg Q^@$ resulta ser *a priori* incoherente y, por ende, inconcebible idealmente. Por otra parte, la teoría también permite ofrecer un sentido en el que $P^{@*} \& \neg Q^@$ parece posible a pesar de no serlo, porque explica por qué podemos confundir cierto mundo posible con uno imposible. La idea es que cuando creemos concebir $P^{@*} \& \neg Q^@$, lo que en realidad hacemos es confundir la representación tipo E' de qualia con presencia de qualia, lo cual conduce a tomar un duplicado físico mínimo de @ en el que no hay ciertos qualia (el cuál no sólo es coherente y posible, sino que bien podría tratarse de @ mismo, porque no hay ningún qualia en @) como si se tratase de un duplicado físico mínimo de @ sin las experiencias fenoménicas relevantes (el cual es incoherente y, por ende, imposible).

Conclusiones

5. Conclusiones

A lo largo de esta investigación he explorado distintas maneras de entender la naturaleza de la consciencia fenoménica, partiendo de una idea compatible con todas ellas: que la consciencia fenoménica es idéntica a presentación tipo E de qualia. Fue un punto de partida neutral porque estaba libre de compromisos sobre qué son en sí los qualia y la presentación tipo E. La estrategia general consistió en estudiar las diversas maneras de entender cada uno de esos dos elementos y las razones que las respaldan, para eventualmente dar con la combinación más plausible. Dicha versión, a mi juicio, fue la que entiende la presentación tipo E como un tipo (E') de representación y a los qualia como propiedades no-físicas. Es decir, terminé concluyendo que la consciencia fenoménica es idéntica a representación tipo E' de ciertas propiedades no-físicas (qualia).

La principal razón para sostener que la presentación tipo E es un tipo de representación fue que la única alternativa (el *acquaintance*) no es naturalizable. Otra razón fue que sostener que los qualia no son físicos me obligó a rechazar su existencia, porque soy físico, pero por definición es imposible tener *acquaintance* con objetos inexistentes. Por otra parte, las principales razones para sostener que los qualia no son físicos fueron que ninguna propiedad física cumple con lo adecuado para ser identificada con qualia y que las descripciones científicas de lo físico parecen no ser aptas para capturar lo que los qualia parecen ser a simple vista.

La postura resultante fue un *representacionismo fuerte, irrestricto, reductivo, impuro y estrecho* que a la vez es un *ilusionismo fuerte*. *Representacionismo fuerte, irrestricto e impuro* porque identifica el tener consciencia fenoménica (del tipo que sea) con tener la propiedad representacional impura de *representar de modo E' qualia*. *Reductivo* porque terminé planteando todo en términos exclusivamente físicos. *Estrecho* porque estoy comprometido con el internismo fenoménico. *Ilusionismo fuerte* porque niega la existencia de los qualia y sostiene que en su lugar sólo hay una muy fuerte ilusión (la representación tipo E') de que están presentes.

Dediqué el último capítulo a detallar la postura en cuestión y respondí a los problemas más obvios que la aquejan: el de explicar cómo es posible representar objetos inexistentes (y con tanto poder de convencimiento sobre su presencia) y el de explicar por qué es que la naturaleza nos dotó de consciencia fenoménica, dado que a fin de cuentas es un autoengaño y, como tal, no parece ser útil para sobrevivir, sino al contrario. Respecto al primer problema, sugerí explicar la representación tipo E' de x en términos de tener una *marcada disposición* de comportarse como si x se encontrase incuestionablemente presente ante uno. Sin embargo, reconocí que hace falta mucho trabajo para determinar con precisión qué disposiciones están involucradas y de qué manera son realizadas físicamente. En cuanto al segundo problema, señalé que algunas ilusiones pueden ser útiles. En particular, la consciencia fenoménica parece ser útil porque lo que sugiere que es el caso es análogo en ciertos aspectos relevantes a lo que de hecho es el caso. Además, es útil que tenga el poder de convencimiento que tiene, porque en la práctica no conviene estar dudando de que sea el caso lo que sugiere que es el caso.

Por último, expliqué cómo responder desde mi postura a los principales argumentos en contra del fisicismo. A grandes rasgos, la idea es que al detallarla lo suficientemente se torna *a priori* que basta con determinados acomodos físicos para obtener representación tipo E' de qualia, que a su vez es idéntica a consciencia fenoménica. Siendo ese el caso, se torna incoherente la idea de que exista la parte física de @ sin sus correspondientes experiencias fenoménicas, lo cual bloquea los argumentos. Lo único que resulta sí ser coherente es que exista la parte física de @ sin los qualia que nuestras experiencias sugieren que hay, pero eso no es un problema para el fisicismo, porque de hecho ese es el caso de @, porque los qualia sencillamente no existen.

Como conclusión, el representacionismo ilusionista que he propuesto resulta ser una postura coherente y con muchas virtudes, a pesar de que puede ser difícil de aceptar, debido a lo fuerte que es la ilusión generada por la representación tipo E'. Espero al menos haber ayudado a hacerla más digerible y tentadora.

Bibliografía

Adams, Frederick R.

(1979) 'A goal-state theory of function attributions', *Canadian Journal of Philosophy*, 1979, 9: 493-518.

Adams, Fred, & Aizawa, Ken

(2010/17) 'Causal Theories of Mental Content', *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), E. N. Zalta (ed.), 2017. URL = <<https://plato.stanford.edu/archives/sum2017/entries/content-causal/>>.

Adriaans, Pieter

(2012/18) 'Information', *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), E. N. Zalta (ed.), 2019.
URL = <<https://plato.stanford.edu/archives/spr2019/entries/information/>>.

Allen, Colin, & Bekoff, Marc

(1995) 'Function, Natural Design, and Animal Behavior: Philosophical and Ethological Considerations', *Perspectives in Ethology 11: Behavioral Design*, N.S. Thompson (ed.), Plenum Press, New York, 1995, 1-47.

Allen, Colin, & Neal, Jacob

(1996/2020) 'Teleological Notions in Biology', *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), E. N. Zalta (ed.), 2020. URL = <<https://plato.stanford.edu/archives/spr2020/entries/teleology-biology/>>.

Amundson, Ron, & Lauder, George V.

(1994) 'Function without purpose: the uses of causal role function in evolutionary biology', *Biology and Philosophy*, 9: 443-469.

Anscombe, G. Elizabeth M.

(1965) 'The Intentionality of Sensation: A Grammatical Feature', *Analytical Philosophy: Second Series*, R. J. Butler (ed.), Basil Blackwell, Oxford, 1965, 158-180.

Aristóteles

(c.350 a. C.) *Acerca del alma*, Gredos, Madrid, 1978.

Armstrong, David M.

(1968) *A Materialist Theory of the Mind*. Routledge, London, 1968.

(1969) 'Color-Realism and the Argument from Microscopes', *Contemporary Philosophy in Australia*, R. Brown & Rollins (Eds.), Allen and Unwin, London, 1969, 119-131.

(1981) 'What is consciousness?', *The Nature of Mind*. J. Heil (ed.), Cornell University Press, New York, 1981, 55-67.

Artiga, Marc

(2011) 'Re-organizing organizational accounts of function', *Applied Ontology*, 2011, 6(2): 105-124.

Artiga, Marc, & Martínez, Manolo

(2016) 'The Organizational Account of Function Is an Etiological Account of Function', *Acta Biotheoretica*, 2016, 64(2): 105-117.

Ayala, Francisco J.

(1970) 'Teleological Explanations in Evolutionary Biology', *Philosophy of Science*, 1970, 37: 1-15.

Aydede, Murat

(1998/2010) 'The Language of Thought Hypothesis', *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), E. N. Zalta (ed.), 2015. URL = <https://plato.stanford.edu/archives/fall2015/entries/language-thought/>.

Balog, Katalin

(2009) 'Phenomenal Concepts', *Oxford Handbook in the Philosophy of Mind*, B. McLaughlin, A. Beckermann & S. Walter (eds.), Oxford University Press, Oxford, 2009, 292-312.

Banks, Erik C.

(2014) *The Realistic Empiricism of Mach, James, and Russell. Neutral Monism Reconsidered*, Cambridge University Press, Cambridge, 2014.

Bayne, Tim, & Montague, Michelle

(2011) *Cognitive Phenomenology*, Oxford University Press, 2011.

Bedau, Mark A.

(1992) 'Goal-Directed Systems and the Good', *The Monist*, 1992, 75: 34-49.

Bennett, Karen

(2003) 'Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it', *Noûs*, 2003, 37(3): 471-497.

Bigelow, John, & Pargetter, Robert

(1987) 'Functions', *Journal of Philosophy*, 1987, 84: 181-96.

Block, Ned J.

(1980) 'Troubles with Functionalism', *Readings in the Philosophy of Psychology* (Volumen 1), N. Block (ed.), Harvard University Press, Cambridge, 1980, 268-305.

(1986) 'Advertisement for a Semantics for Psychology', *Midwest Studies in Philosophy*, 1986, 10: 615-678.

(1990) 'Inverted Earth', *Philosophical Perspectives*, J. Tomberlin (ed.), Ridgeview Publishing Company, Atascadero, 1990, 4: 53-79.

(1994) 'Qualia', *A Companion to Philosophy of Mind*. S. Guttenplan (ed.), Blackwell, Oxford, 1994, 514-520.

(1995) 'On a confusion about a function of consciousness', *Brain and Behavioral Sciences*, 1995, 18(2): 227-247.

(1996) 'Mental Paint and Mental Latex', *Philosophical Issues*, E. Villeneuve (ed.), Ridgeview Publishing Company, Atascadero, 1996, 7: 19-49.

(1998a) 'Is Experiencing Just Representing?', sobre Michael Tye, en un simposio: *Philosophy and Phenomenological Research*, (septiembre) 1998.

(1998b) 'Conceptual Role Semantics', *Routledge Encyclopedia of Philosophy*, E. Craig (ed.), Routledge, 1998, 242-256.

(2003) 'Mental Paint', *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, M. Hahn & B. Ramberg (eds.), MIT Press, Cambridge, MA, 2003, 165-200.

(2007) 'Consciousness, accessibility, and the mesh between psychology and neuroscience', *Behavioral and Brain Sciences*, 2007, 30:481-548.

(2011) 'The higher order approach to consciousness is defunct', *Analysis*, Oxford

University Press, Oxford, 2011, 71(3): 419-431.

Block, Ned J. & Campbell, John

(1987) 'Functional Role and Truth Conditions', *Proceedings of the Aristotelian Society*, 1987, 61: 157-181.

Block, Ned J. & Fodor, Jerry A.

(1972) 'What Psychological States Are Not', *Philosophical Review*, 1972, 81:159-181.

Block, Ned J. & Stalnaker, Robert C.

(1999) 'Conceptual analysis, dualism, and the explanatory gap', *Philosophical Review*, 1999, 108(1):1-46.

Boorse, Christopher

(1976) 'Wright on Functions', *Philosophical Review*, 1976, 85: 70-86.

(2002) 'A Rebuttal on Functions', *Functions*, A. R. Ariew, R. Cummins & M. Perlman (eds.), Oxford University Press, Oxford, 2002, 63-112.

BonJour, Laurence

(2001) 'Toward a Defense of Empirical Foundationalism', *Resurrecting Old-Fashioned Foundationalism*, M. DePaul (ed.), Rowman & Littlefield, Lanham, MD, 2001, 21-38.

(2003) 'A Version of Internalist Foundationalism', *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, L. BonJour & E. Sosa, Blackwell, Malden, MA, 2003, 3-23.

Boyton, Robert M.

(2003) 'Color Science', *Encyclopedia of Physical Science and Technology (Third Edition)*, 2003, 289-313.

Brentano, Franz C. H. H.

(1874) *Psychology from an Empirical Standpoint*, International Library of Philosophy, Londres, 1973.

Brewer, Bill

(2004) 'Realism and the Nature of Perceptual Experience', *Philosophical Issues*, 2004, 14(1): 61-77.

- (2006) 'Perception and Content', *European Journal of Philosophy*, 2006, 14(2): 165–181.
- (2008) 'How to Account for Illusion', *Disjunctivism: Perception, Action, Knowledge*, A. Haddock & F. MacPherson (eds.), Oxford University Press, Oxford, 2008, 168-180.
- (2011) *Perception and its Objects*, Oxford University Press, Oxford, 2011.
- (2017a) 'The Object View of Perception', *Topoi*, 2017, 35(2): 215-227.
- (2017b) 'Consciousness and Content in Perception', *Philosophical Perspectives*, 2017, 31(1):41-54.

Brooks, Rodney A.

- (1987) 'Intelligence without representation', *Artificial Intelligence*, 1991, 47: 139-159.

Burge, Tyler

- (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy*, P. French, T. Uehling & W. Wettstein (eds.), University of Minnesota Press, Minneapolis, 1979, 4: 73-121.
- (1986) 'Individualism and Psychology', *Philosophical Review*, 1986, 95: 3-45.
- (1997) 'Two Kinds of Consciousness', *The Nature of Consciousness: Philosophical Debates*, N. Block, O. Flanagan & G. Guzeldere (eds.), MIT Press, Cambridge, MA, 1997.

Byrne, Alex

- (1997) 'Some like it HOT: consciousness and higher-order thoughts', *Philosophical Studies*, 1997, 86: 103–129.
- (2001) 'Intentionalism Defended', *Philosophical Review*, 2001, 2(110): 199-240.
- (2004/15) 'Inverted Qualia', *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), E. N. Zalta (ed.), 2018. URL = [<https://plato.stanford.edu/archives/win2018/entries/qualia-inverted/>](https://plato.stanford.edu/archives/win2018/entries/qualia-inverted/).

Byrne, Alex, & Hilbert, David R.

- (1997) *Readings on Color: The Philosophy of Color Volume 1*, MIT Press, Cambridge, MA, 1997.

(2003) 'Color Realism and Color Science', *Behavioral and Brain Sciences*, 26(1): 1-44.

Byrne, Alex, & Logue, Heather

(2008) 'Either/Or', *Disjunctivism: Perception, Action, Knowledge*, A. Haddock & F. MacPherson (eds.), Oxford University Press, Oxford, 2008, 57-94.

Byrne, Alex, & Tye, Michael

(2006) 'Qualia ain't in the head', *Noûs*, 2006, 40(2): 241-255.

Campbell, John

(1994) 'A Simple View of Color', *Reality, Representation and Projection*, J. Haldane & C. Wright (eds.), Clarendon Press, Oxford, 1994, 257-269.

(2002a) *Reference and Consciousness*, Oxford University Press, Oxford, 2002.

(2002b) 'Berkeley's Puzzle', *Conceivability and Possibility*, T. S. Gendler & J. Hawthorne (eds.), Oxford University Press, Oxford, 2002, 127-144.

(2005a) 'Précis of *Reference and Consciousness*, and Replies to Neil Manson and Georges Rey', *Philosophical Studies*, 2005, 126: 103-114, 145-153, 155-162.

(2005b) 'Transparency versus Revelation in Color Perception', *Philosophical Topics*, 2005, 33: 105-115.

(2009) 'Consciousness and Reference', *The Oxford Handbook of Philosophy of Mind*, B. McLaughlin, A. Beckermann & S. Walter (eds.), Oxford University Press, Oxford, 2011.

(2011) 'Relational vs. Kantian Responses to Berkeley's Puzzle', *Perception, Causation and Objectivity*, N. Eilan, H. Lerman & J. Roessler (eds.), Oxford University Press, Oxford, 2011, 35-50.

Campbell, John, & Cassam, Quassim

(2014) *Berkeley's Puzzle: what does experience teach us?*, Oxford University Press, Oxford, 2014.

Canfield, John

(1964) 'Teleological explanation in biology', *British Journal for the Philosophy of Science*, 14: 285-95.

Carnap, Rudolf

(1947) *Meaning and Necessity*, The University of Chicago Press, Chicago, 1947.

Carruthers, Peter

(2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge University Press, Cambridge, 2000.

Chalmers, David J.

(1995a) 'Facing Up to the Problems of Consciousness'.

URL = <<http://consc.net/papers/facing.html>>.

(1995b) 'Absent Qualia, Fading Qualia, Dancing Qualia', *Conscious Experience*,

T. Metzinger (ed.), Imprint Academic, 1995.

(1996) *The Conscious Mind*, Oxford University Press, New York, 1996.

(2002a) 'Consciousness and its Place in Nature'.

URL = <<http://consc.net/papers/nature.pdf>>.

(2002b) 'Does Conceivability Entail Possibility?'.

URL = <<http://consc.net/papers/conceivability.html>>.

(2002c) 'On Sense and Intension'. URL = <<http://consc.net/papers/intension.pdf>>.

(2002d) 'The Components of Content'.

URL = <<http://consc.net/papers/content.html>>.

(2003) 'The Content and Epistemology of Phenomenal Belief'.

URL = <<http://consc.net/papers/belief.html>>.

(2004) 'The Representational Character of Experience'.

URL = <<http://consc.net/papers/representation.html>>.

(2005) 'The Two-Dimensional Argument Against Materialism'.

URL = <<http://consc.net/papers/2dargument.html>>.

(2006a) 'The Foundations of Two-Dimensional Semantics'.

URL = <<http://consc.net/papers/foundations.pdf>>.

(2006b) 'Two-Dimensional Semantics'.

URL = <<http://consc.net/papers/twodim.html>>.

(2006c) 'Perception and the Fall from Eden'.

URL = <<http://consc.net/papers/eden.html>>.

(2010) *The Character of Consciousness*, Oxford University Press, Oxford, 2010.

(2018) 'The Meta-Problem of Consciousness'.

URL = <<https://philpapers.org/archive/CHATMO-32.pdf>>.

Chisholm, Roderick

(1957) *Perceiving*, Cornell University Press, New York, 1957.

Choi, Sungho, & Fara, Michael

(2006/18) 'Dispositions', *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), E. N. Zalta (ed.), 2021.

URL = <<https://plato.stanford.edu/archives/spr2021/entries/dispositions/>>.

Christensen, Wayne D., & Mark H. Bickhard

(2002) 'The Process Dynamics of Normative Function', *The Monist*, 2002, 85: 3-28.

Chudnoff, Elijah

(2015) *Cognitive Phenomenology*, Routledge, 2015.

Clark, Andy

(2000) *A Theory of Sentience*, Oxford University Press, Oxford, 2000.

(2015) 'Embodied Prediction', *Open MIND* (7), T. Metzinger y J. M. Windt (Eds.), Frankfurt am Main, MIND group, 2015.

Cohen, Jonathan

(2007) 'Notes on Inferential Role Semantics'.

URL = <<https://pdfs.semanticscholar.org/4278/062b10ed48f1a36f17946fcc6510692d2b.pdf>>.

Crane, Tim

(2001) *Elements of Mind*, Oxford University Press, Oxford, 2001.

(2003) 'The Intentional Structure of Consciousness', *Consciousness: New Philosophical Perspectives*, Q. Smith & A. Jokic, Oxford University Press, Oxford, 2003, 33-56.

Crane, Tim, & French, Craig

(2005/15) 'The Problem of Perception', *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), E. N. Zalta (ed.), 2021. URL =

<<https://plato.stanford.edu/archives/spr2021/entries/perception-problem/>>.

Craver, Carl F.

(2001) 'Role functions, mechanisms, and hierarchy', *Philosophy of Science*, 2001, 68: 53-74.

Cummins, Robert C.

(1975) 'Functional analysis', *Journal of Philosophy*, 1975, 72: 741-765. Reimpresso en D. J. Buller (ed.), *Function, Selection, and Design*, SUNY Press, New York, 1999, 57-83.

(1996) *Representations, Targets and Attitudes*, MIT Press, Cambridge, MA, 1996.

Damasio, Antonio

(2000) *The Feeling of What Happens: Body and Emotions in the Making of Consciousness*, Mariner Books, 2000.

(2010) *Self Comes to Mind: Constructing the Conscious Brain*, Pantheon, 2010.

Davidson, Donald H.

(1967) 'Truth and Meaning', *Synthese*, (septiembre) 1967, 17(3): 304-323.

(1986) 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 1986, 60: 441-458.

Davies, Paul S.

(2001) *Norms of Nature. Naturalism and the Nature of Functions*, MIT Press, Cambridge, MA, 2001.

Dennett, Daniel C.

(1988) 'Quining qualia', *Mind and Cognition*, W. Lycan (ed.), Blackwell, Oxford, 1990, 519-548.

(1991) *Consciousness Explained*, Back Bay Books/Little, Brown and Company, New York, 1991.

(2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, MIT Press, Cambridge, MA, 2005.

(2015) 'Why and how does consciousness seem the way it seems?'

URL = <<https://open-mind.net/papers/why-and-how-does-consciousness-seem-the-way-it-seems/>>.

- (2016) 'Illusionism as the obvious default theory of consciousness', *Journal of Consciousness Studies*, 2016, 23(11-12):65-72.
- Descartes, René
- (1641) 'Meditaciones metafísicas', *Descartes*, Porrúa, México, 2010.
- Donnellan, Keith S.
- (1966) 'Reference and Definite Descriptions', *The Philosophical Review*, (julio) 1966, 75(3): 281-304.
- Dretske, Fred I.
- (1981) *Knowledge and the Flow of Information*, MIT/Bradford Press, Cambridge, 1981.
- (1986) 'Misrepresentation', *Belief: Form, content and function*, R. Bogdan (ed.), Oxford University Press, New York, 1986, 17-36.
- (1988) *Explaining Behavior: Reasons in a World of Causes*, MIT/Bradford, Cambridge, 1988.
- (1995) *Naturalizing the Mind*, MIT Press, Cambridge, MA, 1995.
- (1996) 'Phenomenal Externalism or If Meanings Ain't in the Head, Where are Qualia?', *Philosophical Issues 7: Perception*, E. Villanueva (ed.), Ridgeview, Atascadero, 1996.
- Droege, Paula
- (2003) *Caging the Beast: A Theory of Sensory Consciousness*, John Benjamins Publishing Company, Amsterdam, 2003.
- Edin, Benoni B.
- (2008) 'Assigning Biological Functions: Making Sense of Causal Chains', *Synthese*, 2008, 161: 203-18.
- Egan, Andy
- (2006) 'Appearance properties?', *Noûs*, 2006, 40(3): 495-521.
- Farrell, Brian A.
- (1950) 'Experience', *Mind*, 1950, 59: 170-198.
- Farrell, Jonathan
- (2016) 'What it is Like' talk is no Technical Talk', *Journal of Consciousness*

Studies, 2016, 23 (9-10): 50-65.

Feigl, Herbert

(1958) 'The "mental" and the "physical"', University of Minnesota Press, Minnesota, 1967.

Field, Hartry H.

(1977) 'Logic, Meaning and conceptual role', *The Journal of Philosophy*, (julio) 1977, 74(7): 379-409.

Fish, William C.

(2008) 'Disjunctivism, Indistinguishability, and the Nature of Hallucination', *Disjunctivism: Perception, Action, Knowledge*, A. Haddock & F. MacPherson (eds.), Oxford University Press, Oxford, 2008, 144-167.

(2009) *Perception, Hallucination, and Illusion*, Oxford University Press, Oxford, 2009.

Floridi, Luciano

(2005/15) 'Semantic Conceptions of Information', *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), E. N. Zalta (ed.), 2019.
URL = <<https://plato.stanford.edu/archives/win2019/entries/information-semantic/>>.

Fodor, Jerry A.

(1984) 'Semantics, Wisconsin Style', *Synthese*, 1984, 59: 231-250. También en Fodor (1990).

(1975) *The Language of Thought*, Thomas Y. Crowell, New York, 1975.

(1981) *Representations: Philosophical Essays on the Foundations of Cognitive Science*, MIT Press, Cambridge, 1981.

(1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT/Bradford Press, Cambridge, 1987.

(1990) 'A Theory of Content', *A Theory of Content and Other Essays*, MIT/Bradford Press, Cambridge, 1990.

(1994) *The Elm and the Expert*, MIT/Bradford Press, Cambridge, 1994.

(2008) *LOT2*, Clarendon Press, Oxford, 2008.

Frankish, Keith

- (2016) 'Illusionism as a theory of consciousness', *Journal of Consciousness Studies* (2016), 23(11-12):11-39. Reimpreso en *Illusionism as a Theory of Consciousness*, Frankish, K. (ed.), Imprint Academic, Exeter, 2017.

Frege, F. L. Gottlob

- (1892) 'On Sense and Reference', *The Philosophical Review*, Duke University Press, (Mayo) 1948, 57(3):209-230.

French, Craig, & Gomes, Anil

- (2019) 'How Naïve Realism Can Explain Both the Particularity and the Generality of Experience', *Philosophical Quarterly*, 2019, 69(275): 41-63.

Froese, Tom, & DiPaolo, Ezequiel A.

- (2011) 'The enactive approach: Theoretical sketches from cell to society', *Pragmatics & Cognition.*, 2011, 19(1): 1-36.

Fumerton, Richard

- (1995) *Metaepistemology and Skepticism*, Rowman and Littlefield, Lanham, MD, 1995.
- (2001) 'Classical Foundationalism', *Resurrecting Old-Fashioned Foundationalism*, M. De Paul (ed.), Rowman & Littlefield, Lanham, MD, 2001, .
- (2016) 'The Prospects for Traditional Internalism', *Intellectual Assurance: Essays on Traditional Epistemic Internalism*, B. Coppenger & M. Bergmann (eds.), Oxford University Press, Oxford, 2016.

García-Carpintero, Manuel

- (2002) 'Sense-Data: the Sensible Approach', *Grazer Philosophische Studien*, 2002, 62: 17-63.

Garson, Justin

- (2011) 'Selected Effects Functions and Causal Role Functions in the Brain: The Case for an Etiological Approach to Neuroscience', *Philosophy & Biology*, 2011, 26: 547-565.

Gert, Joshua

- (2006) 'A Realistic Colour Realism', *Australasian Journal of Philosophy*, 2006,

84(4): 565-589.

(2008) 'What Colors Could Not Be: An Argument for Color Primitivism', *The Journal of Philosophy*, 2008, 105(3): 128-155.

Godfrey-Smith, Peter

(1992) 'Indication and Adaptation', *Synthese*, 1992, 92: 283-312.

(1994) 'A Modern History Theory of Functions', *Noûs*, 1993, 28(3): 344-362.

(1996) *Complexity and the Function of Mind in Nature*, Cambridge University Press, Cambridge, 1996.

(2006) 'Mental Representation, Naturalism and Teleosemantics', *New Essays on Teleosemantics*, D. Papineau & G. Macdonald (eds.), Oxford University Press, Oxford, 2006, 42-68.

Gomes, Anil, & French, Craig

(2016) 'On the Particularity of Experience', *Philosophical Studies*, 2016, 173(2): 451-460.

Goodman, H. Nelson

(1968) *Languages of Art: An Approach to a Theory of Symbols*, The Bobbs-Merrill Company, Indianapolis, 1968.

(1972) 'Seven Strictures on Similarity', *Problems and Projects*, Bobbs-Merrill, New York. 1972, 437-446.

(1978) *Ways of Worldmaking*, Hackett Publishing Company, Indianapolis, 1978.

Greenberg, Mark, & Harman, Gilbert

(2006) 'Conceptual Role Semantics', *Oxford Handbook of Philosophy of Language*, E. Lepore & B. Smith (eds.), Oxford University Press, 2008.

Grice, H. Paul

(1948) 'Meaning', *The Philosophical Review*, (Julio) 1957, 66(3): 377-388.

Griffin, David R.

(1998) *Unsnarling the World-Knot: Consciousness, Freedom, and the Mind-Body Problem*, University of California Press, California, 1998.

Griffiths, Paul E.

(1993) 'Functional Analysis and Proper Functions', *British Journal for the*

Philosophy of Science, 1993, 44: 409-422.

Griffiths, Thomas L, Kemp, Charles, & Tenenbaum, Joshua B.

(2008) 'Bayesian Models of Cognition', *The Cambridge Handbook of Computational Psychology*, R. Sun (ed.), Cambridge University Press, Cambridge, 2008, 59-100.

Grote, John

(1865) *Exploratio Philosophica: Rough Notes on Modern Intellectual Science. Part I*, Deighton, Bell, and Company, Cambridge, 1865.

Guillot, Marie

(2017) 'I Me Mine: on a Confusion Concerning the Subjective Character of Experience', *Review of Philosophy and Psychology*, 2017, 8(1): 23-53.

Hacker, Peter M. S.

(1987) *Appearance and Reality: A Philosophical Investigation Into Perception and Perceptual Qualities*, Blackwell Publisher, Oxford, 1987.

Hall, Everett W.

(1961) *Our Knowledge of Fact and Value*, University of North Carolina Press, Chapel Hill, 1961.

Hall, Richard J.

(2007) 'Phenomenal properties as dummy properties', *Philosophical Studies*, 2007, 135(2): 199-223.

Harman, Gilbert

(1973) *Thought*, Princeton University Press, Princeton, 1973.

(1987) '(Nonsolipsistic) conceptual role semantics', *New Directions in Semantics*, E. LePore (ed.), Academic Press, Londres, 1987, 55-81.

(1990) 'The intrinsic quality of experience', *Philosophical Perspectives*, 1990, 4: 31-52.

Hatfield, Gary

(2021) 'Sense Data', *The Stanford Encyclopedia of Philosophy (Fall 2021 Edition)*, E. N. Zalta (ed.), 2021.

URL = <<https://plato.stanford.edu/archives/fall2021/entries/sense-data/>>.

Heil, John

(2013) *Philosophy of Mind. A Contemporary Introduction*, tercera edición, Routledge, Nueva York y Londres, 2013.

Hellie, Benj

(2013) 'The Multidisjunctive Conception of Hallucination', *Hallucination: Philosophy and Psychology*, F. MacPherson & D. Platchias (eds.), MIT Press, Cambridge, MA, 2013, 149-174.

Helmholtz, Hermann L.F. von

(1868) 'The Recent Progress of the Theory of Vision', *Popular Scientific Lectures*, P. H. Pye-Smith (trad), Dover Publications, New York, 1962, 93-185.

Hempel, Carl G.

(1959) 'The Logic of Functional Analysis', *Aspects of Scientific Explanation; And Other Essays in the Philosophy of Science*, The Free Press, Collier-Macmillan Limited, New York, 1965, 297-330.

(1969) 'Reduction: Ontological and Linguistic Facets', *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, S. Morgenbesser, P. Suppes & M. White (eds.), St Martin's Press, New York, 1969, 179-199.

Henderson, Leah

(2018) 'The Problem of Induction', *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), E. N. Zalta (ed.), 2018. URL = <https://plato.stanford.edu/archives/sum2018/entries/induction-problem/>.

Hilbert, David R.

(1987) *Color and Color Perception: A Study in Anthropocentric Realism*, CSLI Publications, Stanford, CA, 1987.

Hill, Christopher S.

(1997) 'Imaginability, conceivability, possibility, and the mind-body problem', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, (Julio) 1997, 87(1): 61-85.

Hill, Christopher S., & McLaughlin, Brian P.

(1999) 'There are fewer things in reality than are dreamt of in Chalmers'

philosophy', *Philosophy and Phenomenological Research*, (Junio) 1999, 59(2): 445-454.

Hintikka, Jaako

(1969) 'On the Logic of Perception', *Models for Modality*, Synthese Library, 1969, 23: 151-183.

Hinton, John M. E.

(1967) 'Visual Experiences', *Mind*, 1967, 76(302): 217-227.

(1973) *Experiences: An Inquiry into Some Ambiguities*, Clarendon Press, Oxford, 1973.

Horgan, Terence

(2011) 'From agentive phenomenology to Cognitive Phenomenology: A guide for the perplexed', *Cognitive Phenomenology*, T. Bayne & M. Montague (eds.), Oxford University Press, 2011.

Horgan, Terence, & Tienson, John L.

(2002) 'The intentionality of phenomenology and the phenomenology of intentionality', *Philosophy of Mind: Classical and Contemporary Readings*, D. Chalmers (ed.), Oxford University Press, Oxford, 2002, 520-533.

(2004) 'Phenomenal intentionality and the brain in a vat', *The Externalist Challenge*, R. Schantz (ed.), Walter De Gruyter, Berlin, 2004, 297-318.

Houghton, George

(2005) *Connectionist models in cognitive psychology*, Psychology Press, Hove, 2005.

Huemer, Michael,

(2004/11) 'Sense-Data', *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), E. N. Zalta (ed.), 2019.

URL = <<https://plato.stanford.edu/archives/spr2019/entries/sense-data/>>.

Hume, David

(1739) *A Treatise of Human Nature*, reimpresso en *Hume's Treatise of Human Nature*, L. A. Selby-Bigge (ed.), Clarendon Press, Oxford, 1888.

Humphrey, Nicholas

(2011) *Soul Dust: The Magic of Consciousness*, Princeton University Press,
Princeton, NJ, 2011.

Husserl, Edmund

(1929) *Las conferencias de París: Introducción a la fenomenología trascendental*,
Instituto de Investigaciones Filosóficas, México, 2009.

Hutto, Daniel D., & Myin, Erik

(2013) *Radicalizing Enactivism: Basic Minds Without Content*, MIT Press,
Cambridge, MA, 2013.

Hyman, John, & Bantinaki, Katerina

(2017) 'Depiction', *The Stanford Encyclopedia of Philosophy* (Summer 2017
Edition), E. N. Zalta (ed.), 2017.

URL = <<https://plato.stanford.edu/archives/sum2017/entries/depiction/>>.

Hyslop, Alec

(2005/14) 'Other Minds', *The Stanford Encyclopedia of Philosophy* (Winter 2018
Edition), E. N. Zalta (ed.), 2018.

URL = <<https://plato.stanford.edu/archives/win2018/entries/other-minds/>>.

Jackson, Frank C.

(1982) 'Epiphenomenal Qualia', *The Philosophical Quarterly*, 1982, 32: 127-136.

(1986) 'What Mary Didn't Know', *The Journal of Philosophy*, (mayo) 1986, 83(5):
291-295.

(1993) 'Armchair Metaphysics', *Philosophy in Mind*, J. Hawthorne & M. Michael
(eds.), Kluwer, Amsterdam, 1994, 23-42.

(1996) 'The Primary Quality View of Color', *Philosophical Perspectives*, 10:
199-219.

(1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Clarendon
Press, Oxford, 1998.

(2003) 'Mind and illusion', *Minds and Persons*, A. O'Hear (ed.), Cambridge
University Press, Cambridge, 2003, 251-272.

(2007) 'Colour for Representationalists', *Erkenntnis*, 2007, 66(1-2): 169-185.

Jacob, Pierre

(2003/19) 'Intentionality', *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), E. N. Zalta (ed.), 2019. URL = [<https://plato.stanford.edu/archives/win2019/entries/intentionality/>](https://plato.stanford.edu/archives/win2019/entries/intentionality/).

James, William

(1884) 'What is an Emotion?', *Mind*, 1884, 9(34): 188-205.

(1890) *The Principles of Psychology*, Holt, New York, 1890.

(1912) *Essays in Radical Empiricism*, University of Nebraska Press, Lincoln, 1996.

Johnson-Laird, Philip N.

(1977) 'Procedural Semantics', *Cognition*, 1977, 5:189-214.

Johnston, Mark

(2004) 'The Obscure Object of Hallucination', *Philosophical Studies*, 103 (120): 113-183.

Kallestrup, Jesper

(2006) 'Physicalism, Conceivability and Strong Necessities', *Synthese*, 2006, 151(2): 273-295.

Kammerer, François

(2021) 'The Illusion of Conscious Experience', *Synthese*, 2021, 198: 845-866.

Kant, Immanuel

(1787) *Crítica de la razón pura*, Biblioteca Virtual Universal, 2003.

URL = [<https://www.biblioteca.org.ar/libros/89799.pdf>](https://www.biblioteca.org.ar/libros/89799.pdf).

Kaplan, David

(1989) 'Demonstratives', *Themes from Kaplan*, J. Almog, J. Perry, & H. Wettstein (eds.), Oxford University Press, New York, 1989, 481-563.

Kelly, Thomas

(2006/14) 'Evidence', *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. N. Zalta (ed.), 2016.

URL = [<https://plato.stanford.edu/archives/win2016/entries/evidence/>](https://plato.stanford.edu/archives/win2016/entries/evidence/).

Kim, Jaegwon

(1993) *Mind and Supervenience*, Cambridge University Press, Cambridge, 1993.

(2005) 'Physicalism, or Something Near Enough', Princeton University Press,

Princeton, 2005.

Kirk, Robert

(1994) *Raw Feeling*, Oxford University Press, Oxford, 1994.

Kitcher, Phillip

(1993) 'Function and design', *Midwest Studies in Philosophy*, 18: 379-397.

Klein, Colin

(2015) *What the Body Commands: The Imperative Theory of Pain*, MIT Press, Cambridge, MA, 2015.

Kraut, Robert

(1982) 'Sensory States and Sensory Objects', *Noûs*, 1982, 16: 277-295.

Kriegel, Uriah

(2002) 'Phenomenal content', *Erkenntnis*, 2002, 57: 175-198.

(2003) 'Consciousness as intransitive self-consciousness: Two views and an argument', *Canadian Journal of Philosophy*, 2003, 33: 103-132.

(2004) 'Consciousness and self-consciousness', *The Monist*, 2004, 87: 185-209.

(2005) 'Naturalizing subjective character', *Philosophy and Phenomenological Research*, 2005, 71: 23-57.

(2006) 'The same-order monitoring theory of consciousness', en Kriegel & Williford (2006), 143-170.

(2009a) 'Self-Representationalism and Phenomenology', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, (Abril) 2009, 143(3): 357-381.

(2009b) *Subjective consciousness: A self-representational theory*, Oxford University Press, Oxford, 2009.

(2011) *The Sources of Intentionality*, Oxford University Press.

(2015a) 'The Character of Cognitive Phenomenology', *Phenomenology of Thinking*, T. Breyer & C. Gutland (eds.), Routledge, 2015.

(2015b) *The Varieties of Consciousness*, Oxford University Press, 2015.

(2017) 'Reductive Representationalism and Emotional Phenomenology', *Midwest Studies in Philosophy*, 2017, 41: 41-59.

- Kriegel, Uriah, & Williford, Kenneth W. (Eds.)
 (2006) *Self-representational approaches to consciousness*, MIT Press, Cambridge, MA, 2006.
- Kripke, Saul A.
 (1972) *El nombrar y la necesidad*, IFFs, UNAM, México, 2005.
- Lange, Carl G.
 (1885) *Om sindsbevægelser: et psyko-fysiologisk Studie*. Traducido como *The emotions*, en I. A. Haupt (Trad.), C. G. Lange & W. James (Eds.), *A series of reprints and translations. The emotions*, Williams & Wilkins Co., Baltimore, 1922, 1:33-90.
- Levine, Joseph
 (1983) 'Materialism and qualia: The explanatory gap', *Pacific Philosophical Quarterly*, 1983, 64: 354-61.
 (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford University Press, Oxford, 2001.
 (2003) 'Experience and Representation', *Consciousness: New Philosophical Perspectives*, Q. Smith & A. Jokic (eds.), Oxford University Press, Oxford, 2003, 57-76.
- Lewis, David K.
 (1979) 'Attitudes *De Dicto* and *De Se*', *Philosophical Review*, 1979, 88: 513-543.
 (1983a) 'Extrinsic Properties', *Philosophical Studies*, 1983, 44: 197-200.
 (1983b) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 1983, 61(4): 343-377.
 (1983c) 'Individuation by Acquaintance and by Stipulation', *Philosophical Review*, 1983, 92: 3-32.
 (1986) *On the Plurality of Worlds*, Blackwell, 1986.
 (1988) 'What experience teaches', *Proceedings of the Russellian Society*, Universidad de Sydney, Sydney, 1988.
 (1994) 'Reduction of Mind', *A Companion to the Philosophy of Mind*, S.

Guttenplan (ed.), Blackwell, Oxford, 1994, 412-431.

Lloyd, Dan

(1987) 'Mental Representation from the Bottom up', *Synthese*, 1987, 70: 23-78.

(1989) *Simple minds*, MIT Press, Cambridge, MA, 1989.

(1991) 'Leaping to Conclusions: Connectionism, Consciousness, and the Computational Mind', *Connectionism and the Philosophy of Mind*, T. Horgan & J. Tienson (eds.), Kluwer, Dordrecht, 1991, 444-459.

Loar, Brian

(1990) 'Phenomenal States', *Philosophical Perspectives 4, Action Theory and Philosophy of Mind*, James Tomberlin (ed.), Ridgeview, Atascadero, 1990, 81-108.

(1997) 'Phenomenal States' (segunda versión), *The Nature of Consciousness*, N. Block, O. Flanagan & G. Güzeldere (eds.), MIT Press, Cambridge, MA, 1997.

(1999) 'David Chalmers' The Conscious Mind', *Philosophy and Phenomenological Research*, (Junio) 1999, 59(2): 464-71.

(2003) 'Phenomenal intentionality as the basis of mental content', *Reflections and Replies*, M. Hahn & B. Ramberg (eds.), MIT Press, Cambridge, MA, 2003, 229-258.

Lockwood, Michael

(1989) *Mind, Brain, and the Quantum*, Oxford University Press, 1989.

Logue, Heather

(2012a) 'What should the naïve realist say about total hallucinations?', *Philosophical Perspectives*, 2012, 26(1): 173-199.

(2012b) 'Why Naïve Realism?', *Proceedings of the Aristotelian Society*, 2012, 112: 211-237.

Lycan, William G.

(1987) *Consciousness*, MIT Press, Cambridge, MA, 1987.

(1995) 'We've Only Just Begun', *Behavioral and Brain Sciences*, 1995, 18: 262-263.

- (1996) *Consciousness and Experience*, MIT Press, Cambridge, MA, 1996.
- (2000/19) 'Representational Theories of Consciousness', *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), E. N. Zalta (ed.), 2019.
 URL = <<https://plato.stanford.edu/archives/fall2019/entries/consciousness-representational/>>.
- (2001) 'The case for phenomenal externalism', *Philosophical Perspectives*, 2001, 15: 17-35.

Lyons, Jack

- (2016) 'Epistemological Problems of Perception', *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), E. N. Zalta (ed.), 2017. URL = <<https://plato.stanford.edu/archives/spr2017/entries/perception-episprob/>>.

MacBride, Fraser

- (2013/19) 'Truthmakers', *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), E. N. Zalta (ed.), 2019.
 URL = <<https://plato.stanford.edu/archives/spr2019/entries/truthmakers/>>.

Mach, Ernst W. J. W.

- (1886) *The Analysis of Sensations and the Relation of Physical to the Psychological*, Dover, Nueva York, 1959.
- (1894) *Popular Scientific Lectures*, The Open Court Publishing Company, La Salle, 1943.
- (1905) *Knowledge and Error*, Reidel, Dordrecht, 1976.

Mally, Ernst

- (1912) *Gegenstandstheoretische Grundlagen der Logik und Logistik*, Barth, Leipzig, 1912.

Martin, Michael G. F.

- (1997) 'The Reality of Appearances', *Thought and Ontology*, M. Sainsbury (ed.), Franco Angeli, Milán, 1997, 77-96.
- (1998) 'Setting Things Before the Mind', *Contemporary Issues in the Philosophy of Mind*, A. O'Hear (ed.), Cambridge University Press, Cambridge, 1998, 157-179.

(2002) 'The Transparency of Experience', *Mind and Language*, 2002, 17: 376-425.

(2004) 'The Limits of Self-Awareness', *Philosophical Studies*, 2004, 120: 37-89.

(2006) 'On Being Alienated', *Perceptual Experience*, T. S. Gendler & J.

Hawthorne (eds.), Oxford University Press, Oxford, 2006, 354-410.

Matthen, Mohan

(1988) 'Biological Function and Perceptual Content', *The Journal of Philosophy*, 1988, 85(1): 5-27.

Margolis, Eric, & Laurence, Stephen

(2005/11) 'Concepts', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), E. N. Zalta (ed.), 2014.

URL = <<https://plato.stanford.edu/archives/spr2014/entries/concepts/>>.

Marshall, Dan, & Weatherson, Brian

(2002/12) 'Intrinsic vs. Extrinsic Properties', *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), E. N. Zalta (ed.), 2018. URL =

<<http://plato.stanford.edu/archives/fall2014/entries/intrinsic-extrinsic/>>.

Martínez, Manolo

(2011) 'Imperative Content and the Painfulness of Pain', *Phenomenology and the Cognitive Sciences*, 2011, 10(1): 67-90.

Maud, Barry

(1997/2018) 'Color', *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), E. N. Zalta (ed.), 2022.

URL = <<https://plato.stanford.edu/archives/spr2022/entries/color/>>.

Maxwell, Grover

(1979) 'Rigid designators and mind-brain identity', *Minnesota Studies in the Philosophy of Science*, 1979, 9: 365-403.

Mayr, Ernst

(1965) 'Cause and effect in biology', *Cause and effect*, D. Lerner (ed.), Free Press, New York, 1965, 33-50.

(1974) 'Teleological and Teleonomic, a New Analysis', *Methodological and Historical Essays in the Natural and Social Sciences* (Boston Studies in the

Philosophy of Science 14), R. S. Cohen & M. W. Wartofsky (eds.), Springer Netherlands, Dordrecht, 91-117.

(1988) 'The multiple meanings of teleological', *Towards a New Philosophy of Biology*, E. Mayr (ed.), Harvard University Press, Cambridge, 38-66.

McClelland, Tom

(2013) 'The neo-Russellian ignorance hypothesis: A hybrid account of phenomenal consciousness', *Journal of Consciousness Studies*, (marzo) 2013, 20(3-4): 125-151.

(2016) 'Can self-representationalism explain away the apparent irreducibility of consciousness?', *Synthese*, (junio) 2016, 193(6): 1755-1776.

McDowell, John H.

(1982) 'Criteria, Defeasibility and Knowledge', *Proceedings of the British Academy*, 1982, 68: 455-79.

(1987) 'Singular Thought and the Extent of Inner Space', *Subject, Thought and Context*, P. Pettit & J. H. McDowell (eds.), Clarendon Press, Oxford, 1987, 137-168.

(1994) 'The Content of Perceptual Experience', *Philosophical Quarterly*, 1994, 44(175): 190-205.

(2013) 'Tyler Burge on Disjunctivism (II)', *Philosophical Explorations*, 2013, 16(3): 259-279.

McGinn, Colin

(1988) 'Consciousness and content', *Proceedings of the British Academy*, 1988, 74:219-39.

(1989) 'Can we solve the mind-body problem?', *Mind*, 1989, 98: 349-66.

(1991) *The Problem of Consciousness*, Blackwell, Oxford, 1991.

(1996) 'Another Look at Color', *The Journal of Philosophy*, 1996, 93(11): 537-555.

(1999) *The Mysterious Flame*, Basic Books, New York, 1999.

McLaughlin, Peter

(2001) *What Functions Explain. Functional Explanation and Self-Reproducing*

- Systems*, Cambridge University Press, Cambridge, 2001.
- (2003) 'The Place of Colour in Nature', *Colour Perception: Mind and the Physical World*, R. Mausfeld & D. Heyer (eds.), Oxford University Press, New York, 2003, 475-502.
- McLaughlin, Brian, & Bennett, Karen
 (2005/18) 'Supervenience', *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), E. N. Zalta (ed.), 2018. URL = [<https://plato.stanford.edu/archives/win2018/entries/supervenience/>](https://plato.stanford.edu/archives/win2018/entries/supervenience/).
- Mehta, Neil
 (2014) 'The Limited Role of Particulars in Phenomenal Experience', *Journal of Philosophy*, 2014, 111(6): 311-331.
- Mehta, Neil, & Ganson, Todd
 (2016) 'On the Generality of Experience: a reply to French and Gomes', *Philosophical Studies*, 2016, 173(12): 3223-3229.
- Meinong, Alexius
 (1904) 'The Theory of Objects', *Untersuchungen zur Gegenstandstheorie und Psychologie*, A. Meinong (Ed.), Barth, Leipzig, 1904.
- Melander, Peter
 (1997) *Analyzing Functions. An Essay on a Fundamental Notion in Biology*, Almqvist & Wiksell International, Estocolmo, 1997.
- Menzel, Christopher
 (2013/16) 'Possible Worlds', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), E. N. Zalta (ed.), 2017. URL = [<https://plato.stanford.edu/archives/win2017/entries/possible-worlds/>](https://plato.stanford.edu/archives/win2017/entries/possible-worlds/).
- Michaelson, Eliot, & Reimer, Marga
 (2003/19) 'Reference', *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), E. N. Zalta (ed.), 2019. URL = [<https://plato.stanford.edu/archives/spr2019/entries/reference/>](https://plato.stanford.edu/archives/spr2019/entries/reference/).
- Mill, John Stuart
 (1843) *A System of Logic, Ratiocinative and Inductive*, Harper & Brothers, Franklin

Square, New York, 1882.

URL = <<https://www.gutenberg.org/files/27942/27942-pdf.pdf>>.

Millar, Alan

(2007) 'What the Disjunctivist is Right About', *Philosophy and Phenomenological Research*, 2007, 74(1): 176-198.

(2008) 'Perceptual-recognitional Abilities and Perceptual Knowledge', *Disjunctivism: Perception, Action, Knowledge*, A. Haddock & F.

Macpherson (eds.), Oxford University Press, Oxford, 2008, 330-347.

Miller, George A., & Johnson-Laird, Philip N.

(1976) *Language and Perception*, Harvard University Press, Cambridge, 1976.

Millikan, Ruth G.

(1984) *Language, Thought, and Other Biological Categories*, MIT Press, Cambridge, 1984.

(1989) 'Biosemantics', *The Journal of Philosophy*, (julio) 1989, 86: 281-297.

(1991) 'Speaking Up for Darwin', *Meaning in Mind: Fodor and his critics*, B. Loewer & G. Rey (eds.), Blackwell, Cambridge, MA, 1991, 151-165.

(1993) 'Propensities, exaptations, and the brain', *White Queen Psychology and Other Essays for Alice*, MIT Press, Cambridge, 1993, 31-50.

(2001) 'What Has Natural Information to Do with Intentional Representation?', *Naturalism, Evolution and Mind*, D. M. Walsh (ed.), Cambridge University Press, Cambridge, 2001, 105-125.

(2002) 'Biofunctions: Two paradigms', *Functions*, A. R. Ariew, R. Cummins & M. Perlman (eds.), Oxford University Press, Oxford, 2002, 113-43.

(2013) 'An Epistemology for Phenomenology?', *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, R. Brown (ed.), de la serie Studies in Brain and Mind, Springer, 2013, 13-26.

(2017) *Beyond Concepts: Unicepts, Language, and Natural Information*, Oxford University Press, Oxford, 2017.

Montero, Barbara, & Papineau, David

(2005) 'A defense of the Via Negativa Argument for Physicalism', *Analysis*, 2005,

65(3): 233-237.

Mossio, Matteo, Saborido, Cristian, & Moreno, Álvaro

(2009) 'An organizational account of biological functions', *British Journal for the Philosophy of Science*, (diciembre) 2009, 60(4):813-841.

Nagel, Ernest

(1961) *The Structure of Science*, Routledge & Kegan Paul, Londres, 1961.

(1977) 'Teleology revisited', *Journal of Philosophy*, 74: 261-301.

Nagel, Thomas

(1974) 'What It Is Like to Be a Bat', *The Philosophical Review*, (Octubre) 1974, 83(4): 435-450.

(2012) *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*, Oxford University Press, Oxford, 2012.

Neander, Karen

(1991) 'Function as selected effects: The conceptual analyst's defense', *Philosophy of Science*, 1991, 58: 168-184.

(1995) 'Misrepresenting and Malfunctioning', *Philosophical Studies*, 79: 109-141.

(1998) 'The division of phenomenal labor: a problem for representational theories of consciousness', *Language, Mind, and Ontology*, J. Tomberlin (ed.), Blackwell, Oxford, 1998.

(2017) *A Mark of the Mental*, MIT Press, Cambridge, MA, 2017.

Neander, Karen, & Schulte, Peter

(2004-20) 'Teleological Theories of Mental Content', *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), E. N. Zalta (ed.), 2021. URL = <https://plato.stanford.edu/archives/spr2021/entries/content-teleological/>.

Nissen, Lowell

(1980) 'Nagel's self-regulation analysis of teleology', *Philosophical Forum*, 1980, 12: 128-138.

Noë, Alva

(2004) *Action in Perception*, MIT Press, Cambridge, MA, 2004.

(2009) *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from*

the Biology of Consciousness, Hill & Wang, New York, 2009.

Orilia, Francesco, & Paolini Paoletti, Michele

(1999/2020) 'Properties', *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), E. N. Zalta (ed.), 2020.

URL = <<https://plato.stanford.edu/archives/win2020/entries/properties/>>.

O'Shaughnessy, Brian

(2003) 'Sense data', *John Searle*, B. Smith (ed.), Cambridge University Press, Cambridge, 2003, 169-188.

Papineau, David

(1984) 'Representation and Explanation', *Philosophy of Science*, 51: 550-572.

(1987) *Reality and Representation*, Blackwell, Oxford, 1987.

(1993) 'Physicalism, consciousness, and the antipathetic fallacy', *Australasian Journal of Philosophy*, 1993, 71:169-83.

(2016) 'Teleosemantics', *How Biology Shapes Philosophy*, D. L. Smith (Ed.), Cambridge University Press, Cambridge, 95-120.

Parsons, Terence

(1980) *Nonexistent Objects*, Yale University Press, New Haven, 1980.

Pautz, Adam

(2007) 'Intentionalism and Perceptual Presence', *Philosophy of Mind (Philosophical Perspectives)*, Hawthorne, J. (ed.), Wiley-Blackwell, Malden, 2007, 21(1): 495-541.

(2010) 'Why Explain Experience in Terms of Content?', *Perceiving the World*, B. Nanay (ed.) 2010, Oxford University Press, New York.

(2014) 'The Real Trouble for Phenomenal Externalists: New Evidence for a Brain-Based Theory of Sensory Consciousness', *Consciousness Inside and Out*, R. Brown (ed.), Springer. Dordrecht, 2014.

(2018) 'Naïve Realism and the Science of Consciousness', *Analytic Philosophy*, por publicarse.

URL = <<https://philpapers.org/archive/PAUNRA-2.pdf>>.

Peirce, Charles S.

(1931/35/58) *The Collected Papers of Charles Sanders Peirce*, 1994.

URL = <<https://colorysemiotica.files.wordpress.com/2014/08/peirce-collectedpapers.pdf>>.

Pereboom, Derk

(2011) *Consciousness and the Prospects of Physicalism*, Oxford University Press, New York, 2011.

Perry, John R.

(2001) *Knowledge, Possibility, and Consciousness*, MIT Press, Cambridge, MA, 2001.

Peacocke, Christopher A. B.

(1983) *Sense and Content*, Oxford University Press, Oxford, 1983.

(1992) *A study of concepts*, MIT Press, Cambridge, MA, 1992.

Pitt, David

(2000/12) 'Mental Representation', *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), E. N. Zalta (ed.), 2018.

URL = <<https://plato.stanford.edu/archives/win2018/entries/mental-representation/>>.

(2004) 'The phenomenology of cognition or what is it like to believe that p?', *Philosophy and Phenomenological Research*, 2004, 69: 1-36.

(2011) 'Introspection, Phenomenality, and the Availability of Intentional Content', *Cognitive Phenomenology*, T. Bayne & M. Montague (eds.), Oxford University Press, 2011.

Pittendrigh, Collin S.

(1958) 'Adaptation, natural selection, and behavior', *Behavior and Evolution*, A. Roe & [G. Gaylord Simpson](#) (eds.), Yale University Press, New Haven, 1958, 390-416.

Prinz, Jesse J.

(2011) 'Is Attention Necessary and Sufficient for Consciousness?', *Attention: Philosophical and Psychological Essays*, C. Mole, D. Smithies & W. Wu (eds.), Oxford University Press, Oxford, 2011.

Putnam, Hilary W.

(1973) 'Meaning and reference', *The Journal of Philosophy*, 1973, 70(19): 699-711.

(1975) 'El significado de 'significado'', Cuadernos de crítica 28, IIFs, UNAM, México, 1984.

(1981) *Razon, verdad e historia*, Tecnos, Madrid, 1988.

(1999) *The Threefold Chord*, Columbia University Press, New York, 1999.

Ramachandran, Vilayanur S., & Hirstein, William

(1998) 'The perception of phantom limbs: The D.O. Hebb lecture', *Brain*, 1998, 121:1603-1630.

Rescorla, Michael

(2015) 'The Computational Theory of Mind', *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), E. N. Zalta (ed.), 2015. URL = <https://plato.stanford.edu/archives/spr2017/entries/computational-mind/>.

Rey, Georges

(1992a) 'Sensational Sentences', *Consciousness: Psychological and Philosophical Essays*, M. Davies & G. Humphreys (eds.), Blackwell, Oxford, 1992.

(1992b) 'Sensational Sentences Switched', *Philosophical Studies*, 1992, 68: 289-331.

(1995) 'Toward a projectivist account of conscious experience', *Conscious Experience*, T. Metzinger (ed.), Ferdinand Schoningh, 1995, 123-142.

(1998) 'A narrow representationalist account of qualitative experience', *Philosophical Perspectives*, 1998, 12: 435-458.

(2007) 'Phenomenal content and the richness and determinacy of colour experience', *Journal of Consciousness Studies*, 2007, 14 (9): 112-131.

Reynoso, Diego

(2017) *Una defensa del materialismo a posteriori en contra de los argumentos apoyados en la concebibilidad de la falsedad del materialismo*, tesis de maestría, UNAM, México, 2017.

Rosenberg, Gregg

(1997) *A Place for Consciousness: Probing the Deep Structure of the Natural*

World, Tesis doctoral, Indiana University, 1997.

Rosenblueth, Arturo, Wiener, Norbert, & Bigelow, Julian

(1943) 'Behavior, Purpose and Teleology', *Philosophy of Science*, 10: 18-24.

Rosenthal, David M.

(1976) 'Mentality and neutrality', *Journal of Philosophy*, , 1976, 73(13): 386-415.

(1986) 'Two concepts of consciousness,' *Philosophical Studies*, 1986, 49: 329-359.

(1993) 'Thinking that one thinks', *Consciousness: Psychological and Philosophical Essays*. M. Davies & G. W. Humphreys (eds.), Blackwell, 1993.

(2002) 'Explaining Consciousness', *Philosophy of Mind: Classical and Contemporary Readings*, D. J. Chalmers (ed.), Oxford University press, Oxford, 2002, 109-131.

(2005) *Consciousness and Mind*, Oxford University Press, Oxford, 2005.

(2010) 'How to Think About Mental Qualities', *Philosophical Issues*, un suplemento de *Noûs*, 2010, 20(1): 368-393.

(2011) 'Exaggerated reports: reply to Block', *Analysis*, 2011, 71: 431-437.

Rupert, Robert D.

(1999) 'The Best Test Theory of Extension: First Principle(s)', *Mind & Language*, 1999, 14: 321-355.

Ruse, Michael

(1971) 'Functional statements in biology', *Philosophy of Science*, 38: 87-95.

Russell, Bertrand A. W.

(1905) 'On Denoting', *Mind*, New Series, Oxford University Press, Oxford, (Octubre) 1905, 14(56): 479-493.

(1910/11) 'Knowledge by Acquaintance and Knowledge by Description', *Proceedings of the Aristotelian Society*, 1910-11, 11: 108-128. También en Russell (1912).

(1912) *Los problemas de la filosofía*.

URL= <<http://www.enxarxa.com/biblioteca/RUSSELL%20Los%20problemas%20de%20la%20filosofia.pdf>>.

(1914) 'The Relation of Sense-Data to Physics', *Scientia*, 1914, 16: 1-27.

(1927) *The Analysis of Matter*, Kegan Paul, Londres, 1927.

Ryder, Dan

(2004) 'SINBAD Neurosemantics: A Theory of Mental Representation', *Mind & Language*, 2004, 19: 211-240.

Ryle, Gilbert

(1949) *The Concept of Mind*, Hutchinson and Co., 1949.

Sayre, Kenneth M.

(1976) *Cybernetics and the Philosophy of Mind*, Humanities Press, Atlantic Highlands, 1976.

Schaffer, Jonathan,

(2003/16) 'The Metaphysics of Causation', *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), E. N. Zalta (ed.), 2016.

URL = <<https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/>>.

Schellenberg, Susanna

(2011) 'Perceptual Content Defended', *Noûs*, 2011, 54 (4): 714-740.

(2018) *The Unity of Perception: Content, Consciousness, Evidence*, Oxford University Press, Oxford, 2018.

Schlosser, Gerhard

(1998) 'Self-Re-Production and Functionality', *Synthese*, 1998, 116: 303-354.

Searle, John R.

(1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Estados Unidos de América, 1983.

(1990) 'Consciousness, explanatory inversion and cognitive science', *Behavioral and Brain Sciences*, 1990, 13: 585-642.

Sebastián, Miguel Ángel

(2012) 'Experiential Awareness: Do you Prefer It to Me?', *Philosophical Topics*, 2012, 40(2): 155-177.

(2013) 'Not a HOT Dream', *Consciousness Inside and Out: Phenomenology*,

Neuroscience, and the Nature of Experience, Springer Studies in Brain and Mind, 2013, 415-432.

- (2015) 'On a Confusion About Which Intuitions to Trust: From the Hard Problem to a Not Easy One', *Topoi* (2015). doi:10.1007/s11245-015-9326-3
- (2017) 'Foundations for a Metasemantic Theory of *de se* Representation' (ponencia), XXXV Simposio Internacional de Filosofía, IIFs, (noviembre) 2017.
- (2018) 'Consciencia', *Enciclopedia de la Sociedad Española de Filosofía Analítica*, 2018. URL = <<http://www.sefaweb.es/consciencia/>>.
- (2019) 'Drop it like it's HOT: a vicious regress of higher order thought theories', *Philosophical Studies*, 2019, 176(6):1563-1572.
- (2022) *El problema de la consciencia: una introducción crítica a la discusión filosófica actual*, Catedra, Madrid, 2022.

Sellars, Wilfrid

- (1956) 'Empiricism and the Philosophy of Mind', *Minnesota Studies in the Philosophy of Science*, 1956, 1:253-329.
- (1967) *Science and Metaphysics*, Routledge and Kegan Paul, Londres, 1967.

Shannon, Claude E.

- (1948) 'A Mathematical Theory of Communication', *Bell System Technical Journal*, (octubre) 1948, 27(3): 379-423, 623-56.

Shea, Nicholas

- (2018) *Representation in Cognitive Science*, Oxford University Press, Oxford, 2018.

Shoemaker, Sydney

- (1980) 'Causality and Properties', *Time and Cause*, P. van Inwagen (ed.), D. Reidel Publishing Company, Dordrecht, 1980.
- (1982) 'The Inverted Spectrum', *The Journal of Philosophy*, 1982, 79(7): 357-381.
- (1991) 'Qualia and Consciousness', *Mind*, 1991, 100: 507-524.
- (1994a) 'Self-Knowledge and Inner Sense; Lecture III: The Phenomenal Character of Experience', *Philosophy and Phenomenological Research*, (junio) 1994, 54(2): 291-314.

- (1994b) 'Phenomenal Character', *Nous*, 1994, 28: 21-38.
- (1998) 'Causal and Metaphysical Necessity,' *Pacific Philosophical Quarterly*, 1998, 79: 59-77.
- (1999) 'On David Chalmers' The Conscious Mind', *Philosophy and Phenomenological Research*, 1999, 59: 539-44.

Siegel, Susanna

- (2010) *The Contents of Visual Experience*, Oxford University Press, Oxford, 2010.

Siewert, Charles

- (1998) *The Significance of Consciousness*, Princeton University Press, Princeton, 1998.
- (2011) 'Phenomenal Thought', *Cognitive Phenomenology*, T. Bayne & M. Montague (eds.), Oxford University Press, 2011.

Skyrms, Brian

- (2010) *Signals: Evolution, Learning and Information*, Oxford University Press, New York, 2010.

Snowdon, Paul F.

- (1980/81) 'Perception, Vision and Causation', *Proceedings of the Aristotelian Society*, 1980/81, 81: 175-92.
- (1990) 'The Objects of Perceptual Experience', *Proceedings of the Aristotelian Society*, 1990, 64: 121-150.
- (2005) 'Some Reflections on an Argument from Hallucination', *Philosophical Topics*, 2005, 33(1): 285-305.

Sober, Elliott

- (1984) *The Nature of Selection*, University of Chicago Press, Chicago, 1984.

Sommerhoff, Gerd

- (1950) *Analytical Biology*, Oxford University Press, Oxford, 1950, 62.
- (1959) 'The abstract characteristics of living organisms', *Systems Thinking*, F. E. Emery (ed.), Harmondsworth, Londres, 1959.

Soteriou, Matthew

- (2009/20) 'The Disjunctive Theory of Perception', *The Stanford Encyclopedia of*

Philosophy (Summer 2020 Edition), E. N. Zalta (ed.), 2020.

URL = <<https://plato.stanford.edu/archives/sum2020/entries/perception-disjunctive/>>.

(2013) *The Mind's Construction: The Ontology of Mind and Mental Action*, Oxford University Press, Oxford, 2013.

Speaks, Jeff

(2010/19) 'Theories of Meaning', *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), E. N. Zalta (ed.), 2019.

URL = <<https://plato.stanford.edu/archives/win2019/entries/meaning/>>.

Sperling, George

(1960) 'The information available in brief visual presentation', *Psychological Monographs: General and Applied*, 1960, 74(11):1-29.

Spinoza, Baruch

(1677) *Ethics*, T. Sharpe (ed.).

URL = <<http://www.gutenberg.org/files/3800/3800-h/3800-h.htm>>.

Stalnaker, Robert

(1984) *Inquiry*, MIT Press, Cambridge, MA, 1984.

(1999) *Context and Content*, Oxford University Press, Oxford, 1999.

Stampe, Dennis W.

(1977) 'Toward a Causal Theory of Linguistic Representation', *Midwest Studies in Philosophy*, P. French, H. K. Wettstein & T. E. Uehling (eds.), University of Minnesota Press, Minneapolis, 1977, 2: 42-63.

Stanford, P. Kyle

(2009/17) 'Underdetermination of Scientific Theory', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), E. N. Zalta (ed.), 2017.

URL = <<https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/>>.

Steiner, Pierre

(2014) 'Enacting anti-representationalism. The scope and the limits of enactive critiques of representationalism', *Avant: Trends in Interdisciplinary Studies*,

2014, 2: 43-86.

Sterelny, K,

(1990) *The Representational Theory of Mind: An Introduction*, Blackwell, Cambridge, MA, 1990.

Stoljar, Daniel

(2001) 'Two concepts of the Physical', *There's Something about Mary: Essays on Frank Jackson's Knowledge Argument*, Y. Nagasawa, P. Ludlow & D. Stoljar (eds.), MIT Press, 2004, 309-331.

(2001/15) 'Physicalism', *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), E. N. Zalta (ed.), 2017.

URL = <<http://plato.stanford.edu/archives/spr2015/entries/physicalism/>>.

(2010) *Physicalism*, Routledge, New York, 2010.

Strawson, Galen J.

(1994) *Mental Reality*, MIT Press, Cambridge, MA, 1994.

(2000) 'Realistic materialist monism', *Toward a Science of Consciousness III*, S. Hameroff, A. Kaszniak & D. Chalmers (eds.), MIT Press, Cambridge, MA, 2000.

(2004) 'Real intentionality', *Phenomenology and the Cognitive Sciences*, 2004, 3(3): 287-313.

(2011) 'Cognitive phenomenology: real life', *Cognitive Phenomenology*, T. Bayne & M. Montague (eds.), Oxford University Press, 2011.

Stubenberg, Leopold

(2005/16) 'Neutral Monism', *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. N. Zalta (ed.), 2016. URL = <<https://plato.stanford.edu/archives/win2016/entries/neutral-monism/>>.

Sturgeon, Scott

(2006) 'Reflective Disjunctivism', *Supplement to the Proceedings of the Aristotelian Society*, 2006, 80: 185-216.

(2008) 'Disjunctivism About Visual Experience', *Disjunctivism: Perception, Action, Knowledge*, A. Haddock & F. MacPherson (eds.), Oxford

University Press, Oxford, 2008, 112-143.

Suárez, Mauricio

(2003) 'Scientific Representation: Against Similarity and Isomorphism',
International Studies in the Philosophy of Science, 2003, 17: 225-244.

Tartaglia, James

(2013) 'Conceptualizing physical consciousness', *Philosophical Psychology*, 2013,
26 (6): 817-838.

Thau, Michael

(2002) *Consciousness and Cognition*, Oxford University Press, Oxford, 2002.

Thomas, Michael S. C., & McClelland, James L.

(2008) 'Connectionist Models of Cognition', *The Cambridge Handbook of
Computational Psychology*, R. Sun (ed.), Cambridge University Press,
Cambridge, 2008, pp. 23-58.

Travis, Charles

(2004) 'The Silence of the Senses', *Mind*, 2004, 113(449): 57-94.

(2013) *Perception: Essays After Frege*, Oxford University Press, Oxford, 2013.

Tye, Michael

(1992) 'Visual Qualia and Visual Content', *The Contents of Experience*, T. Crane
(ed.), Cambridge University Press, Cambridge, 1992, 158-76.

(1994) 'Qualia, Content, and the Inverted Spectrum', *Noûs*, 1994, 28: 159-83.

(1995) *Ten Problems of Consciousness: A Representational Theory of the
Phenomenal Mind*, MIT Press, Cambridge, MA, 1995.

(1997/2017) 'Qualia', *The Stanford Encyclopedia of Philosophy* (Summer 2018
Edition), E. N. Zalta (ed.), 2018.

URL = <<https://plato.stanford.edu/archives/sum2018/entries/qualia/>>.

(1998) 'Inverted Earth, Swampman, and Representationism', *Noûs* 32, 1998, S12:
459-477.

(2000) *Consciousness, Color, and Content*, MIT Press, Cambridge, MA, 2000.

Usher, Marius

(2001) 'A Statistical Referential Theory of Content: Using Information Theory to

Account for Misrepresentation’, *Mind and Language*, 2001, 16: 311-334.

Van Gulick, Robert

(2004/14) ‘Consciousness’, *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), E. N. Zalta (ed.), 2018. URL = [<https://plato.stanford.edu/archives/spr2018/entries/consciousness/>](https://plato.stanford.edu/archives/spr2018/entries/consciousness/).

Varela, Francisco J., Thompson, Evan, & Rosch, Eleanor

(1991) *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA, 1991.

Walsh, Denis M.

(1996) ‘Fitness and function’, *British Journal for the Philosophy of Science*, 47: 553-574.

Walsh, Denis M., & Ariew, André

(1996) ‘A taxonomy of functions’, *Canadian Journal of Philosophy*, 26: 493-514.

Watkins, Michael

(2005) ‘Seeing Red: the Metaphysics of Colours Without the Physics’, *Australasian Journal of Philosophy*, 2005, 83(1): 33-52.

Wittgenstein, Ludwig

(1953) *The Philosophical Investigations*, Basil Blackwell, Oxford, 1986.

Woods, William A.

(1981) ‘Procedural Semantics as a Theory of Meaning’, *Elements of Discourse Understanding*, A. Joshi, B. Webber & I. Sag (eds.), Cambridge University Press, Cambridge, 1981.

Walsh, Denis.

(2012) ‘Mechanism and Purpose: A Case for Natural Teleology’, *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012, 43: 173-81.

White, Stephen

(1982) ‘Partial Character and the Language of Thought’, *Pacific Philosophical Quarterly*, 1982, 63: 347-365.

Williams, George C.

(1966) *Adaptation and Natural Selection: A Critique of Some Current Evolutionary*

Thought, Princeton University Press, Princeton, 1966.

Wilson, Jessica

(2006) 'On Characterizing the Physical', *Philosophical Studies*, 2006, 131: 61-99.

Woodfield, Andrew

(1976) *Teleology*, Cambridge University Press, New York, 1976.

Wouters, Arno G.

(1995) 'Viability Explanation', *Biology & Philosophy*, 1995, 10: 435-57.

(2005) 'The Function Debate in Philosophy', *Acta Biotheoretica*, 2005, 53(2):
123-51.

(2007) 'Design Explanation: Determining the Constraints on What Can Be Alive',
Erkenntnis, 1995, 67: 65-80.

Wright, Larry

(1973) 'Functions', *Philosophical Review*, 1973, 82: 139-68.

(1976) *Teleological Explanations: An Etiological Analysis of Goals and Functions*,
University of California Press, Berkeley, 1976.

Yablo, Stephen

(1992) 'Mental Causation', *The Philosophical Review*, 1992, 101: 245-280.