

El "problema" de la consciencia

La prestidigitación explicada

Autor:

Zárate, Alejandro C.

Tutor:

Pedace, Karina S.

2017

Tesis presentada con el fin de cumplimentar con los requisitos finales para la obtención del título Doctor de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires en Filosofía

Posgrado

EL “PROBLEMA” DE LA CONSCIENCIA: La prestidigitación explicada

TESIS DOCTORAL

Lic. Alejandro Zárate

Directora: Dra. Karina Pedace

Consejera: Dra. Diana Pérez



FILO:UBA
Facultad de Filosofía y Letras

Contenido

INTRODUCCION	5
1. ¿Cuál es el aspecto problemático de la consciencia?	8
2. Hacia una (di)solución del problema de la consciencia	23
CAPÍTULO 1	29
DUPLICADOS ZOMBIS: ¿COMO HERMANOS GEMELOS?	29
1. Introducción	29
2. Filosofía del más allá: el argumento de los zombis.....	31
3. En contra de la concebibilidad de los zombis: round 1.....	41
4. En contra de la concebibilidad de los zombis: round 2.....	52
5. Conclusiones	57
CAPÍTULO 2	59
LA SENSACIÓN DE UN ABISMO INFRANQUEABLE	59
1. Introducción	59
2. El argumento del hiato explicativo	60
3. Objeciones en contra del argumento del hiato explicativo	69
4. Conclusiones	83
CAPÍTULO 3	86
FUNCIONES CANCELADAS EN EL TEATRO CARTESIANO	86
1. Introducción	86
2. Una perspectiva cartesiana de la mente y sus motivaciones.	87
3. 1. Algunos dolores de cabeza para la mente cartesiana.....	99
3.2. Problemas del más allá para la mente cartesiana.	104
3.3. La mente cartesiana y el lenguaje privado.	104
4. Una alternativa expresivista para la asimetría entre la primera persona y la tercera y en contra del carácter “epistémicamente privado” de la experiencia.	105
5. Algunas dificultades potenciales para el expresivismo.....	114
6. Conclusiones	121
CAPÍTULO 4	124
DECIME QUÉ SE SIENTE	124
1. Introducción	124
2. Precisiones en torno a la tesis de la inefabilidad de la experiencia.	126
2.1. La inefabilidad y el lenguaje privado.	132
3. El argumento del conocimiento.	143

3.1. El argumento del conocimiento y el externismo social.	146
3.2. La réplica de Dennett al argumento del conocimiento.....	156
3.3. El argumento del conocimiento recargado.	157
4. Conclusiones	160
CAPÍTULO 5	164
YO, CEREBRO.....	164
1. Introducción	164
2. Los problemas de adoptar la tesis del cerebro creador.....	166
3. Las motivaciones detrás del fantasma cartesiano	173
4. Conclusiones	186
CONCLUSIÓN.....	189
La prestidigitación explicada: desde la (di)solución del “problema” de la consciencia hacia nuevos horizontes	189
1. Introducción	189
2. ¿La consciencia? ¿Qué es eso? La crítica de Hacker	190
3. La respuesta de Janzen a Hacker	199
4. La propuesta de Janzen: ¿la consciencia explicada?.....	204
5. ¡No hay nada como ser un murciélago!.....	212
6. La prestidigitación explicada: desde la (di) solución del “problema” de la consciencia hacia la apertura de nuevos horizontes.....	214
BIBLIOGRAFÍA.....	227

Agradecimientos

Sin dudas, este escrito, que contiene muchos años de laboriosa indagación filosófica, no hubiera sido posible sin el estímulo intelectual, el afecto, y el apoyo de mi directora de tesis y amiga, Karina Pedace. Su confianza en mis esfuerzos, su orientación decidida, y sus comentarios minuciosos y constructivos, fueron imprescindibles para mejorar mis planteos iniciales y para la culminación de esta tesis.

Agradezco también a Diana Pérez, experta en los temas que se tratan en esta tesis, el apoyo y su solícita disponibilidad para discutir muchas de las ideas que aparecen en este escrito, con la esperanza de haber podido capitalizar no sólo sus comentarios, sino también la siempre estimulante e iluminadora lectura de sus originales textos sobre el tema del que aquí me ocupo.

No quiero dejar de mencionar en esta instancia mi gratitud para con la educación pública, en la que me formé en todos sus niveles, en tiempos, como los actuales, en los que la necesidad e importancia para los destinos de un país de fomentar la investigación nacional a partir de políticas estatales acordes parece ser puesta en cuestión, a contramano de la más que probada calidad de los investigadores locales y la de sus proyectos. En mi caso, la culminación de este trabajo no hubiera sido posible sin el sustento económico que me brindó el otorgamiento de una Beca Doctoral UBACYT, precisamente en el marco de las políticas de desarrollo de la investigación y formación de recursos humanos idóneos de la Universidad de Buenos Aires y su Secretaría de Ciencia y Técnica.

Quiero agradecer, en particular, a mis padres y a mis hermanos, y a mi familia toda, la de siempre y la nueva, por su paciencia, comprensión y acompañamiento incondicional durante este trabajoso ciclo que culmina con esta tesis. Y a Lucía, mi compañera incesante, por su aliento, su comprensión y su amor ilimitados durante todo el esfuerzo vertido en las páginas que siguen.

INTRODUCCION

En la actualidad, asistimos a una época de creciente efervescencia en torno al cerebro. Su estudio, se supone, es la clave para elucidar cuestiones relevantes en torno a los temas más variados: desde las capacidades cognitivas y las emociones, hasta el arte y la educación. Incluso, algunas publicaciones recientes explican qué aspectos del cerebro es necesario conocer para ser un buen presidente de la República. En este contexto, gran parte de la opinión especializada avizora que la consciencia, ese santo grial de filósofos y científicos, tendrá pronto una explicación neurológica. Sin embargo, como señala el filósofo Alva Noë (2010), tras varias décadas de esfuerzo común por parte de neurocientíficos, psicólogos y filósofos, solo una de las propuestas sobre cómo el cerebro nos hace conscientes parece no haber sido refutada: que no tenemos ni idea de cómo es que esto ocurre. ¿Cómo es posible que el agregado de miles y miles de millones de células de tipos diferentes que nos conforman y cuyo grado de consciencia individual no supera al de una levadura, genere, o de lugar a, la experiencia subjetiva? Hasta los entusiastas de la neurociencia de la consciencia admiten que, hoy por hoy, nadie tiene una explicación plausible de cómo la experiencia -¡la sensación de rojo!- brota de la acción del cerebro. A pesar de toda la tecnología y los experimentos con animales, no es claro que estemos más cerca comprender este fenómeno. Por este motivo, la consciencia suele presentarse por parte de filósofos y científicos cognitivos de diferentes corrientes y filiaciones teóricas, como un fenómeno que constituye “el mayor obstáculo pendiente en nuestra búsqueda de una comprensión científica del universo” (Chalmers 1999), un misterio, es decir, “un fenómeno para el cual todavía no hallamos una forma de pensar” (Dennett 1995).

En efecto, si bien se reconoce, por ejemplo, que la ciencia física o la biología son aún incompletas y que no todas las preguntas en esos campos cuentan en la actualidad con una respuesta acabada, se considera que tenemos de ellas una comprensión más que razonable y una buena idea de la forma que podrían tener las soluciones a esas preguntas pendientes. Por el contrario, la respuesta a la pregunta de cómo es posible que la consciencia surja, con sus características específicas, de la opaca materia gris del cerebro, aparece en principio como una cuestión

cualitativamente distinta a las preguntas que se plantean en otros campos científicos, configurándose así como un enigma desconcertante para la ciencia cognitiva y la neurociencia. En ocasiones, se compara a la consciencia con un espectáculo de magia, una serie interminable de efectos especiales que desafían toda explicación racional. De hecho, todavía se cita con aprobación la frase del biólogo británico del siglo XIX, T.H. Huxley, que sostiene que el que algo tan notable como un estado de consciencia se produzca como resultado de la irritación del tejido nervioso es tan inexplicable como la aparición del genio cuando Aladino frotó su lámpara en el cuento.

Por mi parte, creo que, tal como sugiere Daniel Dennett, de lo que se trata cuando se aborda el “misterio” de la consciencia desde un punto de vista filosófico es, en gran medida, de explicar la “magia” que la rodea. En esta explicación, sin embargo, no debe esperarse que nos encontremos con una indagación sobre milagros, taumaturgias y poderes sobrenaturales. De lo que se trata es de explicar la magia *de verdad*, es decir, aquella que se logra con un conjunto de trucos y pases de manos. Porque, después de todo, ¿qué es la magia sino la creación de creencias falsas sostenidas con sinceridad respecto de haber presenciado algo maravilloso? El mago no corta a la mujer en dos de verdad, sino que nos hace creer que lo vimos hacer eso.

Con respecto a la consciencia, gran parte de la dificultad para explicarla radica en la existencia de “fuerzas poderosas” que nos hacen creer que es más maravillosa de lo que en realidad es. En ocasiones, el truco que habrá que desentrañar está constituido por una adhesión más o menos implícita a ciertas tradiciones filosóficas anquilosadas, que invocan nociones espurias para ejercer su arte. Ahora bien, el problema es que, como la mayoría de los buenos trucos, termina antes de que el público se dé cuenta de que había empezado. En efecto, se hace referencia a la consciencia inocentemente como “lo mejor conocido”, como si fuera un objeto o un proceso interno y privado, o como si realmente supiéramos con claridad a qué nos referimos cuando la consideramos como un residuo que permanece inexplicado una vez despejados otros fenómenos concomitantes.

Pero entonces, como sostiene Wittgenstein, el truco del prestidigitador ya fue llevado a cabo. Por eso, se hace necesaria una elucidación conceptual, junto con cierta reflexión acerca de la auténtica significación de ciertos resultados empíricos,

que logre ensanchar nuestra imaginación para evitar ser embaucados. Y este es el desafío que me propuse afrontar en este escrito.

Por lo pronto, en lo que sigue inmediatamente, buscaré, en primer lugar, aproximarme a una caracterización del fenómeno problemático tal como se lo discute dentro de la tradición de la filosofía analítica contemporánea. Sin embargo, la relevancia de las discusiones que serán llevadas adelante en este trabajo no queda circunscripta al ámbito de la filosofía porque, como se verá, buena parte de la ciencia cognitiva contemporánea comparte muchos de los supuestos y compromisos conceptuales controvertidos de los filósofos, e incluso adopta una porción de su terminología.

Con respecto a la delimitación del fenómeno en cuestión, se comprobará que el término “consciencia” es polimórfico y se dice de muchas maneras. Sin embargo, no todos los aspectos a los que se hace referencia con él se consideran especialmente problemáticos. Por este motivo, se hace necesario deslindarlo de otros fenómenos cercanos. Luego de esta presentación, voy a enunciar explícitamente dónde radica o por qué surge, a mi juicio, el “problema” de la consciencia.

Básicamente, mi hipótesis es que las dificultades emergen porque se le atribuyen a la consciencia, de forma más o menos explícita, propiedades que resultan incompatibles con las propiedades admitidas por una perspectiva “naturalista” sobre el ámbito de lo mental. Esta apuesta, expresada con esta contundencia, no resulta comúnmente suscripta por quienes participan en las controversias en torno a la consciencia. O, por lo menos, no de forma manifiesta. Por eso, en la última sección de esta introducción ofreceré distintas razones para justificar la adopción de esta propuesta, para mostrar las ventajas que conlleva y para señalar dónde radica su plausibilidad.

Por último, adelantaré de qué manera me propongo resolver (o disolver) el problema (o los problemas) en torno a la consciencia a lo largo de los distintos capítulos y cuáles son mis tesis centrales al respecto. Así, como mostraré a lo largo de este trabajo, a pesar de las sostenidas y a menudo bizarras discusiones que se han desarrollado durante mucho tiempo dentro de la filosofía de la mente, lo que podríamos llamar “el problema” de la consciencia es, en realidad, el efecto de un persistente truco de prestidigitación que puede desenmascarse una vez que se

indaga en los bastidores (conceptuales y filosóficos) que subyacen detrás del “espectáculo”.

1. ¿Cuál es el aspecto problemático de la consciencia?

Para aproximarnos a una caracterización de lo que será el objeto de estudio del presente trabajo, podemos comenzar utilizando la distinción que popularizó el filósofo David Chalmers (1999, 2010), de acuerdo con la cual, existen dos tipos de problemas asociados al fenómeno de la consciencia: los problemas “fáciles” y el problema “difícil”. Entre estos problemas fáciles se encontrarían, entre otros, la explicación de las diferencias entre el estado de vigilia y el sueño, la habilidad, por parte de un organismo o individuo de discriminar, categorizar, y reaccionar ante los diversos estímulos ambientales, de reportar los propios estados mentales, de acceder a sus propios estados mentales, de dirigir el foco de la atención a algún evento interno o externo, de controlar de manera racional la propia conducta en base a la información disponible, entre otros.

En principio, todos estos fenómenos están asociados de algún modo con el concepto de consciencia. Así, por ejemplo, se dice que una persona es consciente de algo cuando atiende a algún objeto o algún acontecimiento de su entorno, o cuando puede reportar aquello que percibe o sus propios estados mentales. En ocasiones, se dice también que una persona es consciente de alguna información cuando tiene la capacidad de reaccionar de modo apropiado en base a esa información, o cuando la utiliza para guiar su conducta de un modo racional y más o menos sofisticado. En otros casos, decimos que una acción es consciente cuando es deliberada, y otras veces se dice que un organismo es consciente como un modo de decir que está despierto.¹

¹ En líneas generales, podríamos decir que estos distintos fenómenos relacionados o identificados con la consciencia, pueden ser catalogados o bien como parte de lo que se llama consciencia *transitiva*, o bien como parte de la consciencia *intransitiva*, de acuerdo con la distinción de Norman Malcom (1984). En efecto, la primera hace referencia al hecho de que una criatura esté simplemente despierta o a aquello que una persona o animal puede perder por un desmayo, por estar anestesiado o en estado de coma y posteriormente recobrar. Por su parte, la consciencia *intransitiva* hace referencia a los casos en que se dice que se es actualmente consciente *de algo* o *de que* algo es de tal o cual manera, como cuando nuestra atención permanece dirigida hacia algún objeto ubicado dentro de nuestro campo perceptivo o como cuando nos percatamos de que no se están cumpliendo las metas inflacionarias proyectadas para el año en curso.

Sin embargo, el problema verdaderamente difícil que se enfrenta a la hora de intentar explicar la consciencia en términos compatibles con el conocimiento científico contemporáneo y con sus compromisos filosóficos y ontológicos, surge cuando se pretende dar cuenta del llamado “aspecto subjetivo” o “cualitativo” de la experiencia, exhibido paradigmáticamente por las sensaciones corporales, las percepciones, los estados de ánimo y las emociones. En general, se suele hacer referencia a este aspecto de lo mental con expresiones como “*qualia*”, “consciencia fenoménica” o “cómo se siente (o cómo es) para un organismo, estar en un determinado estado mental” (de acuerdo con el célebre y extravagante giro del filósofo Thomas Nagel (1974/2003)). Precisamente, este aspecto será objeto de análisis del presente trabajo.

Ahora bien, ¿qué es lo que hace que los problemas fáciles sean fáciles y el problema difícil sea difícil, de acuerdo con esta clasificación? En primer lugar, debe aclararse que, en este contexto, se considera que los problemas “fáciles” con fáciles en relación con “el” problema difícil, pero se reconoce que no tenemos todavía ni siquiera algo cercano a una explicación completa de todos estos fenómenos, y que alcanzarla podría llevarnos algunos siglos. Sin embargo, el punto es que, en estos casos, contaríamos con una idea más o menos clara acerca de cómo podrían explicarse estos fenómenos sin que haya que transgredir o modificar los métodos estándar vigentes en las ciencias cognitivas contemporáneas. Y esto es así porque todo lo que demandan los problemas fáciles es una explicación de ciertas *habilidades* y *funciones* cognitivas, y los métodos de la ciencia cognitiva resultan perfectamente adecuados para afrontar este tipo de cuestiones. En este sentido, podríamos decir también que los fenómenos involucrados en los problemas fáciles resultan perfectamente compatibles con una perspectiva funcionalista de la mente. Veamos esta cuestión con mayor detalle.

Históricamente, el funcionalismo surgió en la segunda mitad del siglo XX como una perspectiva más adecuada que sus antecesoras² para abordar el problema mente-cuerpo desde un espíritu naturalista, y a la vera de una analogía de acuerdo con la cual la mente es al cuerpo humano o (más específicamente) al cerebro, lo que el *software* o los programas de computadora son al *hardware* de una computadora.

² Estoy pensando fundamentalmente en el conductismo (en sus diferentes variantes) y la teoría de la identidad.

Desde esta perspectiva, que es una de las más extendidas en la actualidad y entre quienes se ocupan de la consciencia, la relación entre lo mental y lo físico es vista como una relación entre un programa de computación y su implementación o realización física.

A partir de esta metáfora, se lograrían varias cosas al mismo tiempo. En primer lugar, permitiría otorgar a la psicología un dominio propio de objetos y capacidades a estudiar independiente y autónomo del estudio de las complejidades que rigen el funcionamiento del cerebro (o de los estudios del *hardware*, según la metáfora computacional). En segundo lugar, la mencionada analogía permitiría dar cuenta en forma naturalista de uno de los rasgos que es considerado como característico y distintivo del ámbito mental: el aspecto intencional/representacional. En efecto, en líneas generales, podría decirse que una de las principales nociones a las que se apela en la filosofía analítica contemporánea para dar cuenta de la “marca” de lo mental, es decir, de aquello que se supone que lo separa de lo “físico”, es precisamente su carácter intencional o representacional. Desde esta perspectiva, y como ya lo señalara el filósofo y psicólogo Franz Brentano en el siglo XIX, sólo los estados mentales, están *dirigidos* hacia un objeto o son *acerca de* algo, mientras que los fenómenos físicos, carecen de esta direccionalidad. De acuerdo con este planteo, se dice que en toda creencia o en todo pensamiento hay algo creído o pensado, que en todo temor hay algo temido, y que en todo deseo hay algo deseado.

Ahora bien, siempre había resultado un grave problema el de darle lugar a la mente en el mundo natural (es decir, el mundo de la ciencias naturales). Sin embargo, en el caso de los ordenadores, parece que comprendemos perfectamente cómo es que una máquina física implementa un programa siguiendo reglas que pueden leerse como refiriendo a entidades en el mundo. En efecto, cuando un ordenador ejecuta una tarea y nos da un “resultado” nosotros somos capaces de interpretar ese resultado de tal manera de otorgarle un significado: las manchas que aparecen en el *display* de una calculadora, o las luces que se encienden en una pantalla de una Notebook, no son sólo manchas físicas, sino que, en principio, portan un contenido, *dicen* algo verdadero o falso acerca del mundo, “están en lugar de” otra cosa, es decir, tienen carácter representacional. De esta manera, concebir a la mente humana como una computadora permite entender cómo una máquina “sintáctica”

(como una máquina de Turing) puede tener también estados con propiedades semánticas.

En última instancia, las tesis centrales del funcionalismo podrían resumirse de la siguiente manera.

1) Los estados mentales son (idénticos a) estados funcionales de los organismos.

2) Los estados funcionales se caracterizan en virtud de su rol dentro de la red causal de la que forman parte. Lo característico del dolor, por ejemplo, es que es el resultado de alguna lesión o traumatismo corporal; que provoca angustia, incomodidad y alguna forma de razonamiento práctico destinado a aliviarlo. Y también da lugar a que una persona se intranquilece, se proteja y prodigue cuidados a la zona afectada. La idea es que todo estado que cumpla exactamente esa función es un dolor. En este punto, el funcionalismo considera que el significado de los términos mentales refiere a cierto tipo de estado relacional, que involucra a los *inputs* perceptuales, a los *outputs* conductuales y a otros estados internos del agente.

3) Cada estado funcional se realiza o implementa en algún estado físico. Pero un mismo estado funcional puede realizarse en diversos estados físicos. Esto es lo que se conoce con el nombre de “realizabilidad múltiple” de los estados funcionales y es lo que hace que el funcionalismo sea una teoría de la mente fisicalista no reductiva. En efecto, en la medida en que la caracterización de los estados mentales es funcional, es posible sostener que las clases naturales de la psicología no son idénticas a las clases naturales de las ciencias naturales.

Ahora bien, a partir de esta breve caracterización del funcionalismo, podríamos concluir que esta propuesta se configura como una posición monista, fisicalista y naturalista. En efecto, se trata de una posición *monista* porque, en contra de una perspectiva dualista que sostiene que existen en el mundo (por lo menos) dos tipos de sustancias o propiedades ontológicamente diversas (las mentales y las físicas), el funcionalismo, si bien define a los estados mentales como estados funcionales, considera que estos estados se realizan o instancian en estados físicos. De este modo, puesto que no propone ningún tipo de dualismo ontológico, se trata de una posición monista. Y es *fisicalista* porque, en última instancia, considera que todo lo que hay es físico o se reduce, de algún modo, a lo físico, o bien puede ser explicado sin tener que recurrir a entidades adicionales a aquellas aceptadas por las

teorías físicas³. Por último, podríamos decir que se trata de una postura “naturalista” porque, por un lado, se inserta cómodamente dentro del credo dominante en la filosofía analítica actual según el cual se considera que no es necesario apelar a ninguna entidad o fuerza sobrenatural para explicar lo que ocurre en nuestro mundo.

La hegemonía de este punto de vista se explica en gran medida por la influencia cultural que tuvo y tiene la teoría de la evolución de Darwin. Después de ella, pierde fuerza la idea del alma humana concebida como un “soplo divino” en un trozo de barro, y nuestra especie comienza a considerarse como una más en la historia de la evolución. De este modo, el “alma humana”, mucho más compleja pero no cualitativamente diversa a la de otros animales, se “integra” dentro del mismo “orden” que otros fenómenos y objetos naturales y resulta pasible de ser estudiada con métodos similares a los de las ciencias naturales. Por lo demás, existe un principio metodológico que favorece a toda posición monista frente a una dualista. Se trata del principio conocido como la “navaja de Occam”, que sugiere preferir, entre dos teorías o explicaciones rivales, aquella que resulte más simple.

De esta manera, aun en el caso de que ambas posiciones (la dualista y la fisicalista) pudieran explicar adecuadamente los fenómenos que pretenden, debería elegirse la postura fisicalista porque postula un único tipo de “sustancias” o propiedades, mientras que el dualismo multiplicaría el número de entidades más allá de lo necesario. Por otra parte, se puede decir que el funcionalismo que predomina en la filosofía analítica contemporánea es naturalista porque frecuentemente sus representantes más conspicuos se embarcan en la difícil empresa de dar cuenta del ámbito de lo mental en términos de propiedades no mentales, es decir, en términos de entidades que las ciencias naturales estén dispuestas a admitir. Como dije, existe cierto consenso entre los filósofos que se ocupan del problema de la consciencia en torno a la idea de que alguna versión de la perspectiva funcionalista constituye la apuesta adecuada para dar cuenta de la totalidad del ámbito de lo mental.⁴

³ En el capítulo 1 ofrezco una caracterización más detallada de la noción de “fisicalismo”.

⁴ De acuerdo con el filósofo David Papineau (2002), el principal y más contundente argumento en favor del fisicalismo es el llamado “argumento causal”. De acuerdo con dicho argumento, muchos efectos cuya causa atribuimos a estados conscientes tienen causas físicas (como por ejemplo ir a buscar algo para beber cuando nos percatamos de que tenemos sed y deseamos saciarla). Resultaría absurdo, de acuerdo con este razonamiento, suponer que estos efectos son causados dos veces o que están sobredeterminados

Sin embargo, la estrategia funcionalista para abordar la realidad mental parece encontrar un límite. En efecto, existen, en principio, ciertos aspectos de lo mental que van más allá de lo representacional y de lo funcional: se trata precisamente del llamado carácter o aspecto cualitativo o fenoménico, es decir, de ciertas “propiedades” que estarían presentes con mayor claridad en los estados perceptivos (como las experiencias visuales de rojo) y en las sensaciones (como los orgasmos y los dolores).

En principio, parece claro que algunos episodios de nuestra vida mental no son “acerca de” nada, es decir, no tienen un carácter intencional o representacional: en efecto, es dudoso pensar que, por ejemplo, un dolor de muelas sea *acerca de* la muela que duele. Por otra parte, se supone que el análisis funcional en términos de roles y papeles causales no permitiría capturar el aspecto cualitativo precisamente porque este aspecto parece ir más allá del rol causal de un estado mental. El dolor, en principio, proporcionaría un ejemplo de este carácter dual de (al menos) algunos estados mentales. En efecto, por un lado, podríamos decir que se trata de un tipo de estado que cumple cierto rol causal, es decir, que es aquel que, por ejemplo, es causado a partir de algún daño que recibe el organismo, que tiende a producir reacciones de aversión y deseos de aliviarlo. Sin embargo, el análisis funcional no parece agotar la caracterización completa del fenómeno en cuestión.

(es decir, causados por alguna causa física y, *además*, por algún estado mental consciente). Por lo tanto, los estados conscientes deben ser idénticos a al menos alguna parte de estas causas físicas. De manera esquemática, podríamos decir que el argumento es el siguiente: (1) Los estados conscientes tienen efectos físicos. (2) Todos los efectos físicos son completamente causados por antecedentes puramente físicos. (A esta premisa Papineau la llama “la tesis de la completitud de la física”. La idea detrás de esta premisa es que cualquier conducta estará siempre completamente causada por cierta contracción de nuestros músculos, a su turno causados por “mensajes” eléctricos que viajan por nuestros nervios, en sí mismos debidos a la actividad física de córtex motor, a su vez causado por la actividad física causada en el córtex sensorial y así). (3) Los efectos físicos de los estados conscientes no siempre están sobredeterminados por causas distintas. De esta última premisa se sigue la verdad del fisicalismo. En efecto, las premisas (1) y (2) nos dicen que ciertos efectos (como por ejemplo nuestras conductas) tienen una causa que es un estado mental consciente y una causa física. La premisa 3 nos dice que no tienen dos causas distintas. La única posibilidad que queda es que los estados conscientes mencionados en la premisa 1 sean idénticos con alguna parte de las causas mencionadas en la premisa 2. Con respecto a la posibilidad de la sobredeterminación causal, si bien puede admitirse que algunos eventos son causados de esa manera, (como la muerte de un hombre al cual simultáneamente le disparan y le cae un rayo, este no parece ser el modelo de la causación mental). Después de todo, la sobre determinación causal implica que si una causa hubiera estado ausente, el resultado habría ocurrido de todos modos gracias a la otra causa. Sin embargo, parece incorrecto decir que una persona se hubiera dirigido hacia la heladera si no hubiera sentido sed porque sus neuronas se dispararon o, a la inversa, que podría haberse dirigido a la heladera incluso si sus neuronas nunca se hubieran disparado.

Lo que resulta omitido en este caso es cierta cualidad displacentera, la experiencia subjetiva de dolor. Y esto es precisamente lo que distingue a los problemas fáciles de “el” problema difícil de la consciencia, porque, en definitiva, todo lo que requiere explicarse en el caso de los fenómenos que constituyen parte de los problemas fáciles es la realización de una función o conjunto de funciones por parte de un sistema físico u organismo, es decir, todo lo que debe explicarse son ciertos roles causales.

En contraposición, el problema difícil es realmente difícil porque no se trata de un problema acerca de la explicación de cómo se realiza cierta función.⁵ Por este motivo, para cualquier explicación en términos de procesos puramente físico/funcionales que se proponga, será posible plantear de manera inteligible las siguientes preguntas: (1) ¿por qué ciertos procesos físicos / funcionales están acompañados por experiencias conscientes y no suceden, antes bien, “en la oscuridad”? O bien, ¿cómo pueden surgir de la opaca materia gris, por ejemplo, las multicolores experiencias visuales? (es decir, ¿por qué hay consciencia antes que nada?), y (2) ¿por qué *estos* procesos están acompañados de *ésta*s experiencias conscientes y no de otras, es decir, por qué, por ejemplo, un determinado proceso da lugar a una experiencia de rojo y no a una experiencia del sonido de una campanada?

⁵ La división entre los problemas fáciles y “el” problema difícil de la consciencia, y los respectivos fenómenos a los que se haría referencia con esta distinción, guarda estrecha relación con otra distinción propuesta por el filósofo Ned Block (1997) entre dos tipos de consciencias cualitativamente diversas: la *consciencia de acceso* y la *consciencia fenoménica*. De acuerdo con esta distinción, se dice que un estado es consciente en el primer sentido si su contenido está dispuesto o listo (*poised*) para su uso libre en el control del pensamiento y la acción. Más específicamente, un estado con cierto contenido es consciente en este sentido si, en virtud de tenerlo, una representación con ese contenido está (1) dispuesta para ser usada libremente como premisa en un razonamiento, (2) dispuesta para ser usada libremente para el control de la acción y, en el caso de los organismos usuarios del lenguaje, (3) para dar origen a un reporte verbal sobre dicho contenido. Con respecto al segundo concepto de consciencia, la consciencia fenoménica, Block sostiene que es imposible definirlo de una manera no circular. Por esta razón, una posibilidad de acercarnos al significado de este concepto es “señalando” los estados mentales que se consideran casos paradigmáticos de estados fenoménicamente conscientes: básicamente, se recurre siempre en este punto a las sensaciones corporales y las experiencias perceptivas. En última instancia, el concepto hace referencia al rasgo problemático de la consciencia que viene siendo objeto de crítica en este trabajo: su aspecto “cualitativo”. Nuevamente, la parte más controversial de este último concepto de consciencia es que referiría a un fenómeno que se distingue de toda capacidad cognitiva que involucre el pensamiento, de todo rasgo intencional (es decir, de la propiedad de un estado de ser *acerca* de algo), y de toda propiedad funcional.

En cambio, si, por ejemplo, deseáramos explicar un fenómeno cognitivo como el aprendizaje, *todo* lo que tendríamos que hacer es explicar cómo un sistema físico realiza una función o una serie de funciones, es decir, cómo cumple o realiza ciertos papeles causales. En este caso particular, lo que deberíamos explicar es el modo en que las capacidades conductuales de un sistema son modificadas a la luz de la información acerca del entorno y cómo la nueva información permite una adaptación de la conducta del sistema a dicho entorno. Ahora bien, una vez que logramos explicar estos hechos, no tiene sentido preguntar “¿por qué este funcionamiento está acompañado de aprendizaje?”, porque ese funcionamiento *es* lo que significa aprender.

Dicho en otras palabras, una vez que se acepta que el aprendizaje no es otra cosa que un proceso que puede descomponerse en una serie de relaciones causales, es imposible que algo pueda instanciar esos roles causales sin aprender. Además, en este contexto se suele suponer que para casi *todo* fenómeno natural por encima del nivel de la física macroscópica, debe existir en principio una explicación reductiva, esto es, una explicación completa en términos de entidades más simples (Chalmers 1999: 71), precisamente porque existe la posibilidad de analizar los fenómenos a explicar funcionalmente, es decir, en términos de roles causales. Así, por ejemplo, para explicar el gen, necesitaríamos especificar el mecanismo que guarda y transmite información hereditaria de una generación a la otra. Si se descubre que el ADN realiza esa función, una vez explicada esa función, habremos explicado el gen. Para explicar la vida, en última instancia necesitamos explicar cómo un sistema se reproduce, se adapta a su entorno, metaboliza, etc. Y lo mismo ocurriría para la mayor parte de los problemas de la ciencia cognitiva.

Con todo, si bien en la actualidad hay un relativo consenso en torno a la idea de que los demás aspectos de la vida mental podrán ser incorporados en una pintura fisicalista del mundo, la resistencia del mentado aspecto subjetivo de lo mental a ser incorporado en ese cuadro condujo a filósofos contemporáneos como David Chalmers (1996/1999) y Frank Jackson (1982/2003) a propugnar por la adopción de una versión epifenomenalista del dualismo de propiedades con respecto a la consciencia. En efecto, si bien hoy por hoy nadie sostiene, tal como lo hacían algunos filósofos modernos, un dualismo de *sustancias* (la *res extensa* y la *res cogitans*), todavía persisten dualistas de *propiedades*, para los cuales, si bien no hay ninguna

sustancia diferente de los cuerpos físicos, ciertas entidades se caracterizarían por poseer, además de determinadas propiedades físicas, otras propiedades especiales no físicas.

Entre los distintos tipos de dualismos podemos mencionar el dualismo interaccionista, de acuerdo con el cual la microfísica no está causalmente cerrada y las propiedades fenoménicas o cualitativas juegan un rol causal que afecta al mundo físico. La idea aquí es que existen dos tipos de propiedades o sustancias ontológicamente diferentes que se influyen causalmente de modo recíproco. Por lejos, una de las mayores objeciones que se le dirigen al dualismo de este tipo es que su adopción supone rechazar una tesis que se considera bien fundada: la clausura causal del mundo físico⁶. De acuerdo con esta tesis, todo evento físico tiene una explicación suficiente en términos estrictamente físicos sin necesidad de apelar a principios, leyes o entidades por fuera de ese ámbito. Por este motivo, resulta más atractivo, a los ojos de autores como Chalmers, el dualismo epifenoménico, que sostiene que las propiedades fenoménicas de la consciencia son ontológicamente distintas de las propiedades físicas pero, a diferencia del dualismo interaccionista, considera que dichas propiedades, si bien surgen de lo físico, no tienen ninguna incidencia causal en él.

Con todo, la dificultad más grande que afronta este tipo de dualismo es que tiene que conceder, en contra de fuertes intuiciones, que ciertos estados mentales no tienen incidencia causal en nuestra conducta. En algunas ocasiones se señala incluso que este tipo de dualismo es de algún modo auto-contradictorio, porque si se comprende el conocimiento como un proceso que involucra algún tipo de relación causal, habría que reconocer que si el dualismo epifenoménico fuera cierto, no

⁶ Con todo, el dualismo interaccionista podría intentar armonizar con la tesis de la clausura causal adoptando la tesis de la sobredeterminación causal, según la cual, las propiedades fenoménicas tendrían cierto poder causal sobre ciertos eventos del mundo físico que, al mismo tiempo, también contarían con causas físicas suficientes e independientes, que bastarían en principio para explicar por sí solas los eventos en cuestión sin que hubiera necesidad de postular la existencia de ninguna propiedad fenoménica. A decir verdad, ninguno de los interlocutores que participan en las discusiones en torno a la consciencia suelen considerar seriamente esta posibilidad. Otros, como David Papineau (2002), consideran que si bien puede admitirse la sobredeterminación causal para algunos eventos (como el de la muerte de un hombre al cual simultáneamente le disparan y le cae un rayo), este no parece ser el modelo que aplica a la causación mental). Después de todo, la sobre determinación causal implica que si una causa hubiera estado ausente, el resultado hubiera ocurrido de todos modos gracias a la otra causa. Sin embargo, prosigue el autor, parece incorrecto decir que una persona se hubiera dirigido hacia la heladera si no hubiera sentido sed porque sus neuronas se dispararon o, a la inversa, que podría haberse dirigido a la heladera incluso si sus neuronas nunca se hubieran disparado.

podríamos conocer nuestros propios estados mentales conscientes, es decir, no podríamos saber que somos conscientes. Sin embargo, este parece ser precisamente el tipo de saber más seguro con el que contamos. Si bien no forma parte de mis objetivos generales, en el capítulo 1 desarrollaré esta última crítica dirigida al dualismo junto con otras diferentes que, en conjunto, lo convierten en una opción poco atractiva.

De todos modos, y a pesar de estas dificultades indicadas, suelen presentarse en este contexto una serie de argumentos típicos para desafiar la verdad del fisicalismo / funcionalismo que fueron y son ampliamente discutidos dentro de la literatura especializada: me refiero fundamentalmente al argumento de la concebibilidad o de los zombis, el argumento del conocimiento, y el argumento del hiato explicativo. El peso que tienen estos argumentos dentro de las discusiones filosóficas en torno a la consciencia fenoménica es considerable, ya que la mayor parte de los autores que pretenden defender una caracterización de la consciencia compatible con el fisicalismo considera que forma parte de su tarea ofrecer alguna respuesta a los mismos. En principio, podríamos presentar estos argumentos esquemáticamente de la siguiente manera.

1. El argumento de la concebibilidad o de los zombis

- 1) Es concebible que haya zombis.
- 2) Si es concebible que haya zombis, es metafísicamente posible que haya zombis.
- 3) Si es metafísicamente posible que haya zombis, entonces la consciencia no es física.
- 4) La consciencia no es física.

2. Argumento explicativo (El problema difícil / el hiato explicativo).

- 1) Las explicaciones físicas solo pueden dar cuenta de estructuras y funciones.
- 2) Explicar la estructura y la función no es suficiente para explicar la consciencia.
- 3) Ninguna explicación física puede dar cuenta de la consciencia.

3. El argumento del conocimiento.

1. Hay verdades acerca de la consciencia que no son deducibles de las verdades físicas.

2. Si hay verdades acerca de la consciencia que no son deducibles de las verdades físicas, entonces el materialismo es falso.
3. El materialismo o fisicalismo es falso.

Como señala Chalmers (2010), todos estos argumentos tienen, si se quiere, una estructura común: establecen algún tipo de “gap” epistémico entre lo “físico” y lo “fenoménico” o “subjetivo” (la imposibilidad de explicar, la posibilidad de concebir o la imposibilidad de deducir) para arribar a un hiato ontológico. Frente a estos argumentos suelen ensayarse diferentes respuestas que, al mismo tiempo, permiten trazar un panorama de las diferentes posturas filosóficas en torno a la consciencia.

De acuerdo con lo que, a partir de las distinciones de Chalmers, se denomina “materialismo⁷ de Tipo A”, no habría en rigor un hiato epistémico entre las verdades físicas y las fenoménicas. En todo caso, podría admitirse que hay un hiato aparente pero fácilmente descartable. Como representantes salientes de este tipo de materialismo, en ocasiones se ubica a Daniel Dennett (1988/2003) y David Lewis 1988. De acuerdo con este tipo de materialismo, no hay un problema adicional una vez que se resuelven todos los llamados “problemas fáciles” (es decir, los problemas vinculados a la explicación de las varias funciones cognitivas y conductuales anteriormente señaladas). De acuerdo con Chalmers, este tipo de materialismo puede tomar la forma de un eliminativismo, sosteniendo que la consciencia fenoménica no existe y que no hay verdades fenoménicas o bien puede sostener, que la consciencia existe pero puede ser definida completamente en términos funcionales o conductuales. En ambos casos, se acepta que somos “conscientes” en algún sentido no problemático.

⁷ Si bien a partir de ahora sigo la terminología de Chalmers y comienzo a hablar de fisicalismo y materialismo como si fueran sinónimos, lo cierto es que hoy en día es comúnmente aceptado que la noción de materia resulta difícil de elucidar. En principio, todas las caracterizaciones que el sentido común nos proporciona de la materia, a saber, que es impenetrable, sólida, opaca, que existe independientemente de nosotros, etcétera, son caracterizaciones que no se condicen con las propiedades que la física contemporánea atribuye a los objetos más básicos del mundo. Además, los físicos de hoy en día no sostienen que los materiales básicos de los que está compuesta la realidad sean “trozos de materia” sino que son partículas subatómicas, como electrones o quarks, que difícilmente puedan ser consideradas como entidades “materiales”. Por estas razones, los filósofos contemporáneos prefieren auto etiquetarse como “fisicalistas”, y reemplazan el slogan “todo lo que hay es material” por el de “todo lo que hay es físico”. Sin embargo, este cambio no los exime del arduo problema de elucidar la idea misma de “entidad física” (Pérez 2002b: 112).

De acuerdo con Chalmers, uno de los argumentos más interesantes de los materialistas de Tipo A es aquel según el cual se sostiene que puede darse una explicación fisicalista de nuestras *creencias* acerca del carácter presuntamente problemático de la consciencia, de las creencias de que hay un *explanandum* adicional o de que la consciencia es algo no físico. Luego, se argumenta que una vez que se han explicado estas creencias ya se ha hecho lo suficiente como para explicar el fenómeno en cuestión o para eliminarlo. Con todo, para Chalmers este tipo de materialismo “niega lo evidente” (esto es, la consciencia en sí misma, el verdadero *explanandum*) y, por lo tanto, debe presentar poderosos argumentos para defender su posición. Con respecto a la última estrategia en particular, el materialismo en cuestión se reduciría a afirmar que explicar nuestras disposiciones para hablar acerca de la consciencia (y cosas por el estilo) explica todo lo que hay explicar. Pero este punto es precisamente el que podría objetarse. Curiosamente, Chalmers ubica en este equipo a Fred Dretske (1995) quien, junto con Michael Tye (1995), desarrollan versiones diferentes de una teoría *representacional* de la consciencia fenoménica, bajo la consigna de que, en definitiva, “todos los hechos mentales [tanto los intencionales como los cualitativos] son hechos representacionales” (Dretske 1995: xiii). De acuerdo con esta perspectiva, se supone que los estados conscientes son estados representacionales (aunque no-conceptuales) y, por este motivo, es posible ofrecer una explicación de los mismos compatible con el funcionalismo /fisicalismo. Por criterios similares, debería incluirse dentro de este grupo a quienes consideran que los estados de consciencia fenoménica son estados representacionales de nivel superior, esto es, estados que son *acerca de* otros estados mentales numéricamente distintos. Sin embargo, considerar a estos enfoques como eliminativistas me parece, en el mejor de los casos, confuso.

En efecto, como explicaré con un poco más de detalle en el capítulo 1, sostener que es posible dar cuenta de la consciencia en términos funcionales / representacionales no implica necesariamente ser un eliminativista con respecto a la consciencia fenoménica ni comprometerse con la tesis de que lo único que habría que explicar son ciertas creencias (equivocadas) acerca de ella; en todo caso, lo que implica es un rechazo de cierta *caracterización* de la misma que la convierte, por definición, en un fenómeno “no funcionalizable”.

Por su parte, el llamado “materialismo tipo C”, reconoce que existe un hiato explicativo entre lo fenoménico y lo físico pero considera que podría cerrarse en principio cuando se disponga de los conocimientos y de los conceptos adecuados. Existen diversas versiones de esta posición. Entre ellas, encontramos la de Thomas Nagel (1974), quien sugiere que nuestra situación con respecto a la explicación de la consciencia es similar a la que padecían los pre-socráticos, quienes no podían comprender cómo la materia podía ser, en realidad, energía. En nuestro caso, lo que no resulta todavía posible comprender es cómo la consciencia podría ser algo físico. Sin embargo, una suerte de revolución “conceptual” proporcionaría, eventualmente, la comprensión necesaria. En una línea similar, Patricia Churchland (1997), sostiene que si no podemos concebir cómo es que la consciencia es en rigor un proceso físico, eso se debe simplemente a una limitación de nuestra parte que el progreso científico podría superar. Collin McGinn (1989/2003), con un talante mucho menos optimista, ha sostenido que el problema es irresoluble debido a la presencia de ciertas profundas limitaciones cognitivas propias de la naturaleza humana. Dentro de la terminología que emplea este filósofo, el problema de la consciencia resulta en realidad un *misterio* para seres como nosotros con *estas* capacidades cognitivas. De acuerdo con esta clasificación, mientras que los problemas son aquellas cuestiones (cotidianas y científicas) que un individuo con determinadas capacidades cognitivas puede en principio responder, los misterios son preguntas y problemas acerca de fenómenos naturales cuya resolución está fuera del alcance de cierto tipo de individuos en virtud de sus limitaciones cognitivas. De este modo, la diferencia entre un problema y un misterio no radica ni en la naturaleza de la pregunta que se formula, ni en la porción del mundo por la cual se interroga, sino que depende exclusivamente de las capacidades cognitivas que de hecho posee un individuo o cierta clase de individuos. Por lo tanto, es una distinción relativa, que variará en diferentes especies biológicas en función de sus capacidades. Ahora bien, en la medida en que la relación entre la consciencia y el mundo físico constituye para McGinn un misterio, no es un problema que podamos ni que debamos intentar resolver.

Por su parte, el problema que ve Chalmers en esta estrategia para lidiar con los problemas de la consciencia es que, en última instancia, colapsa con algunas de las otras posiciones mencionadas anteriormente y, por lo tanto, arrastra también las

objeciones que pudieran plantearse a éstas. En el fondo, esta estrategia desembocará en alguna de estas variantes: (a) o bien aceptará que la consciencia es después de todo un fenómeno “funcionalizable” a pesar de lo que podría parecer en principio debido, quizás, a algún tipo de error conceptual, y entonces esta posición equivaldría a adoptar el materialismo tipo A;(b) o bien sostendrá que hay algo más en el mundo físico que estructuras y dinámicas (lo que podría equivaler a adoptar algún tipo de dualismo), o bien (c) considerará que la verdad del fisicalismo no requiere la derivación de los conceptos fenoménicos o acerca de la experiencia consciente de los conceptos físicos, lo que equivaldría a adoptar alguna forma del materialismo de tipo B.

De acuerdo con este último tipo de materialismo, el de tipo B, se reconoce que hay un hiato epistémico entre las verdades fenoménicas y las verdades físicas, pero esto no implica aceptar que haya, además, un hiato ontológico. La forma más común que adopta este tipo de materialismo considera que los estados fenoménicos pueden ser identificados con ciertos estados físicos o físico/ funcionales. Esta identidad sería análoga en cierta medida a identidades como la del agua y H₂O o la de los genes con el ADN. No se trata de identidades que puedan obtenerse a partir de un mero análisis conceptual de los términos involucrados sino que se descubren empíricamente. Además, el concepto de agua es diferente al concepto de H₂O, pero sin embargo ambos refieren a la misma entidad en el mundo. Con respecto a la consciencia ocurriría algo similar: ciertos conceptos acerca de la consciencia son diferentes a los conceptos físicos o funcionales, pero puede descubrirse empíricamente que refieren a la misma cosa en el mundo. De esta manera podría explicarse la presencia de una brecha explicativa sin comprometerse con el reconocimiento de una brecha ontológica. Con todo, esta posición enfrenta una serie de dificultades.

En primer lugar, en el resto de los casos que se consideran análogos a las identidades psicofísicas, todo lo que se requiere explicar es alguna función. Pero no parece ser esto lo que está en juego en el caso de la consciencia. En segundo lugar, no parece posible concebir un mundo idéntico al nuestro en la totalidad de sus aspectos físicos, pero en el cual no haya agua o genes. Por el contrario, sí resultaría en principio posible concebir un mundo zombi, es decir, un mundo que constituya una réplica del nuestro en la totalidad de sus aspectos físicos pero que carezca o

difiera en sus propiedades fenoménicas. Dadas estas razones, la analogía entre las identidades psicofísicas y otro tipo de identidades parece debilitarse.

Pertenece a este materialismo la llamada “estrategia de los conceptos fenoménicos” (Stoljar 2005), que es la respuesta fisicalista de mayor consenso en la actualidad a la hora de enfrentar los argumentos en favor del dualismo. En definitiva, la idea desde esta propuesta consiste en señalar que dado que es posible acceder a una *misma* propiedad en el mundo a partir de conceptos alternativos irreducibles entre sí -para simplificar, mediante conceptos “físicos” y mediante conceptos fenoménicos, los mencionados argumentos no pueden pretender derivar legítimamente una irreducibilidad ontológica de una irreducibilidad epistémica o conceptual. Sin embargo, ninguna de las variantes que adopta esta estrategia está exenta de dificultades. En otras palabras, la idea es que tendríamos dos maneras de conceptualizar la experiencia consciente. Por un lado, podemos concebir nuestra experiencia consciente en “primera persona”, en términos de su carácter cualitativo o subjetivo, de cómo se siente al tenerla, independientemente de toda relación causal, o de otro tipo, con cualquier otra entidad o proceso físico, (como los estímulos que provocan esas experiencias, las conductas que dichos estados producen y cómo interactúan con otros estados mentales) y, en general, independientemente de toda descripción o explicación en términos “físicos” de ese estado o sus relaciones causales.

Con todo, esta propuesta fue objeto de diversas críticas sumamente plausibles que ponen en duda su viabilidad. Por mi parte, si bien no me ocuparé específicamente de estas discusiones en este trabajo, consignaré brevemente en lo que sigue algunas de ellas. En primer lugar, habría que señalar que no existe un consenso en cuanto a la caracterización definitiva de los conceptos fenoménicos. Así, existen autores que los caracterizan como conceptos demostrativos (Loar 1997 y Levin 2007), otros como constituidos por las mismas propiedades a las que refieren (como es el caso de Papineau 2002 y su teoría citacional de los conceptos fenoménicos). Al mismo tiempo, cada una de las diversas caracterizaciones adolece de problemas internos nada despreciables. Uno de los defectos comunes a varias caracterizaciones tiene que ver con que carecen de condiciones de individuación de grado fino, y si se supone que este es un requisito innegociable para considerar a una representación como un concepto, debería concluirse que los conceptos

fenoménicos no son conceptos en absoluto (Pérez 2011, 2013; Tye 2009). En muchos otros casos, señala Tye 2009, la caracterización ofrecida de los conceptos fenoménicos no explica por qué surge cierta perplejidad (o por qué tenemos la sensación de que hay un vacío explicativo) cuando se nos explica, mediante estos conceptos, que una misma propiedad mental puede ser conceptualizada mediante dos modos de presentación diferentes y por qué no surge una perplejidad equivalente cuando se nos muestran otros casos en los cuales una misma propiedad, que no pertenece al ámbito de lo mental, es conceptualizada de dos maneras diferentes.

En otras palabras, mientras que nos veríamos motivados a preguntarnos “¿Cómo puede este estado (caracterizado fenoménicamente), ser el mismo que este otro, (caracterizado mediante conceptos “físicos”)”, no nos invade una incredulidad semejante que nos motive a preguntarnos, por ejemplo, “¿cómo Héspero puede ser Fósforo?”. En principio, el partidario de los conceptos fenoménicos, en cualquiera de sus variantes, no parece poder ofrecer una respuesta sensata a esta exigencia. Frente a este cuadro, el diagnóstico de Levine 2001, 2007 es que si la pregunta en cuestión causa perplejidad o resulta inteligible es porque en algún lugar se está haciendo referencia a dos propiedades distintas. Por último, algunos autores, como Tye 2009 o Ball 2009, han cuestionado que existan verdaderamente conceptos que cumplan con los variados requisitos que se supone que deberían reunir los conceptos fenoménicos para enfrentar con éxito los diversos argumentos en favor del dualismo.

2. Hacia una (di)solución del problema de la consciencia

Frente a este panorama, de acuerdo con el cual, cualquier perspectiva que se adopte parece conducir a un callejón sin salida, mi punto de partida consiste en sostener, en primer lugar, que el problema central a la hora de dar cuenta de la consciencia tiene su origen en la incompatibilidad de algunas de las propiedades que se le atribuyen más o menos implícitamente a dicho fenómeno, y las propiedades de los objetos y entidades admitidos por el resto de las disciplinas científicas y por una perspectiva fisicalista / funcionalista con respecto a lo mental.⁸

⁸En rigor, estas propiedades constituirían un obstáculo insuperable también para enfoques no necesariamente funcionalistas pero sí naturalistas y no dualistas de la mente, como el relativamente

Por este motivo, me propuse, siguiendo en parte el espíritu del trabajo de otros autores (Dennett 1988, Pérez 2002, Ward 2006), identificar con claridad esas propiedades polémicas. Luego, a lo largo de los diferentes capítulos, ofreceré dos estrategias argumentativas diferentes y complementarias: por un lado, con respecto a algunos de estos rasgos problemáticos mencionados, sostendré que, a pesar de lo que puede parecer a primera vista, la consciencia no los posee. O, al menos, que hay buenas razones para pensar que no los posee. Con respecto a otros de estos rasgos, afirmaré que si bien la consciencia los posee, no representan un auténtico problema después de todo.

Entre estos aspectos mencionados se encuentran aquellos que Daniel Dennett (1988/2003) originalmente le atribuye a los *qualia*. “Qualia” es una palabra latina que se usa en el contexto de la filosofía de la mente para designar ciertas propiedades de ciertos estados mentales. En líneas generales, se supone que son las propiedades que permiten distinguir los estados mentales conscientes de aquellos que no lo son y que se diferencian en principio de otras propiedades de los estados mentales como las propiedades intencionales o representacionales. Entre las propiedades de estas propiedades, se suelen considerar, en la literatura especializada, las siguientes (Pérez 2002, Orlando 1997, Dennett 1988, Ward 2006):

(1) *Intrinsecalidad*: con este rasgo se hace referencia al carácter no relacional de estas propiedades, esto es, se trataría de propiedades que un estado mental tiene “en sí mismo” y no en virtud de su relación con otra cosa, ya sea un objeto del mundo externo o un estado interno del individuo que los porta. Como puede verse, es este rasgo de los estados conscientes el que impediría su identificación funcional. En otras palabras, si hay ciertos estados mentales con propiedades cualitativas así concebidas, entonces la tesis funcionalista de que los estados mentales son idénticos a estados funcionales parece ser, en principio, falsa, porque no podrían ser capturados en términos de interacciones causales entre *inputs*, *outputs* y otros estados mentales.⁹

más reciente “enfoque sensoriomotor” de la percepción desarrollado originalmente en O’Regan y Noë 2001. De hecho, en este influyente artículo seminal, los autores dedican un espacio a negar que la consciencia tenga algunos de los rasgos problemáticos que se mencionarán a continuación al mismo tiempo que reconocen, implícitamente, la incompatibilidad de dichos rasgos con su propuesta.

⁹ En este punto cabe aclarar que, aunque no me adentraré en estas cuestiones, lo cierto es que no existe un consenso establecido en cuanto al modo en que habría que definir la noción de “intrinsecalidad”. De hecho, hay una caracterización de índole epistemológica basada en la idea de

(2) *Acceso privado y directo*: con esta propiedad se hace referencia al carácter privado de nuestro conocimiento acerca de los estados mentales conscientes, es decir, se trataría de estados no accesibles a la observación pública sino únicamente mediante introspección. En última instancia, la idea es que, en sentido estricto, sólo cada uno de nosotros puede *saber* qué pasa por su propia mente. En contraposición, se supone que el resto de las personas sólo puede *conjeturar* o realizar *hipótesis* acerca de la presencia y los rasgos de esos estados basándose en algún tipo de evidencia indirecta, como el comportamiento observable, o algún indicio del entorno. Por otra parte, se supone que dicho acceso es, además, *directo*, en la medida en que el sujeto que los instancia tiene un acceso no inferencial a estas propiedades de sus estados conscientes. En algunos casos, se supone que este acceso es infalible o, al menos, más seguro que el conocimiento que puede tenerse acerca de cualquier otro hecho del mundo.

(3) *Inefabilidad*: con esta propiedad se hace alusión a la idea de que el aspecto cualitativo de la experiencia consciente es inexpresable en el lenguaje público e incomunicable. De esta manera, alguien que no ha experimentado un *quale* determinado, o que no ha atravesado una experiencia determinada, no puede comprender qué es estar en ese estado cualitativo a través de una comunicación verbal por muy detalladas que sean las descripciones que se le ofrezcan y por muy imaginativo que sea. Quien no experimentó jamás una cierta sensación, o no tuvo las experiencias perceptivas adecuadas, en rigor no posee, ni puede poseer, el concepto correspondiente a esa sensación o esa percepción (o, al menos, no puede poseerlo de manera completa). La inefabilidad, junto con el carácter privado del acceso a los estados conscientes, son las dos propiedades frecuentemente atribuidas a la consciencia que mantienen a raya un abordaje científico de la misma, puesto que permiten que se argumente que las descripciones de la mente de tercera persona, del tipo del que da la ciencia, deban en principio dejar siempre algo afuera –a saber, la naturaleza “intrínseca” de la experiencia. Y se supone que esto es así porque si la experiencia consciente es privada, entonces la perspectiva científica no puede

que sólo podemos conocer las propiedades que producen efectos en nosotros, esto es, las propiedades relacionales, y por oposición, las propiedades intrínsecas serían las “cosas en sí mismas” que escapan a nuestro conocimiento, pero que están detrás de todos los fenómenos conocidos. Sin embargo, esta interpretación de la “intrínsecidad” no parece ser adecuada como una propiedad de la consciencia porque se supone que la consciencia es “lo mejor conocido” y no algo que está siempre “más allá” de nuestro conocimiento (Pérez 2002: 78).

alcanzarla, y si es inefable entonces, incluso si la ciencia lograra alcanzarla, no podría describirla en una teoría (Ross 1993: 443).

Dentro de este marco, *sostendré la tesis de que el problema de la consciencia (entendido en el sentido explicitado en este apartado) es, en rigor, un falso problema, porque, por un lado, es posible demostrar que no hay buenas razones para creer que la consciencia exhiba alguna de las características problemáticas arriba mencionadas, como la intrinsecalidad, la privacidad y el acceso directo; por el otro lado, porque otra de esas características, como la inefabilidad, no representa un verdadero inconveniente a pesar de lo que puede parecer a primera vista.* Así, en el capítulo 1 me ocuparé de la intrinsecalidad. En el capítulo 2, en el contexto de una crítica al argumento del hiato explicativo, y con relación al carácter intrínseco o no-relacional de la experiencia consciente, reconstruyo las objeciones a dicho argumento que muestran plausiblemente que aun cuando se acepte que ciertos *conceptos* legítimos que poseemos acerca de la experiencia consciente no la caracterizan en términos de ciertos roles causales o funcionales (es decir, en términos no relacionales), esto no constituye, por sí solo, un impedimento para identificar a la propiedad referida por dichos conceptos con una propiedad físico / funcional.

En el capítulo 3, por su parte, me ocupo de la privacidad y del acceso directo. En el capítulo 4, de la inefabilidad y, en parte, otra vez de la privacidad. Como veremos, gran parte de la tentación de atribuirle estas propiedades a la consciencia se explica por la adopción más o menos tácita de una perspectiva de la mente que podríamos llamar “cartesiana” la cual, además de asumir ciertos compromisos ontológicos y epistemológicos controvertidos y disputables, realiza una interpretación inadecuada de los conceptos psicológicos ordinarios o de sentido común. Por lo demás, como mostraré en el capítulo 5, el fantasma cartesiano (junto con las dificultades que entraña), goza de buena salud no sólo dentro de la filosofía sino también en buena parte de la neurociencia contemporánea. Por eso, en este mismo capítulo me encargaré de desestimar las razones que se esgrimen en dicho ámbito para mantener con vida al fantasma en la máquina.

Por último, en el capítulo 6 muestro, a partir del análisis de la célebre frase que utiliza Nagel (1974) para referirse al aspecto cualitativo de la experiencia consciente que, o bien no resulta inteligible de qué estamos hablando cuando hablamos del “aspecto cualitativo” de la experiencia, o bien que, si resulta inteligible,

no aparece como un aspecto especialmente problemático de nuestra vida mental. Como en el resto del presente trabajo, el propósito general de este capítulo no consiste en negar la realidad de la experiencia consciente, sino en rechazar una caracterización inadecuada de la misma como la que le atribuye de manera más o menos explícita los cuatro rasgos problemáticos mencionados.

De este modo, aunque no voy a comprometerme con ninguna teoría específica acerca de la consciencia, lo cierto es que si mis razonamientos son adecuados, se verá que la empresa de desarrollar una teoría de este tipo desde una perspectiva fisicalista, funcionalista y naturalista, deja de ser una tarea imposible por principio.¹⁰

Ahora bien, en este punto debo admitir que no todos los que participan en las discusiones en torno a la experiencia consciente dentro de la filosofía contemporánea estarán dispuestos a aceptar la estructura de mi planteo. De hecho, no suele ser ésta la manera habitual en la que se configura la discusión de estas cuestiones. Sin embargo, mi propuesta no resulta arbitraria y se justifica por diversas razones. En primer lugar, permite comprender con claridad en qué medida la consciencia es un fenómeno problemático o de difícil armonización con los conocimientos científicos aceptados en la actualidad y las posiciones filosóficas coherentes con ellos. Esta dificultad se hace visible al cotejar la incompatibilidad de los rasgos problemáticos mencionados y las premisas de las que parte toda perspectiva fisicalista / funcionalista / naturalista acerca de lo mental.

En segundo lugar, como me encargaré de mostrar a lo largo de los diversos capítulos, muchos autores adscriben estos rasgos, aunque no siempre de manera explícita, a la experiencia consciente.

En tercer lugar, dichos rasgos parecen estar, de una u otra manera, a la base de los principales argumentos que se esgrimen en la filosofía de la mente actual para evidenciar el carácter problemático de la consciencia: me refiero al argumento de

¹⁰ En este punto cabe aclarar que si bien mi trabajo tiene en buena medida un espíritu disolutivo y terapéutico con respecto al problema del que se ocupa (en el sentido en que se suele decir que las consideraciones de Wittgenstein con respecto a ciertos problemas filosóficos son terapéuticas), no deseo de ningún modo comprometerme con la idea de que *todos* los problemas filosóficos deban ser tratados con el mismo espíritu. De hecho, nada de lo que digo en este trabajo me obliga a aceptar esa posición extrema. Sólo me limito a sostener que este tratamiento es adecuado para *este* “problema” en particular. Una vez disuelto, confío en que se avizore claramente en qué sentido la tarea aludida más arriba dejará de ser, pues, imposible por principio.

los zombis y al argumento del conocimiento. Voy a sustentar esta afirmación en los capítulos 1, 3 y 4. Al mismo tiempo, mi posición frente a estos rasgos problemáticos permite dar respuestas atendibles a los argumentos en cuestión.

Por último, si de hecho se acepta, al menos en principio y de manera intuitiva, que, por diferentes motivos, resulta plausible considerar que la experiencia consciente exhibe las mencionadas propiedades controvertidas, entonces cualquier teoría acerca de la consciencia que se ofrezca dentro del marco del funcionalismo estará condenada inexorablemente al fracaso, porque jamás podrá dar cuenta de estos aspectos.

Por este motivo, resulta perentorio, para cualquier alternativa teórica que se proponga, enfrentarse de alguna manera con estos rasgos si quiere evitar que se la acuse de no “tomarse en serio la consciencia” o de dejar de lado precisamente aquello que debería explicar. Por otra parte, también se podría pensar que si se concluye que hay buenas razones para pensar que estas propiedades no corresponden a ningún aspecto de nuestra vida mental, o que, cuando lo hacen, no resultan verdaderamente problemáticas después de todo, la razón de ser de las teorías que se ocupan del aspecto cualitativo de lo mental, se desvanece.

En cualquier caso, la consideración de estas propiedades y de las observaciones y argumentaciones desarrolladas en este trabajo, resultan prioritarias para cualquier teoría o enfoque de la consciencia que se proponga.

CAPÍTULO 1

DUPLICADOS ZOMBIS: ¿COMO HERMANOS GEMELOS?

1. Introducción

Como se dijo en la introducción de este trabajo, si se considera que la intrinsecidad es una de las propiedades definitorias de la consciencia, este fenómeno se convierte en un verdadero problema para las teorías fisicalistas ortodoxas acerca de la mente. En efecto, si la consciencia cuenta con propiedades intrínsecas, entonces parece imposible que estas propiedades puedan ser explicadas en términos aceptables para el funcionalismo, es decir, en términos de las interacciones causales entre *inputs* (estímulos sensoriales), *outputs* (respuestas conductuales) y otros estados mentales, y en términos de las relaciones causales de éstos últimos con objetos del entorno, porque si hay aspectos intrínsecos en los estados conscientes, no pueden ser reducidos a relaciones de ningún tipo.

Ahora bien, si esto es así, es decir, si existen propiedades intrínsecas de la experiencia que no pueden individuarse por las relaciones que entablan con otros estados mentales, estímulos, conductas y objetos, y por los roles causales que ocupan en esas relaciones, se sigue que dos estados mentales funcionalmente idénticos podrían estar asociados o relacionados con diferentes estados conscientes o aspectos cualitativos de los mismos. De este modo, resultaría en principio metafísicamente posible y concebible que dos personas sean exactamente iguales en su constitución física y funcional, en sus disposiciones para actuar y en sus relaciones con un entorno también idéntico, pero que difieran en sus experiencias conscientes. Así, por ejemplo, cuando una de estas personas dice algo así como “estoy teniendo una sensación de rojo” ante una puesta de sol, mientras tiene efectivamente una sensación de rojo, es posible que la otra, cuando emite la misma

frase o una frase similar en un contexto idéntico o similar, esté teniendo, en realidad, una sensación de verde. En otras palabras, ¿cómo sé que usted y yo vemos el mismo color *subjetivo* cuando miramos algo (o, para el caso, como sé que usted experimenta algún color en absoluto)? Habida cuenta de que aprendimos las palabras para los colores mostrándonos objetos coloreados, nuestras conductas verbales y discriminatorias coincidirán aún en el caso de que experimentemos colores subjetivos completamente distintos.

Esta es la famosa intuición de la “inversión del espectro” propuesta por primera vez por el filósofo John Locke, según la cual, no hay un lazo necesario entre los estados cualitativos y funcionales de la mente, ni con los estados físicos en general, y que parece estar a la base de diferentes argumentos célebres dentro la filosofía de la mente, como por ejemplo el argumento de los *qualia* ausentes o zombis.

En este panorama, me gustaría retomar el espíritu de algunos argumentos que desarrolla el filósofo Michael Tye 2006 para rechazar, entre otras cosas, la idea de que la experiencia consciente tenga propiedades intrínsecas y “más allá” de lo físico. Siguiendo en parte su estrategia, podríamos decir, como primera premisa, lo que ya se dijo en el párrafo anterior: que si la experiencia consciente es intrínseca y no física, debería ser metafísicamente posible y, en principio, concebible, que haya dos individuos idénticos en cuanto a sus propiedades físicas y sus relaciones con un entorno también idéntico, pero que difirieran en sus propiedades cualitativas e intrínsecas.

En otras palabras, si la consciencia contara con propiedades intrínsecas y no físicas, entonces debería ser concebible y metafísicamente posible un “escenario zombi”, es decir, un escenario física y funcionalmente idéntico a nuestro mundo pero en el cual no hay experiencia consciente. Ahora bien, si se acepta este planteo, debería admitirse que si un escenario como el descrito no fuera concebible, entonces se pone en duda que se trate efectivamente de una posibilidad metafísica y, en última instancia, que la experiencia consciente cuente verdaderamente con propiedades intrínsecas y no físicas. Precisamente, mi objetivo en este capítulo consistirá en impugnar esta premisa crucial del argumento de los zombis, esto es, la idea de que estas criaturas sean realmente concebibles.

Para ello, estructuraré mi trabajo de la siguiente manera. En la sección siguiente (2) voy a reconstruir el argumento de los zombis de acuerdo con una de sus versiones más discutidas, la que ofrece el filósofo David Chalmers. Al mismo tiempo, consignaré brevemente las impugnaciones más frecuentes que recibe y señalaré dentro de qué tipo podría ser incripta mi propia impugnación. Por último, en la sección 3 mostraré por qué creo, a pesar de lo que puede parecer a primera vista, que no es claro que la noción de zombi resulte coherente.

2. Filosofía del más allá: el argumento de los zombis

El argumento de los zombis, o de la concebibilidad, es una ficción filosófica diseñada con el objetivo de poner en primer plano ciertos problemas en torno a la conciencia y su incierta conexión con el mundo físico. En particular, pretende constituirse como una amenaza para el fisicalismo, es decir, para la tesis a la que suscribe la mayor parte de los filósofos de la mente y científicos contemporáneos según la cual, todo lo que hay en el mundo es, en última instancia, físico. Así, si el argumento de los zombis resultara exitoso, deberíamos aceptar que la conciencia no forma parte del mundo físico.

Desde el punto de vista histórico, el término ‘zombi’ fue introducido en este tipo de discusiones por Robert Kirk en un artículo de 1974, pero alcanzó su momento de mayor popularidad a fines del siglo XX y principios del siglo XXI a partir del trabajo de David Chalmers 1996. Sin embargo, la estructura del argumento encuentra un antecedente distinguido en el argumento clásico de Descartes a favor de la “distinción real” entre mente y cuerpo. En efecto, en la Sexta Meditación cartesiana encontramos un argumento que, como el de los zombis, está basado en la concebibilidad. En particular, ahí se nos dice que el cuerpo es tal que puede ser concebido como divisible en partes, pero la mente es tal que debe concebirse como indivisible; luego, la mente es distinta del cuerpo.

Por su parte, el argumento de los zombis intenta llegar, a través de medios similares, a una conclusión parecida pero con respecto al estatus ontológico de la conciencia. Ahora bien, para comprender una de las versiones más célebres y discutidas del argumento, la que propone Chalmers, debemos primero introducir algunas nociones que el autor utiliza para desarrollarlo como la de “fisicalismo” y “superveniencia”.

Básicamente, decía, el fisicalismo sostiene que todo lo que existe en el mundo es físico, o que los hechos físicos agotan en cierto sentido todos los hechos acerca del mundo, o que los hechos más básicos acerca del mundo son hechos físicos y que todos los demás hechos dependen de ellos. En este contexto, el concepto filosófico de *superveniencia* intenta formalizar precisamente la idea de que un conjunto de hechos puede determinar por completo otro conjunto de hechos.

Específicamente, la superveniencia es un tipo de relación que existe entre dos conjuntos de propiedades dados. Al primer *relata* lo llamamos *base* de superveniencia, mientras que al segundo lo llamamos *propiedad superveniente*. La relación es tal que, si fijamos el tipo de propiedades de la base superveniente estaremos fijando las propiedades de lo superveniente. Así, de acuerdo con la definición de Chalmers (1999) de la que me voy a servir en este capítulo, “Las propiedades B *supervienen* a las propiedades A si ningún par de situaciones posibles es idéntico respecto de sus propiedades A pero difiere en sus propiedades B”.

Establecida esta definición, debemos reconocer distintos tipos de superveniencia que varían en función del tipo de noción de “posibilidad” que consideremos en juego. Para los fines que persigo, será conveniente distinguir entre dos tipos: la posibilidad lógica y la posibilidad natural. De acuerdo con esta distinción, si hablamos de posibilidad en este mundo, con *estas* leyes naturales que nos rigen, hablamos de superveniencia *natural*; si, en cambio, hablamos de posibilidad en cualquier tipo de mundo, con cualquier tipo de leyes, hablamos de superveniencia *lógica*. Así, cuando hablemos de superveniencia natural estaremos pensando en situaciones naturalmente posibles, es decir, situaciones que no desafían las leyes de la naturaleza que rigen en nuestro mundo; y cuando hablemos de superveniencia lógica estaremos pensando en situaciones lógicamente posibles, es decir, básicamente, en situaciones no contradictorias o incoherentes, pero que pueden ser incompatibles con las leyes naturales conocidas.

Así, por ejemplo, de acuerdo con Chalmers la presión que ejerce un mol de gas depende sistemáticamente de su temperatura y volumen según la ley $pV = KT$. En el mundo real, cuando hay un mol de gas a una temperatura y volumen dados, su presión estará determinada: es empíricamente imposible que dos moles distintos de gas supervengán a su temperatura y volumen. De acuerdo con la definición de superveniencia que vimos, podríamos decir en este caso que la presión superviene

naturalmente a la temperatura y el volumen de un mol de gas, es decir, es natural o empíricamente imposible, dadas las leyes de la naturaleza que rigen en nuestro mundo, que un mol de gas con una temperatura y volumen determinados puedan tener una presión diferente.

Sin embargo, no es lógica o conceptualmente imposible que esto no sea así de la misma manera en que no lo es un universo sin gravedad. En cambio, para Chalmers hay una relación de superveniencia lógica entre los hechos físicos de nuestro mundo y los hechos biológicos, de modo tal que no habría espacio lógico para que haya un mundo físicamente idéntico al nuestro pero que sea biológicamente distinto porque una vez que fijemos todos los hechos físicos acerca del mundo, habremos fijado también “la forma macroscópica de todos los objetos en el mundo, el modo como se mueven y funcionan, el modo como interactúan físicamente” (Chalmers 1999:63). La noción de un mundo físicamente idéntico al nuestro pero biológicamente diferente es, para Chalmers, tan incoherente como la noción de “zorra macho”.

En general, “cuando las propiedades B supervienen lógicamente a las propiedades A, podemos decir que los hechos A implican a los hechos B, donde un hecho implica lógicamente a otro si es lógicamente imposible que el primero sea verdadero y el segundo no” (Chalmers 1999: 63).

Por el contrario, puede considerarse que parte de lo que intentaron probar muchos de los argumentos más discutidos dentro de la filosofía de la mente de los últimos años con respecto a la consciencia (como el “argumento del conocimiento” de Frank Jackson 1982 o el argumento de los zombis), es precisamente que la consciencia no superviene lógicamente a lo físico. El argumento de los zombis en particular, sostiene como una de sus premisas que es posible concebir, sin incurrir en ningún tipo de incoherencia, un “mundo zombi”, es decir, un escenario en el cual las propiedades físicas sean idénticas a las de nuestro mundo pero en el que la experiencia consciente está completamente ausente (o es diferente).

En otras palabras, para los defensores del argumento de los zombis, es posible concebir, por ejemplo, un individuo que exhiba todas las conductas verbales y no verbales asociadas al dolor en nuestro mundo, que se encuentre en el estado neurofisiológico en el que se encuentran los individuos de nuestro mundo cuando sienten dolor, pero que no sienta dolor en lo absoluto. Sin embargo, la posibilidad

del escenario zombi es contraria a la verdad del fisicalismo. En efecto, si utilizáramos la noción de superveniencia para definir al fisicalismo podríamos decir que el fisicalismo es verdadero si todos los hechos acerca del mundo supervienen lógicamente a los hechos físicos o, de manera alternativa, si todos los hechos físicos implican a todos los hechos positivos acerca de nuestro mundo.

Ahora bien, si esto es así, es decir, si el fisicalismo es verdadero, los hechos acerca de la experiencia consciente (en tanto parte del conjunto de los hechos verdaderos acerca de nuestro mundo) deberían estar implicados lógicamente por los hechos físicos, de manera tal que no debería ser posible pensar sin contradicción que los últimos son verdaderos y los primeros no (de la misma manera en que se supone que nos llevaría a una contradicción intentar imaginar un mundo que tuviera las mismas propiedades físicas que el nuestro pero que no instanciara al mismo tiempo las mismas propiedades biológicas de nuestro mundo).

Sin embargo, un escenario de estas características parece, en principio, perfectamente concebible, es decir, parece posible concebir un mundo zombi idéntico al nuestro en lo que respecta a las propiedades físicas pero que carezca de experiencia consciente, sin incurrir en ningún tipo de contradicción. Por lo tanto, el fisicalismo es falso. En base a estas distinciones, podríamos reconstruir el argumento de la siguiente manera.

(1) *Hay experiencia consciente en nuestro mundo.* Esto parece ser un hecho indiscutido. (2) *Los escenarios 'zombis' son concebibles* (es decir, no se incurre en ningún tipo de contradicción al concebir un escenario física y funcionalmente idéntico a nuestro mundo pero en el cual no hay experiencia consciente).

(3) *Si un escenario zombi es concebible, entonces es metafísicamente posible.* (El tipo de posibilidad que está acá en juego no es una posibilidad natural o referente al mundo en el que habitamos con las leyes naturales que lo rigen; sino antes bien, a una posibilidad lógica.

(4) *Si un escenario zombi es metafísicamente posible, el fisicalismo o materialismo es falso.*

En efecto, como vimos, el fisicalismo se compromete con la tesis de que los hechos microfísicos determinan todos los hechos del mundo, es decir, que cualquier mundo que sea exactamente idéntico con el nuestro con respecto a todas sus propiedades físicas (hasta el más mínimo detalle de la partícula más ínfima), debería

ser igual a nuestro mundo en *todos* los aspectos (es decir, debería contener idénticas montañas, lagos, glaciares, árboles, rocas, ciudades y *criaturas conscientes*). Pero si esto es así, contrariamente a la premisa 2, un escenario zombi no debería ser realmente concebible.

En otras palabras, los zombis son, por definición, exactamente iguales a nosotros desde el punto de vista físico (con todo lo que veremos que eso implica). Se supone, además, que habitan en un entorno idéntico al nuestro y mantienen con él el mismo tipo de relaciones. Ahora bien, si dos criaturas idénticas físicamente pueden diferir con respecto a sus propiedades fenoménicas, entonces esto parece indicar que dichas propiedades son algo que está “más allá” del mundo físico, porque si la conciencia fuera idéntica con ciertas propiedades físicas, resultaría imposible para una criatura poseer las mismas propiedades físicas que los seres conscientes pero carecer de conciencia y para nosotros concebirla, a menos que nos veamos envueltos en algún tipo de confusión conceptual.

Antes de avanzar, prestemos atención a las diversas capacidades cognitivas con las que cuentan los zombis ya que esto resultará importante en la sección siguiente. En efecto, a diferencia de los clásicos zombis del cine o de series recientes como *The Walking Dead*, los zombis de los filósofos resultarían mucho más difíciles de detectar, ya que no exhiben necesariamente un rostro pálido o notablemente deteriorado, no se mueven con un andar peculiar y poco elegante, no usan ropa andrajosa ni tiene hábitos caníbales. Como se dijo, un zombi replica de manera exacta la constitución física y el comportamiento de su gemelo consciente “molécula por molécula”. Además, habita en un entorno también idéntico.

En tercer lugar, es un duplicado exacto desde el punto de vista *funcional* y desde el punto de vista *psicológico*. Esto quiere decir que los zombis filosóficos poseerán todas aquellas capacidades y habilidades cognitivas y de cualquier tipo, junto con los estados mentales, que puedan ser identificados cabalmente por sus conexiones causales con los estímulos internos o externos, otros estados mentales y la conducta. Así, por ejemplo, un zombi será capaz de aprender cosas, si por “aprender” se entiende la capacidad de adaptar las conductas de un modo apropiado en respuesta a ciertos tipos de estimulación del entorno. En general, un zombi será capaz de discriminar, categorizar y reaccionar adecuadamente ante los estímulos externos, podrá reportar sus estados mentales (definidos en los términos

antedichos) y podrá usar el acceso a alguna información (ya sea de sus propios estados, o de estado externos) para controlar racional y/o voluntariamente su conducta. Así las cosas, ni nuestras madres podrían distinguirnos a nosotros de nuestros duplicados de ultratumba.

Sin embargo, se supone que cualquiera de estos estados y capacidades carece de esa “cualidad subjetiva” que (presuntamente) acompaña a nuestras percepciones (como aquella que suscita el particular aroma del café recién hecho), a nuestras sensaciones corporales (como la cualidad completamente única que tienen los orgasmos o la sensación de golpearse el nervio en el codo), y a nuestras emociones y pasiones (como el fulgor de un arranque de furia o la melancolía por el paso del tiempo): todo estará silencioso y oscuro dentro de estas criaturas. En otras palabras, aun cuando, luego de lastimarse, por ejemplo, reaccione apropiadamente, tanto desde el punto de vista físico en general como desde el punto de vista conductual en particular, el zombi no *sentirá* dolor realmente.

Con todo, Chalmers considera que el resultado de este argumento no debe conducir necesariamente a la aceptación de un dualismo ontológico de sustancias “*a la*” Descartes, donde se postula un reino de lo mental (una sustancia inmaterial o *res cogitans*) que ejerce algún tipo de influencia causal sobre los fenómenos físicos. Para Chalmers, además de innecesaria, esta opción choca contra una tesis a su juicio bien fundamentada por la ciencia contemporánea a la que se conoce como la tesis de la “clausura causal” del mundo físico. De acuerdo con esta tesis, todo evento físico tiene una explicación suficiente en términos estrictamente físicos sin necesidad de apelar a principios, leyes o entidades por fuera de ese ámbito. En cambio, el autor propone adoptar lo que él llama un dualismo de propiedades *naturalista* que no tendría nada de anti-científico porque, en primer lugar, no supondría la violación de la clausura causal del mundo físico (con lo cual, resultaría compatible con la ciencia contemporánea) y, en segundo lugar, porque admitiría y propiciaría la formulación de leyes psicofísicas que especificarían cómo las propiedades de la experiencia consciente dependen de, o supervienen *naturalmente* a, las propiedades físicas. De este modo, se mantendría, a pesar de lo que puede parecer en principio, el “espíritu” científico que busca explicar todos los hechos mediante leyes y propiedades básicas o fundamentales las cuales, a su vez, no pueden explicarse en términos de propiedades y leyes más básicas. En todo caso, de lo que se trataría es de reconocer

que a las explicaciones basadas en las propiedades y leyes básicas que postula la física (que, por lo demás, permiten explicar casi todos los hechos del mundo), debe complementárselas, si se quiera lograr una explicación verdaderamente total de la realidad, con la postulación de ciertas propiedades adicionales y leyes básicas que están “más allá” de las características y leyes físicas del mundo¹¹. Según Chalmers, “no es necesario que haya nada especialmente trascendente acerca de la consciencia; sólo es *otro* fenómeno natural. Lo único que ocurrió es que nuestra imagen de la naturaleza cambió” (Chalmers 1999:172. La cursiva es mía.).

Frente a esta propuesta, cabe señalar brevemente algunos motivos centrales que existen para rechazar cualquier tipo de dualismo en general y el dualismo de Chalmers en particular. En definitiva, tal como señala Kriegel (2007), el principal motivo que suele postularse para rechazar el dualismo sigue siendo uno similar al que expresó el filósofo J.J.C. Smart hace más de medio siglo: de acuerdo con esta idea, parece que la ciencia nos ofrece, cada vez de manera más acabada, un punto de vista desde el cual los organismos son pasibles de ser vistos como complejos mecanismos físico-químicos. En este contexto, resultaría por lo menos curioso que la consciencia se destacara del resto de las entidades que pueblan la naturaleza.

Por su parte, un dualismo como el de Chalmers en particular, resulta permeable a la siguiente crítica. Si, tal como hace el autor, asume la clausura causal del mundo físico, se enfrenta con un dilema con respecto a las propiedades que no supervienen a las propiedades físicas (es decir, las propiedades fenoménicas o cualitativas de la experiencia): o bien acepta que estas propiedades son causalmente inertes, es decir, que no tienen ningún efecto causal en el mundo físico, o bien acepta la tesis un tanto implausible de la sobre determinación causal, según la cual, las propiedades fenoménicas tendrían cierto poder causal sobre ciertos

¹¹ De hecho, el filósofo australiano sostiene que la necesidad de realizar una ampliación en la ontología aceptada hasta el momento surgió en otras instancias de la historia de la ciencia. De acuerdo con Chalmers, en la física, en el siglo XIX, resultó que los procesos electromagnéticos no podían ser explicados en términos de los procesos mecánicos a los que hasta ese momento las teorías físicas apelaban. Por lo tanto, Maxwell y otros introdujeron las cargas electromagnéticas y las fuerzas electromagnéticas como componentes fundamentales adicionales de la teoría física. Para explicar el electromagnetismo, la ontología de la física tuvo que ser expandida. Nuevas propiedades básicas y nuevas leyes básicas se necesitaron para dar una explicación satisfactoria del fenómeno. De la misma manera, la consciencia debería considerarse como fundamental a juicio de Chalmers en el sentido apuntado, porque la totalidad de la ciencia física es compatible con la ausencia de la consciencia. Ahora bien, en la medida en que esta posición propone agregar nuevas entidades a las entidades postuladas por la física, califica como un tipo de dualismo ontológico.

eventos del mundo físico que, al mismo tiempo, también contarían con causas físicas suficientes e independientes, que bastarían en principio para explicar por sí solas lo eventos en cuestión sin que hubiera necesidad de postular la existencia de ninguna propiedad fenoménica.

Ahora bien, más allá de estas observaciones generales, lo cierto es que la discusión en torno al argumento de los zombis dentro de la filosofía de la mente contemporánea ha sido realmente profusa. Con su estilo habitual, Daniel Dennett (2006) compara estas extendidas y omnipresentes discusiones con el trillado estereotipo de las acaloradas discusiones filosóficas acerca de cuántos ángeles podían bailar en la cabeza de un alfiler. Entre los intentos de bloquear este argumento encontramos principalmente dos opciones. Por un lado, están quienes ponen en duda la concebibilidad del escenario propuesto, es decir, quienes rechazan la premisa (2) del argumento en cuestión. Dentro de este grupo de réplicas encontramos, entre otros, los trabajos de Kirk 2008; Tye 2006; Marcus 2004 y Dennett 1995.

En particular, Dennett sostiene que la noción de “zombi” es en algún sentido auto-contradictoria y que los filósofos que afirman que los zombis son concebibles subestiman la tarea imaginativa que el experimento mental solicita y terminan imaginando algo que viola sutilmente la definición de zombi. Por mi parte, debo confesar que simpatizo con esta idea y, en gran medida, los argumentos que voy a desarrollar en la sección siguiente para justificarla, se basan en los trabajos de los autores anteriormente citados.¹²

Con todo, la táctica más ensayada para resistir el argumento de los zombis es la que busca socavar la legitimidad de la premisa (3), es decir el paso de la concebibilidad a la posibilidad metafísica. Dentro de quienes optan por esta opción, se encuentran los partidarios de una estrategia relativamente reciente: los defensores de la llamada “estrategia de los conceptos fenoménicos” (Stoljar 2005).

¹² Por lo demás, cabe aclarar que determinar si algo resulta concebible o no, no siempre es una tarea que pueda ser cumplida mediante una simple inspección superficial. Una situación, un escenario o un concepto podrían parecer imaginables o concebibles y sin embargo ocultar dentro de sí elementos contradictorios. Un ejemplo de este fenómeno se encuentra en muchos teoremas de la matemática. Antes de la prueba, parece concebible que el teorema pudiera ser falso. Sin embargo, una vez que la prueba fue realizada, vemos que en realidad la afirmación de su falsedad conduce a contradicciones. En filosofía, muchas nociones que se consideraron no problemáticas –como la del “libre albedrío” o la noción de Dios- han sido denunciadas como completamente incoherentes.

Como señala Diana Pérez (2013), la base de la línea argumentativa de quienes siguen esta estrategia “es la distinción entre propiedades y conceptos”. En efecto, la idea es que dado que es posible acceder epistémicamente a una misma propiedad del mundo mediante dos “modos de presentación” diferentes o, de otra manera, mediante conceptos alternativos irreducibles entre sí (básicamente, mediante los conceptos fenoménicos y los conceptos “físico/funcionales”) no resulta lícito inferir un dualismo ontológico a partir de un dualismo conceptual.

En otras palabras, tenemos dos maneras de conceptualizar la experiencia consciente. Por un lado, podemos concebir nuestra experiencia consciente en “primera persona”, en términos de su carácter cualitativo o subjetivo, de cómo se siente al tenerla, independientemente de toda descripción o explicación de ese estado en términos “físicos”, de sus relaciones causales con otros estados o elementos físicos (como los estímulos que provocan esas experiencias, las conductas que dichos estados producen, o cómo interactúan con otros estados mentales). Por otro lado, tenemos una manera de caracterizar estos mismos estados en términos físicos / funcionales.

Ahora bien, gracias a esta doble conceptualización, se explica que podamos concebir sin contradicción un mundo zombi que, en rigor, es metafísicamente imposible, dado que si bien existen dos conceptos acerca de las experiencias conscientes que son irreducibles entre sí, ambos refieren a las mismas propiedades en el mundo.

Por mi parte, si bien no me ocuparé de evaluar de forma pormenorizada esta estrategia en este trabajo, mencionaré algunas de las críticas que enfrenta. Tomadas en conjunto, considero que constituyen razones atendibles para, por lo menos, buscar una posición alternativa a esta táctica. En este sentido, lo primero que habría que señalar es que no existe un consenso en cuanto a la caracterización definitiva de los conceptos fenoménicos. Así, existen autores que los caracterizan como conceptos demostrativos (Loar 1997 y Levin 2007), otros como constituidos por las mismas propiedades a las que refieren (como es el caso de Papineau 2002 y su teoría citacional de los conceptos fenoménicos).

Al mismo tiempo, cada una de las diversas caracterizaciones adolece de problemas internos nada despreciables. Uno de los defectos comunes a varias caracterizaciones tiene que ver con que carecen de condiciones de individuación de

grado fino para estos conceptos, y si se supone que este es un requisito innegociable para considerar a una representación como un concepto, debería concluirse que los conceptos fenoménicos no son conceptos en absoluto (Pérez 2011, 2013; Tye 2009). En muchos otros casos, señala Tye 2009, la caracterización ofrecida de los conceptos fenoménicos no explica por qué surge cierta perplejidad (o por qué tenemos la sensación de que hay un vacío explicativo) cuando se nos explica, mediante estos conceptos, que una misma propiedad mental puede ser conceptualizada mediante dos modos de presentación diferentes y por qué *no* surge una perplejidad equivalente cuando se nos muestran otros casos en los cuales una misma propiedad que no pertenece al ámbito de lo mental es conceptualizada de dos maneras diferentes.

En otras palabras, mientras que nos veríamos motivados a preguntarnos “¿Cómo puede este estado (caracterizado fenoménicamente), ser el mismo que este otro, (caracterizado mediante conceptos “físicos”)”, no nos invade una incredulidad semejante que nos motive a preguntarnos, por ejemplo, “¿cómo Héspero puede ser Fósforo?”. Y el partidario de los conceptos fenoménicos, en cualquiera de sus variantes, no parece poder ofrecer una respuesta sensata a esta exigencia. Por otro lado, algunos autores, como Tye 2009 o Ball 2009, han cuestionado que existan verdaderamente conceptos que cumplan con los variados requisitos que se supone que deberían reunir los conceptos fenoménicos para enfrentar con éxito los diversos argumentos en favor del dualismo.

Los partidarios de los zombis, por su parte, tienen réplicas listas para cada una de las objeciones que acabo de esbozar. Con respecto a esta última, por ejemplo, suelen argumentar que el modo en que se explica con éxito que podamos creer sin incurrir en contradicción que ciertas identidades son falsas cuando de hecho no lo son (como la de Agua = H₂O), no permite explicar de manera igualmente exitosa la posibilidad de concebir sin incurrir en contradicción que la conciencia no sea algún estado físico / funcional.

De todos modos, considero que un punto en donde justamente se exhiben con mucha claridad las dificultades que pueden presentarse al intentar concebir el “escenario zombi”, es en la intersección entre la conciencia y el resto de las capacidades cognitivas. Me refiero específicamente a la relación entre los juicios o creencias fenoménicas, esto es, básicamente, los juicios y pensamientos *acerca* de

nuestros estados mentales conscientes y la consciencia. Entre la minoría (casi diría, de uno) que actualmente en la filosofía de la mente defiende una posición dualista, es el filósofo David Chalmers quien se ha ocupado de las “paradojas” involucradas en estas cuestiones con gran consistencia y rigurosidad. Por eso se transformará a partir de ahora en el principal interlocutor de este capítulo.

En particular, abordaré estos problemas a partir de la reconstrucción que hace Chalmers (2010) de un argumento de Sydney Shoemaker (1975) contra la posibilidad de los “*qualia* ausentes”, es decir, en contra de lo que podríamos considerar una “versión” del argumento de los zombis. Luego de presentar el argumento en cuestión, me ocuparé de las dos respuestas con las que Chalmers busca rechazarlo, ya que mi crítica a la noción de zombi se basará en estas respuestas.

3. En contra de la concebibilidad de los zombis: round 1

Hay un capítulo crucial de la serie *The Walking Dead* en el cual nos enteramos que, dentro de ese apocalipsis zombi que se nos cuenta, todos los seres humanos están infectados, es decir, que después de su muerte, cualquiera fuera su causa, todos los seres humanos se convertirán en esas desagradables y peligrosas criaturas. Por su parte, si se acepta el argumento de Shoemaker en contra de la posibilidad de los *qualia* ausentes, deberíamos llegar a una conclusión semejante: todos estamos o podríamos estar infectados con el virus zombi. Más precisamente, la idea es que si se aceptan las nociones involucradas en el argumento de los zombis, no tendríamos manera de determinar, ni siquiera en nuestro propio caso, si somos zombis o no. Una reconstrucción posible del argumento de Shoemaker puede ofrecerse en los siguientes términos:¹³

(1) Si los zombis son posibles (tal como afirma el argumento de los zombis), entonces tienen las mismas creencias (o juicios) fenoménicas (es decir, acerca de las experiencias conscientes) que sus gemelos conscientes, formadas por los mismos mecanismos.

¹³ Para simplificar la presentación y concentrar la atención en aquellos aspectos que son relevantes para mis propósitos, en esta reconstrucción del argumento modifiqué deliberadamente la reconstrucción que hace Chalmers en 2010 pp. 293-294. Sin embargo, esta modificación no afecta en ningún sentido la fuerza de las objeciones que Chalmers le dirige.

(2) Las creencias de los zombis son falsas e injustificadas (después de todo, como carecen de experiencia consciente, lo más natural es pensar que cuando afirman, por ejemplo, “Tengo un fuerte dolor en el cuello”, no pueden estar diciendo algo verdadero si entendemos “dolor” en un sentido fenoménico).

(3) Ahora bien, si es cierto que las creencias fenoménicas de los zombis son falsas e injustificadas, y las creencias de los seres conscientes están formadas por los mismos mecanismos que las de estas desdichadas criaturas, entonces, si se acepta que los procesos de formación son los que justifican las creencias, debe concluirse que las creencias fenoménicas de los seres conscientes también carecen de justificación.

(4) Si se acepta, además, que el conocimiento presupone o implica justificación, debería reconocerse entonces que nosotros, los seres presuntamente conscientes, no tenemos conocimiento en sentido estricto de nuestras propias experiencias conscientes; es decir, que ¿no sabemos realmente si somos zombis o no!

El resultado de este argumento resulta inaceptable para un dualista como Chalmers que cree que la consciencia es aquello sobre lo cual tenemos un conocimiento que es más directo y seguro que el podemos llegar a tener sobre cualquier otra cosa en el mundo. Sin embargo, antes de avanzar con este argumento, necesito realizar algunas aclaraciones terminológicas. Empecemos con la noción de “creencia fenoménica”.

En este contexto, vamos a llamar juicios fenoménicos o creencias fenoménicas a aquellas que refieren a la experiencia consciente y sus características. En particular, en este trabajo capítulo voy a ocuparme principalmente de dos tipos de creencias fenoménicas. El primer tipo hace referencia a las experiencias y no a los posibles objetos de la misma. En efecto, cuando observo o percibo algo rojo, se supone que tengo una experiencia fenoménica de “rojedad”, además del estado perceptivo acerca de un objeto del mundo. Un ejemplo de este tipo de creencia o pensamiento se expresaría en una afirmación como la siguiente: “estoy teniendo una sensación de rojo en este momento”. El segundo tipo de creencias fenoménicas son aquellas que tenemos cuando reflexionamos sobre el hecho de que tenemos experiencia consciente y acerca de sus características. Y un ejemplo de éste clase de

creencias podría expresarse en una afirmación que dijera “las sensaciones son misteriosas”.

Ahora bien, una vez que acepta la validez del argumento de Shoemaker, Chalmers considera dos opciones (no necesariamente excluyentes) para resistir su conclusión: la primera consiste en negar la idea de que los zombis tengan exactamente los mismos conceptos que sus pares conscientes y, por lo tanto, que puedan tener las mismas creencias fenoménicas. La segunda opción consistiría en conceder que los zombis tienen exactamente los mismos conceptos y creencias que sus gemelos conscientes pero rechazar la idea de que el tipo de justificación de estas creencias sea la misma en ambos casos. En este punto, no me interesa tanto evaluar la respuesta de Chalmers al argumento de Shoemaker como mostrar de qué manera las dos respuestas que considera el autor sirven para poner en duda la concebibilidad del escenario zombi y, consiguientemente, sobre la legitimidad de atribuirle un carácter intrínseco a la consciencia.

Comencemos por la segunda opción. En general, en el marco de las discusiones en torno a la experiencia consciente, se suele admitir que las creencias y los pensamientos son estados físico/ funcionales.¹⁴ Por lo tanto, de esto se seguiría que los zombis (que, por definición, son idénticos desde el punto de vista físico /funcional a sus pares conscientes), deberían compartir creencias con sus duplicados conscientes. Aceptando este punto de partida, en su libro de 1999/1996, Chalmers sostiene que si bien tanto los zombis como sus gemelos conscientes comparten las mismas creencias y los mecanismos de formación de las mismas, no son éstos últimos los que *justifican* a las primeras ni los que las transforman en auténtico conocimiento. Aquello que cumple ese rol en el caso de los seres conscientes es la propia experiencia.

Por ejemplo, el propio hecho de tener una experiencia de rojo en este momento proporciona justificación a la creencia de que estoy teniendo una

¹⁴ En efecto, en líneas generales se suele aceptar que los estados mentales a los que se categoriza bajo el rótulo de “actitudes proposicionales” o estados intencionales (como los deseos y las creencias), pueden ser explicados cabalmente en términos funcionales e identificados de manera completa por sus relaciones causales con otros estados y eventos del mundo. Así, de acuerdo con este enfoque, creer que una proposición es verdadera es estar en un estado en el que actuamos del modo apropiado si fuese verdadera, un estado que tiende a producirse por ser el caso de que es verdadera la proposición en cuestión, y en el cual nuestra propia dinámica cognitiva del razonamiento refleja la interacción apropiada de la creencia en cuestión con otros estados mentales como otras creencias y deseos.

experiencia de rojo (Chalmers 1999: 255). Sucede que, desde esta perspectiva, tener una experiencia es automáticamente estar en alguna relación epistémica íntima con la misma, una relación que Chalmers llama, siguiendo en parte a Russell, de “familiaridad” (*acquaintance*).¹⁵ Ni siquiera habría una posibilidad conceptual de que un sujeto pueda tener una experiencia de rojo sin tener ningún contacto epistémico con ella (Chalmers 1999: 256). Sería precisamente esta “familiaridad” que el sujeto tiene con su propia experiencia consciente lo que proporcionaría evidencia a nuestras creencias y juicios fenoménicos antes que cualquier relación causal. Así, la respuesta de Chalmers consiste en señalar que la justificación de los juicios o creencias fenoménicas no depende únicamente -o en absoluto- de los procesos (causales o de otro tipo) de formación de los mismos, sino que involucra de un modo crucial a la propia experiencia consciente.

Ahora bien, el problema es que Chalmers también dice que el único motivo legítimo por el cual creemos que hay en el mundo algo tan maravilloso y misterioso como la consciencia es nuestro acceso en primera persona (“privado”) a nuestros propios estados conscientes. En efecto, según Chalmers, puesto que todo el resto de los hechos del mundo supervienen lógicamente a sus propiedades físicas, nuestras razones para creer en la consciencia se derivan exclusivamente de nuestra propia experiencia de ella. En efecto, desde esta perspectiva,

Nuestras razones para creer en la conciencia se derivan exclusivamente de nuestra propia experiencia de ella. Aun si conociésemos cada detalle acerca de la física del universo –la configuración, causalidad y la evolución de todos los campos y partículas en el continuo espaciotemporal- esa información no nos llevaría a postular la existencia de la experiencia consciente. [...] Es mi experiencia de primera persona lo que me fuerza a enfrentar el problema. [...] De todos los hechos de bajo nivel acerca de las configuraciones físicas y la causalidad, podemos en principio derivar toda clase de hechos de alto nivel acerca de los sistemas macroscópicos, su organización y la causalidad entre ellos. Podríamos determinar todos los hechos acerca del funcionamiento biológico, la conducta humana y los mecanismos cerebrales que la causan. Sin embargo, nada en esta vasta historia causal llevaría a quien no lo haya experimentado directamente a creer que debería haber *una conciencia*.” [...] (Chalmers 1999:141)

¹⁵ En el capítulo 3 haré algunas críticas a la noción de “familiaridad”.

Es más, según el autor, “el eliminativismo acerca de la experiencia consciente [es decir, la posición que niega que exista algo así como la consciencia en el sentido fenoménico] es una posición irrazonable *sólo* debido a nuestro conocimiento de ella. Si no fuese por este conocimiento directo, la consciencia podría seguir el camino del espíritu vital” (Chalmers 1999: 142). Como vimos, desde esta perspectiva, es sólo nuestra experiencia de primera persona lo que nos fuerza a enfrentar el problema (Chalmers 1999:141). Pero en este punto es precisamente en donde comienzan las dificultades para el argumento de los zombis.

En efecto, de las anteriores declaraciones, se seguiría, en mi opinión, que (1) si nuestros gemelos zombis son realmente iguales a nosotros en todos los aspectos físicos, y (2) que la capacidad para evaluar y sopesar la evidencia con la que cuentan es también un aspecto en última instancia “físico” (tal como cabría esperar dado que, en principio, el argumento de los zombis sólo pretende mostrar la irreducibilidad de la consciencia), entonces (3) debería admitirse que éstas criaturas desdichadas carecen de la única evidencia relevante para plantearse el “problema de la consciencia” y que, por este motivo, luego de considerar las pruebas relevantes, es razonable suponer que en su mundo no habría publicaciones ni reuniones científicas acerca del “problema difícil de la consciencia” (y que el gemelo zombi de Chalmers no tendría mucho trabajo en ese mundo alternativo). Después de todo, ¿qué motivación tendrían los zombis para postular una entidad tan misteriosa como la consciencia?

En el fondo, su situación epistémica sería similar a la de quien observa su conducta “desde afuera”, porque el zombi tampoco cuenta con ninguna evidencia relevante que lo obligue a aceptar la existencia de los *qualia*. En sentido estricto, desde una perspectiva como la de Chalmers, todo lo que pueda saberse sobre los estados mentales conscientes de las personas “desde afuera”, es compatible con la ausencia de la consciencia y de sus rasgos característicos.

Es verdad que ciertos hechos físicos acerca del mundo podrían proporcionar alguna evidencia indirecta en favor de la existencia de la consciencia. Por ejemplo, a partir de estos hechos podríamos averiguar que hay muchos organismos que *afirman* ser conscientes y dicen tener misteriosas experiencias subjetivas. Con todo, esta evidencia sería bastante inconcluyente, y podría ser muy natural extraer una conclusión eliminativa: que de hecho no hay ninguna

experiencia presente en estas criaturas, sólo mucha charla (Chalmers 1999: 141-142).

Ahora bien, si se admite este razonamiento, debe admitirse que a menos que se atribuya alguna diferencia cognitiva entre los zombis y sus gemelos conscientes que la noción de zombi prohíbe, debería concluirse que (4) habría al menos *una* diferencia entre nuestros procesos cognitivos y nuestra conducta y las de ellos, que excede en principio un mera diferencia entre la presencia y ausencia de propiedades fenoménicas. Pero, si esto es así, entonces el escenario zombi no resultaría realmente concebible después de todo. En otras palabras, no es tan claro que puedan concebirse dos criaturas que diverjan en sus propiedades fenoménicas pero que no lo hagan también en sus propiedades físicas.

Este argumento puede plantearse de otra manera apelando a ciertas observaciones que hace Chalmers a propósito del argumento de Shoemaker. En ese contexto, el autor, a la inversa de lo que venimos suponiendo, nos propone entender la *diferencia* entre las creencias de un zombi y su gemelo consciente a partir de una modificación del argumento del conocimiento de Jackson (1982). Si bien este argumento será tratado especialmente en el capítulo 4, conviene reconstruirlo brevemente ahora para poder comprender la posición de Chalmers.

Concretamente, el argumento del conocimiento nos pide que imaginemos la historia de una súper-neurocientífica, llamada Mary, que conoce todos los hechos físicos (es decir, todos los hechos que la Física, la Biología, la neurofisiología y la Química ideales o acabadas nos dicen que hay) involucrados en la percepción humana del color. Sin embargo, por haber sido confinada desde su nacimiento a una prisión monocromática, Mary no ha tenido nunca experiencias de color. Según Jackson, a pesar de toda su erudición Mary aprenderá algo nuevo cuando sea liberada de su encierro: aprenderá, por ejemplo, cómo es ver algo rojo cuando se enfrente con algún objeto que tenga ese color. Ahora bien, (1) si el fisicalismo es verdadero, entonces si alguien sabe todo lo que se puede saber acerca de la naturaleza física del mundo, sabe todo lo que hay para saber sin más. Sin embargo, dado que, como ilustraría el caso de Mary, (2) es posible poseer todo el conocimiento físico pero de todos modos aprender algo nuevo, (3) debería concluirse que eso nuevo que se conoce no es físico sino de otra naturaleza. De esto se seguiría que el

fisicalismo es falso. Y lo sería en un doble sentido: ni todo el conocimiento acerca de lo real es conocimiento físico (tesis epistemológica), ni todo lo real es físico o reducible a lo físico (tesis ontológica).

Con este argumento de trasfondo, Chalmers nos pide considerar la historia de “Mary zombi”, donde se recombina los argumentos del conocimiento y de los zombis en un sentido obvio. ¿Qué ocurre cuando Mary zombi es liberada? Bajo la hipótesis que venimos considerando, Mary zombi no puede adquirir los mismos conceptos y creencias fenoménicas que su gemela Mary. De hecho, el propio Chalmers considera que podría pensarse que “cuando Mary zombi ve por primera vez una flor, puede adquirir ciertas habilidades para reconocer y discriminar, aunque incluso estas habilidades serán severamente restringidas puesto que no pueden involucrar experiencias.” (Chalmers 2010: 295). En principio, suponer que Mary zombi cree (y afirma) que su liberación le permitió adquirir un conocimiento fenoménico inesperado, implica atribuirle una comprensión errada de su situación epistémica que no tiene un correlato en su gemela consciente Mary. De esta manera, este caso imaginario comienza a develar las debilidades de su planteo y de la coherencia pretendida de la noción de zombi.

Supongamos ahora que el argumento del conocimiento falla y en realidad la experiencia visual consciente no incluye ninguna propiedad “especial” no física. No pido imaginar un mundo en el cual no haya propiedades fenoménicas en absoluto sino simplemente uno en el cual este tipo de propiedades no se instancian cuando ocurren estados perceptivos visuales. En este caso, tanto Mary como su gemela Mary Zombi aceptarán que, en el mejor de los casos, sólo adquirieron una habilidad para realizar ciertas discriminaciones luego de su liberación y no habrá, en el escenario supuesto, una diferencia entre ambas de ningún tipo (más allá de una diferencia numérica, puesto que Mary bis y su gemela no son una única entidad).

Ahora bien, dado que tampoco hay diferencias entre la Mary zombi del ejemplo de Chalmers y la Mary zombi de mi ejemplo (llamémosla “Mary zombi bis” para evitar confusiones) ni en su constitución física interna ni en su relación con el entorno, y puesto que ambas carecen de experiencia consciente, ¿por qué deberíamos aceptar que Mary zombi, luego de ser liberada, cree y afirma haber adquirido, además de ciertas habilidades discriminativas, un conocimiento proposicional totalmente nuevo acerca de ciertas propiedades misteriosas que

están “más allá” de las propiedades físicas de la visión que ya conocía? ¿Qué cosa tan especial ha sucedido en su mundo o en su interior, con respecto a los sucesos correlativos del mundo de Mary zombi bis, que justifique esta divergencia en la opinión? Respuesta: ¡Nada! Para decirlo de otra manera: los zombis están encerrados para siempre en la prisión monocromática de Mary: ¡y a Mary jamás se le hubiera ocurrido allí adentro que había hechos sorprendentes acerca de la percepción visual del color que no podría explicar con su vasta ciencia!

Suponer que Mary zombi cree (y afirma) que su liberación le permitió adquirir un conocimiento fenoménico inesperado, implica, otra vez, atribuirle una comprensión errada de su situación epistémica que no tiene un correlato en su gemela consciente Mary. Pero esta diferencia parece ser una diferencia que excede una mera diferencia en la ausencia o presencia de ciertas propiedades fenoménicas. De ser así, Mary y su gemela zombi no son iguales tampoco desde el punto de vista físico / funcional.¹⁶

En este punto, es probable que un adversario como Chalmers pensara en apelar nuevamente a la distinción mencionada entre la *justificación* de los juicios fenoménicos y los *mecanismos* (causales) por los cuales se forman. Como vimos, de acuerdo con esta distinción, si bien la formación de los juicios podría en principio explicarse cabalmente en términos compatibles con el fisicalismo, la justificación de los mismos depende exclusivamente de la presencia de la experiencia relevante. Así, si bien un zombi podría tener la misma creencia acerca de su experiencia y realizar los mismos juicios que su gemelo consciente, dado que, por definición, carece de experiencias, sus juicios y creencias fenoménicas no estarán justificados y serán falsos. Si aplicamos ahora esta distinción al argumento anterior, podríamos decir que si bien los zombis podrían plantearse “el problema de la consciencia”, sus creencias acerca de la existencia del fenómeno que generaría el problema (la consciencia) resultarían falsas y estarían absolutamente injustificadas.

¹⁶ En todo caso, si Mary Zombi (y Mary Zombi bis), luego de liberarse de su encierro, insisten con la idea de que tienen un peculiar conocimiento nuevo (conocimiento acerca de presuntas propiedades que no pueden ser explicadas en términos físicos), estaríamos tentados a sugerirles algún tipo de terapia filosófica que les haga ver el tipo de confusión conceptual en el que incurrir o que les muestre que tienen una comprensión equivocada de su situación epistémica. Pero no es nada que no pueda remediarse con algún tipo de análisis más profundo. Una vez clarificadas estas cuestiones, parece que el dualismo no debería tener demasiado éxito en el mundo zombi.

Esta réplica, sin embargo, trae otras dificultades al partidario de los zombis. En efecto, dado que se supone que los juicios, creencias y pensamientos son el tipo de cosa que en principio resultan físicamente reducibles, debería ser posible explicar en términos compatibles con el fisicalismo por qué *creemos*, o podemos ser llevados a *creer* que la conciencia no superviene lógicamente a lo físico sin tener que reconocer su existencia (o, mejor dicho, sin reconocer la existencia de algo que no superviene a lo físico). De acuerdo con Chalmers, esta explicación sería análoga a aquella que busca explicar la creencia en la existencia de Dios o de los extraterrestres, sin postular la existencia de una divinidad o de seres de otros planetas, en términos de fenómenos sociales y psicológicos. Sin embargo, continúa Chalmers, la analogía falla, porque la conciencia, a diferencia de Dios, es ella misma un *explanandum*, y no algo que postulamos para explicar otra cosa, como por ejemplo nuestros juicios y creencias acerca de ella.

En contraposición, adoptar un enfoque que sostenga que lo único que hay que explicar son nuestros juicios y creencias acerca de la conciencia es, para Chalmers, una posición eliminativista que desconoce que la “conciencia no es una construcción explicativa postulada para ayudar a explicar la conducta o los sucesos en el mundo. Más bien, es un *explanandum* primitivo, un fenómeno por derecho propio que necesita explicación.” (Chalmers 1999: 245). De hecho, para Chalmers *todo* lo que probaría la posibilidad de explicar en términos compatibles con el fisicalismo los juicios y creencias acerca de la experiencia es precisamente que los juicios y creencias pueden explicarse de esta manera. Pero lo que realmente querríamos explicar no es esto, sino la propia conciencia.

Con todo, no creo que esta potencial respuesta funcione. En particular, considero que hay en Chalmers una cierta ambigüedad en el uso de la idea de “explicar los juicios fenoménicos en términos compatibles con el fisicalismo”. Por un lado, puede interpretarse, como hace Chalmers, como una posición eliminativista que sostiene que lo único que hay que explicar son los juicios y creencias fenoménicos y nada más, porque de hecho no hay nada más que explicar. Sin embargo, la idea puede también interpretarse como sosteniendo que es posible explicar los juicios y creencias fenoménicas en términos compatibles con el fisicalismo sin necesidad de negar que, además, exista un fenómeno adicional (el *explanandum*) al que estos juicios refieren. En todo caso, lo que se podría estar

negando es cierta *caracterización* del fenómeno en cuestión (en particular, una caracterización que presente a la consciencia como un hecho que no superviene lógicamente a lo físico) mostrando, por ejemplo, qué rasgos de nuestra experiencia podrían conducirnos a conclusiones dualistas como las de Chalmers.

En efecto, tal como señalan Carruthers y Schier (2014), “hay un número de personas que dan explicaciones estructurales/funcionales de la consciencia fenoménica. Al hacer eso no niegan que la consciencia fenoménica exista, en lugar de eso niegan la caracterización de Chalmers de ella [...], i.e. como siendo necesariamente inexplicable en términos de estructura y función de estados mentales.” (Carruthers y Schier 2014: 5). Pretender que aceptemos sin reparos que la consciencia es físicamente irreducible con el pretexto de evitar ser acusados de eliminativistas constituiría, en el contexto de esta discusión, una petición de principio.

Ahora bien, si esto es así, es decir, si se reconoce que es posible explicar por qué creemos que la consciencia es algo que está “más allá” de lo físico sin convertirnos en eliminativistas, es decir, sin negar, que exista la consciencia, pero, al mismo tiempo, sin poner en riesgo la compatibilidad de nuestra explicación con los compromisos ontológicos del fisicalismo / funcionalismo, entonces ya no resulta tan claro que la premisa (3) del argumento de los zombis, esto es, aquella que establece que un mundo zombi es concebible, sea cierta. Por las razones antedichas, si Chalmers deseara objetar esta estrategia, no le bastaría con replicar que quien procede de esa manera no se toma en serio la consciencia o que deja de lado el *explanandum* fundamental. En todo caso, lo que debería mostrar es por qué la caracterización alternativa a la suya del fenómeno en cuestión es insuficiente o inadecuada. Mientras eso no ocurra, seguimos teniendo razones para pensar que el escenario zombi no es realmente concebible después de todo.

Por lo demás, existe un motivo adicional por el cual un filósofo como Chalmers podría preferir rechazar la idea de que los zombis y sus pares conscientes comparten los mismos conceptos y creencias fenoménicas para hacer frente al argumento de Shoemaker. En particular, creo que no lo convence demasiado el tipo de justificación de las mismas que consideramos en esta sección. En efecto, de acuerdo con Chalmers en un trabajo posterior al de 1999:

Para que una creencia inferida esté *prima facie* justificada, existen tres requerimientos centrales: [...] Primero, el contenido epistémico de la creencia debe estar relacionado apropiadamente con el contenido de aquella creencia de la que es inferida. Segundo, la creencia debe estar apropiadamente *causada* por la creencia que la justifica. Tercero, la creencia que justifica debe estar ella misma justificada.

En la justificación *prima facie* de una creencia fenoménica directa por medio de una experiencia, hay tres factores del mismo tipo: Primero, [...] el contenido epistémico de una creencia fenoménica directa debe reflejar [*mirror*] la cualidad de la experiencia. Segundo, una conexión natural: la creencia fenoménica debe estar apropiadamente *constituida*¹⁷ por la experiencia. Y tercero, el estatus epistémico: el sujeto debe estar en una relación de familiaridad con la cualidad [o con el aspecto cualitativo de la experiencia] que justifica [a la creencia]. (Chalmers 2010: 287. La cursiva es mía).

Como puede verse, dentro de este nuevo modelo de justificación, es el segundo elemento mencionado (el que expresa la conexión *natural* entre experiencia y creencia o entre evidencia y creencia) el que está ausente en el primer modelo que vimos arriba. En efecto, de acuerdo con éste, la mera *presencia* de una propiedad fenoménica y una relación de familiaridad con la misma, bastaban para justificar las creencias en cuestión. Ahora bien, si se acepta que este segundo elemento es necesario para que haya conocimiento *conceptual* acerca de la experiencia, entonces debería admitirse que, de acuerdo con el primer modelo de justificación propuesto por Chalmers, las creencias y juicios fenoménicos de los seres conscientes, no expresarían verdadero conocimiento.¹⁸

¹⁷ En principio, si bien la noción de “constitución” no recibe en Chalmers un desarrollo demasiado extenso, podría decirse que lo que está pensando en resumidas cuentas es que, en el caso particular de los conceptos fenoménicos, parte de su significado está compuesto, en un sentido literal, por su referente (i.e. la propia experiencia consciente), como si éste se introdujera en aquél.

¹⁸ En mi opinión, el motivo por el cual Chalmers adopta el segundo tipo de justificación tiene que ver con una adhesión más o menos implícita al enfoque tripartito tradicional del conocimiento, según el cual, el conocimiento equivale a una creencia verdadera justificada. De acuerdo con este enfoque, se dice que un sujeto S sabe que P cuando: (i) P es verdadera; (ii) S juzga que P y (iii) S está justificado en juzgar que P. En este marco, la postulación de la tercera condición busca impedir que sean considerados como casos de efectivo conocimiento aquellos juicios o creencias que carecen de justificación o que son sostenidos por motivos inadecuados aun cuando, por una azarosa coincidencia, resulten finalmente verdaderos. Ahora bien, si se tiene en cuenta este análisis, podría pensarse que la primera justificación que esboza Chalmers para los juicios fenoménicos no cumple con el tercer requisito. En efecto, mientras que el enfoque falibilista o causal del conocimiento podría dar cuenta de la diferencia entre tener una creencia verdadera y tener una creencia verdadera

4. En contra de la concebibilidad de los zombis: round 2

En esta sección voy a considerar las dificultades que persiguen al partidario de los zombis si opta por la segunda opción apuntada arriba para desestimar el argumento de Shoemaker, esto es, negar que zombis y seres conscientes compartan las mismas creencias punto por punto. De hecho, Chalmers también explora esta variante. Después de todo, es razonable pensar que alguien que no hubiera atravesado por las experiencias relevantes, tampoco podría tener los conceptos ni las creencias correspondientes. En este sentido, Chalmers afirma, por ejemplo, que “no hay ninguna razón para aceptar que los zombis tienen [las mismas creencias fenoménicas] que sus gemelos conscientes y toda razón para creer que no” (Chalmers 2010:294). Y esto es así, básicamente, porque “el contenido de una creencia fenoménica directa de un ser consciente está parcialmente constituido por [las] cualidades fenoménicas [de la experiencia] subyacentes”. En efecto, de acuerdo con Chalmers, los conceptos fenoménicos directos (que posteriormente componen las creencias fenoménicas correspondientes) se suscitan “cuando un sujeto atiende a la cualidad [o aspecto cualitativo] de una experiencia y forma un concepto completamente basado en este acto de atención, “llevando” la cualidad dentro del concepto” (Chalmers 2010: 267).

Así, dado que el contenido de las creencias fenoménicas directas está conformado en parte por las mismas propiedades a las que éstas refieren, los zombis que, por definición, no tienen experiencias fenoménicas, jamás podrán formar los conceptos necesarios para tener al menos ciertas creencias acerca de la experiencia

justificada, afirmando que una creencia verdadera está justificada si y sólo si fue formada a través de los mecanismos cognitivos fiables o adecuados -es decir, por aquellos que tienden a producir creencias verdaderas-, la propuesta de Chalmers no puede dar cuenta de esta distinción o no lo hace explícitamente. ¿Por qué? En principio, podría pensarse que estar familiarizado con una experiencia basta para hacer verdadero al juicio o a la creencia fenoménica correspondiente pero no para justificarlo, precisamente porque, en este escenario propuesto, no hay ningún tipo de conexión entre la experiencia y la creencia o juicio acerca de ella. Pero, si esto es así, puesto que la justificación es condición necesaria para el conocimiento, se sigue que los juicios fenoménicos no constituirían efectivo conocimiento porque no estarían debidamente justificados. Así, lo que faltaría en este enfoque es proponer algún tipo de vinculación entre la evidencia para sostener una creencia y la creencia misma. De lo contrario, parece haber lugar para considerar que no es posible distinguir entre tener una evidencia de la existencia de una experiencia fenoménica y juzgar que se está teniendo dicha experiencia, y juzgar que se está teniendo una experiencia *en virtud* de la evidencia disponible. Ahora bien, el hecho de que el segundo modelo de justificación propuesto por Chalmers recurra a la idea de *constitución*, le trae otros problemas que serán examinados en la sección siguiente. De todos modos, aun cuando mi lectura acerca de los motivos por los cuales Chalmers rechaza el primer modelo de justificación resulte inadecuada, mis argumentos en contra de la concebibilidad de los zombis no depende de ella.

consciente. El problema es que, en mi opinión, esta concesión resulta fatal para el zombista porque lo obliga a reconocer que los zombis no pueden ser física / funcionalmente idénticos a sus gemelos conscientes. Veamos por qué.

Los conceptos y creencias fenoménicas (ahora) alternativas de los zombis deben ser explicables completamente en términos de propiedades físicas (porque, por definición, los zombis carecen de otro tipo de propiedades). Pero si esto es así, parece que sus gemelos conscientes, que son idénticos desde el punto de vista físico, también deberían formar estos conceptos y creencias fenoménicas alternativas. De este modo, deberían pensar y decir cosas extrañas como “Poseo dos conceptos muy similares acerca de la experiencia consciente”. En cambio, sus contrapartes zombis no deberían realizar estas extrañas afirmaciones porque ellos no poseen los conceptos y creencias fenoménicas auténticas. Considerar que las harían supondría otra vez atribuirles una cierta incapacidad para reconocer semejanzas y diferencias entre conceptos y creencias y, en definitiva, implicaría admitir una diferencia considerable e ilegítima, es decir, no permitida por las reglas del argumento de los zombis. Ahora bien, la enunciación de elocuciones como las señaladas por parte de las criaturas conscientes y su ausencia en los gemelos zombis, supone una diferencia entre ambos que va más allá de una diferencia en la presencia o ausencia de propiedades fenoménicas.

Siguiendo esta línea de razonamiento, podríamos comenzar a preguntarnos si en este segundo escenario propuesto los zombis podrían siquiera plantearse el “problema difícil” de la consciencia. En efecto, si los conceptos fenoménicos de los seres conscientes son necesarios para desarrollar los argumentos en favor del dualismo (como el de los zombis o incluso el del conocimiento), entonces los zombis que, de acuerdo con Chalmers, carecen de estos conceptos en absoluto o poseen otros distintos, jamás podrían plantearse esas cuestiones. Pero ¿por qué sería necesario poseer estos conceptos para comprender / desarrollar los argumentos en favor de dualismo? Es necesario, por ejemplo, para poder determinar si un escenario zombi es concebible, es decir, si es conceptualmente posible un mundo en donde nuestros conceptos y creencias fenoménicas no se aplican con verdad en absoluto pero en el que sí se aplican todo el resto de los conceptos y creencias “físicas”. Ahora bien, si los conceptos relevantes fueran diferentes en ambos escenarios, no es claro que en el mundo zombi debieran existir discusiones filosóficas en torno al problema

de la consciencia, porque no habría razones ni herramientas conceptuales para pensarlos. Pero esto constituiría una diferencia entre nuestro mundo y el mundo zombi que, otra vez, parece que debería reflejarse en la conducta de los habitantes de ambos mundos y que, por lo tanto, excede una diferencia a nivel fenoménico.

Dicho de otra manera: asumidas estas diferencias entre las creencias de los zombis y las de sus pares conscientes, la única manera en que Chalmers puede conservar la coherencia interna de su planteo y de la noción de “zombi” que utiliza es sosteniendo, al mismo tiempo, que estas diferencias (es decir, las de la posesión o carencia de ciertas propiedades fenoménicas junto con la diferencia en las creencias acerca de dichas propiedades), no arrastran consigo una diferencia físico / funcional entre ambos pares de gemelos. Pero este punto es justamente el que puede ser puesto en duda.

En primer lugar, porque, como sostiene Pérez 2009, parece ir en contra de los mismos puntos de partida del autor. En efecto, de acuerdo con su planteo, se supone que (i) los conceptos son elementos constitutivos de creencias y pensamientos y que (ii) las creencias y pensamientos son estados funcionales. Ahora bien, la aceptación de (i) y (ii) parece contradictoria con la aceptación de que (iii) no hay diferencia funcional entre los zombis y sus gemelos conscientes a pesar de que poseen diferentes conceptos y creencias. En principio, la aceptación de (iii) no es compatible con la aceptación de (i) y (ii). Vinculado con estas cuestiones, deberíamos agregar que, tanto si se acepta que las creencias fenoménicas están constituidas por las experiencias mismas como si no, lo cierto es que Chalmers se compromete con una posición según la cual, no sólo no es posible explicar reductivamente (i.e. en términos compatibles con el fisicalismo /funcionalismo) las experiencias conscientes sino que, desde su perspectiva, tampoco es posible dar cuenta, en términos compatibles con el fisicalismo / funcionalismo, de otras capacidades en principio no problemáticas. En efecto, dado que, como vimos, en cualquiera de sus versiones la justificación de las creencias fenoménicas depende de la existencia misma de la experiencia consciente, tampoco es posible explicar reductivamente el *conocimiento* de la experiencia, las *propiedades semánticas o referenciales* de las creencias en cuestión o la *memoria* de dichas experiencias. Como puede verse, el argumento de los zombis, de ser exitoso, prueba mucho más de lo que se supone que pretende.

En segundo lugar, si, como quiere Chalmers, las creencias entre los pares de gemelos son realmente diferentes con respecto a su contenido y referente, cabría esperar una diferencia en las inferencias que puedan realizarse a partir de ellas, y en las relaciones de oposición, compatibilidad, incompatibilidad, semejanza y diferencia que puedan mantener con otras creencias. Otra vez, este tipo de diferencias parece ser del tipo de diferencias que implicarían una diferencia incluso a nivel conductual, a menos que se sostenga que estas diferencias y los conceptos y creencias involucradas en ellas, pueden ser *pensadas* pero no *expresadas* mediante algún tipo de reporte verbal (es decir, a menos que se considere que se trata de diferencias que no tienen ni pueden tener un correlato en la conducta, ni verbal ni de ningún tipo). ¿Pero por qué deberíamos aceptar esto último? ¿No la expresa acaso Chalmers en su respuesta al argumento que reproduce al inicio de la sección anterior cuando *describe* la distinción entre las creencias y conceptos de los seres conscientes y las creencias y conceptos de los zombis? Este mismo punto puede mostrarse de otra manera: supongamos que tenemos otro par de gemelos: un ser consciente y un “zombi part-time”. De acuerdo con este experimento mental que propongo, ambas criaturas son exactamente idénticas desde el punto de vista físico / funcional y fenoménico hasta que un mal día todo se apaga para el zombi parcial de forma tal que, a partir de ese momento, pasa a convertirse en un zombi hecho y derecho. Por hipótesis, se supone que este cambio no debería traer aparejado un cambio a nivel físico/funcional. Sin embargo, cuando el zombi *recuerda* (tal vez con nostalgia) sus viejas experiencias y las *compara* con su estado actual, debería percatarse de que se produjo un cambio en él, porque sus conceptos acerca de sus experiencias actuales y pasadas, ya no son iguales. Incluso, el zombi part-time debería poder comunicar su desencanto a sus camaradas y afirmar cosas tales como: “No sé qué pasa muchachos, pero la cerveza ya no sabe como hace unos días”.¹⁹

¹⁹ En realidad, ni siquiera es necesario apelar a algún tipo de *cambio* en las propiedades cualitativas de las experiencias entre dos seres. Imaginemos dos seres que tienen experiencias cromáticas invertidas de modo tal que las experiencias de cosas rojas permiten formar conceptos fenoménicos de tipo R en uno de ellos y de tipo V en el otro. En principio, se supone que, a pesar de las diferencias mencionadas, deberían poder ser física y funcionalmente idénticos. Por lo menos, esta posibilidad debería ser admitida por quien echa mano del argumento de los zombis y acepta que la posesión de conceptos fenoménicos no trae aparejadas necesariamente diferencias de este tipo. Ahora bien, ¿qué ocurre cuando comparan sus conceptos de los colores invertidos con los conceptos de otros colores? ¿No debería arrojar resultados diferentes esta comparación en cada uno de ellos o dar lugar a inferencias distintas? ¿Y no deberían poder pensar estas diferencias conceptuales y expresarlas? De ser así, nuevamente estaríamos en presencia de diferencias que exceden las diferencias a nivel

Ahora bien, si esto es así, habría diferencias físico / funcionales entre ambos gemelos después del cambio en las propiedades fenoménicas que están prohibidos por las reglas del argumento de los zombis. Por otro lado, si se optara por afirmar que el zombi no se percató de esas diferencias, también se estarían atribuyendo, paradójicamente, diferencias físico / funcionales entre ambos gemelos, porque, en este caso, el zombi part-time tendría serios problemas de discernimiento acerca de su situación epistémica que no tienen un correlato en su gemelo consciente, porque estas capacidades de discernimiento y percatación serían, por hipótesis, capacidades apresables en términos físico / funcionales. En otras palabras, de una u otra manera, otra vez se estarían adjudicando diferencias cognitivas entre ambos gemelos que, en principio, exceden las diferencias fenoménicas. Así las cosas, volvemos por otra vía a la conclusión de que la coherencia del escenario zombi no es del todo clara.²⁰

En resumen, parece que, tanto si se acepta, por diversos motivos, que las creencias de los zombis son, o bien idénticas, o bien diferentes a las de sus duplicados conscientes, resulta inevitable desembocar en la sospecha de que, debido precisamente a las diferencias fenoménicas entre ambos, los zombis no pueden ser funcionalmente idénticos a sus gemelos conscientes.

fenoménico. Por otro lado, sino atribuyéramos estas diferencias, ¿no estaríamos admitiendo una diferencia en la capacidad de comparar conceptos y realizar inferencias a partir de dichos conceptos que también sobrepasa lo meramente fenoménico?

²⁰ Si se optara por sostener que el zombi no puede realizar las comparaciones mencionadas porque para eso debería poder conservar *todos* los conceptos fenoménicos con los que contaba antes de su transformación fatídica pero que, sin embargo, ciertos conceptos acerca de la experiencia están de algún modo “atados” a ellas de manera tal que duran lo que duran dichas experiencias y se desvanecen cuando éstas desaparecen, entonces nos estaríamos comprometiendo con ciertas consecuencias a mi juicio completamente inaceptables. En efecto, la idea de postular conceptos cuya posesión dura únicamente el tiempo que dura el objeto al que refieren es difícil de aceptar. Llevada al extremo, esta propuesta nos lleva a la conclusión de que, en rigor, jamás podríamos determinar si estamos teniendo *ahora* una experiencia que ya tuvimos en un tiempo *anterior*, *pero que ya no perdura en el presente*, porque esto supondría admitir la posesión del concepto fenoménico correspondiente a la experiencia anterior para poder realizar la comparación necesaria. Pero esto último es justamente lo que impide la tesis de los conceptos fenoménicos efímeros. En este sentido, no tenemos forma de saber si lo que ayer sabía como un jugo de naranja ahora sabe como una coliflor. Esta imposibilidad no se debe a un fallo en la memoria sino al escaso “tiempo de vida” del que gozarían estos peculiares conceptos que impide justamente la comparación requerida.

5. Conclusiones

En el inicio de este capítulo apunté que si se admite que los estados mentales conscientes tienen aspectos intrínsecos, parece imposible que puedan ser explicados en términos aceptables para el funcionalismo, que es una de las perspectivas que cuenta con mayor consenso en la actualidad a la hora de enfrentar el problema mente-cuerpo. Básicamente, la cuestión es que si existen propiedades así, no podrían ser individuadas por las relaciones que entablan con otros estados mentales, estímulos, conductas y objetos, y por los roles causales que ocupan en esas relaciones, como si ocurriría con el resto de los aspectos de la realidad mental. Ahora bien, de la imposibilidad de “funcionalizar” este aspecto cualitativo de lo mental se sigue que dos estados mentales funcionalmente idénticos podrían estar asociados o relacionados con diferentes estados conscientes.

En otras palabras, a partir del carácter “intrínseco” o “no-relacional” de cierto aspecto de las experiencias conscientes, se llega rápidamente a la idea de que es concebible y, en principio, metafísicamente posible, que dos personas sean exactamente iguales en su constitución física, en sus disposiciones para actuar y en sus relaciones con un entorno también idéntico, pero que difieran en sus experiencias conscientes. Esto equivale a decir que un “escenario zombi” debería ser concebible y metafísicamente posible. Ahora bien, si se acepta este planteo, debería admitirse también que si un escenario como el descrito no fuera concebible, entonces se arrojarían dudas sobre su posibilidad metafísica y, de este modo, se debilitarían fuertemente las razones para creer que nuestras experiencias conscientes cuentan con propiedades intrínsecas y no físicas.

Precisamente, parte de mi objetivo en este capítulo consistió en impugnar esta premisa crucial del argumento de los zombis. En particular, a partir de un examen pormenorizado de la conexión entre las experiencias conscientes y ciertos estados mentales en principio no problemáticos para una perspectiva funcionalista de la mente, esto es, los juicios y creencias acerca de dichas experiencias, busqué probar que el argumento de los zombis se ve obligado a atribuir a estas criaturas, subrepticia e ilegítimamente, algunas incapacidades cognitivas a la hora de evaluar su propia situación epistémica que no están presentes en sus réplicas conscientes y que conforman la base para marcar una diferencia entre ellos que va más allá de una diferencia con relación a las propiedades fenoménicas de la experiencia consciente.

Por lo tanto, si mi argumentación es correcta, debe admitirse que los zombis no son (no pueden ser) réplicas exactas de sus “gemelos” conscientes tampoco desde el punto de vista físico / funcional. Ahora, si esto es así, debemos concluir que, a pesar de lo que puede parecer a primera vista, no resulta tan claro que sea posible concebir dos seres que tengan diferencias en sus estados conscientes sin que, al mismo tiempo, tengan también una diferencia en sus estados físico / funcionales. En cualquier caso, creo que el problema de fondo que evidencia mi argumento es la improductividad de cierta manera de pensar acerca de la experiencia consciente. En efecto, a modo de conclusión, se podría sostener, tal como señala Diana Pérez (2013), que las dificultades indicadas en este trabajo al momento de imaginar la situación epistémica en que se encuentra los zombis surgen a partir de la separación tajante entre la cognición y la experiencia, o entre los conceptos y la experiencia, que parece anidar implícitamente en la misma noción de “zombi”. Así, “tratar de separar artificialmente estas dos cosas, solo genera situaciones paradójicas, es decir, escenarios no tan claramente consistentes, y por lo tanto no tan claramente concebibles como requiere el argumento de los zombis” (Pérez 2013: 240).

CAPÍTULO 2

LA SENSACIÓN DE UN ABISMO INFRANQUEABLE

1. Introducción

En el capítulo anterior sostuve que si la consciencia tuviera propiedades intrínsecas, entonces mundos idénticos a nuestro mundo en todos sus aspectos físico / funcionales pero diferentes en sus aspectos fenoménicos deberían ser, en principio, metafísicamente posibles. Sin embargo, puse en duda esta última afirmación cuestionando la idea de que un escenario como el descrito fuera claramente concebible. En cambio, en este capítulo voy a mostrar, entre otras cosas, que aun cuando se acepte que ciertos *conceptos* legítimos que poseemos acerca de la experiencia consciente no la caracterizan en términos de ciertos roles causales o funcionales, esto no constituye, por sí solo, un impedimento para identificar a la propiedad referida por dichos conceptos con una propiedad físico / funcional. Para eso, me ocuparé de otro de los argumentos que típicamente se postulan como una amenaza para el fisicalismo y de las respuestas más contundentes que recibió en la literatura relevante: me refiero al argumento del “hiato explicativo” o de la “brecha explicativa” diseñado por Joseph Levine.

Para cumplir con mis objetivos, estructuré el capítulo de la siguiente manera: en la sección siguiente voy a reconstruir con detalle lo que podría considerarse la versión estándar del argumento basada en los artículos clásicos (Levine, 1983 y 1993). En la sección tercera voy a consignar las objeciones más importantes que se le dirigen y que nos conducen finalmente a desestimar la fuerza del argumento (al menos en su versión original). Dentro de esta sección, se exhibirán las razones para fundamentar la afirmación mencionada arriba, esto es, la idea de que la existencia de conceptos legítimos acerca de la experiencia consciente que no la caracterizan en

términos de roles causales o funcionales, no es un obstáculo para identificar a su referente con propiedades físico / funcionales.

2. El argumento del hiato explicativo

En su versión más célebre, la expresión “brecha explicativa” o “hiato explicativo” (*explanatory gap*) es introducida por el filósofo Joseph Levine en un artículo de 1983 en el marco un intento de dar sustento argumentativo a la intuición de que incluso cuando sepamos todo lo que la ciencia pueda decirnos acerca del cerebro, su relación con la consciencia o el aspecto cualitativo de lo mental, siempre permanecerá misteriosa. Más específicamente, se trata de una estrategia tendiente a mostrar las *limitaciones explicativas* ineludibles que enfrenta lo que podríamos llamar, en un sentido amplio, una “teoría fisicalista” para dar cuenta de “por qué cuando ocupamos ciertos estados físicos-funcionales experimentamos el carácter cualitativo del tipo que experimentamos” (Levine 1997: 548).

En otras palabras, como puede verse, lo que está en juego es la capacidad para explicar el carácter cualitativo de la experiencia, por qué se siente cómo se siente ver rojo o sentir dolor. En este sentido, el objetivo de Levine se desmarca del de otros argumentos clásicos -como el argumentos de Descartes (1982), el de Kripke (1980), el de Jackson (1982)-, que parten de ciertas premisas epistémicas (acerca de lo que puede concebirse o conocerse) para llegar a una conclusión ontológica (acerca de la existencia de ciertos rasgos peculiares de lo mental) y favorecer así algún tipo de dualismo metafísico u ontológico entre el reino de la materia y el de lo mental. Levine, en cambio, considera que quienes dan el paso desde una posibilidad epistemológica a una posibilidad metafísica, siguen en el fondo a Descartes en su idea de que cuando tenemos una idea “clara y distinta” (acerca de la distinción entre dos cosas) tenemos acceso a cómo la cosas son en realidad, desde el punto de vista metafísico. Sin embargo, de acuerdo con Levine, si rechazamos este modelo cartesiano de acceso a la realidad y en cambio aceptamos que nada nos asegura que nuestras ideas (incluso aquellas claras y distintas) se correspondan necesariamente con la estructura real del mundo, la concebibilidad no resulta una guía confiable para resolver la cuestión de “qué es y qué no es metafísicamente posible” (Levine 1997:543-545). En todo caso, lo que ciertamente revelarían para Levine esta clase

de argumentos es el mencionado déficit explicativo inherente a toda teoría fisicalista de la mente.

Por su parte, el argumento del hiato explicativo se desarrolla a partir de la consideración de una presunta asimetría entre dos tipos de enunciados de identidad: el primer tipo, a los que en 1983 llama “completamente explicativos” (*fully explanatory*), estarían ejemplificados por enunciados como (I) “Agua = H₂O” o “calor = movimiento de moléculas”; el otro, por enunciados como (II) “dolor = disparo de fibras C”, donde aparece un término (en nuestro ejemplo, “dolor”) que hace alusión a una experiencia consciente.

En definitiva, esta diferencia radicaría en lo siguiente: por un lado, con respecto a (II) “es igualmente concebible que pudiera existir dolor sin disparo de fibras C, y disparo de fibras C sin dolor”. En otras palabras, es posible concebir sin contradicción un escenario en el cual se instancie la sensación de dolor sin que se instancie el disparo de las fibras C y viceversa. Sin embargo, de acuerdo con Levine, si bien es posible concebir otra sustancia que no sea H₂O que exhiba las propiedades macro del agua, es inconcebible que se instancie H₂O sin que se manifiesten dichas propiedades (Levine, 1997/1993:548). En este caso, el carácter contingente de la identidad no funciona en ambas direcciones.

Al mismo tiempo, esta diferencia refleja o tiene su origen en una “diferencia epistemológica importante” entre las pretendidas reducciones del agua a H₂O y del dolor a disparo de fibras C, a saber: que la teoría química del agua explica lo que necesita ser explicado, mientras que la teoría fisicalista de los *qualia* sin embargo ‘deja algo afuera’. Y es precisamente por este motivo por el cual resulta concebible “que una criatura pudiera ocupar el estado físico o funcional relevante [es decir, aquel que se supone que se identifica con algún estado fenoménico] y sin embargo no experimentar el carácter cualitativo [del dolor]” (Levine, 1997/1993:548).

En resumen: en el primer caso, todo lo que hay que explicar acerca del calor o del agua, es explicado mediante una explicación reductiva y, como veremos, podría decirse que las identidades en cuestión aparecen como conclusión de una explicación de corte deductivo. Por esta razón, en estos casos la explicación en cuestión no “deja nada crucial afuera”. En particular, como veremos, lo que es y necesita ser explicado en los casos considerados, de acuerdo con Levine, son las propiedades macro físicas ordinarias atribuidas al agua y al calor. Por el contrario,

en el segundo tipo de enunciados de identidad, sí quedan residuos inexplicados. En el caso particular del dolor, si bien puede ser posible explicar sus causas y efectos físicos, lo que no parece ser posible es explicar el carácter cualitativo del dolor, es decir, por qué el dolor se siente como se siente. Veamos ahora por qué esto es así.

Empecemos con la identidad “temperatura = movimiento de moléculas o mke”. De acuerdo con Levine, el significado de nuestro concepto ordinario o de sentido común de “calor” se *agota* (este punto es clave) en la caracterización de cierto rol causal. Por lo tanto, una paráfrasis del concepto podría ser la siguiente: “el fenómeno que experimentamos a través de sensaciones de calor y frío, que es responsable de la expansión y la contracción del mercurio en los termómetros, que causa que algunos gases aumenten y otros disminuyan, etc.,” (Levine, 1983: 355). Al mismo tiempo, “nuestro conocimiento de química y física hace inteligible cómo algo como el movimiento de moléculas podría jugar el rol causal que asociamos con la temperatura”, y dado que, como dijimos, según Levine, dicho rol causal agota nuestro concepto de “temperatura”, “una vez que entendemos cómo este rol causal se lleva a cabo no hay nada más que necesitamos comprender” (Levine, 1983: 357). Así, una historia similar podría contarse con respecto al agua.

Como puede verse, el carácter explicativo de las identidades de tipo (I) se basa en dos afirmaciones. La primera, (1) que los conceptos en consideración (“calor” y “agua”) se agotan en la especificación de cierto rol causal. La segunda, (2) que la física y la química hacen inteligible precisamente cómo ciertas sustancias satisfacen ese rol o bien cómo dan lugar a ciertas propiedades macro físicas de los fenómenos en cuestión expresadas mediante oraciones del lenguaje cotidiano. En este sentido, Levine define explícitamente los dos pasos fundamentales que deben llevarse a cabo para realizar una explicación reductiva exitosa: “La etapa 1 implica el (¿relativo? ¿cuasi?) proceso *a priori* de poner el concepto de la propiedad que va a ser reducida ‘en forma’ para la reducción por medio de la identificación de rol causal para el cual estamos buscando el mecanismo subyacente [que lo realiza]. La etapa 2 implica el trabajo empírico de descubrir justamente cuáles son esos mecanismos subyacentes” (Levine, 1997/1993: 551).²¹

²¹Estos pasos que aparecen explícitamente detallados en el artículo del 93, aparecen implícitamente en su artículo del 83 tal como se desprende de las citas en el cuerpo del texto.

De esta manera, para cualquier *explanandum* propuesto, esto es, para cualquier propiedad o rol causal específico que se pretenda explicar de un fenómeno dado, podrá derivárselo del *explanans* de la siguiente forma. Supongamos que queremos explicar que el agua hierve a los 100° C. Este será nuestro *explanandum*. Veamos ahora los pasos de la explicación.

- 1) Mediante análisis conceptual *a priori* sabemos cuál es el rol causal del agua (hervir a los 100° C, congelarse a los 0° C, ser líquida a temperatura ambiente, etc.).
- 2) Mediante investigación (química, en este caso), sabemos que H₂O satisface el rol causal asignado al agua.
- 3) La sustancia que satisfaga el rol casual del agua es agua (porque nuestro concepto de “agua” se agota en la realización de ciertos roles causales).
- 4) De 2 y 3 se sigue que H₂O es agua.
- 5) H₂O hierve a los 100° C. (Como anticipé, la identidad “agua = H₂O” aparece como la conclusión de las premisas anteriores).
- 6) Por lo tanto, de 4 y 5 se sigue que el agua hierve a los 100° C. Y así llegamos a dar cuenta del *explanandum* propuesto.

En contraposición, identidades como dolor = disparo de fibras C (y, en general, todas las que incluyan en uno de sus costados una propiedad fenoménica) no resultarán jamás “completamente explicativas”. En primer lugar, porque si bien asociamos al dolor (o la experiencia fenoménica relevante), con diversos roles causales, (como por ejemplo, su capacidad de ocasionar reacciones tendientes a evitar aquellas situaciones que se cree que traerán dolor), y si bien consideramos también que “el rol causal del dolor es crucial para nuestro concepto de él”, y que es algo que querríamos y que en principio podríamos explicar, el punto es que dicho rol causal *no agota* el significado del concepto: “hay más en nuestro concepto de dolor que su rol causal, está su carácter cualitativo, cómo se siente; y lo que queda inexplicado por el descubrimiento del disparo de las fibras C *es ¡por qué el dolor debería sentirse del modo en que se siente!* (Levine, 1983:357. La cursiva es del autor).

Por este motivo, es decir, por la imposibilidad de cumplir con el paso (1) de toda explicación reductiva exitosa que consigné arriba (esto es, “la funcionalización” del concepto del fenómeno a explicar), y bajo la consideración de que todo lo que

puede ser explicado en términos de las teorías físicas y químicas son roles causales, una explicación fisicalista de los *qualia*, similar a la que esboqué de cierta propiedad del agua, estará siempre condenada al fracaso: “Lo que parece ser responsable del hiato explicativo [...] es el hecho de que nuestros conceptos del carácter cualitativo no representan, [...] roles causales” (Levine, 1997:553).

Al mismo tiempo, esto explica la asimetría entre ambos tipos de enunciados que indiqué más arriba y que constituía el punto desde el cual partía el argumento del hiato explicativo (esto es, el hecho de que si bien podemos concebir “que algo diferente a H₂O manifieste las propiedades superficiales macro del agua, [...] no es concebible que H₂O no tuviera las propiedades macro del agua” y sí en cambio, “es igualmente concebible que pudiera existir el dolor sin disparo de fibras C, y disparo de fibras C sin dolor”). Esta misma idea Levine la expresa de la siguiente manera:

...dada una elaboración suficientemente rica [...] es inconcebible que H₂O no hierva a los 212° F al nivel del mar (asumiendo, otra vez, que mantenemos el resto del mundo químico constante). Pero ahora contrastemos esta situación con una reducción física o funcional de algún estado sensorial consciente. No importa cuán rica sea la historia de procesamiento de información o neurofisiológica, siempre parecerá coherente imaginar que todo continuara igual sin que haya algo que sea como estar en ese estado en cuestión. Sin embargo, si la historia física o funcional realmente explicara el carácter cualitativo, no debería ser tan claramente imaginable que los *qualia* estuvieran ausentes (Levine, 1997/1993: 549)²²

Como se desprende de estas citas, Levine considera que en una “explicación reductiva” adecuada, si se acepta la verdad del *explanans*, debe ser inconcebible que el *explanandum* sea falso. “En otras palabras, alcanzamos una comprensión cuando

²² Esta misma idea es expresada de una forma ligeramente distinta en el artículo de 1983 con respecto al calor:

“Si no hay nada que podamos determinar acerca del disparo de las fibras C que explique por qué estos disparos tienen el carácter cualitativo que tienen —o, para ponerlo de otra manera, si lo que particularmente se siente tener los disparos de las fibras C no es explicado, o hecho inteligible, por la comprensión de las propiedades físicas o funcionales de los disparos de las fibras C— se vuelve inmediatamente imaginable que haya disparo de fibras C sin la sensación de dolor, y viceversa. No tenemos esa intuición correspondiente en el caso del calor y el movimiento de moléculas [...] porque lo que sea que haya que explicar acerca del calor es explicado por ser el movimiento de moléculas.” (Levine, 1983:359). Como puede verse, otra vez la idea es que la falta de una conexión explicativa adecuada entre el presunto *explanans* y el *explanandum* para el caso de las explicaciones fisicalistas de las experiencias conscientes, permite concebir fácilmente escenarios en los cuales se da el primero pero no el segundo.

podemos ver por qué, dada la información citada en el *explanans*, el fenómeno citado en el *explanandum tiene que darse*" (Levine, 2001: 74. *La cursiva es mía*). Además, Levine sostiene (explícitamente en su artículo del 1993) que la relación entre ambos *relata* es deductiva: "la explicación se supone que involucra una relación deductiva entre *explanans* y *explanandum*" (Levine, 1997/1993: 550). Así, parte del esfuerzo argumentativo de Levine consiste precisamente en mostrar que las teorías fisicalistas de la consciencia no pueden satisfacer estas exigencias. En particular, en el caso de la identidad psicofísica, la relación entre el *explanans* (disparo de fibras C y la porción de la teoría neurofisiológica relevante) y el *explanandum* (el aspecto cualitativo del dolor) no es aquella que debería ser de acuerdo con los criterios explicativos aceptados por el autor.²³

²³ Para ser justos, debe señalarse que en la primera versión de Levine del argumento del hiato explicativo en su artículo de 1983, no sostiene que la relación entre el *explanans* y el *explanandum* deba ser deductiva. Lo que dice es que la presencia del fenómeno descrito en el primero debe "hacer inteligible" el surgimiento o la presencia del fenómeno descrito en el segundo. Sin embargo, el propio autor reconoce que no ofrece una definición clara de este término. En este sentido, podría objetarse que hasta que no se brinde una caracterización más o menos precisa de este concepto, no resulta claro por qué una teoría fisicalista que pretendiera dar cuenta de la experiencia consciente no podría satisfacer esa exigencia.

Otra cuestión que no resulta del todo clara en la presentación "original" del argumento del hiato explicativo en el artículo mencionado es la relación de este argumento con el argumento de Kripke en contra de la teoría de la identidad psicofísica. De acuerdo con Levine, su propósito en ese texto es transformar el argumento metafísico de Kripke en un argumento epistemológico (Levine, 1983: 354). En otras palabras, la idea es intentar ofrecer un argumento que muestre las limitaciones explicativas del materialismo o fisicalismo a la hora de dar cuenta de la consciencia sin que, al mismo tiempo, se favorezca algún tipo de dualismo metafísico.

Por mi parte, considero que el argumento del hiato explicativo es independiente del argumento de Kripke y el cuerpo del texto ofrezco mi reconstrucción del mismo. De todos modos, para comprender la relación que Levine traza entre ambos argumentos, observemos en primer lugar la reconstrucción que hace del argumento de Kripke: "Todos los enunciados de identidad que tienen designadores rígidos [es decir, expresiones que refieren a una misma entidad en todos los mundos posibles en los que esa entidad existe, y que no designan nada en aquellos mundos en los que no existe] en ambos lados del signo de identidad, si son verdaderos, son verdaderos en todos los mundo posibles donde los términos tienen referencia; segundo, las afirmaciones de identidades psicofísicas pueden concebirse como falsas, y por lo tanto, por la primera afirmación, son realmente falsas." (Levine, 1983:354). Consideremos ahora la diferencia entre los siguientes enunciados de identidad.

(1) Dolor es disparo de fibras C.

(2) Calor igual movimiento de moléculas.

Si ambas son identidades verdaderas, son necesariamente verdaderas. Pero, si son necesariamente verdaderas, significa que "no hay un mundo posible en el cual sean falsas" (Levine, 1983:355). Por lo tanto, concebir un mundo en el cual hay temperatura pero no movimiento de moléculas o un mundo en el cual hay dolor sin disparo de fibras C, implicaría concebir un mundo lógicamente imposible. Sin embargo, estos escenarios parecen perfectamente coherentes. Es decir, parece perfectamente posible concebir que sean falsos sin caer en ningún tipo de contradicción. Pero si esto es así, y si se acepta la afirmación kripkeana de que si una identidad es verdadera, es necesaria, deberíamos concluir entonces que ambas son falsas.

Sin embargo, Kripke considera que la "contingencia aparente" o la "ilusión de contingencia" del enunciado (2) puede ser disipada una vez que nos damos cuenta de que, en realidad, el enunciado parece contingente (es decir, no necesariamente verdadero) cuando, en lugar de la temperatura, lo

que imaginamos es, por ejemplo, algún fenómeno que afecta nuestros sentidos en el modo en que lo hace la temperatura, pero que no es el movimiento de moléculas que constituye auténticamente la referencia del concepto. Este último escenario sí es realmente contingente. En otras palabras, “lo denotado efectivamente por el término de <<calor>> es (necesariamente) <<energía cinética media de las moléculas>>. Pero nuestro mundo podría haber sido distinto, en el sentido de que lo que nos produce la sensación de calor podría haber sido la luz [...]. Y esta contingencia explica la ilusión de contingencia que se produce en el caso del enunciado necesario <<calor = energía cinética media de las moléculas>>. Hay un hiato (contingente) entre la esencia del término de clase natural y aquellas manifestaciones de esa esencia que dependen del hecho de que tengamos el aparato perceptivo que tenemos; esto nos hace pensar que la esencia era contingente, cuando lo contingente son nuestras capacidades cognitivas y perceptivas” (Pérez, 2013: 200).

En contraposición, no es posible dar una explicación similar para el caso de (1) debido a que no podemos hacer la distinción entre el modo en que se nos aparece el dolor y el fenómeno mismo, porque, de acuerdo con Kripke, la experiencia de dolor, la sensación de dolor, es el dolor mismo. Es decir, en el caso del dolor, a diferencia del de la temperatura, lo que conocemos es el dolor mismo: no hay en este caso una distancia entre la base evidencial en virtud de la cual atribuimos dolor y el dolor mismo. Por lo tanto, en la medida en que no podemos dar cuenta de la intuición de que (1) es contingente, habría que concluir que (1) es falsa, esto es, que no representa una verdadera identidad y que es por eso que nos resulta contingente. En este punto, Levine señala que se puede desestimar este argumento una vez que reconocemos el abismo que existe entre una posibilidad conceptual y una posibilidad metafísica. Después de todo, es el propio Kripke quien nos advierte acerca de lo temerario de dar ese paso. “Puesto que una posibilidad epistemológica no es suficiente para una posibilidad metafísica, el hecho de que lo que es intuitivamente contingente resulte ser metafísicamente necesario no debería molestarnos terriblemente. Por esto es que pienso que el argumento de Kripke no es enteramente exitoso.” (Levine, 1983:356).

Sin embargo, Levine considera que hay otra diferencia entre ambos tipos de enunciados que sí debería preocupar a los materialistas y que, a su juicio, explicaría la vulnerabilidad de los enunciados de tipo (1) frente a los argumentos como el de Kripke. Esta diferencia no es otra que aquella a la que hace referencia Levine cuando sostiene que las identidades de tipo (2), a diferencia de las identidades de tipo (1) son “completamente explicativas”. Esta diferencia es la que detallé en el cuerpo del texto y, como vimos, consiste básicamente en que en el caso de las identidades no psicofísicas, es posible explicar las propiedades macro y roles causales atribuidos al agua o al calor a partir de la porción relevante de la teoría química o física de manera tal que una vez dado el *explanans* es inconcebible que no se dé el *explanandum*. Por el contrario, este tipo de explicación no estaría disponible para las propiedades cualitativas de la experiencia.

Ahora bien, como dije, Levine afirma que el carácter contingente de ambos tipos de enunciados que señala Kripke, así como también el hecho de que esa presunta contingencia no pueda “disiparse” para el caso de las identidades psicofísicas, se debe precisamente a que éstas últimas no son “completamente explicativas” (en el sentido apuntado). Por mi parte, no estoy seguro de que esta explicación de Levine sea correcta. De acuerdo con las distinciones mencionadas del planteo de Kripke, el enunciado “el rol causal del calor = movimiento de moléculas”, es un enunciado *realmente* contingente, en la medida en que, para Kripke, la descripción involucrada en él en el lado izquierdo no expresa la esencia del fenómeno en cuestión (calor o agua) sino ciertas propiedades macro (contingentemente) asociadas con él. Una identidad como esta resultaría verdadera para Kripke porque el referente es el mismo para los dos conceptos a los costados del signo de identidad pero no porque sea necesario que el movimiento de moléculas tenga las propiedades macro que tiene, ni porque satisfaga ciertos roles causales específicos. Por este motivo, es igualmente concebible, para Kripke, un escenario en el cual se instancia H₂O y no tales y tales propiedades macro del agua y un escenario en el cual se instancian tales y tales propiedades macro del agua sin que haya H₂O. En este sentido, este enunciado de identidad así entendido está a la par con el de “dolor = disparo de fibras C”. Pero si esto es así, no veo cómo las diferencias que señala Levine entre ambos enunciados explican las diferencias que señala Kripke entre ambos enunciados ya que, como dije, de acuerdo con el autor de *El nombrar y la necesidad*, los enunciados “tales y tales roles causales atribuidos al calor = movimiento de moléculas” y “dolor = disparo de moléculas” son igualmente contingentes, mientras que para Levine, el primero, y no el segundo, constituye una identidad “completamente explicativa”. Sencillamente, parece que la diferencia que indica Levine (sea adecuada o no) no es la que señala Kripke ni puede explicarse la última a partir de la primera. En su artículo del 93 Levine reformula estas cuestiones y admite que si bien de acuerdo con Kripke el enunciado “La sustancia que

En este último caso, la relación entre ambos resulta “ininteligible” o “arbitraria”. Por este motivo, aun cuando aceptemos, por ejemplo, que la sensación de dolor se identifica con los disparos de las fibras C, (es decir, aun cuando se acepte algún tipo de identidad entre los diferentes tipos de experiencias conscientes y ciertos estados físicos / funcionales), no queda claro por qué dichos estados están asociados o se identifican con determinados estados fenoménicos y no con otros, esto es, por qué, pongamos por caso, el disparo de fibras C *es o está asociado con* el dolor y no con la experiencia de oír una campanada.

En este punto, Levine considera una posible objeción a su planteo. De acuerdo con ella, se podría intentar responder al argumento sosteniendo que la identidad en cuestión (dolor = disparo de fibras C) expresa un hecho primitivo del universo, tan “arbitrario” como el valor de la constante gravitacional, que no es derivado de ninguna ley más básica. Simplemente, así es como son las cosas. Ahora bien, si esto es así, ¿por qué no aceptar, de la misma manera, que el dolor = disparo de fibras C es un hecho bruto, primitivo? ¿Por qué exigir explicaciones adicionales?

Según Levine, esta respuesta no resulta adecuada, ya que si bien es correcto adoptar esta actitud con respecto a la constante gravitacional, no ocurre lo mismo con el caso de las experiencias conscientes. La razón de esto es que la consciencia surge a nivel macroscópico, y el materialismo, a su juicio, implica un reduccionismo explicativo al menos de este tipo mínimo: que para cada fenómeno no describible en términos de magnitudes de la física fundamental (cualesquiera resulten ser), hay un mecanismo que es describible en términos de las magnitudes físicas fundamentales de manera tal que las apariciones del primero son inteligibles en términos de las apariciones del segundo (Levine, 1983:359). Así, se supone que no podría haber “hechos brutos” en el nivel macroscópico de la psicología (volveré sobre este punto en la sección siguiente).

Ahora bien, una vez presentada esta reconstrucción del argumento “original” del hiato explicativo, creo que puede decirse que concuerda bastante bien con “la historia convencional” (Papineau, 2011) del mismo, es decir, con aquella “versión” aceptada por la mayoría de los filósofos contemporáneos en sus discusiones al

manifiesta tales y tales propiedades macro del agua = H₂O” (similar al del calor) y “dolor = disparo de fibras C” son igualmente contingentes, hay sin embargo una diferencia entre ambos (vinculada con los diferentes tipos de concebibilidad mencionados en el cuerpo del texto) que definitivamente no es la que señala el autor de *El nombrar y la necesidad*.

respecto. De acuerdo con ella, se supone que el hiato explicativo surge porque no podemos *derivar* los hechos acerca de los estados conscientes *a priori* de los hechos físicos, del mismo modo en que podemos derivar, por ejemplo, hechos acerca del agua, y hechos acerca de la temperatura desde los hechos físicos:

En toda la literatura sobre el hiato explicativo, se asume que la explicación faltante en cuestión es deductiva (Taylor, 2015: 2).

No es posible deducir *a priori* el hecho de que hay experiencia visual de rojo en un lugar específico del hecho de que el estado cerebral B está en ese lugar. Por eso es que hay un hiato explicativo (Schroer, 2010: 506).

Con todo, esta relación de “deducción *a priori*” no supone que podamos saber *a priori* que el agua es H₂O a partir de un mero análisis de los conceptos involucrados. La idea es que si tuviéramos el conocimiento *a posteriori* de todos los hechos físicos, incluyendo los hechos físicos acerca de H₂O, y si poseyéramos además el concepto ordinario de agua, al que se considera como un concepto descriptivo o que hace referencia a su objeto a partir de la atribución de ciertos roles causales, entonces podríamos ser capaces de determinar o concluir que el agua es H₂O sin precisar ninguna información adicional, en la medida en que nos percatamos que H₂O satisface las descripciones y roles causales atribuidos al agua por nuestro concepto ordinario o pre-teórico.

En resumen, podríamos sintetizar la versión estándar o convencional del argumento del hiato explicativo de la siguiente manera. La diferencia entre un enunciado (1) “agua = H₂O” y (2) “la experiencia visual de rojo = estado cerebral B” es que es posible deducir *a priori* el hecho de que hay agua en un lugar específico del hecho de que hay H₂O en ese lugar porque el enunciado (3) “agua = la sustancia que tienes tales y tales propiedades macro-físicas” (como por ejemplo su punto de ebullición a cierta temperatura) es una verdad pre-teórica *a priori* (es decir, nuestro concepto de “agua” es un concepto que se agota en la sumatoria de ciertos roles funcionales y causales y, en general, de ciertas descripciones).

Así, si se determina que hay H₂O en determinado lugar (información que se obtiene a partir de algún tipo de indagación empírica), y se sabe además (3), entonces, dado que H₂O satisface las descripciones y roles causales que constituyen nuestro concepto de agua, entonces se sigue *a priori* que hay agua en ese lugar. En cambio, en el caso de la experiencia visual / estado cerebral, no hay una verdad

análoga porque no hay una verdad *a priori* de la forma “experiencia visual de rojo = D”, en donde “D” es una descripción de ciertas propiedades o de roles causales. O, en todo caso, dichos roles causales no agotan nuestro concepto de “experiencia visual”. Como consecuencia de esto, se sigue que no es posible derivar *a priori* que hay una experiencia visual de rojo en un lugar determinado del hecho de que se da el estado cerebral B. Y esto explica la presencia de una “brecha explicativa” en un enunciado como (2) que está ausente en enunciados como (1) (en esta reconstrucción de la “historia convencional de la brecha explicativa” seguí a Schroer, 2010).

3. Objeciones en contra del argumento del hiato explicativo

Frente a la historia estándar, se dirigen, como es de esperar, diversas críticas. En esta sección voy a consignar aquellas que, a mi juicio, resultan más contundentes. Consideradas en conjunto, nos permiten concluir que, al menos en su formulación convencional, el argumento analizado en este capítulo no logra mostrar que haya un problema explicativo particular con respecto a la consciencia.

Por otra parte, aunque las objeciones que voy a mostrar están íntimamente entrelazadas, voy a dividir las en tres grupos o tipos. El primer tipo de críticas que voy a considerar pone en duda que la relación de implicación que parece suponer Levine entre el *explanans* y el *explanandum* deba sostenerse para cualquier descripción del fenómeno a explicar. El segundo tipo de objeción que voy a desarrollar constituye, a mi parecer, la dificultad más importante que enfrenta el argumento del hiato explicativo. De acuerdo con esta crítica, la presunta asimetría entre los diferentes tipos de enunciados de identidad que señala el argumento de Levine no se sostiene cuando se las examina más de cerca. Si esta crítica es correcta, debería concluirse que no hay una diferencia sustancial entre ambos tipos de enunciados y, por lo tanto, todo el argumento de Levine se desmorona. En última instancia, como veremos, el partidario del argumento del hiato explicativo se enfrenta a una especie de dilema: o bien debe aceptar que hay un “hiato explicativo” en ambos casos, o bien debe reconocer que en ambos casos el hiato puede ser cerrado de la misma manera. Como resulta obvio, cualquiera de las opciones desbarata el argumento de Levine. Por último, mostraré aquellas objeciones que señalan que no es necesario adoptar el criterio por el cual, de acuerdo con el argumento del hiato explicativo, se justifica la aceptación de los enunciados de

identidad. Más específicamente, este tipo de objeción niega que el análisis conceptual que figura como el primer paso que debería darse para realizar una explicación reductiva exitosa sea realmente necesario. Este último punto es relevante para los objetivos generales de este trabajo.

En efecto, como dije en la introducción de este capítulo, aun cuando se acepte que hay conceptos legítimos acerca de la experiencia consciente que no la definen en términos de roles causales u otras propiedades en principio no problemáticas, esto no impide identificar al referente de dichos conceptos con propiedades físico / funcionales. Veamos ahora en detalle cada una de estas críticas mencionadas.

Como vimos, de acuerdo con el argumento de Levine, parte de la dificultad que había a la hora de intentar ofrecer una explicación de la consciencia, tenía que ver con el modo en que se caracterizaba a este fenómeno a través de lo que podríamos llamar nuestros conceptos de sentido común u ordinarios o de la psicología folk. En particular, la dificultad residía en el hecho de que los conceptos acerca de la experiencia consciente no parecían agotarse en la descripción de un determinado rol causal. Ahora bien, suele ocurrir, tal como apunta Carruthers 2004, que en las explicaciones científicas la caracterización original del fenómeno a explicar cambia drásticamente a lo largo del proceso de investigación. Es más: en muchos casos, el modo en que se plantea originalmente la pregunta que desata la investigación es lisa y llanamente eliminado porque se descubre que hace algún uso inapropiado de presuposiciones falsas o de conceptos confusos. En estos casos, simplemente se cambian las preguntas originales por otras.

Por ejemplo, en astronomía se comenzó con un problema que podría ser expresado a través de la siguiente pregunta “¿por qué las estrellas y el sol se mueven a través del cielo de la manera en que lo hacen?”. Sin embargo, la explicación dada por la ciencia no responde finalmente esa pregunta tal como es planteada al inicio. En lugar de eso, se da una explicación de la rotación de la Tierra, y de sus movimientos, así como también de los del sol, los otros planetas y las estrellas y, posteriormente, es en términos de estas cuestiones que se explica por qué el sol y las estrellas *parecen* moverse a través del cielo de cierto modo (Carruthers, 2004:164-165).

Sin ir más lejos, el propio Levine reconoce un escenario similar al recién mencionado en una nota al pie de su artículo del 93: “En algunos casos, por ejemplo,

con las propiedades tales como la liquidez y la masa, sería mejor pensar en sus articulaciones teóricas en la teoría física y química más como una cuestión de incorporación y refinamiento de los conceptos teóricos ordinarios que como una cuestión de reducción” (Levine, 1993/1997: 554).

Ahora bien, si estos casos de “refinamiento” e “incorporación” y, en definitiva, modificación, de los conceptos del lenguaje ordinario son tolerados como parte de explicaciones adecuadas, ¿por qué debería rechazarse de antemano una propuesta similar en torno a los conceptos ordinarios acerca de la experiencia consciente que morigere el carácter presuntamente insoluble del hiato explicativo? Quizás gran parte del problema con respecto a las cuestiones que estamos considerando es que se toma demasiado en serio la caracterización que se hace del fenómeno desde lo que se considera “el sentido común”, como si se tratara de una fuente inapelable.

Por otra parte, también puede objetarse que una teoría fisicalista esté comprometida necesariamente con la idea de que los hechos acerca de la experiencia consciente se deriven (en el sentido apuntado en la sección anterior) de una descripción de los hechos físicos relevantes. De acuerdo con esta objeción, debe reconocerse, en primer lugar, que una relación de deducción se da siempre entre enunciados y no entre hechos y, en segundo lugar, que un mismo hecho puede ser conceptualizado y descrito de muy diversas maneras. Por ejemplo: la oración “Más de la mitad de la superficie de la Tierra está cubierta de agua” expresa el mismo hecho que la oración “Más de la mitad de la superficie de la Tierra está cubierta de H₂O”. Sin embargo, mientras que resulta legítimo exigir que una explicación científica dé cuenta de la experiencia consciente en términos del vocabulario “físico”, no es razonable exigir que dé cuenta de ese mismo hecho bajo cualquier descripción; incluso bajo una descripción en términos de nuestro lenguaje ordinario. Una vez reconocida esta distinción en los conceptos involucrados, podrá solicitarse que se trace algún tipo de puente entre ambos que permita decir justificadamente que ambas descripciones refieren al mismo hecho, y que efectivamente se esté realizando una explicación de los mismos hechos, pero el punto que señalo sigue en pie.

En otras palabras, una manera de responder al argumento del hiato explicativo consiste en indicar que *toda* teoría explica un hecho dependiendo de cómo este hecho sea conceptualizado. En efecto, una explicación (sea o no

deductiva) es una relación entre proposiciones y, por lo tanto, no una relación entre ítems extra-lingüísticos y extra-conceptuales que no pueden entrar en relaciones de deducibilidad entre sí. Pero esto implica que aunque sea posible explicar un hecho cuando es caracterizado en términos de conceptos propios de la neurofisiología, pongamos por caso, no sea posible sin embargo derivar del *explanans* descripciones del mismo hecho expresadas en términos de conceptos fenoménicos. Por este motivo, no puede inferirse un hiato explicativo (ni mucho menos un hiato ontológico) de la imposibilidad de deducir y explicar ciertos hechos conceptualizados de cierta manera pero no de todas las maneras posibles, porque esto es común a toda explicación científica. En parte, esta es la respuesta que siguen algunos de los partidarios de la llamada estrategia de los conceptos fenoménicos como Papineau 1998, 2002, 2007, Levin 2008 o Carruthers 2004.

Vinculado con esta crítica, existe un problema todavía más grave para el argumento del hiato explicativo que pone en cuestión la presunta asimetría entre los dos tipos de enunciados de identidad (“agua = H₂O” y “dolor = disparo de fibras C”) de la que éste parte. Como vimos, dicha asimetría consistía, de acuerdo con Levine, en que mientras que resultaba posible concebir “que algo diferente a H₂O manifestara las propiedades superficiales macro del agua, [...] no es concebible que H₂O no tuviera las propiedades macro del agua” (Levine, 1997:548). En cambio, “es igualmente concebible que pudiera existir el dolor sin disparo de fibras C, y disparo de fibras C sin dolor”. Esta asimetría, proseguía el argumento, era el resultado “una diferencia epistemológica importante” entre ambos enunciados, a saber: “que la teoría química del agua explica lo que necesita ser explicado, mientras que una teoría fisicalista de los *qualia* todavía ‘deja algo afuera’” (Levine 1997/1993:548). Al mismo tiempo, esta diferencia epistemológica explicaba a su turno la imposibilidad de cumplir con los dos pasos que debían cumplir las explicaciones reductivas adecuadas: el análisis *a priori* de los roles causales que agotan el concepto del fenómeno que se pretende reducir / explicar y luego el descubrimiento *a posteriori* del mecanismo que satisface esos roles causales.

Ahora bien, el problema es que una vez que se reconoce que “la teoría química [en la cual se describe cómo cierto mecanismo subyacente realiza ciertos roles causales o satisface ciertas descripciones] y la teoría de sentido común (*folk psychology*) [en la cual se describe el fenómeno a explicar /reducir] no tienen un

vocabulario idéntico” y, además, se reconoce la necesidad de introducir “principios puente”, ya no resulta obvio que la presunta asimetría con respecto a la concebibilidad antes mencionada se sostenga.

En efecto, supongamos que quiero explicar por qué el agua hierve, o se congela, a las temperaturas en que lo hace. Para obtener una explicación de estos hechos [expresados con los conceptos propios del lenguaje ordinario] necesitamos una definición de ‘hervir’ y ‘congelar’ que lleve a estos términos al vocabulario registrado de las teorías a las que se apela en la explicación [es decir, al vocabulario de la química y la física] (Levine, 1997/1993:550-551).

Ahora bien, si esto es así, llegamos al siguiente resultado: en primer lugar, observamos que lo que realmente es explicado o deducido de la teoría química relevante no es el enunciado “el agua hierve”. Porque ni el concepto de “agua” ni el concepto de “hervir” pertenecen al vocabulario de la química. En todo caso, lo que explica la teoría en cuestión es cómo o por qué H₂O se comporta de cierta manera, por ejemplo, por qué muestra cierta conducta molecular B en cierta energía cinética E. Sin la introducción de los “principios puente” nunca llegaríamos desde nuestro conocimiento acerca de H₂O a la descripción de alguna de las propiedades macro del agua realizada en términos del lenguaje ordinario.

Pero una vez que se reconoce la diferencia de conceptos utilizados en el *explanandum* (“el agua hierve a los 100°”) y el *explanans* (los aspectos relevantes de la teoría química / física), parece posible concebir, no sólo un escenario en el cual se instancian las propiedades macro del agua sin que haya H₂O, (tal como admitía Levine): también parece perfectamente posible concebir, en contra de lo que sostenía Levine en el párrafo citado, un escenario en el cual hay H₂O pero en el que no se instancian las propiedades macro del agua *descriptas con los conceptos del lenguaje ordinario*, de la misma manera en que podemos concebir dolor sin disparo de fibras C y viceversa. El único modo en que resulta inconcebible que se dé H₂O sin que se instancien también las propiedades macro del agua descriptas con los conceptos del lenguaje ordinario es si se aceptan los principios puente relevantes que permitan vincular los dos vocabularios (así, por ejemplo, estos principios podrían decir cosas como las siguientes: “conducta molecular B es hervir”, “energía cinética molecular es calentar”, etc.). Pero si esto es así, también se podrían

establecer principios puente análogos para el caso de la identidad “dolor = disparo de fibras C” y se obtendrían los mismos resultados que para el caso del agua.

En efecto, supongamos que queremos explicar una propiedad del dolor F. La explicación referirá a alguna propiedad G del disparo de las fibras C. Como en el caso del agua/ H₂O, tendríamos un “hiato explicativo” hasta que introdujéramos los principios puente, esto es, podríamos concebir F y \neg G y viceversa. Sin embargo, una vez que se determine el principio puente F = G, el hiato explicativo se cierra. Pero si esto es correcto, no hay asimetría entre los dos tipos de enunciados de identidad y entonces el argumento del hiato explicativo no brinda buenas razones para pensar que la explicación de la consciencia encierre algún tipo de dificultad epistemológica especial.

El partidario de dicho argumento podría intentar responder a esta objeción argumentando, o bien que los principios puente, en el caso de las identidades no psicofísicas, se derivan de algún modo de los enunciados micro físicos de las teorías científicas relevantes, o bien que el análisis *a priori* de los conceptos no psicológicos del lenguaje ordinario permite trazar esos vínculos. Sin embargo, cualquiera de estas salidas resulta sospechosa. La primera opción podría descartarse justamente porque los conceptos ordinarios no forman parte del vocabulario científico ni parecen ser deducibles de éste. Y los principios en cuestión combinan ambos tipos de conceptos. Con respecto a la segunda opción, es claro que tampoco resulta plausible considerar que un concepto macro expresado en el lenguaje ordinario pueda ser analizado *a priori* en otros términos que no sean otros conceptos macro del lenguaje ordinario, y mucho menos en términos del lenguaje científico. De lo contrario, bastaría un análisis de sillón para obtener descubrimientos científicos.²⁴ En principio, lo más razonable parece suponer que los principios puente se aceptan

²⁴ Chalmers y Jackson 2001 defienden una idea similar a esta opción. De acuerdo con Elpidorou (2013) estos autores adhieren a lo que él llama “Principio de competencia conceptual”, según el cual, “para cualquier descripción microfísica *D* de un mundo *w*, un sujeto que está en posesión de, y es competente [en el uso de] un concepto macroscópico *C* es capaz de determinar a priori la extensión de *C*.” (Elpidorou 2013: 2). En pocas palabras, este principio afirma que un sujeto que posee, por ejemplo, el concepto macro de “agua”, es capaz de deducir *a priori* la extensión o referente de dicho concepto, si se le da suficiente información del contexto, e incluso si esa información les es suministrada en términos moleculares o atómicos (es decir, en el vocabulario científico). Así, este principio jugaría el mismo papel que juegan los principios puente de Levine y permitiría vincular ambos lenguajes, el cotidiano y el científico, sin caer en la crítica mencionada en el cuerpo del texto. Para una crítica convincente de esta posición sostenida por Chalmers y Jackson ver (Díaz-León 2011 y Elpidorou 2013).

en virtud de la observación de ciertas relaciones de correlación entre los fenómenos descriptos en el lenguaje científico y el fenómeno descrito, a nivel macro, en el lenguaje cotidiano (Hüttemann, 2004: 69).

Pasemos a considerar ahora los reparos que suscitan los criterios según los cuales deberían aceptarse los enunciados de identidad de acuerdo con el planteo del argumento del hiato explicativo. Como apunté, el tipo de explicación reductiva que desarrolla Levine, parece suponer implícitamente que si una propiedad F puede ser reductivamente explicada en términos de G, entonces esto justifica que asumamos que F y G son idénticas. Por el contrario, si F no puede ser explicada reductivamente por G, parece que no tenemos razones para creer que son idénticas:

lo que nos justifica en basar la identificación del agua con H₂O en la responsabilidad causal de H₂O por el comportamiento típico del agua es que nuestro mismo concepto de agua es el de una sustancia que juega tal y cual rol causal. [...] Cuando encontramos la estructura que en este mundo ocupa ese rol, entonces tenemos el referente de nuestro concepto (Levine, 1997: 550).

Así, el análisis “funcional” de los conceptos de los fenómenos a explicar y reducir aparece como una condición necesaria para llevar adelante esa empresa y para justificar las identidades en cuestión. La imposibilidad de cumplir con este requisito sería lo que impediría (por lo menos en parte) tener explicaciones adecuadas de las experiencias conscientes. Sin embargo, este punto es precisamente otro de los supuestos al que muchos de los críticos del argumento del hiato explicativo dirigen sus ataques.

A partir de los trabajos de Papineau 1998 y Block y Stalnaker 1999 entre otros, se comienza a hacer una distinción, dentro de las discusiones en torno al problema del hiato explicativo, entre dos preguntas en relación con los enunciados de identidad. De acuerdo con esta distinción, podemos preguntar lícita e inteligiblemente, por un lado, por los *motivos* o las *razones* por los cuales creemos que una identidad es verdadera. Y en este caso, lo que se estaría solicitando básicamente es una justificación para sostener una creencia o una afirmación. Por otro lado, podemos preguntar por qué F es idéntico a G. Según los autores mencionados, esta última pregunta no tiene sentido porque equivale a preguntar por qué algo es idéntico a sí mismo. La única respuesta posible a esta pregunta sólo puede consistir simplemente en afirmar que así es como son las cosas, y eso es todo.

Papineau y Block y Stalnaker expresan esta idea diciendo que las identidades no necesitan explicación. Por este motivo, de acuerdo con esta posición, una vez que se acepta un enunciado de identidad hay ciertas preguntas que parecerían quedar clausuradas. Por ejemplo, si se asume que dolor = disparo de fibras C, “Podemos pedir una explicación de qué es lo que muestra que los dolores son idénticos a los disparos de fibras C” [...]. “Pero no tiene sentido pedir una explicación adicional acerca de por qué los dolores surgen del disparo de fibras C. Si son disparos de fibras C, no “surgen” de ellos, son ellos, y eso es todo” (Papineau, 1998: 379).

En definitiva, desde la perspectiva de los autores señalados, cuando preguntamos por qué $F = G$, lo que podemos querer saber en realidad son las siguientes cosas:

(a) qué *razón* tenemos para pensar que la identidad es verdadera; qué es lo que *muestra* que es verdadera.

Así, por ejemplo, alguien que no supiera que el nombre del humorista gráfico e historietista Quino es un pseudónimo de Juan Salvador Lavado Tejón y un día le informaran que en realidad se trata de la misma persona, podría solicitar alguna evidencia que pruebe que esto es realmente así. Como respuesta a este requerimiento podría señalársele, entre otros hechos, que siempre aparecen en el mismo lugar al mismo tiempo o que todos los hechos de la vida de ambos coinciden. Esta situación se aplica especialmente para el caso de dos *nombres propios* diferentes.

En estos casos, Block y Stalnaker 1999 consideran que la aceptación de ciertas identidades se justifica, no por ser derivadas de otros enunciados, sino por su poder explicativo. Para mostrarlo, en el artículo mencionado nos piden que imaginemos una reunión de dos grupos de historiadores de un futuro distante que estudian, respectivamente, la vida de Mark Twain y la de Samuel Clemens, sin saber que se trata en realidad de la misma persona. Sin embargo, al poco tiempo comienzan a darse cuenta con sorpresa que los eventos de la vida de Mark Twain y de Samuel Clemens coinciden punto por punto. Y se preguntan qué podría explicar semejante coincidencia. La respuesta, se dan cuenta, es que Mark Twain es Samuel Clemens. En este caso, la postulación de la identidad permite explicar la correlación y coincidencia exacta de dos series (presuntamente distintas) de eventos. Así, la conclusión que deberíamos sacar es que la identidad no es ella misma explicada sino

que *brinda o posibilita* una explicación. En esta línea, Block y Stalnaker consideran que es precisamente este poder explicativo el que justifica la aceptación de ciertas identidades en la ciencia y no, como cree Levine, el hecho de que pueda deducírselas *a priori* de la totalidad de los hechos físicos y de cierto análisis conceptual de nuestros conceptos de sentido común.

¿Por qué suponemos que el calor = energía cinética molecular? [...] Supongamos que el calor = energía cinética molecular, presión = transferencia de energía cinética molecular, y hervir = un cierto movimiento molecular [...] Entonces tenemos una descripción de cómo calentar el agua produce el hervor. Si aceptáramos correlaciones en lugar de identidades, solo podríamos tener una descripción de cómo algo correlacionado con el calentamiento causa algo correlacionado con el hervor. Además, podríamos desear saber cómo es que incrementar la energía cinética molecular de una bolsa de agua causa que hierva. Las identidades permiten una transferencia de fuerza explicativa y causal no permitida por las meras correlaciones. Asumir que calor = mke, que presión = transferencia de energía cinética, etc. nos permite explicar hechos que no podríamos explicar de otra manera. Por lo tanto, estamos justificados por el principio de inferencia a la mejor explicación a inferir que estas identidades son verdaderas.

Si creyéramos que el calor está correlacionado pero no es idéntico a la energía cinética molecular, consideraríamos como legítima la pregunta de por qué la correlación existe [...]. Pero una vez que nos damos cuenta que el calor es la energía cinética molecular, preguntas como estas serán vistas como equivocadas (Block y Stalnaker, 1999:22-23).

Como puede verse, este tipo de movida argumentativa intenta probar, en contra del modelo de explicación reductiva de Levine, que no es necesario un análisis funcional de los conceptos relevantes ni aparecer como la conclusión de un argumento deductivo (del modo en que lo mostré antes) para que pueda aceptarse la verdad de un enunciado de identidad: al parecer, los enunciados de este tipo se aceptan, al menos en ciertas ocasiones, por su capacidad para facilitar explicaciones, o como el resultado de una inferencia a la mejor explicación. Ahora bien, si esto es así, la resistencia que ofrecía el argumento del hiato explicativo para la aceptación de las identidades psicofísicas (como dolor = disparo de fibras C) se debilita.

Por otra parte, el último párrafo citado del artículo de Block y Stalnaker muestra de qué manera podría superarse el escollo que representaba la diversidad de lenguajes involucrados en una explicación (el científico y el del sentido común) a la hora de aceptar los enunciados de identidad que mencioné con anterioridad. Como vimos, resultaba dudoso considerar que alguien pudiera establecer, *únicamente* a través del conocimiento científico relevante y del análisis *a priori* de los conceptos del sentido común, enunciados de identidad que involucraran ambos tipo de conceptos. Sin embargo, si se acepta la propuesta de estos autores, los enunciados de identidad pueden justificarse *a posteriori* por su capacidad explicativa.²⁵

(b) En otras ocasiones, cuando preguntamos por qué $F = G$, lo que podemos querer saber en realidad es por qué se puede decir que una entidad u objeto satisface una cierta descripción o cumple con cierto rol causal. En cierta forma, otra vez lo que se pide es una justificación para aceptar una identidad, sólo que en este caso en el costado derecho del signo “=” aparece una descripción o la expresión de cierto rol causal. De acuerdo con Papineau 1998, tiene sentido preguntar por qué, por ejemplo, Tony Blair fue el primer ministro británico. Y se podría responder esta pregunta diciendo que lo fue porque recibió más votos en la elección, o porque el partido Laborista pensó que era el candidato con mayor posibilidad de vencer a los Conservadores. Nuevamente, la idea es que esta pregunta es inteligible no porque a veces tenga sentido preguntar por qué algo es idéntico a sí mismo, sino que cuando a uno de los costados del signo “=” aparece una descripción, “podemos pedir una

²⁵ En una línea de argumentación similar, podríamos agregar algunas observaciones que hace el filósofo Ansgar Beckermann (2012) en un artículo acerca de la relación entre los enunciados de identidad y las explicaciones reductivas, donde critica las razones por las cuales, de acuerdo con Levine, se justifica la aceptación de la identidad “agua = H₂O”. Según Beckermann, dado que en realidad “agua” es un término de clase natural que refiere a una sustancia química, y que las sustancias químicas se definen por su estructura molecular, bastaría con descubrir que todas las muestras de agua están constituidas por H₂O para identificar los referentes de ambos conceptos, independientemente de si es posible o no explicar todas las propiedades superficiales del agua tal como parece suponer Levine (Beckermann, 2012: 72). Para este autor, la afirmación de la identidad entre agua y H₂O es más parecida a un enunciado como el siguiente: “las nubes son extensas cantidades de gotas de agua en el cielo formando una masa blanca, gris o negra” debido a que “agua” no remite, a su juicio, a un rol causal. De este modo, o bien debe aceptarse que la analogía entre las identidades psicofísicas con el caso del agua no es adecuada y entonces debe reformularse el argumento del hiato explicativo para que muestre algún inconveniente auténtico para las teorías fiscalistas de la mente, o bien debería aceptarse que de la misma manera que en el caso del agua, deberíamos acordar con que es posible justificar la aceptación de la identidad “dolor = disparo de fibras C” si descubrimos, por ejemplo, que dicho disparo ocurre cada vez que se instancia una experiencia de dolor.

explicación de por qué la entidad en cuestión satisface esa descripción, y a menudo estamos en posición de proveer una respuesta satisfactoria.” (Papineau, 1998: 379).

Ahora bien, de acuerdo con Papineau, parte de los problemas que suscita el argumento del hiato explicativo es que exige que los enunciados de identidad que involucran conceptos fenoménicos sean analizados como si fueran descripciones o como si refirieran a su objeto a través de descripciones. Si esto fuera así, deberíamos ser capaces de explicar por qué ciertos estados físicos satisfacen las descripciones asociadas *a priori* con nuestros conceptos de todos los días acerca de los estados conscientes. Pero como, al parecer, no podemos hacer esto, entonces se concluye que hay algo inexplicado de la relación entre el cerebro y la consciencia.

Sin embargo, para autores como Papineau (y muchos de los que adhieren a la llamada “estrategia de los conceptos fenoménicos”), la razón de que no podamos dar este tipo de explicaciones reside simplemente en que “no hay descripciones asociadas *a priori* con nuestros conceptos ordinarios de los estados conscientes” (Papineau, 1998: 381). De esta manera, la objeción al argumento de Levine consiste en parte en señalar que la comparación entre identidades que involucran descripciones (o conceptos que se agotan en la descripción de ciertos roles causales), como lo serían “agua = H₂O”, no son comparables con enunciados como “dolor = disparo de fibras C”, en los cuales se incluye algún concepto fenoménico. Para Papineau (con ciertas precisiones que voy a mencionar más adelante) una analogía más adecuada para este último tipo de enunciado de identidad se encuentra en aquellos enunciados que involucran dos nombres propios, como “Mark Twain = Samuel Clemens”, en los cuales los términos que se colocan a los costados del signo “=” son nombres propios.

(c) Esto nos conduce a una tercera pregunta que puede hacerse con respecto a los enunciados de identidad. En este caso, lo que se pregunta es cómo dos conceptos diferentes, que no tienen conexiones *a priori*, pueden referir a un mismo objeto.

Como se dijo al comienzo de este capítulo, una posibilidad conceptual no implica necesariamente una posibilidad metafísica. En efecto, una misma propiedad u objeto puede caer bajo diferentes conceptos, o dicho de otra manera, puede tener diferentes modos de presentación (mediante conceptos diferentes). Y eso basta en principio para explicar cómo lo que es metafísicamente imposible es, sin embargo, conceptualmente posible. Podemos pensar acerca de una misma propiedad u objeto

bajo dos modos de presentación diferentes, y entonces puede parecer posible que la propiedad presentada por uno de esos modos no sea la misma que la propiedad presentada por otros a pesar de que eso no sea metafísicamente posible. De la misma manera, el mero hecho de que no sepamos *a priori* que dos conceptos seleccionan la misma propiedad en nuestro mundo, no implica que no lo hagan. Algo similar ocurre con las llamadas identidades *a posteriori*.

La identidad “La estrella matutina es la estrella vespertina” es una identidad verdadera y, por lo tanto, no es metafísicamente posible que la estrella matutina no sea la estrella vespertina. Sin embargo, en la medida en que ambas descripciones refieren a su objeto, Venus, a través de modos de presentación diferentes (en este caso, mediante la propiedad de aparecer a la mañana por un lado, y mediante la propiedad de aparecer en las últimas horas de la tarde por el otro), podemos pensar que la identidad es falsa y solicitar así una explicación que nos aclare cómo los dos conceptos refieren al mismo objeto o cómo la misma cosa (en nuestro ejemplo, Venus) tiene diferentes propiedades que originalmente pudimos haber pensado que eran instanciadas por diferentes objetos.

En este punto, sin embargo, los anti-fisicalistas argumentan típicamente que la distinción concepto /propiedad no puede ayudar a sus rivales. En efecto, al parecer, en cada caso donde dos conceptos distintos (no conectados *a priori*) seleccionan la misma propiedad, lo hacen por medio de otras propiedades que son distintas y que sirven como modos de presentación del referente común. En efecto, al parecer, la explicación más natural acerca de cómo estos caminos conceptuales diferentes convergen en el mismo objeto o propiedad, es que lo hacen adscribiendo *distintas* propiedades contingentemente relacionadas, las cuales, en este mundo, son satisfechas por el mismo objeto / propiedad. Ahora bien, si esto es así, se estaría reconociendo, contrariamente a las pretensiones del fisicalismo, un dualismo ontológico de propiedades, no ya entre los referentes de los conceptos fenoménicos y de los conceptos no fenoménicos, sino entre las *propiedades que seleccionan al mismo referente* en cada concepto. De esta manera, lo único que conseguiría el fisicalista por esta vía es aplazar por algún tiempo un desenlace inevitable.

Por este motivo, autores como Papineau o Loar apelan a un modelo diferente. Así, de acuerdo con la propuesta de Papineau de los conceptos fenoménicos, -que sostiene sin mayores diferencias en distintos trabajos (Papineau 1998, 2002, 2007,

2011)-, es precisamente el carácter peculiar de los mismos lo que explica que nos resistamos a aceptar cualquier identidad psicofísica propuesta. En particular, el autor afirma que dichos conceptos son similares a los nombres propios en el hecho de que no involucran descripciones, pero difieren de éstos en que hacen referencia a las experiencias conscientes por medio de una suerte de “simulación” de las mismas.

En efecto, según la versión de 1998, existen dos usos de estos conceptos. Por un lado, se supone que los utilizamos para clasificar (de manera directa y sin apelar a evidencia física o conductual) nuestras propias experiencias conscientes. En este caso, no se trata de que identifiquemos introspectivamente alguna propiedad fenoménica, sino que simplemente ciertos mecanismos subpersonales comparan la experiencia presente con una suerte de “plantilla” creada por experiencias previas del mismo tipo que “disparan” directamente el veredicto “aquella experiencia otra vez” (Papineau, 1998:383). En segundo lugar, utilizaríamos estos conceptos también para imaginar fenoménicamente las experiencias conscientes. En este caso, cuando imagino que alguien siente dolor o tiene una experiencia de rojo se “re-activan” algunas de las partes del cerebro que son activadas por la experiencia relevante y se “proyecta” ese estado a quienquiera que sea en quien estemos pensando.

En definitiva, y por extraño que suene, Papineau considera que los conceptos fenoménicos en su uso imaginativo comparten parte del carácter cualitativo de las experiencias a las que refieren. Así, por ejemplo, imaginar fenoménicamente que estamos probando chocolate se siente parecido a probar chocolate realmente. Ahora bien, precisamente este rasgo peculiar de los conceptos fenoménicos explica, para Papineau, por qué nos sentimos poco inclinados a aceptar las identidades mente – cerebro.

Cuando usamos nuestros conceptos fenoménicos imaginativamente, traemos a la mente, en un sentido literal, una instancia de la propiedad de la experiencia acerca de la que estamos pensando. Cuando usamos conceptos no-fenoménicos [como por ejemplo, “disparo de fibras C”], esto no ocurre. Y esto hace que nos parezca que los conceptos no fenoménicos no pueden [...] denotar las mismas propiedades de la experiencia que son seleccionadas por nuestros conceptos fenoménicos (Papineau, 1998: 385).

Desde luego, hay un sentido en el cual los conceptos no fenoménicos “dejan afuera” la experiencia consciente, pero es un sentido inocuo que no pone en peligro la verdad de las identidades psicofísicas. La diferencia crucial entre los dos modos de presentación de los estados conscientes es que cuando pensamos en ellos utilizando conceptos fenoménicos *usamos* un estado similar al estado en el que estamos pensando para referirnos a él. En el caso de los conceptos no fenoménicos no referimos de esta manera a nuestras experiencias conscientes. Pero el hecho de que no lo hagamos de esta manera no implica que no estemos mencionando o refiriendo a los mismos estados que aquellos a los que referimos con los conceptos fenoménicos. De esta manera, tendríamos un modelo que explica los diferentes “modos de presentación” de una misma propiedad que, en principio, supera la crítica mencionada con anterioridad, que suponía que un enunciado de identidad verdadero que conecta dos conceptos independientes uno de ellos selecciona a su referente por medio de connotar una propiedad contingente del mismo.²⁶

²⁶ Con todo, la propuesta específica de Papineau enfrenta dificultades considerables que la vuelven sumamente implausible. Veamos las más importantes. Como vimos, Papineau sostiene que toda vez que un concepto fenoménico forma parte de los juicios acerca de la experiencia, de alguna manera están presentes las experiencias fenoménicas mismas en ellos. Ahora bien, parece sencillamente falso que cada vez que pensamos mediante conceptos fenoménicos en el dolor hay algo doloroso en los mismos o algo que se parezca al dolor. En segundo lugar, si todo pensamiento acerca de la propia experiencia incluye de algún modo a la experiencia misma, no podríamos tener pensamientos negativos acerca de la experiencia como “Yo no estoy teniendo la experiencia E ahora” porque el hecho de tener este pensamiento supone automáticamente instanciar la propiedad en cuestión. En tercer lugar, y este es un defecto común a varias propuestas que recurren a la idea de “conceptos fenoménicos” señalado por varios críticos (Crane 2005, Tye 2009, Pérez 2010, 2011), los conceptos fenoménicos no individualizan en grado fino a su referente. En efecto, de acuerdo con la propuesta de Papineau, que pertenece a las versiones constitutivas de los conceptos fenoménicos, cuando tengo una experiencia de un cierto matiz de rojo, esa misma experiencia también puede ser conceptualizada como de rojo, de escarlata o de color. Sin embargo, no es posible individualizar los diferentes conceptos EXPERIENCIA DE ROJO, EXPERIENCIA DE ESCARLATA, EXPERIENCIA DE COLOR, porque no hay una diferencia entre ellos. La misma experiencia los constituye a todos. Este problema es fundamental. En efecto, si se piensa que exhibir este rasgo es indispensable para que algo sea considerado un concepto, ninguna propuesta que no lo satisfaga podrá aceptarse. Por último, hay otra dificultad considerable que complica a la posición de Papineau (y también a la mayoría, sino a todas, las diversas caracterizaciones de los conceptos fenoménicos). En principio, a través de esta perspectiva es posible explicar cómo es que enunciados como “dolor = disparo de fibras C” sean informativos a pesar de expresar identidades: los conceptos a los costados del signo de identidad usan modos de presentación diferentes de una misma propiedad. Esto explicaría también que los enunciados que contengan conceptos fenoménicos no puedan ser derivados *a priori* de enunciados que contengan únicamente en conceptos “físicos” (no fenoménicos). Sin embargo, la historia que describe Papineau no da cuenta de la perplejidad que sentimos cuando se nos dice que nuestra experiencia de dolor es cierto estado cerebral. Esta perplejidad motiva que nos preguntemos: ¿cómo es que ese estado es el mismo que este otro? En efecto, de acuerdo con la propuesta de Papineau, “si el dolor es el estado cerebral B, entonces el estado que estoy experimentando es el estado cerebral B, y al representarlo experimentándolo, todo lo que hago es usarlo como caso del estado cerebral B. Así, utilizo el estado cerebral B para representar el estado cerebral B, el mismo estado que también

En resumen, el argumento del hiato explicativo supone en su formulación como condición necesaria para que podamos aceptar justificadamente un enunciado de identidad, que resulte posible realizar cierto análisis de los conceptos a reducir / explicar. En particular, de acuerdo con dicho argumento se requeriría un análisis del concepto del fenómeno a explicar / reducir en términos de una explicitación de los distintos roles causales que agotarían la noción pre – teórica u ordinaria que tenemos del mismo. En esta sección, independientemente de si se acepta o no la caracterización de Papineau de los conceptos fenoménicos o la idea de que éstos refieran como lo hacen los nombres propios, se han mostrado casos en los cuales se exhibe con plausibilidad, cómo se aceptan ciertas identidades sin llevar adelante previamente el análisis conceptual mencionado.

Ahora bien, si estos casos se consideran adecuados, entonces el defensor del argumento del hiato explicativo ya no puede esgrimir como una razón para rechazar las identidades psicofísicas la imposibilidad de realizar cierto análisis de los conceptos fenoménicos involucrados porque existirían otro tipo de identidades (consideradas legítimas dentro de la ciencia) que tampoco se adoptan este procedimiento. Por lo tanto, si verdaderamente hay un hiato explicativo aquí, la causa no parece deberse a la imposibilidad de realizar dicho análisis conceptual. En todo caso, se requeriría señalar alguna diferencia adicional entre estos enunciados de identidad que explique por qué las identidades psicofísicas son problemáticas o por qué demandarían necesariamente un análisis conceptual como el mencionado. Pero esto constituiría un argumento diferente al argumento que me propuse evaluar en este capítulo.

4. Conclusiones

En este capítulo enfoqué mi atención en otro de los argumentos clásicos en contra de las posiciones fisicalistas en torno a la consciencia: el argumento de Levine del “hiato explicativo” o de la “brecha explicativa”. Como señalé, el objetivo de este

estoy representando de una manera descriptiva.” (Tye 2009: 119) Frente a esta situación, Tye se pregunta qué es lo que hace que este modo de representar el estado cerebral en cuestión nos genere alguna perplejidad. Y su respuesta es que, desde una perspectiva como la de Papineau, no hay nada que pueda explicarlo, dado que, desde esta perspectiva, debería presentárenos una perplejidad similar en otros casos en los cuales usamos la propiedad a la que queremos referirnos para representarse a sí misma: “Si digo, por ejemplo, que la expresión tipeada ‘esta expresión tipeada’ debe entenderse como refiriéndose a ella misma, ¿esto hace dificultoso ver cómo el referente de esta expresión puede ser una entidad física localizada en una posición del espacio determinada? Claramente no.” (Tye 2009: 119).

argumento consiste en mostrar que las teorías fisicalistas no pueden dar cuenta adecuadamente del carácter cualitativo de la experiencia consciente. Para demostrar su punto, Levine señalaba una asimetría entre dos tipos de enunciados de identidad: uno representado por enunciados como “agua = H₂O” y “calor = movimiento de moléculas”, y otro tipo representado por enunciados como “dolor = disparo de fibras C”.

Como vimos, dicha asimetría consistía, de acuerdo con Levine, en que mientras que resultaba posible concebir que algo diferente a H₂O manifestara las propiedades superficiales macro del agua, no era posible concebir que H₂O no tuviera las propiedades macro del agua. En contraposición, parecía igualmente concebible que pudiera existir el dolor sin disparo de fibras C, y disparo de fibras C sin dolor. Esta asimetría, proseguía el argumento, era el resultado “una diferencia epistemológica importante” entre ambos enunciados, a saber: que la teoría química del agua explicaba reductivamente todo lo que necesitaba ser explicado, mientras que una teoría fisicalista de los *qualia* siempre dejaba “algo afuera”. Del desarrollo de cada una de las explicaciones en juego, se revelaba la idea de Levine de que en una explicación adecuada, si el *explanans* es verdadero, el *explanandum* no puede ser falso.

Al mismo tiempo, esta diferencia epistemológica se explicaba en gran medida por la imposibilidad de cumplir con los dos pasos que, de acuerdo con el autor, debían cumplir las explicaciones reductivas adecuadas: el análisis *a priori* de los roles causales que agotan el concepto ordinario del fenómeno que se pretende reducir / explicar y el descubrimiento *a posteriori* del mecanismo que satisface esos roles causales. Frente a este escenario, en la sección segunda de este capítulo, desarrollé las críticas más importantes que se le dirigen al argumento del hiato explicativo. Allí señalé, en primer lugar, que hay fundamentos para poner en duda que la relación de implicación que parece suponer Levine entre el *explanans* y el *explanandum* deba sostenerse para cualquier descripción del fenómeno a explicar. De hecho, la idea en este punto consistió en mostrar que ninguna explicación podría cumplir con ese requisito. Por este motivo, que no haya una relación de implicación entre la descripción del fenómeno a explicar (en este caso, la consciencia) en términos de la teoría física relevante y la descripción del mismo fenómeno en

términos del lenguaje ordinario o de sentido común, no puede considerarse en principio un defecto de las explicaciones fisicalistas.

En segundo lugar, y esta es la crítica más dura que puede asestársele al argumento de Levine, puede mostrarse que la asimetría entre las explicaciones reductivas del agua y del dolor no se sostiene. De acuerdo con esta objeción, se concluye que, o bien debe aceptarse que hay un “hiato explicativo” en ambos casos (cuando no se introducen los principios puente), o bien que en ambos casos el hiato puede ser cerrado de la misma manera (a saber: introduciendo los principios puente adecuados). Como resulta obvio, cualquiera de las opciones desbarata el argumento de Levine.

Por último, consigné aquellas objeciones que señalan que no es necesario adoptar el criterio por el cual, de acuerdo con el argumento del hiato explicativo, se justifica la aceptación de los enunciados de identidad. Más específicamente, este tipo de objeción muestra que el análisis conceptual que figura como el primer paso para realizar una explicación reductiva exitosa y la mencionada identificación, no siempre es necesario o, al menos, no parece serlo en diferentes ámbitos científicos. Por lo tanto, la imposibilidad de darlo a la hora de dar cuenta de la experiencia consciente, no constituye por sí sólo un impedimento para identificar legítimamente dichas experiencias con ciertos estados físicos / funcionales. Como señalé en la introducción de este capítulo, esta última conclusión desestima otra de las razones que suelen ofrecerse para pensar que la consciencia sea un fenómeno particularmente inaccesible para una perspectiva fisicalista de la mente.

CAPÍTULO 3

FUNCIONES CANCELADAS EN EL TEATRO CARTESIANO

1. Introducción

En este capítulo voy a ocuparme de otro de los aspectos que torna problemática a la experiencia consciente: el presunto carácter privado de nuestro conocimiento acerca de ella, esto es, la idea de que, en sentido estricto, sólo cada uno de nosotros puede *saber* qué pasa por su propia mente, (al menos en lo que respecta a gran parte de sus estados mentales). En contraposición, se supone que el resto de las personas sólo puede *conjeturar* o realizar *hipótesis* acerca de la presencia y naturaleza de los rasgos de esos estados basándose en algún tipo de evidencia indirecta, como el comportamiento observable, o algún indicio del entorno.

Por mi parte, me propongo justamente *negar* que haya un conocimiento de este tipo. Para cumplir con este objetivo, desarrollaré una argumentación en dos etapas. La primera de ellas (sección 2 y 3) consistirá, por un lado, en mostrar las consecuencias inadmisibles que se siguen después de adoptar la tesis del conocimiento privado, entre las cuales se destacan: la imposibilidad de tratar los estados mentales conscientes desde una perspectiva científica, el compromiso con el “problema de las otras mentes”, el compromiso con la aceptación de escenarios de dudosa coherencia interna (como el escenario zombi) y con una concepción del lenguaje que, en la medida en que supone que los significados de los términos mentalistas están constituidos por entidades privadas, hace difícil comprender cómo es que lograríamos comunicarnos unos con otros, tal como parece que de hecho hacemos, a través de él.

En segundo lugar, mostraré (en la sección 4) que, si se adopta el enfoque expresivista de filósofos como Dorit Bar-On (2004, 2010), es posible dar cuenta de

algunos de los hechos que podrían motivar la aceptación de la tesis del conocimiento privado, sin verse obligados a postular un conocimiento de esas características ni asumir sus consecuencias perniciosas.

Entre las motivaciones mencionadas, sostendré (en la sección 2) que una de ellas es la constatación de cierta asimetría entre las atribuciones mentalistas en primera persona de tiempo presente y las atribuciones mentalistas en tercera persona según la cual, las primeras contarían con un grado considerable de certeza o de “presunción de verdad” que se no se le asigna a las segundas. La segunda motivación para postular la existencia de un conocimiento privado la adjudicaré a otras peculiaridades de las atribuciones psicológicas, a saber: que mientras que aquellas que se realizan en tercera persona se basan fundamentalmente en la observación previa de la conducta de la persona u organismo, para realizar las auto-atribuciones mentalistas en primera persona, no nos basamos en este tipo de evidencia. Así, de acuerdo con una perspectiva “cartesiana” que buscaré rechazar, estas asimetrías y peculiaridades de las atribuciones psicológicas se explicarían en última instancia en virtud de una *asimetría epistémica*, según la cual, contamos con un conocimiento directo, seguro (y acaso infalible) y privado con respecto a nuestros estados mentales conscientes que se contrapone con el conocimiento indirecto e irremediabilmente expuesto al error que tenemos con respecto a los estados mentales de las demás personas. En relación con este punto, veremos también que el cartesianismo acepta un tanto acríticamente la idea de que la auto-atribución de estados mentales conscientes se asemeja a la descripción de fenómenos físicos, en donde primero se daría un acto de observación “interna” de dichos fenómenos y luego, con base en esa evidencia, se daría la posibilidad de llevar adelante la mencionada descripción, de la misma manera en que una afirmación acerca del desorden que domina en mi habitación se basa o puede basarse en una observación previa de la misma y de su estado.

2. Una perspectiva cartesiana de la mente y sus motivaciones.

Si bien no siempre suele reconocerse explícitamente el presunto carácter privado del acceso que tendríamos a nuestros propios estados mentales, el fenómeno parece, sin embargo, bastante extendido entre filósofos y científicos.

[La consciencia es un] fenómeno de primera persona, completamente privado que ocurre como parte de un proceso de primera persona, privado, que llamamos mente (Damasio 1999: 12).

Mi estado presente de consciencia es una característica de mi cerebro, pero sus aspectos conscientes son accesibles para mí de un modo en que no son accesibles para ti. Y tu estado presente de consciencia es una característica de tu cerebro y sus aspectos conscientes son accesibles para ti de un modo en que no son accesibles para mí (Searle 1984: 25).

Estrictamente hablando, cada individuo tiene certeza únicamente de que él mismo es consciente (Crick 1995: 107).

La experiencia consciente, como una percatación de alguna cosa o algún evento, es accesible directamente sólo al individuo que tiene esta experiencia, no a un observador externo (Libet 1993: 368).

Nuestras razones para creer en la consciencia se derivan exclusivamente de nuestra propia experiencia de ella [...]. Es verdad que ciertos hechos físicos acerca del mundo podrían proporcionar alguna evidencia indirecta en favor de la existencia de la consciencia. [...] Con todo, esta evidencia sería bastante inconcluyente [...] (Chalmers 1999: 141-142).

Por su parte, autores como Hacker (1990 y Bennett y Hacker 2003) y Schroeder (2013) creen que la idea de que tenemos un acceso privado y exclusivo a nuestros propios estados mentales, se apoya en gran medida en la suposición de que las experiencias son algo así como una suerte de “propiedad privada” inalienable de cada persona que las atraviesa. De acuerdo con este planteo, si bien tu experiencia puede ser *cualitativamente* idéntica a la mía (como tu libro puede ser cualitativamente idéntico al mío), siempre resultará *numéricamente* distinta. Así, cada experiencia, cada estado consciente, necesita un propietario (dado que es ridículo pensar que pudiera haber dolores que nadie siente o pensamientos que nadie tiene cruzando por la avenida) y, al mismo tiempo, cada experiencia tiene *un sólo* propietario. Ahora bien, puesto que sólo yo podría en sentido e tener o poseer *mis* experiencias, se concluye por esta vía que es imposible que otra persona además de mí pueda conocerlas en sentido estricto (ya sea de manera segura o no).

En este capítulo no voy a ocuparme en particular de este aspecto del carácter privado de la experiencia consciente. De lo que sí me voy a ocupar es del supuesto carácter privado de nuestro *acceso* a los estados mentales conscientes. Para eso, tal como anticipé, en primer lugar enumeraré algunas de las principales razones que podrían postularse para desembocar en esa idea.

(1) Entre esas razones, debe contarse el hecho de que habitualmente se suele aceptar que al menos ciertas afirmaciones y pensamientos que hacemos acerca de nuestros propios estados mentales son mucho más seguros y confiables que las afirmaciones que podemos hacer acerca de los estados mentales de los demás, o de cualquier otro hecho del mundo.

En efecto, ¿puedo verdaderamente dudar de que estoy teniendo un dolor de cabeza intenso en este momento? En este sentido, podríamos decir que cada uno de nosotros tiene una suerte de “autoridad especial” en torno a ciertos hechos mentales y que, por este motivo, existe una asimetría entre el primer tipo de afirmaciones y el segundo. En efecto, en circunstancias habituales, si alguien quiere saber lo que pienso, lo que siento o lo que pretendo, la mejor opción es que me lo pregunte a mí (por lo menos si la información que quiere obtener es acerca de mis estados mentales *conscientes*). Cuando los demás no aceptan mis auto-atruciones mentalistas tal y como yo las presento, suele ser más porque consideran que no soy sincero, que porque piensen que estoy equivocado.

Desde luego que esto no nos obliga a aceptar que otras personas que nos conozcan mucho, como nuestras parejas, nuestros amigos o nuestros familiares, no puedan realizar consideraciones más o menos confiables acerca de nuestra condición psicológica. Sin embargo, en este último caso estaríamos dispuestos a sostener que se trataría de información de “segunda mano”, y en caso de que se suscitara una discrepancia entre mi opinión acerca de ellos y la de los demás, se preferiría mi palabra. En contraposición, la atribución de estados mentales a otras personas no goza de la misma seguridad, dado que, entre otras cosas, siempre persistiría la posibilidad de que esas personas no estén siendo honestas con nosotros o que sencillamente no demuestren sus pensamientos y sensaciones. Así, en lo que respecta a las atribuciones psicológicas dirigidas a otras personas, las cosas no siempre son lo que parecen.

Con todo, tampoco consideramos que la autoridad de primera persona rija con respecto a cualquier estado mental que posea o atraviese el sujeto. Por un lado, la autoridad en cuestión no parece capaz de sustraerse a los eventuales fallos en la memoria. Por lo tanto, no se sostendría en principio en aquellas atribuciones en las que ésta esté involucrada (como la adscripción de estados mentales pasados). Por otra parte, tampoco aplicaría en aquellas ocasiones en las que las auto-atribuciones mentalistas se hacen en base a algún tipo de evidencia externa (como cuando me miro al espejo y por el aspecto de mi cara me percato de que estoy cansado; o cuando, luego de analizar mi comportamiento, me doy cuenta de que estoy enamorado de alguien; o cuando descubro que albergo un rencor hacia mi hermano porque mi analista, en quien confío por diversas y buenas razones, me convenció de ello). En estos casos, la credibilidad de mis auto-atribuciones se sustenta enteramente en la credibilidad que tengan las evidencias que presente en su favor. Pero, en principio, (al menos en los ejemplos mencionados), cualquiera podría disponer de la misma evidencia y hacer afirmaciones con respecto a mi condición psicológica con el mismo tipo de autoridad que yo. De hecho, podría suceder, como en los casos en que se está celoso de alguien, que otras personas adviertan antes que nosotros mismos cuáles son nuestros estados mentales. De todos modos, en este capítulo me ocuparé fundamentalmente de un tipo particular de estado mental sobre el que parece incontrovertible que tenemos una autoridad especial: las sensaciones presentes de dolor. En casos como este, parece sencillamente absurdo creer que es necesario considerar algún tipo de evidencia externa como las mencionadas para ser capaz de atribuirme este tipo de estado mental o que podemos equivocarnos acerca de su presencia.

(2) En segundo lugar, existen otra serie de razones conectadas con lo anterior que pueden conducir a pensar que cada uno de nosotros posee un acceso epistémico privado y privilegiado a (por lo menos algunos de) sus estados mentales. En este caso, las motivaciones se apoyan en el hecho de que, a diferencia de las atribuciones psicológicas dirigidas a otras personas, muchas auto-adcripciones no se basan en la observación de la conducta propia o de algún elemento del entorno. Este aspecto de las atribuciones psicológicas puede derivar en otros más controvertidos, si se supone que nuestros conceptos o términos mentalistas refieren a procesos o

eventos mentales de manera similar a la que un nombre refiere a su portador, sin ningún tipo de conexión conceptual con ciertas manifestaciones conductuales (tal como parecer indicar el hecho de que no nos basemos en nuestras conductas para atribuirnos estados mentales), y que la “gramática” que rige los reportes acerca de los estados mentales propios es la misma que domina la descripción de objetos físicos en donde el modelo constaría de dos pasos: (i) observación de los hechos y (ii) descripción de los mismos en virtud de los resultados de (i). De acuerdo con esta última idea, una afirmación como (a) “La pintura en el lateral de la casa empieza a levantarse”, es similar a una afirmación como (b) “La muñeca empieza a dolerme”, en la medida en que se supone que en ambos casos informa de algo que previamente observé (de alguna manera), detecté o averigüé.

Ahora bien, frente a este escenario que plantea esta autoridad de primera persona, junto con las peculiaridades de las atribuciones psicológicas mencionadas, muchos filósofos se han dado a la tarea de intentar dar cuenta de ella, de determinar si nuestra confianza está verdaderamente justificada y, de ser así, cuál es su alcance y origen. Así, por ejemplo, si seguimos la clasificación de David Finkelstein (2010), muchos de los filósofos y psicólogos adhieren a lo que el autor llama “detectivismo”, según el cual, la autoridad acerca de nuestra propia vida mental se explica gracias a que contamos con un proceso en algún sentido seguro que nos permite *averiguar* acerca de ella, detectando nuestros propias esperanzas, miedos, emociones y sensaciones, y haciendo de esta manera que se tornen conscientes. Así, un detectivista es aquel que considera que ese acceso peculiar que tenemos a nuestros propios estados mentales involucra algún tipo de observación interna más ventajosa en relación con la posición que pudieran tener otras personas con respecto a nuestros estados mentales.

Por su parte, otros filósofos, etiquetados como “constitutivistas”, han defendido otra explicación en torno a la autoridad de primera persona, de acuerdo con la cual, ésta residiría en el hecho de que cuando alguien *juzga* que quiere, cree, anhela o teme algo, con ello *hace* que aquello que juzga sea el caso. Así, por ejemplo, cuando Max juzga que tiene la intención de ir al recital de Dylan *hace que sea el caso* que esa sea su intención.

Por otro lado, de acuerdo con una posición en ocasiones atribuida al filósofo Ludwig Wittgenstein, y a la que podríamos etiquetar con el nombre de

“expresivismo simple” (siguiendo ahora a Bar-On 2004 en la terminología), la autoridad que venimos considerando se basaría en el tipo de *acto* que realizamos al efectuar el tipo de aserciones que estamos analizando. En lugar de considerarse como descripciones, las auto-atribuciones mentalistas serían en realidad *expresiones* de nuestros estados mentales, de una manera similar a la que podemos decir que un grito de dolor, o una sonrisa expresan, respectivamente, un dolor o un estado de alegría.

Ahora bien, de acuerdo con este tipo de expresivismo, en la medida en que estas afirmaciones (como “Siento un fuerte dolor ahora”), son expresiones y no son aseveraciones, reportes o descripciones de ciertos hechos, no tendrían valores de verdad, del mismo modo en que un gesto de dolor no podría tenerlo. De acuerdo con Finkelstein, esta interpretación no recoge fielmente las intuiciones de Wittgenstein al respecto. En principio, porque “insistir en que determinada emisión es una expresión o es parecida a una expresión no exige negar que también sea una aserción” (Finkelstein 2010: 32). En todo caso, de acuerdo con este autor y con el neo-expresivismo de Dorit Bar-On (2004) del que me ocuparé más adelante, cuando alguien sinceramente afirma que tiene un terrible dolor de cabeza no solo expresa su estado mental sino que al mismo tiempo *dice* algo verdadero acerca de su condición psicológica.

Dentro de estas opciones teóricas, es una *versión* de lo que Finkelstein llama “detectivismo” lo que conduce a los problemas que mencioné al comienzo de este capítulo. Por mi parte, prefiero utilizar el rótulo de “cartesianismo” para referirme a esta perspectiva, dado que contiene algunas tesis que comúnmente se le atribuyen al pensamiento de Descartes (pero que, como veremos, no se restringen a él)²⁷. Desde este punto de vista, la autoridad de primera persona (1) se explica gracias a la existencia de una *asimetría epistémica*, es decir, gracias a que poseemos un conocimiento directo y mucho más seguro o certero acerca de (al menos cierta clase de) estados mentales propios que el que podemos alcanzar con respecto a los estados mentales de las demás personas y a cualquier otro hecho del mundo. Estas ideas se justifican por los siguientes motivos. Por un lado, dentro de una tradición filosófica que podemos rastrear hasta la filosofía racionalista moderna, se considera

²⁷ En esta caracterización del cartesianismo sigo en gran medida las notas que Pérez (2007) le atribuye a esta forma de pensamiento.

que el acceso que tenemos a los objetos externos se da a través de nuestro contacto con objetos internos, es decir, a través de nuestras sensaciones, impresiones, o representaciones que se supone que son causadas por los primeros. De esta manera, la existencia y carácter de los objetos externos tiene siempre un estatus conjetural, o de algo así como una hipótesis que explicaría la presencia de las representaciones o impresiones en nuestra mente, del mismo modo en el que las entidades teóricas postuladas por los científicos postulan dan cuenta del comportamiento observable de ciertos objetos.

En otras palabras, desde esta perspectiva, cuando veo o toco un objeto físico, tengo consciencia de su existencia sólo *indirectamente*, en la medida en que *infero* que tal objeto podría ser la causa más probable de los datos sensoriales (visuales o táctiles) que tengo en ese momento. Así, de acuerdo con este esquema, los objetos externos (los autos y los árboles que veo a través de mi ventana) son hipótesis explicativas, es decir, existencias que postulamos como una manera de explicar los datos de los sentidos. Al mismo tiempo, y como una consecuencia inevitable de este planteo, el acceso a las otras mentes presentará algunos problemas adicionales, ya que supone *dos* inferencias controvertidas. En efecto, por un lado, para afirmar que existen otros seres conscientes se necesita establecer, en primer lugar, que el cuerpo de los otros existe (más allá de mis representaciones de él) y, en segundo lugar, se necesitará apelar a algún tipo de razonamiento inductivo-analógico para establecer que los movimientos de esos cuerpos conjeturados están acompañados por los mismos tipos de pensamientos y sensaciones que yo tengo cuando mi cuerpo se mueve de forma parecida.

En definitiva, desde esta perspectiva cartesiana, sólo cada uno de nosotros puede tener un conocimiento *directo* de sus propios estados mentales. El resto puede únicamente *creer* en, o *conjeturar* la presencia y el carácter de mis estados mentales a partir de una inferencia basada en evidencia indirecta como la conducta. Esto constituiría un primer elemento que me interesa destacar de la perspectiva cartesiana.

Un segundo elemento que quiero señalar es que la idea de que tenemos un conocimiento directo de nuestros propios estados mentales en ocasiones es acompañada, dentro de una perspectiva cartesiana, por la idea de que este conocimiento (además de ser directo) es sumamente seguro o incluso infalible. En

este sentido, la propuesta suele consistir en afirmar que no hay posibilidad de cometer un error al momento de atribuirse (al menos ciertos) estados mentales a uno mismo. Si me *parece* que siento dolor, entonces *siento* dolor, porque en el reino de lo mental no hay diferencia entre cómo las cosas se nos aparecen y cómo son realmente. Por el contrario, mi conocimiento acerca de los objetos externos y de los hechos del mundo en general, no goza de dicha prerrogativa: si me parece que hay un árbol en la calle porque creo estar viéndolo, existe la posibilidad de que un vecino con mucho tiempo libre haya reemplazado por la noche al árbol en cuestión por una fotografía que resulta visualmente indistinguible del árbol real desde mi ventana. Pero además, no sólo es posible equivocarse con respecto al hecho de que el árbol está ahí afuera; también es posible equivocarse con respecto a la *naturaleza* de ese objeto: puedo pensar que se trata de una acacia blanca cuando en realidad es un arce. Por el contrario, si bien, como afirma Descartes, es posible equivocarse con respecto a si realmente estamos viendo un trozo de cera derritiéndose en el fuego, no podemos equivocarnos acerca de que creemos, pensamos o nos parece, que vemos un trozo de cera derritiéndose junto al fuego en la habitación en que desarrollamos nuestras meditaciones metafísicas. Tampoco podríamos equivocarnos, desde esta perspectiva cartesiana, con respecto a la *naturaleza* de al menos algunos estados mentales: si algo se me aparece como un dolor punzante, entonces *es* un dolor punzante lo que tengo, aun cuando se trate de un dolor “fantasma” al que no le corresponde ninguna lesión corporal real.²⁸

En resumen: para la perspectiva cartesiana la asimetría señalada al comienzo de esta sección con respecto a la seguridad de ciertas auto-atribuciones mentalistas en primera persona, halla parte de su explicación a través de la distinción entre conocimiento directo e indirecto, en donde el primero solo es posible con respecto a los estados mentales propios. En este sentido, el conocimiento de “primera mano” de mis propios estados mentales es *privado*: sólo yo puedo tener un acceso así. En ocasiones, se considera no solo que este conocimiento es directo sino que también es infalible. Este último rasgo reforzaría la asimetría entre las atribuciones psicológicas en primera y en tercera persona.

²⁸ Dentro de la tradición de la filosofía analítica, el filósofo Bertrand Russell de *Los problemas de la filosofía* encuadra perfectamente en esta caracterización del paradigma cartesiano.

Por otra parte, y tal como indiqué con anterioridad, la pertinencia del esquema “directo / indirecto” explicaría otros hechos cotidianos, como la posibilidad siempre latente de ser engañados por otras personas acerca de sus verdaderos pensamientos y sensaciones. En efecto, de acuerdo con una perspectiva cartesiana como la que presenté el engaño es inevitable porque nuestro acceso a las otras mentes siempre se da a través de elementos externos a ella. Además, el par directo / indirecto daría cuenta del hecho de que las auto-adcripciones psicológicas no se basen en evidencia conductual de ningún tipo. En efecto, en el propio caso no necesitamos recurrir a este tipo de evidencia porque contamos con información de “primera mano” acerca de nuestra mente.

Al mismo tiempo, desde esta perspectiva se considera que la auto-adcripción de estados mentales conscientes responde a la misma lógica de las descripciones de los objetos externos. En efecto, así como los reportes perceptivos en torno a los objetos del medio se basan en la observación previa de los mismos, “la capacidad de una persona para hablar de sus propios estados mentales [...] se puede explicar apelando a un proceso que le permite *averiguarlos*” (Finkelstein 2010: 38. *La cursiva es del autor*), o “detectarlos”. De esta manera, se distinguen dos formas fundamentales de obtener conocimiento empírico mediante dos tipos de “observación”, una interna y otra externa: por *introspección* obtenemos conocimiento acerca de los objetos de nuestro reino interior, como las sensaciones, emociones y pensamientos entre otros. Por medio de los sentidos externos, obtenemos conocimiento acerca de los objetos externos y públicos, incluyendo nuestros cuerpos.

Con el objetivo de examinar los problemas que entraña la adopción de una perspectiva cartesiana de la mente, quisiera considerar con cierto detalle la posición de un representante contemporáneo de dicha perspectiva, el filósofo David Chalmers. De acuerdo con él:

La experiencia consciente es, al mismo tiempo, lo más familiar del mundo y lo más misterioso. De ninguna otra cosa tenemos un conocimiento más directo que de la consciencia, pero no es claro en absoluto cómo reconciliarla con todo el resto de los que sabemos (Chalmers 1999: 25).

Como se desprende de esta cita, Chalmers piensa que el conocimiento que tenemos de nuestros estados mentales conscientes es un conocimiento peculiar. Específicamente, considera, en línea con el paradigma cartesiano, que tener una experiencia consciente es automáticamente estar en alguna relación epistémica íntima con la misma, una relación que el autor llama, siguiendo en parte a Russell, de “familiaridad” (*acquaintance*). Desde esta perspectiva, ni siquiera habría una posibilidad conceptual de que un sujeto pueda tener una experiencia de rojo sin tener ningún contacto epistémico con ella (Chalmers 1999: 256). Sería precisamente esta “familiaridad” que el sujeto tiene con su propia experiencia consciente lo que proporcionaría evidencia a nuestras creencias y juicios fenoménicos y lo que los transformaría en auténtico conocimiento.²⁹

De acuerdo con Chalmers, existen razones poderosas para pensar que el conocimiento que tenemos de nuestros estados conscientes no puede estar basado simplemente en una mera relación causal o en el funcionamiento de algún mecanismo fiable. Como veremos a partir de la cita siguiente, el rechazo de la teoría causal del conocimiento no tiene que ver con que Chalmers considere que una relación natural, como lo es una relación causal, no pueda eventualmente justificar (aunque no por sí sola) una creencia o un juicio. Las razones, en cambio, son las siguientes:

Si la única fuente de justificación de una creencia acerca de X [en este caso, acerca de una propiedad fenoménica o una determinada experiencia consciente], es una conexión causal con X, entonces un sujeto no puede saber con seguridad que la conexión causal existe [ni que X existe, es decir, que haya conciencia del otro extremo]. El único modo en el que podría saber esto con seguridad sería si tuviese algún acceso independiente a X o a la cadena causal, pero esto implicaría conocimiento basado en algo más que la propia cadena causal. Siempre habrá un escenario escéptico en el cual todo le parece igual al sujeto, pero en el que la conexión causal está ausente y X no existe; de modo que el sujeto no puede saber con seguridad acerca de X [...]. En el caso del

²⁹ Cabe aclarar en este punto que Chalmers no sostiene que tener una experiencia implique automáticamente *saber* acerca de ella si el “saber” requiere creencia, porque a su juicio tenemos muchas experiencias de las que no tenemos creencias y que por lo tanto no conocemos en este sentido. Tampoco se compromete con la idea de que nuestras creencias acerca de la experiencia sean indubitables, debido a que dichas creencias podrían formarse por cualquier tipo de razones y a veces lo hacen por razones equivocadas.

conocimiento perceptual, por ejemplo, podemos construir una situación en la cual la conexión fiable [o causal] esté ausente –por ejemplo, una situación en la cual el sujeto es un cerebro en un tanque, digamos- y todo le parecerá igual [...]. Pero nosotros sabemos con seguridad que somos conscientes: de modo que una concepción causal de este conocimiento es inapropiada (Chalmers 1999: 253-254).

Por otra parte, Chalmers muestra con claridad su adhesión a otros aspectos del paradigma cartesiano cuando asume que el conocimiento en torno a nuestros propios estados mentales conscientes es, no sólo infalible o seguro, sino también *epistémicamente privado*. En contraposición, los demás sólo pueden conjeturar la presencia y el carácter de nuestros estados mentales a partir de la observación de cierta evidencia indirecta como la conducta. Pero, en sentido estricto, todo lo que pueda saberse sobre los estados mentales de las personas mediante este tipo de evidencia indirecta, es compatible con la ausencia de la consciencia y de sus rasgos característicos.

Nuestras razones para creer en la consciencia se derivan exclusivamente de nuestra propia experiencia de ella [...].

Es verdad que ciertos hechos físicos acerca del mundo podrían proporcionar alguna evidencia indirecta en favor de la existencia de la consciencia. Por ejemplo, a partir de estos hechos podríamos averiguar que hay muchos organismos que *afirman* ser conscientes y dicen tener misteriosas experiencias subjetivas. Con todo, esta evidencia sería bastante inconcluyente, y podría ser muy natural extraer una conclusión eliminativa: que de hecho no hay ninguna experiencia presente en estas criaturas, sólo mucha charla.” (Chalmers 1999: 141-142).

De esta manera, se desemboca rápidamente en lo que se conoce como el “problema de las otras mentes”, dado que, de acuerdo con Chalmers y la concepción cartesiana en general, “aun cuando conozcamos todas las características físicas de otras criaturas, no sabremos con seguridad que son conscientes, o cuáles son sus experiencias [...]” (Chalmers 1999: 142). Así, esta asimetría entre los distintos tipos de accesos que tenemos a los estados mentales conscientes propios y ajenos, le permite construir a Chalmers un argumento en favor de un dualismo ontológico: el argumento de la asimetría epistémica.

De acuerdo con él, la única evidencia disponible para postular la existencia de la consciencia proviene de mi familiaridad de primera persona con mis propios estados conscientes, dado que aun si conociéramos todos los hechos que hay para conocer acerca de las propiedades físicas fundamentales, las funciones realizadas por ese sistema, y cómo éstas y los mecanismos neurales dan lugar a la conducta observable, nada de esto nos permitiría determinar si estas criaturas son conscientes. La presencia de todos los hechos acerca de su composición física es compatible con la suposición de que ninguna de ellas es consciente. Pero el punto del argumento no reside en demostrar que nuestra capacidad para detectar estados conscientes de otras personas tenga algún tipo de limitación natural. La idea es que no importa cuán detallada sea la información física disponible y cuán perfectas sean nuestras capacidades para aprehenderlas, nunca resultará posible establecer o deducir con certeza, basándose únicamente en esa información, la presencia y los rasgos de los estados conscientes de las demás personas.

De acuerdo con Chalmers, de esto se seguiría que la consciencia no superviene lógicamente a lo físico, dado que a partir del conocimiento completo de los hechos físicos de bajo nivel de una criatura podemos derivar o establecer con seguridad, la presencia o ausencia de todos los hechos positivos acerca de ella sin más (por ejemplo, acerca de su funcionamiento biológico, su conducta y los mecanismos cerebrales que la causan; e incluso del resto de sus estados mentales no conscientes). En estos casos, el rol del conocimiento en primera persona no es crucial.

En resumen,

La asimetría epistémica en el conocimiento de la consciencia hace evidente que la consciencia no puede ser lógicamente superviniente [a lo físico]. Si lo fuese, no existiría esta asimetría epistémica; una propiedad lógicamente superviniente puede detectarse fácilmente sobre la base de la evidencia externa (Chalmers 1999: 142)

Ahora bien, independientemente de la plausibilidad del argumento de la asimetría epistémica, lo que me interesa desatacar es que dicho argumento *supone que es posible concebir sin contradicción un escenario zombi*, puesto que sostiene que la consideración y conocimiento de todos los hechos físicos disponibles y posibles acerca de las personas y del resto del universo, descritos hasta sus más ínfimos

detalles, no obligan a aceptar o a rechazar, ni hace posible deducir, que esas personas sean seres conscientes o que haya consciencia en absoluto.

En otras palabras, de acuerdo con el argumento de la asimetría epistémica, no hay contradicción en aceptar la verdad de toda la información física y, al mismo tiempo, negar que haya experiencia consciente. Y esta es precisamente la definición de un “escenario zombi”.

Concluida esta descripción de ciertos aspectos centrales del modelo cartesiano de la mente, en la sección siguiente voy a explicitar las consecuencias perniciosas que enfrenta.

3. 1. Algunos dolores de cabeza para la mente cartesiana.

En primer lugar, voy a sostener que la caracterización que el filósofo australiano hace del conocimiento por familiaridad que presuntamente tenemos acerca de nuestros estados conscientes, enfrenta una suerte de dilema. En efecto, si se lo analiza detenidamente, podría concluirse que, o bien no puede considerarse como un caso de conocimiento en sentido propio, o bien que no es un conocimiento infalible como se pretende. En segundo lugar, tal como anticipé, voy a recordar la idea, sustentada en el capítulo 1, de que la concebibilidad del escenario zombi que el argumento de la asimetría epistémica supone, no es clara ni indiscutible. Ahora bien, si el argumento de la asimetría implica o supone algún tipo de contradicción, podemos comenzar a sospechar que algo anda mal con él.

Para desarrollar el primer argumento voy a basarme en una serie de consideraciones que realiza Dorit Bar-On (2004: 27-54), apoyándose, a su vez, en el célebre argumento de Wittgenstein del lenguaje privado, justamente con el propósito de desestimar una explicación de raigambre cartesiana para dar cuenta de la asimetría entre las atribuciones mentalistas en primera y en tercera persona.

De acuerdo con Bar-On, con respecto a las expresiones que refieren a hechos o estados del mundo, existe lo que la autora llama ‘distancia semántica’ entre la aplicación de la expresión en cuestión y el blanco putativo o referente. Así, una expresión de estas características tiene una extensión: es decir, un rango de objetos o cosas a las que se supone que aplica. Y algunas cosas caerán bajo la extensión del concepto y otras no. Y debe ser siempre en principio posible que la expresión no sea adecuadamente aplicada. Esto es, debe ser posible usar la expresión para referir

equivocadamente a algo que cae fuera de su extensión. Esto supone que siempre que un hablante usa una expresión para referirse a algo, debe existir una distinción conceptual entre que el sujeto *piense* que aplicó el término correctamente y el hecho de si *efectivamente* aplicó el término correctamente. Por ejemplo, tomemos el término “silla”, que se usa para hacer referencia a ciertos objetos (las sillas). Decir esto es decir que se supone que el término se aplica de manera correcta únicamente a las sillas. Pero todo lo que esto puede significar es que resultará incorrecto aplicarlo a algo que no sea una silla (como cuando alguien dice, refiriéndose a una hamaca, “¡Ey, no te sientes en mi silla!”). Y esto, a su vez, implica que es en principio *posible* aplicar el término a algo que no es una silla, incluso si nunca nadie en la historia de la humanidad pasada, presente y futura, cometiera de hecho ese error.

Ahora bien, si esta posibilidad no existiera en principio, es decir, si no existiera siquiera la posibilidad lógica o conceptual de errar el blanco a la hora de hacer referencia a un objeto, se perdería la “distancia semántica” necesaria para que podamos hablar de referencia de un término en absoluto. Para decir que un término o un concepto refiere a un objeto, y poder decir que el concepto o el término se aplican correcta o incorrectamente, los objetos referidos por dichos términos deben tener condiciones de identidad independientes del éxito de cualquier individuo en identificarlos. Pero es precisamente esta posibilidad la que queda descartada en el caso del conocimiento por familiaridad del que supuestamente gozamos con respecto a nuestros propios estados conscientes, y entonces un argumento análogo puede aplicarse en este caso. En efecto, como vimos, el “huésped” de una sensación puede detectarla, o “saber” de ella, de manera infalible: de acuerdo con Chalmers, una sensación es el tipo de cosa que, si está presente, su presencia será inmediatamente ‘reconocida’ (en algún sentido) por su huésped. “Ni siquiera hay una posibilidad conceptual de que un sujeto pueda tener una experiencia de rojo [...] sin tener ningún contacto epistémico con ella: tener una experiencia es estar relacionado con ella de ese modo” (Chalmers 1999: 256).

Ahora bien, el planteo de Chalmers parece querer asegurar la infalibilidad del conocimiento acerca de nuestros propios estados conscientes descartando la “posibilidad conceptual” de que éstos se instancien y el sujeto no lo “sepa” (de algún modo), es decir, “atando” esencialmente su presencia a ciertos actos de “reconocimiento” o “saber” por parte del sujeto. Sin embargo, la idea misma de

reconocimiento de un objeto (sea un objeto físico o una propiedad fenoménica) parece suponer conceptualmente una cierta distancia entre la propiedad (re)conocida y el acto de reconocimiento o conocimiento de la misma por parte del sujeto, es decir, condiciones de identidad independientes para el objeto reconocido (en este caso, las propiedades fenoménicas de la experiencia) y el acto de reconocerlas por parte del sujeto. De lo contrario, ni siquiera parece tener sentido hablar de “reconocimiento” o de “saber”. Ahora bien, si para evitar estos inconvenientes se habilita la *posibilidad lógica* de estar equivocado con respecto a la presencia de las propiedades fenoménicas (aun cuando, de hecho, nunca lo estemos), se requiere entonces una explicación que justifique la presunta infalibilidad de dicho conocimiento, ya que se vuelve a abrir una distancia epistémica entre los sujetos y sus estados conscientes y con ésta la posibilidad de error.³⁰

Por otra parte, existen otros motivos para poner en duda que el tipo de relación epistémica que, de acuerdo con el planteo de Chalmers, tendríamos acerca de nuestras propias experiencias conscientes, merezca calificar como auténtico conocimiento en algún sentido razonable de ese concepto. En primer lugar, porque se supone que es posible estar familiarizado con algo sin conocer ninguna verdad o ningún hecho acerca de ese algo (Crane 2012: 193). O, al menos, sin que éstos se exprese en términos conceptuales: en efecto, de acuerdo con Chalmers, “tenemos muchas experiencias de las que no tenemos creencias y por lo tanto no las conocemos. Más aún, podríamos tener una experiencia sin conceptualizar la experiencia de modo alguno” (Chalmers 1999: 256). En mi opinión, esta caracterización de la relación epistémica de familiaridad pone en duda la plausibilidad de considerarla como auténtico conocimiento. Consideremos el siguiente interrogante para sustentar esta idea. ¿Estaríamos dispuestos a decir que conocemos a alguien que vimos cara a cara pero acerca del cual no sabemos ningún hecho (ni su nombre, ni su edad, ni sus preferencias estéticas o sus convicciones políticas)? En un caso como el descrito, lo más natural sería decir que no tenemos

³⁰ Como veremos más adelante, el expresivismo parece ofrecer una alternativa que, sin comprometerse con dualismo ontológico alguno, explicaría el hecho de que, en el caso de las autoatribuciones mentalistas en tiempo presente, hacemos referencia a nuestros propios estados mentales conscientes de manera segura porque somos invulnerables a caer en ciertos errores epistémicos, ya sea a la hora de identificar al sujeto que los instancia, como al momento de identificar el tipo de estado mental que se instancia.

conocimiento alguno de dicha persona, sino simplemente que la *vimos*. En líneas generales, podríamos decir que habitualmente consideramos que para poder decir que conocemos un objeto, no basta con haber tenido simplemente un contacto con él. En principio, tendríamos que poder aplicar, además, algunas creencias a la cosa o la persona con la cual estuvimos en contacto. Creo que algo análogo puede decirse con respecto al “conocimiento” que representaría la familiaridad de la que habla Chalmers. En este caso, estaríamos admitiendo que es posible tener algún tipo de “saber” acerca de la propia experiencia sin necesidad de atribuir ninguna creencia o aplicar algún concepto a dicha experiencia: como por ejemplo si se trata de un dolor o una experiencia cromática.

Por otra parte, en la medida en que la relación epistémica de familiaridad parece suponer necesariamente para darse la presencia de la experiencia con la cual el sujeto mantiene dicha relación, estaríamos en presencia de un saber efímero que persiste únicamente durante el tiempo que persiste la experiencia en cuestión. Por el contrario, el conocimiento acerca de una persona o de una ciudad, no es algo que consideremos que desaparezca (o, por lo menos, no necesariamente) cuando dormimos, cuando no estamos pensando en esas cosas, o cuando dejamos de verlas o visitarlas. Además, dado el carácter fugaz de este presunto conocimiento, parece imposible que pudiera utilizárselo en un instante posterior a la desaparición de la experiencia en juego para determinar si, por ejemplo, una experiencia actual es del mismo tipo que una experiencia pasada.

Señaladas estas primeras dificultades, ahora quisiera detenerme brevemente a considerar las consecuencias que se siguen para la concepción cartesiana de la mente una vez que acepta la tesis de que el conocimiento de los propios estados conscientes es “epistémicamente privado”, es decir, de la idea de que sólo podemos tener acceso directo, acerca de los estados conscientes *proprios*.

En primer lugar, de esta situación se deduce inmediatamente uno de los motivos por los cuales se puede pensar que la consciencia constituye, tal como sostienen muchos científicos y filósofos, un obstáculo para nuestra comprensión científica del universo. En efecto, el hecho de que ciertos aspectos o propiedades de la experiencia consciente no puedan ser accedidos desde un punto de vista objetivo o de tercera persona, impide su estudio científico, porque el punto de vista de “tercera persona” es precisamente el punto de vista propio de la ciencia. Siguiendo

en esta dirección, podríamos comenzar a indagar de qué manera podría desarrollarse una ciencia de la consciencia desde esta perspectiva. Simplificando un poco las cosas, podríamos decir que, de acuerdo con Chalmers, existe una creencia muy extendida de que la física proporciona un catálogo completo de los elementos y las leyes fundamentales del universo. Sin embargo, a lo largo de distintos trabajos y por medio de diferentes argumentos (como el de los zombis y el de la asimetría epistémica), Chalmers intenta probar (junto con otros autores), que la consciencia no parece poder derivarse de leyes y entidades físicas más fundamentales. Por este motivo, considera que en lugar de intentar reducirla infructuosamente a las propiedades físicas, una teoría que pretendiera dar cuenta de “todo” debería incluir un componente extra fundamental (es decir, un componente primitivo no-físico y no derivado de otro más simple), a saber: la consciencia. Una vez reconocido este elemento “adicional”, restaría ponerse a trabajar para establecer cuáles son las distintas “leyes puente” que rigen las relaciones entre el mundo físico y el mundo fenoménico y que mostrarán, a juicio de Chalmers, cómo depende el último del primero. En definitiva, una teoría completa tendrá dos componentes fundamentales: las leyes físicas, que nos dirán lo que les pasa a los sistemas físicos, desde los infinitesimales hasta los cosmológicos, y las que podríamos llamar leyes psicofísicas, que nos dirán cómo se relacionan los procesos físicos (o al menos algunos de ellos) con los distintos tipos y grados de experiencia consciente (o bien cómo los primeros dan lugar a las últimas).

Ahora bien, admitiendo que existan tales leyes psicofísicas, ¿cómo podríamos descubrirlas? Como habíamos visto, la consciencia es privada o subjetiva; no hay, por lo tanto, manera de observarla en otros. Por lo tanto, la única salida consistiría en advertir las regularidades entre la experiencia y los estados físicos o funcionales en nuestro propio caso, y luego extenderlas a las demás personas. El problema es que, como puede verse, la formulación de estas leyes estará basada, en última instancia, en una inferencia inductiva cuya base es extremadamente pobre, ya que para realizar las generalizaciones o descubrir las regularidades se basa en un único caso: la propia experiencia consciente. Sin embargo, la credibilidad de una inferencia inductiva basada en un único caso puede resultar por lo menos dudosa. En otras palabras, cuando “miro” dentro de mí, según la concepción cartesiana, “veo” la consciencia. Pero ¿no es una irresponsabilidad el

generalizar desde mi propio caso el de los demás? No puedo mirar dentro de los estados de los demás: constituye la esencia de la introspección el ser algo que todo el mundo debe hacer por sí mismo. ¿Puedo realizar una inferencia causal a partir de la conducta de otras personas? No puedo empezar a establecer una correlación entre la conciencia de otras personas y sus conductas cuando el primer término de esa correlación es, en principio, inobservable. Desde luego, puedo creer que observo la correlación en mi propio caso; pero es precisamente esa correlación lo que parece temerario generalizar.

Según esta concepción, en mí mismo observo la conducta y la consciencia, pero en los demás sólo observo la conducta. La muestra que observo es ridículamente pequeña para permitir extrapolación alguna (Kenny 2000: 32). Por otro lado, en caso de que se suscitara una controversia entre dos investigadores acerca de algunas de las regularidades, no habrá manera en principio de zanjar la cuestión, porque no será posible tener acceso a los datos relevantes, es decir, a los estados mentales de las otras personas.

3.2. Problemas del más allá para la mente cartesiana.

En la sección precedente se dijo que alguien que, como Chalmers, aceptara el argumento de la asimetría epistémica junto con las ideas implicadas en él, como la del conocimiento privado de los estados mentales conscientes propios, se comprometería también con la aceptación de la coherencia del escenario zombi. Sin embargo, como mostré en el capítulo 1, esta coherencia puede ser puesta en entredicho. Ahora bien, si esto es así, el argumento de la asimetría epistémica y sus supuestos también comienzan a tambalear, precisamente porque implican escenarios incoherentes.

3.3. La mente cartesiana y el lenguaje privado.

Otra de las dificultades que enfrenta el enfoque cartesiano de la mente se observa cuando se propone explicar de qué manera adquirimos el vocabulario y los conceptos mentalistas ordinarios. Como se dijo, mientras que la evidencia o fundamento para atribuir estados mentales conscientes a otras personas o animales es fundamentalmente su comportamiento, las auto-atribuciones equivalentes no se basan en este tipo de evidencia. Si una persona se lastima y comienza a gritar o a

quejarse, consideramos ese tipo de conducta como un fundamento o criterio para atribuirle dolor. Pero no necesitamos observarnos en un espejo para ver nuestras lágrimas y advertir entonces que estamos sufriendo. Y parece obvio que debemos saber qué significan los conceptos que utilizamos para atribuirnos estados mentales a nosotros mismos. En este sentido, puede caerse en la tentación de suponer que dicho significado es *independiente de su expresión conductual asociada*, puesto que una persona no aplica esas expresiones a sí misma sobre la base de su conducta. De esta manera, la explicación más plausible parece ser entonces que “sabemos qué significan esas palabras en virtud de asociarlas con las experiencias que tenemos, y con respecto a las cuales tenemos un acceso privilegiado” (Bennett y Hacker 2003: 87).

Ahora bien, como señalan Bennett y Hacker (2003), uno de los problemas que genera esta explicación es que si efectivamente a nuestros términos psicológicos se les asignara un significado por medio de una “observación interna” o algo así como una ostensión privada (esto es, una suerte de señalamiento de un evento mental equivalente al de utilizar el dedo índice para indicar que el color rojo es aquel que exhibe esa remera del Che Guevara), nadie podría saber qué significan los términos mentalistas de las demás personas, porque lo que una persona significa por la palabra ‘dolor’, por ejemplo, es definido supuestamente por referencia a la experiencia que tiene (la cual, de acuerdo con el modelo favorecido es privada y no puede ser tenida por otros), y no puede en principio saber lo que otras personas tienen cuando dicen que tienen un dolor (dado que, en el modelo favorecido, es inaccesible a los demás). Por todo lo que él sabe, los demás podrían tener una sensación diferente, que sólo está vinculada con el llanto y la queja (Bennett y Hacker 2003: 98). En cualquier caso, podría pensarse que el rol que podría llegar a cumplir el significado privado dentro de las prácticas comunicativas es equivalente al del escarabajo en la caja de Wittgenstein, es decir, ninguno.

4. Una alternativa expresivista para la asimetría entre la primera persona y la tercera y en contra del carácter “epistémicamente privado” de la experiencia.

Como vimos, para la idea cartesiana de la mente que delineamos en este capítulo, existen ciertos hechos acerca de los cuales tengo un acceso epistémico privilegiado: mis estados conscientes actuales. De acuerdo con esta perspectiva, se supone que

conozco estos estados de manera directa, inmediata, no inferencial (e, incluso, infalible). Los conozco de un modo en que nadie más puede conocerlos y de una manera en que no puedo conocer otra cosa. Como contrapartida, se asume que la atribución de estados mentales conscientes en tercera persona está basada en la observación de la conducta de las personas que, en el mejor de los casos, nos permite inferir o conjeturar el estado “escondido” detrás de los actos.

La filósofa Dorit Bar-On (2004) se aparta de esta presuposición (compartida también por posiciones fisicalistas) según la cual tenemos un acceso epistémico privilegiado a nuestros estados mentales presentes, y parte en cambio de la premisa menos controvertida que atribuye una peculiar seguridad a las adscripciones psicológicas en primera persona del singular en tiempo presente o, en inglés, *avowals*. Esta seguridad especial se refleja en el hecho de que si alguien afirma de manera espontánea y sincera “Siento un dolor muy fuerte en este momento”, no consideramos que sea adecuado preguntarle qué fundamentos o evidencias tiene para hacer esa afirmación, ni cuestionamos normalmente su verdad. En este sentido, Bar-On sostiene que sobre estas afirmaciones existe una “presunción de verdad”. En este escenario, parte de la tarea que asume la autora es explicar esta seguridad (que daría lugar a la asimetría mencionada arriba), sin recurrir a ningún acceso epistémico privilegiado ni favorecer algún tipo de dualismo ontológico.

De acuerdo con esta perspectiva, la seguridad de la que gozan este tipo de auto-atribuciones está justificada en parte por la inmunidad o invulnerabilidad a cometer ciertos errores al realizarlas. Veamos. Desde el punto de vista semántico, dichas atribuciones informan acerca de estados de cosas contingentes; específicamente, identifican a un individuo –el que las emite- y le adscriben cierto estado mental. Sin embargo, la idea es que cuando nos atribuimos estados mentales de esta manera no podemos equivocarnos ni en la identificación del sujeto al que se le atribuyen (quién realiza la atribución), ni en la identificación del tipo de estado que se adscribe. Sin embargo, esta imposibilidad de cometer ciertos errores no supone un funcionamiento seguro o exitoso de algún proceso cognitivo sumamente fiable o infalible. Por el contrario, la idea de Bar-On es que la imposibilidad de cometer estos errores de identificación se basa en que no llevamos adelante ninguna tarea de identificación a la hora de realizar las adscripciones mentalistas en primera persona en tiempo presente. En particular, no se trata de que, al realizar estas auto-

atribuciones, primero me percato de que alguien, por ejemplo, siente un dolor, y después, en un segundo momento, reconozco que ese alguien soy yo (como sucedería a alguien que dijera: “Hay alguien que siente dolor ¿pero soy yo el que lo siente?”). Tampoco sucede que primero identifico que siento algo y luego reconozco que eso que siento es un dolor de cabeza. (Como si dijera, “siento algo, ¿pero es un dolor de cabeza o de panza?). De todas maneras, todo esto no implica que para Bar-On la verdad de estas atribuciones mentalistas esté necesariamente garantizada o que crea que son infalibles. Para la autora, existe la posibilidad de que sean falsas. Lo que sí podríamos afirmar desde esta perspectiva es que si hay algún tipo de error, ese error no puede deberse a un error por identificación o algún tipo de error de corte epistémico.

Al mismo tiempo, esta doble invulnerabilidad al error se explica, según Bar-On, por el carácter *expresivo* de estas atribuciones en tanto *actos*. En efecto, de acuerdo con la propuesta de la autora, la asimetría entre los dos tipos de atribuciones psicológicas que venimos considerando descansaría en el hecho de que las auto atribuciones en primera persona en tiempo presente constituyen la forma lingüística que tenemos de *expresar* nuestros estados mentales, de “dar voz” a nuestra mente (*speaking our mind*).³¹

Para desarrollar su posición, Bar-On hace una distinción entre tres sentidos distintos de “expresar”:

EXP1 en el sentido de la acción: una persona expresa un estado propio haciendo algo intencionalmente. Por ejemplo cuando intencionalmente doy un abrazo o digo “¡Qué bueno verte!”, intencionalmente hago algo y doy expresión a mi alegría de verte.

EXP2 el sentido causal: una emisión o conducta expresa un estado subyacente por ser la culminación de un proceso causal que empieza con dicho estado. Por ejemplo, una mueca involuntaria o las manos temblorosas pueden expresar, en el sentido causal, el dolor o el nerviosismo, respectivamente.

³¹ Existen otras versiones (a mi juicio menos elaboradas) de esta perspectiva expresivista en la literatura filosófica contemporánea en los trabajos de Hacker 2005 y Finkelstein 2010. Por lo demás, la perspectiva expresivista tiene su origen en cierta lectura de los aportes de Wittgenstein acerca de los *avowals* y de la relación de la conducta con los estados mentales.

EXP3 el sentido semántico: e.g. una afirmación expresa una proposición abstracta, un pensamiento o un juicio por ser una representación (convencional) de la misma. Por ejemplo, la frase “Está lloviendo afuera”, expresa en el sentido semántico la proposición de que está lloviendo afuera en el momento de su emisión en la vecindad del emisor (Bar-On 2004: 248).

De esta manera, Bar-On sostiene que las atribuciones mentalistas en primera persona en tiempo presente se asemejan, en tanto *acciones*, a otras conductas expresivas que incluyen expresiones naturales no lingüísticas (como los abrazos); pero en tanto *productos*, son oraciones de un lenguaje que expresan³ proposiciones autoadscriptivas. Así, por ejemplo, supongamos que me siento contento de ver a un amigo después de mucho tiempo sin hacerlo y digo (o pienso) “¡Estoy tan contento de verte!”. En este caso, mi auto-adscripción expresa³ una proposición que es verdadera si efectivamente es cierto que estoy contento de ver a mi amigo. Así, en la medida en que mi atribución proviene, según Bar-On, “directamente” desde mi estado mental de estar contento de ver a mi amigo, podemos decir que esta auto-atribución expresa² el sentimiento de alegría, como podría hacerlo una expresión facial (una sonrisa amplia) o un gesto. Pero la auto-atribución mentalista, al igual que la afirmación “¡Qué bueno verte!” (que no es auto-adscriptiva) y el abrazo que podemos darle a mi amigo, no son un simple acto reflejo de mi cuerpo. Es algo que *yo hago*; un acto intencional de mi parte. De hecho, el abrazo, la afirmación no auto-adscriptiva, y la auto-adscriptiva, son todos actos intencionales genuinos que sirven para expresar mi alegría pero no meramente en un sentido causal.

Ahora bien, el *producto* de cada uno de estos actos sí es diferente. En particular, Bar-On considera que sólo las auto-atribuciones mentalistas como las que venimos considerando tienen propiedades semánticas que expresan³ una *proposición* acerca de nuestros estados mentales. Dicho de otra manera, en tanto *actos*, los auto-atribuciones mentalistas expresan¹ los estados mentales del individuo que los emite. Y en tanto *actos lingüísticos*, expresan³ en qué estado mental se encuentra el individuo que los realiza. De esta manera, el expresivismo rechaza la idea cartesiana de que (al menos algunas de) las auto-atribuciones psicológicas sean una suerte de reporte de los estados mentales basados en la “observación” de los objetos y sucesos internos que pasan por nuestra mente. Además, como enfatiza Pérez (2014), el planteo de Bar-On la lleva a “cuestionar una

de las consecuencias tradicionales más ampliamente aceptadas de las explicaciones epistémicas clásicas de la asimetría entre la primera y la tercera persona: la idea de que sólo tenemos acceso a nuestros propios estados psicológicos, y que a los estados psicológicos ajenos sólo accedemos por inferencia” (Pérez 2014: 13). En efecto, para Bar-On,

Mientras que el comportamiento expresivo puede considerarse, en cierto sentido, igualmente abierto a la observación tanto por parte del observador como del sujeto, para el sujeto el comportamiento servirá para dar voz a, o dar aire a, su actual estado mental, mientras que para el observador es la manifestación del estado psicológico de otra persona. El sujeto está *en* el estado mental y se está expresando a través de su comportamiento, el observador percibe ese estado. No hay que asumir que sólo aquel que está en un estado lo puede ver con un ‘ojo interno’, mientras que los otros que observan su comportamiento a lo sumo pueden conjeturar al respecto. Todo lo contrario [...]. Si seguimos al sentido común, deberíamos considerar a la conducta natural expresiva como permitiéndonos ver la furia, el temor, la vergüenza, etc. de los sujetos. *Pace* la perspectiva tradicional introspeccionista (y sus herederos contemporáneos), no percibimos meramente la conducta e inferimos la presencia y el carácter de los estados internos del sujeto como la mejor explicación de la conducta que percibimos; en lugar de eso, percibimos sus estados mentales presentes. [...].

En este sentido, la conducta expresiva *es transparente-a- la-condición-mental del sujeto*. (Bar-On 2004: 278-279).

De acuerdo con este esquema, los estados mentales son estados expresables de los diferentes individuos que muy frecuentemente se hacen perceptibles para sus congéneres mediante diferentes tipos de conducta expresiva. Esto significa que no se piensa a los estados mentales como condiciones o estados esencialmente privados y ocultos dentro de la mente de los sujetos, que serían los únicos que capaces de captarlos. Y las diferentes manifestaciones conductuales, no serían meros signos o síntomas que sirven como indicadores más o menos fiables de sus causas (i.e. los estados mentales de los sujetos). Por el contrario: para Bar-On la conducta expresiva es suficiente para *mostrar* o, como dijimos, para “hacer perceptibles” los estados mentales de los sujetos que la realizan. De hecho, “si [...] pensamos a las expresiones conductuales como meros efectos de condiciones internas, seguiría siendo desconcertante por qué deberíamos tratar a [las auto-

adscripciones psicológicas] [...] como proveyendo más que pistas contingentemente seguras para realizar inferencias causales por parte de los observadores” (Bar-On 2004: 418).

Por último, habíamos observado que otra de las dificultades que asolaba al enfoque cartesiano era su compromiso con la existencia de un lenguaje privado. En efecto, en virtud de la constatación de ciertas particularidades de la atribución psicológica en primera persona, el cartesiano se inclinaba por considerar que lo que una persona significa por la palabra “dolor”, por ejemplo, se define por referencia a la experiencia que tiene (es decir, que dicha experiencia constituiría el significado del término). Sin embargo, puesto que de acuerdo con el modelo favorecido esta experiencia es privada, no resultaría posible comprender lo que otras personas dicen cuando afirman que sienten un dolor, dado que esta sensación resulta inaccesible para ellas.

En contraposición, el enfoque expresivista contiene ciertos recursos para dar cuenta (en parte) del significado y adquisición del lenguaje mentalista ordinario sin comprometerse con la existencia de un lenguaje privado. En particular, Hacker considera que parte de la clave en torno al significado de estos términos se aclara si prestamos atención al modo en que un niño aprende a utilizar este lenguaje mentalista. De acuerdo con este autor, no enseñamos a un niño el uso del término ‘dolor’ consiguiendo que identifique una sensación y que luego la nombre. Lo que ocurriría en cambio es que el niño aprende a sustituir ciertas reacciones naturales, como sus llantos y quejidos, por ciertas expresiones lingüísticas. Aprende a gritar, por ejemplo, “¡Ay!”; después, quizás, a decir “¡Duele!”, y después aprende a auto-adscribirse estados mentales y dice cosas tales como “Siento dolor”. De este modo, el uso y significado de las afirmaciones mentalistas es aprendido como una extensión “cultural” de una *conducta expresiva*. Así, como no puede decirse que un chico grita o se queja del dolor arbitrariamente, tampoco resultaría correcto afirmar que aprende a dar expresión verbal a su dolor *arbitrariamente*, ni tampoco que necesite fundamentos epistémicos, ‘introspectivos’ o conductuales, para exclamar “¡Me duele!” o “¡Qué dolor!”, por los mismos motivos por los que no solemos considerar que los necesite cuando grita de dolor. Todavía más tarde según Hacker, el niño aprenderá a usar “Tengo dolor” como un reporte “frío” (cuando el dolor no sea muy severo, y el doctor le pregunte cómo se siente). Junto con esto, el niño

también aprenderá que así como gritar es una manera de que sus padres lo asistan, calmen su dolor, etc., decir “Tengo dolor” también provee a los padres (y a otros) de una razón para asistirlo, un fundamento para la comprensión y la compasión. Al mismo tiempo, el niño también aprende que cuando su madre se lastima y exclama ‘¡Ay!’ o ‘Me lastimé’, esas expresiones constituyen un fundamento para utilizar la oración “Mami se lastimó”. Así, el chico aprende simultáneamente las dos facetas complementarias del uso del término “dolor”: su uso expresivo en primera persona y su uso descriptivo en tercera persona, y la manera compleja en que ese uso de la palabra en oraciones se integra dentro de la conducta humana natural y cultural (Bennett y Hacker 2003: 101-102).

Por su parte, Bar-On (2004) ofrece una explicación similar (en sus palabras, “una reconstrucción racional” no empírica) con respecto a los orígenes del lenguaje mentalista. De acuerdo con la autora, como seres humanos estamos diseñados para reaccionar apropiadamente frente a los distintos tipos de conducta expresivas: enfrentados con un chico que llora luego de sufrir algún tipo de lesión, sus padres o cuidadores no sólo buscarán reconfortarlo sino que emitirán exclamaciones como “¡Ay!” o “¿Te duele mucho?”. De la misma manera, cuando come su alimento haciendo ciertos sonidos que denotan placer, sus padres o cuidadores no sólo le ofrecen más comida sino que también podrían decirle cosas como “Mmmm.. ¡qué rico!”

En mi historia expresivista, la respuesta verbal, a menudo sirve para poner en manos del aprendiz del lenguaje *nuevas formas de conducta expresiva* –formas que pueden suplementar, y después suplantar, la conducta expresiva no lingüística que el hablante [maduro] presencia. En particular, los términos mentalistas no son vistos como ofreciendo al niño una etiqueta lingüística para una causa hipotetizada de su estado [...]. En lugar de eso, el hablante experto está nombrando la condición que percibe en la conducta naturalmente expresiva. Y transmite el término mental como una nueva adquisición para la bolsa de trucos del aprendiz (Bar-On 2004: 288).

Esta historia, que resultaría más convincente si contara con algún tipo de respaldo empírico, se contrapone con lo que Bar-On describe como versiones “introspeccionistas” de la adquisición del lenguaje mentalista, que se asemejan a las explicaciones que Hacker rechazaba. En efecto, de acuerdo con la perspectiva introspeccionista, el habla mentalista empieza cuando el aprendiz identifica dentro

suyo un estado o evento que resulta observacionalmente destacado (en un sentido peculiar de “observacional”) únicamente para él y que sólo puede ser inferido o hipotetizado por los demás. “Podemos glosar esto diciendo que los términos mentalistas funcionan como términos observacionales en sus aplicaciones en primera persona, mientras que funcionan como términos teóricos en sus aplicaciones en tercera persona” (Bar-On 2004: 290).

Pero si esta fuera el caso, nos enfrentamos a los problemas indicados arriba con respecto a la posibilidad de comunicarse a través del lenguaje, porque para el introspeccionismo sólo el aprendiz tiene acceso “observacional” al ítem interno mientras que el que enseña el lenguaje no está jamás en posición de observarlo, y sólo puede, en el mejor de los casos, conjeturar su presencia y sus características.

En resumen: ¿por qué mis auto-atribuciones mentalistas en primera persona en tiempo presente se consideran mucho más seguras que las atribuciones mentalistas en tercera persona y que las afirmaciones acerca de otros hechos del mundo? O, dicho de otra manera, ¿por qué soy la persona más indicada para contestar cuáles son mis propios estados mentales? Para el (neo) expresivismo hacer estas preguntas es similar a preguntarse por qué el rostro de una persona puede ser la mejor opción si se pretende determinar al menos algunos de sus estados mentales. Si estoy, digamos, furioso o alegre, es posible que mi furia o mi alegría, se haga visible, patente, o manifiesta en mi expresión facial. Y algo similar ocurriría con ciertas auto-adcripciones mentalistas en primera persona, en la medida en que a menudo hacen manifiestas las condiciones psicológicas del sujeto que las realiza. Por este motivo, no es necesario postular la existencia de un conocimiento privado y seguro para dar cuenta de la seguridad que caracteriza a los *avowals*. Al mismo tiempo, esta similitud también explicaría por qué no se requiere evidencia conductual o de algún otro tipo para realizar este tipo de auto-atribuciones: básicamente, por el mismo motivo por el cual no necesitamos ni reclamamos este tipo de fundamentación epistémica cuando expresamos dolor gritando o llorando.

Sin embargo, esto no supone que dichas atribuciones mentalistas, que en tanto *actos* son similares a ciertas expresiones naturales de los estados mentales, no tengan, además, valores de verdad, puesto que en tanto *actos lingüísticos* son

afirmaciones que dicen algo acerca de ciertos hechos del mundo.³² En este sentido, el neo-expresivismo intenta socavar el prejuicio filosófico según el cual, en palabras de Bar-On, la epistemología de los *avowals* debe reflejar su semántica (Bar-On 2010: 58), es decir, la idea de que la producción intencional de una auto-adscripción con significado (y que haga referencia a cierta condición psicológica) deba estar respaldada por algún tipo de auto-observación.

Por último, una vez que el expresivismo esclarece la lógica de las atribuciones mentalistas, la motivación para decir que la evidencia conductual para la adscripción de estados mentales conscientes es indirecta e inevitablemente insuficiente pierde su fuerza porque no existe (o no es necesario pensar que exista) una evidencia *directa* para el caso de las auto-atribuciones en primera persona con la cual contraponerla.

³² Este último aspecto de su posición le permite a Bar-On distinguirla del llamado expresivismo simple, según el cual, las mencionadas auto-atribuciones, al igual que los gritos o las sonrisas, no dicen ni afirman literalmente nada. En efecto, de acuerdo con este tipo de expresivismo las auto-atribuciones mentalistas en primera persona no deben considerarse ni como descripciones ni como afirmaciones; en rigor, serían *únicamente* expresiones de nuestros estados mentales. De esta manera, el expresivismo simple podría explicar la asimetría entre las atribuciones psicológicas en primera y tercera persona sin postular ninguna clase de conocimiento privado y certero. Al mismo tiempo, en la medida en que estas emisiones lingüísticas no se consideren como aseveraciones, reportes o descripciones de ciertos hechos, tampoco tendrían valores de verdad, del mismo modo en que un gesto de dolor no podría tenerlo. Sin embargo, conceder este punto implica abrazar una serie de afirmaciones poco atractivas. Por un lado, si nos tomamos en serio el expresivismo simple debemos negar algo que resulta en principio evidente y sumamente intuitivo, a saber: que las auto-atribuciones mentalistas en primera persona parecen *informar* acerca de estados de cosas contingentes. En particular, parecen atribuir cierta condición psicológica a un sujeto (el que los emite). Al mismo tiempo, también resulta razonable considerar que los *avowals* serán verdaderos en las mismas circunstancias en las que las atribuciones psicológicas relevantes en tercera persona lo son. En otras palabras, la afirmación “Siento dolor ahora” pronunciada por Juan, parece tener las mismas condiciones de verdad que la afirmación “Juan siente dolor” pronunciada por otra persona. Sin embargo, no podríamos mantener estas ideas si adoptamos el expresivismo simple. Por el mismo motivo, tampoco podrían considerarse como casos de desacuerdo genuino aquellos en los cuales un agente se auto-adscribe un estado mental en el presente (“Siento un terrible dolor en la espalda”) y otro agente rechaza la veracidad de esa atribución (“Estás mintiendo porque no quieres ayudarme a hacer la mudanza”), porque aun cuando fuera cierto lo que dice el agente acerca de su condición psicológica, su emisión no sería, en rigor, una afirmación en absoluto. Sin embargo, en una situación como la descrita quisiéramos poder decir que ambos interlocutores están en desacuerdo acerca de ciertos hechos del mundo (es decir, de cierto estado mental de uno de ellos). En segundo lugar, y otra vez por las mismas razones, el expresivismo simple tampoco nos permitiría aceptar que los *avowals* puedan servir como premisas en ciertas inferencias cotidianas como “Me siento nervioso, y vos también, por lo tanto, somos dos los que estamos nerviosos”, ya que, de acuerdo con el expresivismo simple, no deberíamos considerar que la frase “Me siento nervioso” sea una afirmación en absoluto. Por estos motivos, el desafío que asume Bar-On es desarrollar una posición expresivista que no asegure la asimetría entre las atribuciones psicológicas en primera y en tercera persona al precio de negar sus propiedades semánticas.

5. Algunas dificultades potenciales para el expresivismo.

En esta sección voy a consignar algunas objeciones que pueden plantearse al expresivismo tal como lo describimos en este trabajo y luego examinaré qué respuestas podrían ofrecerse desde esta perspectiva para hacer frente a estos desafíos.

En primer lugar, si bien un adversario como Chalmers podría llegar a admitir la explicación de la seguridad de la que gozan ciertas auto-adscripciones psicológicas de acuerdo con el esquema que ofrece Bar-On, insistiría con la idea de que, esta explicación no impide que exista un conocimiento por familiaridad acerca de nuestra propia experiencia como el que describimos arriba, que no es conceptual o que es pre-conceptual y que, por este motivo, no es abarcado por la explicación expresivista.

Frente a esta posible objeción, creo que alguien como Bar-On podría ensayar una respuesta “terapéutica”. En particular, podría afirmar que el hecho de que Chalmers piense que gozamos de un conocimiento de esas características tiene su origen precisamente en una interpretación inadecuada de un hecho menos controvertido: la especial seguridad que poseen las atribuciones mentalistas en primera persona en tiempo presente.

En otras palabras, lo que sería cierto, podría sostener un expresivista, pero que aparece completamente distorsionado en la imagen cartesiana del auto-conocimiento y de la auto-percatación, es que mi palabra acerca de algunos de mis estados mentales tiene un estatus “privilegiado”. Sin embargo, no es necesario aceptar que esto se debe a que tengo un acceso *sui generis* a un espectáculo privado que me permite luego describir o reportar lo que veo en él mientras que los demás tienen el acceso vedado a estas funciones. En lugar de eso, lo que afirmamos acerca de algunos de nuestros estados mentales es especialmente seguro porque constituye una expresión o manifestación de nuestra experiencia, mientras que lo que otros dicen no lo es.³³ De esta manera, es posible mantener la intuición de que

³³ En este punto de la esgrima dialéctica habitual, Chalmers podría insistir con la idea de que lo que queremos explicar no son los juicios, atribuciones o creencias acerca de la consciencia y sus características, sino el fenómeno mismo, a saber: la propia experiencia consciente. En el capítulo 1 indiqué de qué manera podemos responder a este tipo de declaraciones sin ser acusados de eliminativistas o de que “no nos tomamos en serio la consciencia”. Básicamente, allí mi propuesta consistió en distinguir entre *dos* maneras de “explicar” *reductivamente* (de acuerdo con la terminología de Chalmers) nuestros juicios o creencias acerca de nuestros estados mentales conscientes. Una opción consistiría en adoptar una postura verdaderamente eliminativista y pensar

hay algún tipo de diferencia importante entre las afirmaciones que hacemos acerca de nuestros propios estados mentales y las que hacemos acerca de los estados mentales de los demás (y del resto de los hechos del mundo en general), que justifica la confianza de la que gozan, pero que no precisa comprometerse, ni con las dificultades que entraña el dualismo, ni con la existencia de un conocimiento de carácter privado y misterioso inaccesible a la ciencia.

Una segunda objeción que podría plantearse al expresivismo se exhibe en un artículo de David Robjant y Stowey Nether (2012) dedicado a refutar la tesis wittgensteniana, según la cual, de acuerdo con el parágrafo #246 de las *Investigaciones Filosóficas*, “no puede decirse de mí (excepto quizás como una broma) que sé que tengo un dolor”. De acuerdo con esta objeción, si bien es cierto que normalmente no nos basamos en evidencia conductual a la hora de afirmar si sentimos o no dolor, es falso que jamás utilicemos este tipo de evidencia como una suerte de fundamento para afirmar si tenemos o no cierta sensación. En principio, esto parece atentar contra la verdad del expresivismo en la medida en que éste aseguraba que parte de lo que convertía a los *avowals* en afirmaciones seguras era que para realizarse no se basaban en ningún tipo de fundamento epistémico y mucho menos en la observación de la conducta propia. De ahí la imposibilidad de cometer ciertos errores al realizarlos. Sin embargo, Robjant y Stowey parecen hallar un contraejemplo para esta tesis.

¿...no hay un tipo de dolor, que podríamos llamar ‘emocional’ [...], en donde podría ofrecer mi conducta a mí mismo como un fundamento para alcanzar una conclusión acerca de si tengo dolor? [...] [Por ejemplo, imaginemos que] me he convencido a mí mismo equivocadamente de que ya no siento ningún dolor [...], y estoy forzado a reconocer este error por cierto patrón de conducta. Tal vez encuentro que rompo en llanto. ¿No sería esa conducta un fundamento para una conclusión? Si estuve terriblemente afectado por N previamente, y ahora encuentro que estoy llorando, podría tratar estos dos hechos como fundamentos para pensar que todavía siento dolor? (Robjant y Nether 2012: 273).

que lo *único* que habría que explicar serían precisamente nuestros juicios y creencias acerca de la consciencia y su naturaleza. La otra opción consistiría en reconocer que hay un fenómeno adicional a explicar además de juicios y creencias, pero sin admitir cierta *caracterización* fraudulenta del mismo. En este caso, lo que me propuse negar es que hubiera un conocimiento privado de nuestras experiencias, pero no las experiencias mismas.

Esta última pregunta que plantean los autores es retórica. Ellos consideran que la respuesta es obviamente afirmativa. Sin embargo, no creo que pueda considerarse que este argumento representa una objeción realmente seria en contra de la plausibilidad del enfoque expresivista. Por un lado, porque Bar-On no piensa que *todas* las auto-atracciones mentalistas sean *expresiones*, en sentido estricto, de determinados estados mentales. De modo que no contradice al expresivismo verse obligado a reconocer que *algunas* auto-adscripciones mentalistas podrían ser el fruto de algún tipo de inferencia, e incluso una basada en la observación del comportamiento propio. De hecho, di algunos ejemplos de ese tipo de auto-atracciones mentalistas al comienzo de este capítulo. Podría, por ejemplo, describirme a mí mismo como estando furioso con una persona después de haber observado mi comportamiento hacia ella, o de que otro me lo haya hecho notar. En estos casos, cuando conozco mi estado mental por medio de algún tipo de observación o inferencia, no estoy, en sentido estricto, *expresando* mi estado mental, y la credibilidad de mi atribución estará dada por la credibilidad de la evidencia que presente en su favor.

En otras palabras, con respecto a estos estados mentales atribuidos por estos medios, la palabra del sujeto no goza de ninguna presunción de verdad *per se*. De hecho, dado que en principio cualquiera puede acceder a la misma evidencia que él (a saber, su conducta), las afirmaciones de terceros acerca de sus estados mentales son tan seguras como las del propio sujeto. Por otro lado, no debe olvidarse que parte de lo que convertía a la posición cartesiana en una postura inviable, era que postulaba la existencia de un conocimiento *privado* acerca de nuestros propios estados mentales, es decir, acerca de ciertos hechos que sólo el sujeto podía conocer. En este caso que estamos considerando ahora, en cambio, la evidencia en cuestión es claramente pública.

Existen sin embargo otras cuestiones que pueden resultar más preocupantes para el neo-expresivismo de Bar-On. Como se dijo arriba, la autora no supone que los auto-atracciones mentalistas en primera persona en tiempo presente sean infalibles o inevitablemente verdaderas. Lo único que parecía asegurarnos el carácter expresivo era la invulnerabilidad a cometer cierto tipo de errores epistémicos. Sin embargo, esto no nos evade de la posibilidad de cometer errores de otra clase. De hecho, Bar-On considera que hay tres tipos de situaciones en los cuales

las atribuciones psicológicas en primera persona en tiempo presente resultarían erróneas o desacertadas y por lo tanto falsas.

A. Casos de expresión silenciosa o ausente, como cuando el sujeto se encuentra en cierta condición mental expresable pero sin expresarla. En este caso ninguno de sus potenciales observadores podrá determinar su condición.

B. Casos que involucran la simulación, como las actuaciones que ejecuta un actor en un escenario o alguien que nos miente. En estos casos, Bar-On considera que el sujeto expresa un estado o condición mental pero no la suya propia. Aquí parecerá que el individuo está en cierto estado cuando en realidad no lo está.

C. Por último, están los casos de expresión “fuera de lugar”, en donde la conducta expresiva no concuerda con el estado en que se encuentra el sujeto, como el caso de la persona que va al dentista y precavido de todo lo que son capaces estos profesionales, comienza a gritar antes de que el médico utilice el torno. En este caso el sujeto estaría expresando una condición (de dolor, por ejemplo) que no es la que de hecho atraviesa.

Ahora bien, una vez que se reconoce que la conducta expresiva no es una condición *necesaria* para considerar que efectivamente estamos en cierto estado mental (como mostraría el caso de la simulación o de la expresión ausente), ni tampoco *suficiente* (como probaría el caso del mentiroso y de aquel que expresa “mal” el estado en que se encuentra), ¿no se abre la posibilidad para plantear ciertos escenarios escépticos en los cuales las conductas expresivas resultan sistemáticamente inadecuadas con respecto a los estados que se supone que deberían expresar y en donde es imposible determinar por observación de la conducta en qué condición se encuentran nuestros semejantes? ¿Podrían los propios agentes determinar en su propio caso la no concordancia de sus actos con sus estados si no poseen algún tipo de conocimiento “interno” acerca de estos últimos que les permita comparar?

En algunos pasajes de su producción, Bar-On (2004, 2007) sostiene que el hecho de que en ocasiones podamos equivocarnos con respecto a la autenticidad de la conducta expresiva, -y tomemos, por ejemplo, una simulación como si fuera una conducta sincera- no impide que pueda decirse que, cuando esto no sucede, efectivamente seamos capaces de percibir los estados mentales de las demás personas por medio de su conducta expresiva, de la misma manera en que el hecho

de que podemos en ocasiones tener la impresión de que estamos viendo un árbol no significa que a veces no veamos realmente un árbol en lugar de inferir su presencia indirectamente a través de “apareceres” visuales.

Con todo, no creo que esta afirmación alcance para mitigar las dudas que señalé en el párrafo anterior. En efecto, lo que intenté señalar ahí es que no es claro que, desde la perspectiva expresivista, una vez que se acepta que las conductas expresivas no son condiciones necesarias ni suficientes para que existan los estados mentales que expresan, se pueda despejar de plano la posibilidad de concebir escenarios escépticos en los cuales, puesto que no contamos con *criterios* adecuados, no somos capaces de reconocer que las conductas de nuestros congéneres están “desajustadas” con respecto al estado mental que se supone que deberían expresar. Es cierto que para ser capaz de fingir acerca de ciertas condiciones psicológicas se requiere haber desarrollado previamente algunas destrezas específicas que no adjudicaríamos, por ejemplo, a los niños muy pequeños, no porque sean especialmente honestos sino precisamente porque no cuentan con los recursos que consideramos necesarios para que pueda decirse que realizan actos verdaderos de impostura.

Por otra parte, dominar el concepto de “fingimiento”, es decir, ser capaz de fingir y de considerar ciertas conductas de los demás como falsas, parece requerir necesariamente ser capaz de considerar otras como casos de expresiones *genuinas*. Además, si vemos a un hombre en llamas gritando y revolcándose en el piso, no resulta natural suponer que podríamos poner en duda la autenticidad de su desesperación y sufrimiento. En conjunto, todas estas consideraciones apuntan a señalar que la objeción que sostiene que no siempre es posible separar los casos en que la conducta “coincide” con el estado mental relevante de los que no, tendría, al menos, algunas limitaciones más o menos claras. Sin embargo, nada de esto parece eficaz para disipar los problemas que se suscitan cuando enfrentamos casos del tipo C, es decir, en aquellas ocasiones en que si bien el sujeto no está fingiendo, su conducta expresiva no resulta adecuada. ¿Cómo distinguiríamos un mundo en el cual esto sucede todo el tiempo de un mundo en el que no? Al parecer, otra vez carecemos de un criterio relevante. En todo caso, se podría intentar ofrecer una explicación alternativa del caso del hombre que va al dentista y grita antes de que se encienda el torno. En lugar de suponer que expresa equívocamente una sensación

de dolor (por el momento) inexistente, podría pensarse que su grito responde a un estado de alarma producto de aquello que intuye que puede ocurrir. De todos modos, no es seguro que esta alternativa pueda extenderse a otros casos del mismo tipo.

Por otro lado, en la sección anterior se dijo que, de acuerdo con el enfoque expresivista, ciertas conductas expresan estados mentales en el sentido de que de algún modo los “muestran” o los “hacen perceptibles”. De acuerdo con esta propuesta, la idea era que, más o menos literalmente, vemos, pongamos por caso, el dolor en la cara de un sujeto. Sin embargo, no resulta muy claro de qué manera las conductas mencionadas muestran o hacen perceptibles los estados mentales de los sujetos y, sobre todo, de qué modo las auto-atribuciones mentalistas lo hacen (si es que, como parece inclinada a opinar Bar-On, no lo hacen de la misma manera que las conductas).

Con todo, si este enfoque pretende dar cuenta de la especial seguridad o de la presunción de verdad que se le asigna a las auto-adcripciones mentalistas en primera persona en tiempo presente, es claro que no puede plantearse que la relación entre el estado mental y la conducta expresiva sea una relación similar a la que se da entre una causa y su efecto respectivamente. Esto es así porque, como se dijo, “si [...] pensamos a las expresiones conductuales como meros efectos de condiciones internas, seguiría siendo desconcertante por qué deberíamos tratar a [las auto-adcripciones psicológicas] *avowals* (entendidos como expresiones) como proveyendo más que pistas contingentemente seguras para realizar inferencias causales por parte de los observadores” (Bar-On 2004: 418).

En otras palabras, si esta fuera la propuesta expresivista, entonces no se distinguiría esencialmente de otras alternativas rechazadas, porque en este caso, la conducta expresiva, devenida ahora en una suerte de efecto de cierta condición interna (sea material o no), resultaría simplemente un síntoma más o menos confiable del estado mental “oculto” que presuntamente lo causa pero que, en sentido estricto, siempre permanece “escondido” detrás de la conducta como una hipótesis, y de ningún modo podría considerarse como algo que es “percibido” de forma directa.

En este contexto, y aunque coquetea con esa idea, Bar-On no parece estar del todo convencida con la sugerencia de que las conductas posibilitan la percepción

porque son algo así como un *componente característico* de los estados mentales que expresan (Bar-On 2004: 297). De acuerdo con esta propuesta, así como puedo ver un árbol al ver una de sus ramas -que contaría como un componente característico del mismo-, podría ver tu tristeza al ver tu llanto. Al mismo tiempo, se supone que un componente característico podría tener un origen, digamos, cultural o social, (en lugar de tener un origen “natural” o “innato”), de manera que expresiones lingüísticas como “¡Ay!”, al convertirse en una “segunda naturaleza” para los sujetos, se transforman también en un componente característico de cierto estado mental. Además, dicho componente podría ser un elemento característico exclusivamente de determinados estados mentales en ciertas culturas, comunidades o sujetos.

Por último, estos componentes característicos no son necesariamente *esenciales*, es decir, no necesitan estar presentes para que pueda afirmarse que está presente el objeto o evento del que forman parte, de la misma manera en que el árbol seguiría siendo el tipo de objeto aun cuando perdiera la rama por la cual lo vemos. De este modo, esta noción admite que el estado mental puede producirse sin que deba producirse obligatoriamente alguna conducta expresiva característica (esto habilitaría el tipo de situaciones descritas antes en el punto A y B por lo menos).

La ventaja de esclarecer la idea de mostrar o hacer perceptible un estado mental a través de la noción de “componente característico” reside en el hecho de que nos permite por lo menos vislumbrar una explicación más informativa en torno al carácter seguro de los *avowals*. En efecto, en la medida en que se constituirían como parte integrante (aunque no necesaria ni esencial) de los estados mentales que expresan (o al menos eso es lo que parece implicar el concepto de “componente característico”), cuando el sujeto realiza una auto-atribución de este tipo de forma sincera, no existe la posibilidad de que se dé la conducta lingüística expresiva sin que se dé el estado mental que es expresado por ella, y ese estado es justamente el que hace verdadera la auto-adscripción mentalista.

Sin embargo, Bar-On no considera que para que su postura resulte explicativamente exitosa, necesite adoptar obligatoriamente la idea de “componente característico”. Según la autora, la tesis que resulta innegociable para el neo-expresivismo es la que afirma (más débilmente) que lo que vuelve distintivamente seguros a los *avowals* es el hecho de que cuando realizamos este

tipo de emisiones *hacemos lo mismo que hacemos cuando decimos que expresamos nuestra condición psicológica mediante una emisión no auto-adscriptiva* (Bar-On 2010: 55-56).

En otras palabras, realizar este tipo de auto-atribuciones implica usar un vehículo lingüístico auto- adscriptivo (como por ejemplo la oración “Estoy muy contento de verte”) para hacer lo mismo que podemos hacer cuando usamos un vehículo lingüístico no auto-adscriptivo (como por ejemplo la oración “¡Qué bueno verte!”, que puede ser usada para expresar mi estado de ánimo). Con todo, esto y no mucho más es lo que Bar-On nos dice acerca de la noción central de su propuesta. Pero precisamente por el peso que tiene en la economía argumentativa de su filosofía y porque “expresar” se dice de muchas maneras, se vuelve perentorio realizar un esfuerzo para intentar ofrecer una caracterización más amplia e informativa de la misma.

6. Conclusiones

En este capítulo procuré desestimar otra de las notas que convertiría a la consciencia en un fenómeno de difícil armonización con el resto de los conocimientos científicos contemporáneos: el carácter privado de nuestro acceso a ella. Con este objetivo, desarrollé una argumentación en dos etapas.

La primera de ellas consistió, por un lado, en mostrar las consecuencias inadmisibles que podrían seguirse después de adoptar la tesis del conocimiento privado, entre las cuales se destacan: la imposibilidad de tratar los estados mentales conscientes desde una perspectiva científica, el compromiso con el “problema de las otras mentes”, el compromiso con la aceptación escenarios de dudosa coherencia interna (como el escenario zombi) y con una concepción del lenguaje que, en la medida en que supone que los significados de los términos mentalistas están constituidos por entidades privadas, hace difícil comprender cómo es que lograríamos comunicarnos unos con otros, tal como parece que de hecho hacemos, a través de él.

En segundo lugar, mostré que, si se adopta el enfoque expresivista de filósofos como Dorit Bar-On (2004, 2010), es posible dar cuenta de algunos de los hechos que podrían motivar la aceptación de la tesis del conocimiento privado, sin verse obligados a postular un conocimiento de esas características ni asumir sus

consecuencias perniciosas. Entre las motivaciones mencionadas, consideré que una de ellas era la constatación de cierta asimetría entre las atribuciones mentalistas en primera persona de tiempo presente y las atribuciones mentalistas en tercera persona según la cual, las primeras contarían con un grado considerable de certeza o de “presunción de verdad” que se no se le asigna a las segundas. La segunda motivación para postular la existencia de un conocimiento privado la adjudiqué a otras peculiaridades de las atribuciones psicológicas, a saber: que mientras que aquellas que se realizan en tercera persona se basan fundamentalmente en la observación previa de la conducta de la persona u organismo, para realizar las auto-atribuciones mentalistas en primera persona, no nos basamos en este tipo de evidencia. Así, de acuerdo con una perspectiva que llamé “cartesiana”, estas asimetrías y peculiaridades de las atribuciones psicológicas se explicarían en última instancia en virtud de una asimetría epistémica, según la cual, contamos con un conocimiento directo, seguro (y acaso infalible) y privado con respecto a nuestros estados mentales conscientes que se contrapone con el conocimiento indirecto e irremediamente expuesto al error que tenemos con respecto a los estados mentales de las demás personas. En relación con este punto, el cartesianismo acepta un tanto acriticamente la idea de que la auto-atribución de estados mentales conscientes se asemeja a la descripción de fenómenos físicos, en donde primero se daría un acto de observación “interna” de dichos fenómenos y luego, con base en esa evidencia, se daría la posibilidad de llevar adelante la mencionada descripción, de la misma manera en que una afirmación acerca del desorden que domina en mi habitación se basa o puede basarse en una observación previa de la misma y su estado.

Por mi parte, y a pesar de algunas dificultades mencionadas en la sección 5 de este capítulo, indiqué que el expresivismo (o neo-expresivismo) de Bar-On (2004, 2010) ofrece el inicio de una alternativa al paradigma cartesiano, por lo menos en lo que respecta a las atribuciones psicológicas. De acuerdo con esta posición, la seguridad especial que identifica a las mencionadas atribuciones se podría justificar por el carácter expresivo de los *actos* que las producen en lugar de hacerlo postulando algún tipo de acceso epistémico privilegiado o infalible al espectáculo “privado” que se proyecta en nuestra mente. Al mismo tiempo, este aspecto expresivo explica que por el mismo motivo por el cual no necesitamos ningún tipo

de fundamento epistémico cuando expresamos, por ejemplo, nuestro dolor gritando o llorando, tampoco necesitamos -ni es legítimo reclamar- una evidencia conductual o de algún tipo para realizar determinadas auto-atribuciones mentalistas en primera persona.

Sin embargo, como vimos, esto no impide en principio que dichas atribuciones mentalistas digan, además, algo acerca de ciertos hechos del mundo y que, por lo tanto, tengan valor de verdad. En este contexto, el enfoque expresivista se destaca como una opción que, aunque su presencia no resulta nada frecuente en las discusiones en torno a la consciencia, merece ser explorada porque representa una ruptura con respecto a ciertas ideas de toda una larga y problemática tradición en la filosofía y la psicología en torno al tipo de acceso que tenemos a nuestras mentes y a las mentes de los demás. Con respecto a este último punto, el expresivismo considera que así como aquellos que están adecuadamente situados y poseen el conocimiento necesario, *ven* el significado de las palabras en una página, si tenemos las sensibilidades adecuadas y el conocimiento relevante, *vemos* (por lo menos determinados) fenómenos psicológicos en el comportamiento de las personas. Por este motivo, la idea de que la evidencia conductual constituye siempre un signo indirecto y sistemáticamente poco confiable de los estados mentales de las demás personas pierda gran parte de su fuerza porque, si el expresivismo representa una opción aceptable, no hay un conocimiento *directo* que sirva como base a las auto-adscripciones mentalistas en primera persona con el cual contraponerlo.

CAPÍTULO 4

DECIME QUÉ SE SIENTE

1. Introducción

El saber popular suele afirmar que la experiencia es la mejor maestra, y que ni el aula ni las bibliotecas pueden sustituir a la vida real. En términos generales, si alguien quisiera saber cómo es atravesar experiencias nuevas (como viajar a un lugar totalmente distinto al que suele habitar, entablar una amistad o disfrutar de una obra de arte), lo más sensato que podríamos sugerirle es que salga a la calle y se disponga a *tener* esas experiencias. En este punto, nadie objetaría que ninguna cantidad de descripciones y relatos, por muy detallados que sean, podrían de algún modo igualar a la vivencia personal de estas experiencias.

Ahora bien, cuando esta verdad inofensiva y trivial contenida en el saber popular se traslada a la filosofía de la mente, se producen algunos desconciertos metafísicos difíciles de sortear. En efecto, una de las notas que parece convertir a la experiencia consciente en un misterio, es su presunto carácter inefable, es decir, el hecho de que ninguna cadena de oraciones de una lengua natural o perteneciente a alguna parte del corpus científico resulta capaz de expresar de manera adecuada cómo se siente beber una cerveza fría en una noche calurosa de enero, o cómo se siente la experiencia táctil de los labios de otra persona. Más específicamente, la inefabilidad, junto con el carácter privado del acceso a los estados conscientes, son las dos propiedades frecuentemente atribuidas a la experiencia consciente que, tal como señala el filósofo Don Ross 1993, mantienen a raya el abordaje científico de la consciencia, puesto que permiten que los filósofos argumenten que las descripciones de la mente de tercera persona, del tipo que da la ciencia, deban en principio dejar siempre algo afuera –a saber, la naturaleza “intrínseca” de la experiencia. Y se supone que esto es así porque si la experiencia consciente es privada, entonces la perspectiva científica no puede alcanzarla, y si es inefable, entonces incluso si la ciencia lograra alcanzarla no podría describirla en una teoría (Ross 1993: 443).

Como puede preverse, en este capítulo voy a abordar este aspecto del problema de la consciencia. Para eso, voy a proceder de la siguiente manera: en la sección siguiente voy a indagar en qué sentido puede afirmarse con legitimidad que las experiencias conscientes son inefables. En este contexto, intentaré bombear la intuición de que el único tipo de inefabilidad que parecería razonable atribuirle a las experiencias perceptivas, no resulta problemático en ninguno de los sentidos apuntados. En particular, sostendré que el grano de verdad que subyace a la tesis de la inefabilidad radica en el hecho de que, en condiciones normales, no podemos tener una maestría completa de los conceptos acerca de las experiencias perceptivas sin haber atravesado las experiencias relevantes y sin contar con nuestras facultades perceptivas funcionando en perfecto orden. En la sección 3 (y sus distintas subsecciones), utilizaré las enseñanzas obtenidas en los apartados anteriores, junto con otros recursos argumentativos, para contestar al reconocido argumento de Frank Jackson en favor del dualismo ontológico con respecto a la experiencia consciente: el argumento del conocimiento. De acuerdo con este argumento, es posible contar con *toda* la información física ideal acerca de la percepción consciente del color y sin embargo ignorar todavía ciertas verdades en torno a dichas experiencias. En principio, de esto se seguiría que existen verdades acerca de la percepción consciente del color que refieren a hechos que están más allá de los hechos físicos.

Sin embargo, mostraré que si se aceptan las consideraciones de lo que llamaremos “externismo social” con respecto a ciertos conceptos acerca de la experiencia consciente, se seguiría, en línea con mi propuesta, que el tipo de “saber” que no podría en sentido estricto derivarse de ninguna cantidad de información teórica no podría ser un saber proposicional o conceptual. Para alcanzar esta última conclusión por otra vía, y reforzar de este modo mi punto, usaré un aspecto de la réplica de Dennett 1995 al argumento de Jackson y posteriormente desestimaré una reformulación de dicho argumento a cargo de Nida-Rümelin 1998 que podría leerse como una crítica a mi propuesta en torno al carácter no proposicional o no conceptual del conocimiento faltante.

Por último, si se acepta que la experiencia consciente es inefable en el único sentido que parece sensato atribuirle esta propiedad, es posible explicar por qué no se puede obtener cierto conocimiento no proposicional o conceptual acerca de los

estados conscientes sin comprometerse con ningún tipo de dualismo ontológico tal como pretende el argumento de Jackson. Por lo demás, en el contexto de este capítulo resulta relevante responder a este argumento porque podría creerse que de algún modo presupone y avala el carácter inefable de la experiencia. En efecto, de acuerdo con Dennett, algunos filósofos estarían a afirmar que la imposibilidad por parte de María de derivar ciertas verdades acerca de la experiencia consciente de la información física completa en torno a la percepción, se debe precisamente al carácter inefable del aspecto subjetivo de la experiencia (Dennett 2006: 135).

2. Precisiones en torno a la tesis de la inefabilidad de la experiencia.

El carácter inefable de la experiencia consciente o de algunas de sus propiedades resulta más o menos aceptado por autores de diversas afiliaciones teóricas.

La inefabilidad es el hecho de que es imposible comunicar a otras personas cómo qué es tener sensaciones (*feelings*). ¿Cómo le explicas a alguien cómo se ve el rojo para ti? ¿Cómo puedes estar seguro que cuando escuchas una nota musical particular tu compañero escucha las mismas cosas? (O'Regan 2010:7)

Nadie puede decirle a otro, por elocuente que sea uno y por cooperativo e imaginativo que resulte su interlocutor, la manera exacta en que uno está actualmente viendo, saboreando, oliendo y demás, los *qualia* [es decir, ciertas propiedades de la consciencia] son *inefables*. (Dennett 1988/2003: 219).

El carácter que los *qualia* tienen para el sujeto no puede ser expresado con precisión en palabras. Podríamos ser capaces de aludir o señalar al tipo de *qualia* que estamos teniendo, pero ninguna cantidad de descripción metafórica o verbal estará cerca de una caracterización completa de la experiencia. No importa cuán elocuentemente describa el sabor del vino que estoy tomando, a mi descripción siempre le faltará algo de la riqueza y particularidad de mi experiencia (Ward 2006:108).

...se suele presuponer que los qualia son: [...] inefables: inexpresables en el lenguaje público e incommunicables." (Orlando 1997: 2-3).

No tenemos ningún lenguaje independiente para describir las cualidades fenoménicas. [...] existe algo inefable en ellas (Chalmers 1999: 47).

En definitiva, la idea detrás de la tesis de la inefabilidad parece ser “que alguien que no ha experimentado un *quale* determinado no puede comprender qué es estar en ese estado cualitativo a través de una comunicación verbal con otras personas que sí han experimentado el estado en cuestión. Es imposible dar una descripción exhaustiva de una experiencia a quien no la ha experimentado” (Pérez 2002: 70). Así, quien no experimentó jamás una cierta sensación, o no tuvo las experiencias perceptivas adecuadas, en rigor no posee ni puede poseer el concepto de la sensación o de la percepción en cuestión (o, al menos, no de manera completa o de la manera en que sí lo posee quien ha tenido dichas experiencias). Ahora bien, como se mencionó en la introducción, la verdad de la tesis de la inefabilidad, por un lado, impide (junto con la tesis del carácter privado) un abordaje científico del fenómeno en cuestión. Por otro lado, como veremos más adelante, parece estar a la base de uno de los argumentos más famosos en favor del dualismo ontológico. Volveré más adelante sobre este asunto.

Antes de avanzar en esa dirección, quisiera precisar en qué sentido puede afirmarse que la experiencia consciente es realmente inefable. Después de todo, es un hecho que, como señala Pérez (2002), “nuestro lenguaje ordinario posee una cantidad de recursos bastante grande que permiten describir bastante adecuadamente nuestra vida mental, y que permiten la comunicación” (Pérez 2002: 79). En esta línea, Bennett y Hacker 2003 apuntan que de hecho describimos tanto las *experiencias* como los *objetos* de dichas experiencias. Así, por ejemplo, no habría ninguna dificultad en describir una experiencia de oler el café recién molido, o de oír el sonido de un oboe o de ver el cielo de verano del Mediterráneo. La primera experiencia es placentera (al menos para la mayoría de los bebedores de café), la segunda puede ser absolutamente maravillosa, si quien lo toca es estupendo y la pieza tocada es sublime, pero puede ser insoportable si quien la toca es un principiante y la pieza es mediocre. Por su parte, la experiencia de ver el cielo de verano del Mediterráneo puede ser embriagadora y fascinante o algo que no tiene nada especial en absoluto (como le parecería a alguien que estuviera ya familiarizado con esa experiencia). Como puede verse, las propiedades de

experiencias como éstas, o al menos las propiedades que interesan en los casos señalados, son principalmente propiedades hedónicas (Bennett y Hacker 2003: 286-287).

De manera semejante, tampoco parece haber inconveniente en describir los *objetos* de esas experiencias, si por dichos objetos se entiende cosas tales como “el café que uno huele, el oboe que uno oye, o el cielo mediterráneo que uno ve. En este sentido, podría decirse, por ejemplo, que el café está recién molido, que tiene un aroma de café recién tostado, que es marrón oscuro o color moca. Del color del cielo del Mediterráneo puede decirse que es profundamente azul, claro y brillante. (op.cit.).

En principio, no se ve con claridad que estos enunciados que consideramos no constituyan descripciones legítimas de las experiencias. Y si esto es así, debemos buscar en otra parte aquello que podría sustentar la tesis de la infabilidad de la experiencia consciente. En todo caso, de acuerdo con Bennett y Hacker, parte de la insatisfacción que puede provocar la afirmación de que estas descripciones de las diferentes experiencias son adecuadas, se vincula con el hecho de que dichas descripciones no responden al mismo paradigma con el cual se describen ciertos objetos. De acuerdo con los autores, “describimos una mesa como siendo circular, hecha de caoba, de tres pies de diámetro, etc., así como describimos una taza de café como siendo negra, caliente y amarga. Describir un objeto material es dar sus propiedades o cualidades. Pero cuando nos piden describir las propiedades o cualidades, en lugar del objeto que tiene esas propiedades o es cualificado por estas cualidades, este paradigma de descripción no aplica.” (Bennett y Hacker 2003: 287).

Ahora bien, si esto es todo lo que está en juego en esta discusión, no queda para nada claro que esta diferencia implique que no podamos describir en absoluto, *si bien de otra manera*, las propiedades y cualidades en cuestión, ni que ese tipo de descripción sea, en principio, ilegítimo o incompleto.

Por otro lado, tampoco puede considerarse como una carencia de las descripciones de las experiencias conscientes el hecho de que no constituyan algo así como un sustituto de la experiencia misma. Es obvio que una descripción de una taza de café o de una silla no son sustitutos para una taza de café y para una silla respectivamente. Al mismo tiempo, la impresión que provoca una descripción puede ser muy diferente de la impresión que provoca la percepción de aquello que

describe. Una descripción de la Capilla Sixtina, por ejemplo, no importa cuán refinada sea, no nos impresionará jamás del modo en que puede hacerlo el *ver* la Capilla Sixtina. Una descripción del amor o del duelo posterior a una pérdida, incluso si fuera detallada de la manera más refinada y “artística” posible, no nos afectará de la misma manera en que sentir amor o pesar podría hacerlo. Pero esto de ningún modo indica algún tipo de limitación en nuestras capacidades descriptivas con respecto a la experiencia consciente. Lo único que puede concluirse de estas consideraciones es que las descripciones no son equivalentes o idénticas a aquello que describen. Y no hay nada de sorprendente o inusual en ello, porque esto aplica para todo tipo de descripciones, y no sólo para las descripciones de las experiencias.

Por otra parte, si prestamos atención a otros casos discutidos por Daniel Dennett a lo largo de algunas de sus publicaciones, tampoco encontraremos allí un motivo para probar el carácter inefable de la experiencia. De acuerdo con el autor, en ocasiones atribuimos inefabilidad a nuestras experiencias conscientes porque resultan inefables en la *práctica*, pero no en principio, debido al alto grado de “informatividad” que portan. Por esta razón, resultaría imposible (siempre en la práctica) describirlas con palabras de manera exhaustiva y precisa. Para ilustrar este punto, Dennett suele apelar al ejemplo de dos espías rusos que diseñan un inteligente sistema de identificación que consiste en romper en dos partes una caja de gelatina y en conservar, cada uno de ellos, uno de estos trozos. Así, cada trozo se convierte en un “detector” único de su contraparte. El éxito de este sistema depende de que los perfiles de los bordes de las partes cortadas de la caja resulten tan “informativamente complejos” que sea imposible generar bordes idénticos mediante un corte deliberado. Así, “el borde irregular de un trozo de cartón se convierte en un dispositivo de reconocimiento de patrones prácticamente único de su pareja; es un aparato o un traductor para detectar la propiedad formal M , donde M se instancia sólo en su pareja” (Dennett 1995: 388). Cualquier intento de ofrecer una descripción exhaustiva que sirva como receta para fabricar detectores de M fracasará precisamente por la complejidad del serrado de los bordes. La manera más económica de decir qué es la propiedad M , consiste en indicar, poco informativamente, que es la propiedad que es detectada por *este* detector.

Ahora bien, para Dennett “eso es precisamente lo que hacemos cuando parecemos ostentar, con el dedo mental de la intención interna, un *qualia* o un complejo de *qualia* en nuestra experiencia. Nos referimos a una propiedad –una propiedad pública de fronteras indefinidas- haciendo referencia a nuestra capacidad personal e idiosincrásica para responder a ella [es decir, a sus detectores]” (Dennett 2003:253). Ahora bien, si este fuera el tipo de imposibilidad que impide la comunicación de nuestras experiencias, Dennett habría logrado ofrecer una explicación de por qué al menos *algunas* de ellas son incomunicables (en la práctica) sin necesidad de invocar misteriosas propiedades internas como los *qualia*: de acuerdo con esta teoría, la creencia de que nuestra experiencia es incomunicable tiene su base en el hecho de que en situaciones normales es difícil (*muy* difícil) contar con los recursos expresivos para caracterizarlas porque hay mucho (muchísimo) para decir.

Adicionalmente, Dennett sostiene que, en otros casos, la inefabilidad puede deberse al carácter *temporalmente* inanalizable de ciertas experiencias, que hace que sean consideradas como propiedades “simples” u homogéneas incapaces de ser descompuestas en unidades más simples y, por lo tanto, imposibles de ser descritas. Sin embargo, esta inefabilidad podría disiparse si el sujeto se expone a una buena cantidad de experiencias similares y se adiestra adecuadamente, como sucedería en los casos en los que se “educa” el paladar (como lo hacen, por ejemplo, los catadores de vino) o en los casos en los que se “entrena” el oído (como lo hacen los músicos). En estos ejemplos, se mostraría que lo que antes del entrenamiento parecía una propiedad “atómica” o “inanalizable”, se revela como un fenómeno compuesto y capaz de ser descrito por enumeración de sus partes. Según Dennett, no hay nada que detenga el refinamiento ulterior de nuestra capacidad para describir estas complejidades que parecían inefables (Dennett 2003:257).

Por último, existe una tercera razón por la cual podemos pensar erróneamente que la experiencia consciente es inefable. De acuerdo con este nuevo razonamiento, la enorme dificultad o imposibilidad de hecho de *imaginar* ciertas experiencias por las que no transitamos, podría llevarnos a concluir equivocadamente que estas dificultades se explican precisamente por el carácter inefable de ciertas propiedades de la experiencia consciente. Según esta idea, el “bloqueo” de la imaginación tiene que ver con la imposibilidad de abandonar, *de*

hecho, todo lo que sabemos y/o de reponer todas las asociaciones (incluso emocionales) que las experiencias evocan. Así, por ejemplo, me puede resultar difícil imaginar qué se siente leer *La Guerra y la Paz* a los veintiún años sin haber leído jamás nada de ficción rusa con anterioridad porque leí a Dostoievsky a los diecinueve y ahora paso los treinta. La razón de esta dificultad para imaginar la situación descrita no involucra la posesión, por parte de las experiencias conscientes descritas, de ciertas propiedades “intrínsecas” e “inefables”, sino en la imposibilidad empírica de olvidar todo lo que sé y de modificar el orden en que lo aprendí. Y esto es así, en última instancia, porque el orden en que adquirimos determinadas creencias, datos y ciertos gustos o deseos, impacta sobre el contexto “psicológico”, por decirlo así, en que se insertan posteriores creencias, deseos y gustos.

En el ejemplo de Dennett, el autor sostiene que resultaría prácticamente imposible imaginar hoy lo que debe haber sido para un pío y devoto luterano de Leipzig escuchar una de las cantatas corales de Bach en 1725. Y esto sería así, por un lado, por la tremenda influencia de la cultura que hace que nuestro mundo psicológico sea muy diferente del suyo, por ejemplo con respecto a las expectativas que tendríamos al escuchar una cantata de Bach por primera vez (Dennett 1995:398). Así, no podemos llevar a cabo en toda su extensión la tarea imaginativa de ponernos en el lugar de aquellos hombres porque no podemos olvidar o abandonar todo lo que sabemos. Por otro lado, para atravesar esas experiencias anteriores necesitaríamos prepararnos para responder a la música con “las mismas angustias, emociones y oleadas de nostalgia”, es decir, contar con todas las posibles asociaciones mentales que la experiencia musical podía provocar en aquella época y en cada hombre. Estas dificultades podrían llevarnos a creer que la imposibilidad fáctica de llevar adelante semejante ejercicio imaginativo se vincula con la presencia de un cierto “no sé qué” inenarrable presente en nuestras experiencias conscientes. Sin embargo, según Dennett, “no habría ningún misterio; solamente una experiencia que podría describirse con bastante precisión, pero que no podríamos disfrutar a menos que dedicáramos un esfuerzo ridículo en reconstruir nuestras estructuras de disposiciones personales” (Dennett 1995:400).

Ahora bien, todo lo dicho hasta acá no significa que no haya circunstancias en las cuales la comunicación entre las personas falla y nos vemos entonces llevados a

considerar que, en algún sentido, los demás pueden ser un enigma después de todo. En efecto, como señala Tero Vaaja (2013) siguiendo a Wittgenstein, esto podría suceder cuando los interlocutores no comparten un contexto de intereses comunes, metas y motivaciones para la acción, o un trasfondo de al menos algunas experiencias de vida relevantes familiares a ambos. Pero estos “cortocircuitos” en la comprensión de los otros no son el resultado de algún tipo de inefabilidad profunda de la experiencia subjetiva, sino de la incapacidad de compartir ciertos contextos de acción con otros. Esta falta de contextos y trasfondos compartidos es lo que impediría que, al decir de Wittgenstein, aun si un león pudiera hablar, de todos modos no seríamos capaces de comprenderlo (Vaaja 2013: 121).

Frente a este escenario ¿qué motivos podríamos encontrar entonces para sostener que la experiencia consciente es realmente inefable en algún sentido problemático? Una posibilidad consistiría en adherir a alguna versión de la tesis del lenguaje privado. En la sección siguiente, me ocuparé de desestimar esta propuesta.

2.1. La inefabilidad y el lenguaje privado.

De acuerdo con la tesis del lenguaje privado, los nombres o términos que hacen referencia a ciertos estados mentales como las sensaciones y las percepciones, son dotados de significado en virtud de ser asociados con las experiencias en cuestión, mediante una suerte de ostensión interna por parte de quien posee (de manera privada) dichos estados mentales. Si esto fuera así, se explicaría entonces por qué quien no tuviera las experiencias relevantes, no comprendería ni los términos ni los conceptos acerca de dichas experiencias por mucho que otras personas se esforzaran en explicárselos: quien no tuvo las experiencias en cuestión, no puede tener acceso a las entidades privadas que constituyen el significado y contenido de dichos términos y conceptos.

Como sostuve en el capítulo anterior, uno de los motivos por los cuales podría desembocarse en una tesis de estas características es la verificación de un hecho menos controvertido, a saber, que existe una asimetría en la atribución de estados perceptivos a otras personas y a uno mismo. En el primer caso, las atribuciones se basan en la observación de diversas conductas. Así, por ejemplo, decimos que alguien posee la facultad de percibir visualmente porque constatamos su capacidad de discriminar objetos visibles y sus cualidades, su competencia para moverse sin

llevarse por delante muebles, objetos o personas en condiciones de iluminación adecuadas, por encontrar cosas al mirar a su alrededor, por su capacidad de seguir objetos o seres en movimiento con la vista, orientando su cabeza y su cuerpo de manera apropiada, y por su capacidad de responder a cosas tales como luces, destellos y flashes.

En líneas generales, los tipos de conductas que manifiestan la posesión de una facultad perceptual dada consisten en la relativa eficiencia para realizar diferentes tipos de discriminación, reconocimiento y discernimiento; por la capacidad de perseguir fines y explorar el entorno y, en el caso de los seres humanos adultos, por la capacidad de comunicar de forma lingüística sus percepciones. En contraposición, no necesitamos observar nuestra propia conducta para afirmar que estamos percibiendo ni que estamos de hecho percibiendo algo. Al mismo tiempo, parece claro que cuando procedemos de esta manera, comprendemos el significado de los términos o conceptos que estamos utilizando. Y también parece que dicho significado o contenido fuera completamente independiente de su conducta expresiva asociada. Por lo tanto, la explicación más plausible acerca de cómo estos predicados o conceptos son dotados de contenidos supondría la postulación de algún tipo de asociación entre dichos conceptos y palabras con las experiencias mismas, sobre las que tendríamos un acceso epistémico privilegiado.

Sin embargo, habíamos visto que no debemos apresurarnos a explicar esta asimetría en la atribución de estados perceptivos en primera y en tercera persona (como “Veo un tomate rojo” y “Juan ve un tomate rojo” respectivamente) por medio de la postulación de una asimetría *epistémica*, de acuerdo con la cual, se presume que cada uno de nosotros tendría algún tipo de conocimiento privilegiado y privado acerca de sus propios estados perceptivos. Al mismo tiempo, desde esta perspectiva se suponía que este conocimiento directo se contrastaba con el conocimiento indirecto que tendríamos acerca de los estados mentales de las demás personas.

En contraposición, de acuerdo con el neo-expresivismo desarrollado en el capítulo anterior, se supone que, en ocasiones, cuando afirmamos que sentimos dolor, nuestra auto-atribución psicológica no es el resultado de una suerte de “observación” que realizamos con el fin de “averiguar” (1) si nosotros (en oposición a alguien más) (2) nos encontramos en un estado que satisface algún criterio que permita catalogarlo como perteneciendo a alguna clase de estado mental (como un

dolor, por ejemplo). Por el contrario, la idea del neo-expresivismo consistía en argumentar que dicha auto-atribución, si bien tenía valor de verdad, constituía al mismo tiempo una *expresión* del estado mental en cuestión, análoga a otras expresiones naturales de emociones o sensaciones como pueden ser los llantos y las sonrisas. Al mismo tiempo, esta estrategia permitía explicar por qué las atribuciones en primera persona, a diferencia de las atribuciones correspondientes en tercera persona, gozan de cierta seguridad, es decir, de cierta “presunción de verdad”, en la medida en que, cuando se realizan, no es posible cometer errores epistémicos.

Continuando por esta línea, Michael Hymers (2017) considera que estas apreciaciones podrían extenderse a la auto-atribución de estados mentales perceptivos, como los que están en juego en este capítulo, en la medida en que comparten algunas de las propiedades de las auto-atribuciones de sensaciones como las que vimos antes. Por un lado, en el caso de las auto-adscripciones de estados perceptivos, tampoco parece plausible sostener que ponemos en marcha algún mecanismo para determinar quién es el sujeto que posee el estado perceptivo en cuestión. En segundo lugar, tampoco parecemos basarnos en ningún criterio para establecer qué tipo de estado perceptivo estamos teniendo y cuál es su objeto. Con todo, no deben olvidarse algunas diferencias entre las atribuciones mentalistas en primera persona de sensaciones y las de estados perceptivos. En efecto, si bien no parece adecuado considerar que podemos equivocarnos con respecto al acto de percepción que nos auto-atribuimos (si se trata de un acto de percepción visual o auditiva), puede haber dudas razonables en torno al *objeto* de mi percepción. De acuerdo con Hymers,

sería absurdo preguntar si estoy seguro que *yo* –por oposición a *alguien más*– escucho un intervalo musical particular, y sería igualmente absurdo preguntar si *escucho* ese intervalo –en lugar de *verlo* u *olerlo*– así como sería absurdo preguntar si *veo* un cormorán – en lugar de *probarlo* y *sentir* sus plumas. Sin embargo, puede ser una buena pregunta si realmente escucho el *intervalo* que digo escuchar –podría ser sordo al tono. Y podría perfectamente ser una buena pregunta si realmente veo un *cormorán*, y no *otra ave*. Tal vez no sé mucho acerca de aves marinas, o tal vez no soy un observador muy atento, o quizás las condiciones de iluminación no son muy buenas. (De hecho, en circunstancias especiales puede ser una buena pregunta si veo un pájaro en absoluto o si estoy alucinando) (Hymers 2017: 142).

Con todo, podrían surgir algunas sospechas con respecto a la posibilidad de extender de manera inmediata el enfoque expresivista al tratamiento de las atribuciones psicológicas que involucran estados perceptivos. En principio, no es claro que existan expresiones conductuales naturales para los estados perceptivos equivalentes a las conductas naturales que expresan otros estados psicológicos como las sensaciones o las emociones que luego, de acuerdo con la reconstrucción expresivista, serían reemplazadas por expresiones lingüísticas del modo esquemático en que se mencionó en el capítulo anterior.

Sin embargo, Bar-On 2004 aclara que su propuesta no necesita postular una conducta expresiva natural como antecedente para cada auto-atribución psicológica correspondiente, porque las auto-atribuciones mentalistas en primera persona en tiempo presente pueden reemplazar expresiones lingüísticas previas no auto-adscriptivas. Así, no resultaría indispensable que los niños cuenten con comportamientos naturales no verbales que funcionen como expresiones de, pongamos por caso, su experiencia de oler vainilla. En estos casos, es suficiente para Bar-On que el niño esté en posición de ofrecer una caracterización verbal de su experiencia sin nombrarla. De este modo, podemos imaginar una situación en la cual un niño huele y dice: “¡Mmm, vainilla!”. Tomando esto como una expresión verbal de la experiencia olfatoria del niño, sus padres podrían responderle diciendo algo como: “¡Ah! Estás oliendo vainilla. ¡Muy bien! Papi está haciendo galletitas”. Posteriormente, en una circunstancia similar el niño articula la auto-adscripción, “¡Papi, huelo vainilla!”, la cual puede interpretarse como la expresión auto-adscriptiva de un estado perceptivo que reemplaza a lo que antes era una expresión lingüística no auto-adscriptiva de dicho estado (Bar-On 2004: 294-295). Por su parte, Hymers considera que, sin embargo,

Hay expresiones naturales claramente espontáneas de ciertos tipos de percepciones. Un perro levanta sus orejas con el sonido de un ruido extraño. Un gato cambiará su postura y mirada cuando otro gato se aproxime a su territorio. Un niño pequeño señalará o estirará su mano hacia la mariposa que aterriza en una flor cercana. Estas reacciones naturales, espontáneas, son aumentadas y reemplazadas cuando aprendemos a hablar un lenguaje. “¡Mira!” digo, señalando con entusiasmo un borrego cimarrón al costado del camino. “¡Gavagai!” exclamas cuando un conejo pasa corriendo [...]. (Hymers 2017:176).

Con todo, con respecto a algunos de estos casos mencionados por Hymers, como el del niño estirando su brazo, resulta más intuitivo considerarlos como expresiones de deseos que de estados perceptivos. En todo caso, Hacker (2013) sostiene que cuando realizamos atribuciones como “Veo un conejo” o “Me pareció que era un conejo”, no estamos reportando algún hecho interno que funcionaría como una suerte de “evidencia” para nuestras aserciones (el “aparecer” fenoménico en el que se basarían nuestras auto-atribuciones). A juicio del autor, lo que muestra el primer caso (“veo un conejo”) es que somos capaces de decir qué percibimos y mediante qué facultad sensitiva estamos obteniendo cierta información. De acuerdo con esta perspectiva, aquellos que poseen las habilidades conceptuales y lingüísticas adecuadas, son capaces de decir mediante qué facultades sensoriales accedieron a cierta información, y de responder antes preguntas como la siguiente, “¿cómo sabés que hay un conejo?”, “(Porque) lo veo”. Pero el punto central es que las condiciones que justifican agregar el prefijo “Yo veo...” o “Yo escucho...” no difieren de las condiciones que deben cumplirse habitualmente para afirmar qué es lo que se percibe. En este sentido, no se trata de que estemos considerando cierta evidencia adicional (nuestra experiencia perceptiva) a la “evidencia” que sirve de base para afirmar qué objeto o propiedad estamos percibiendo. El prefijo sólo indicaría a otros cómo se sabe que las cosas son como una las describe (Bennett y Hacker 2003: 348). Al mismo tiempo, en la medida en que las facultades perceptivas son facultades cognitivas resultan, como todas las facultades cognitivas, capacidades falibles. Por lo tanto, existe la posibilidad de acertar y también de errar al utilizarlas.

De este modo, cuando nos damos cuenta de que las cosas no son como decimos que las percibimos, ya sea porque no estaban dadas las condiciones óptimas de trasfondo, o porque se dio una falla temporal o permanente en nuestros órganos perceptuales, se utilizan expresiones para “retirar lo dicho”, como, por ejemplo, “Me pareció como si....” o “Pensé que...”, “Visualmente se me apareció como...”, y así. Pero, en principio, Hacker considera que en ningún caso se está haciendo referencia a una suerte de evidencia interna en la que se basaría la atribución psicológica. Por otra parte, si bien Hacker no explica cómo es que somos capaces de atribuirnos estados perceptivos sin criterios y, en general, sin equivocarnos, más allá de reafirmar la validez de la tesis expresivista también para

estos casos, sostiene que lo importante es no confundir la capacidad de *decir* lo que uno percibe, con la capacidad de *percibir la propia percepción* (Hacker 2007: 46).

De cualquier modo, podría en este punto recordarse que el hecho de que otra persona tenga un dolor, no implica que yo también lo tenga; y que es posible percibir la conducta de otra persona sin poder determinar qué está percibiendo. Ahora bien, si esto es así, parecería que hay algún fundamento para creer que, efectivamente, no podemos percibir sus percepciones ni sus dolores mientras que él sí puede. Pero esta consideración es equívoca. Por un lado, sí podemos llegar a percibir exactamente lo que el otro percibe (si adoptamos la posición adecuada y poseemos un aparato visual funcionando adecuadamente). Por otra parte, si aceptamos que ni percibir ni sentir consisten en *percibir* una experiencia perceptiva o una sensación, nadie (ni nosotros, ni los demás) puede *percibir* de hecho sus experiencias conscientes. Sencillamente, no es este el tipo de relación que tenemos con nuestras experiencias perceptivas o sensoriales. Al menos es esta idea la que intenté defender en el capítulo anterior. En efecto, si se adopta una perspectiva expresionista, las auto-atruciones psicológicas en primera persona deberían ser vistas como una expresión de un estado mental y no como el producto de un acto de auto-observación o algo por el estilo. Por otro lado, habíamos visto también que, de acuerdo con el expresivismo, tenemos un acceso todo lo directo que puede ser a al menos algunos de los estados mentales de las demás personas en virtud de las manifestaciones o expresiones conductuales de las mismas.

Al mismo tiempo, existen ciertas dificultades adicionales para quien pretenda sostener la tesis del lenguaje privado que no mencioné en el capítulo anterior. Como se dijo, de acuerdo con esta perspectiva, los nombres o términos que hacen referencia a ciertos estados mentales como las sensaciones y las percepciones, son dotados de significado en virtud de ser asociados con las experiencias en cuestión, mediante una suerte de ostensión privada por parte de quien posee dichos estados mentales o de un acto en el cual el sujeto presta atención a su estado mental. Sin embargo, existen algunas diferencias notables entre un procedimiento y el otro (i.e. el de dar una definición ostensiva utilizando un objeto público o muestra y el de dar una definición ostensiva de un presunto objeto interno), que arrojan algunas dudas en torno a la posibilidad de desarrollar un lenguaje de esas características. En todo caso, estas diferencias trasladan la carga de la prueba de modo tal que sea aquel que

quiere sostener una posición semejante el que debe convencernos de su consistencia. En particular, en función de las diferencias que señalaré a continuación, es el partidario del lenguaje privado el que deberá demostrarnos que el procedimiento que describe puede ser considerado lo suficientemente similar a los casos no controvertidos y cotidianos de “nombrar” un objeto o de definirlo ostensivamente como para considerarlo como miembro de una misma clase.

Como señalan Bennett y Hacker 2003, el tipo de definición relevante para esta discusión es aquella en la cual se determina parte del significado de un término a partir de un gesto corporal adecuado que apunta hacia un objeto (como una manzana) o una propiedad de un objeto (como su color), al mismo tiempo que se pronuncian frases como “Esto es una manzana” o “Esto es el color rojo”. De esta manera, se intenta establecer que todo objeto que sea semejante al objeto señalado en algún aspecto específico, puede ser legítimamente considerado como parte de una clase, como por ejemplo la de las manzanas o de aquellas entidades que tienen la propiedad de ser rojas. Al mismo tiempo, el objeto señalado en la definición ostensiva se considera como una *muestra* o un *paradigma* para la aplicación del término o concepto en cuestión, funcionando así como un estándar de corrección para dicha aplicación. De acuerdo con autores como Hacker (y tal vez Wittgenstein), este modo de asignar significado a un término, de explicarlo y/o de determinar la correcta aplicación del mismo, es un procedimiento habitual usado constantemente en la enseñanza del significado de los términos tanto a niños como adultos. Todo el punto es si puede asignársele a ciertas expresiones psicológicas un significado realizando un procedimiento análogo pero dirigido a un objeto (presuntamente) interno (como una sensación o una experiencia visual).

En principio, se observan algunas diferencias notables entre un caso y el otro. En primer lugar, no es posible señalar una sensación o un acto perceptivo de la misma manera en que decimos que señalamos un color en las muestras que encontramos en las pinturerías. En el mejor de los casos, lo que podemos hacer es prestar atención o atender al estado mental en cuestión, pero no señalarlo literalmente. Pero aun así, también en este caso podemos constatar diferencias importantes. En efecto, la atención volcada hacia un objeto externo supone el uso de alguna de las facultades perceptivas las cuales, a su vez, dependen, para su ejercicio, del funcionamiento adecuado de ciertos órganos, de las condiciones de observación

(como por ejemplo, las condiciones de iluminación para el caso de la percepción visual)³⁴ y, en algunos casos, de las habilidades perceptivas del sujeto (en efecto, hay capacidades de discriminación perceptiva que pueden mejorarse con entrenamiento). En estos contextos, es posible rectificarse de una observación preliminar luego de un exámen u observación más profunda y detenida. Por el contrario, la introspección (por nombrar de alguna manera a ese acto con el cual dirigiríamos nuestra atención hacia nuestros estado mentales) no involucra ningún tipo de órgano perceptual –es decir, no se utilizan los ojos, los oídos, u otro órgano para comunicar o percatarse de qué es lo que está pensando o se está intentando hacer, ni tampoco parece haber algo análogo a malas o buenas condiciones de observación, ni a la posibilidad de que las capacidades introspectivas fueran más o menos eficaces o mejoradas con la práctica y el entrenamiento. Al mismo tiempo, mi ubicación con respecto a un objeto externo y público establece la perspectiva que tenga de él. Cuando examino muestras de tela para mi nuevo sofá, mis ojos están situados a cierta distancia variable de esas muestras y puedo llevarlas a otra habitación o a otra parte para apreciar cómo luce en diferentes condiciones lumínicas. Además, puedo cambiar el ángulo de observación del objeto en cuestión o pedirle al vendedor que me dé su opinión acerca de cómo combinar el color.

Todo esto tampoco parece tener algún tipo de correlato en el caso de los actos de ostensión que llevaría a cabo el lingüista privado. En el caso de las muestras privadas, por ejemplo, no habría algo análogo a comparar una de ellas con otro objeto interno en el sentido en que podemos hacer una comparación del color de las paredes de nuestra casa con el de la muestra de la pinturería. En este último caso, podemos llevarnos la muestra del color a nuestra casa y yuxtaponerla con las paredes para determinar si hay coincidencia. Sin embargo, no podemos yuxtaponer de esta manera el dolor que sentimos en un momento determinado o con su recuerdo con una sensación de dolor actual para determinar si efectivamente estamos sufriendo el mismo tipo de padecimiento. Además, la infalibilidad con la que se supone que somos capaces de atribuirnos estados perceptivos o sensaciones luego del “aprendizaje privado” pone en duda que dichas atribuciones se basen efectivamente en algún tipo de observación interna, porque si lo que estuviera en

³⁴ Por ejemplo, si uno tiene mala vista como yo, puede ponerse anteojos para mejorar su visión, y si está todo oscuro, puede encender las luces.

juego fuera algo análogo al ejercicio de una facultad cognitiva, no se explica qué es lo que la haría infalible o más segura que la percepción de objetos externos.

Sostener que dicha infalibilidad se explicaría por la presencia de un conocimiento *sui generis* de propiedades que están más allá de lo físico, no parece suficiente por sí solo para dar cuenta de esa supuesta infalibilidad o seguridad, a pesar de lo que suelen creer los dualistas.³⁵ En todo caso, lo único que se logra con esta respuesta es *posponer* la explicación, porque una diferencia ontológica no explica en principio la supuesta diferencia epistemológica entre el conocimiento acerca de los estados mentales y del resto de los hechos del mundo.

En cualquier caso, concuerdo con Hymers (2017) en que aquello que parece funcionar como un supuesto necesario para soslayar el carácter por lo menos sospechoso de postular actos de ostensión interna como el mecanismo por el cual se instituiría el lenguaje acerca de las sensaciones y estados perceptivos, o se formarían los conceptos fenoménicos pertinentes, es la aceptación previa, más o menos explícita, de una metáfora según la cual los estados mentales “internos” son de algún modo análogos a los objetos del mundo externo. De esta manera, así como no resulta nada extraordinario señalar o atender sin problemas a los objetos externos dentro del espacio físico, ni realizar definiciones ostensivas como las mencionadas anteriormente, tampoco habría nada especialmente problemático a la hora de “señalar” o “atender” a los “objetos” del espacio fenoménico interno, enfocando nuestra atención en ellos.

Sin embargo, es esto precisamente lo que podría ponerse en cuestión. En particular, una vez que se presta atención a las cuestiones mencionadas en este apartado, y se contemplan además las razones que aduje en el capítulo anterior para desestimar la idea de que estemos obligados a aceptar esta manera de concebir nuestra relación con ciertos estados mentales, las motivaciones para considerar plausible la postulación de un lenguaje privado se debilitan fuertemente.

Por otro lado, la adopción del neo-expresivismo tiene consecuencias para el argumento que analicé en el capítulo 1, esto es, el argumento de los zombis. En efecto, a partir de las observaciones realizadas en este apartado y en el capítulo

³⁵ En el capítulo anterior argumenté, siguiendo a Bar-On 2004, que si en este punto el dualista opta por decir que en este tipo de atribuciones el error está excluido por principio, corre el riesgo de tener que reconocer que entendidas de esa forma las atribuciones no expresan en rigor conocimiento alguno.

anterior, estamos ahora en condiciones de sumar un nuevo argumento en contra de la concebibilidad del escenario zombi. Veamos. Habíamos dicho que, para Chalmers, existen dos conceptos diferentes acerca de ciertos estados mentales como el dolor. Por un lado, está el concepto que él llama “psicológico”, que recoge los aspectos funcionales de esta sensación, es decir, su relación con ciertos *inputs* y *ouputs* (como, por ejemplo, las heridas que suelen provocarlo y las actitudes reactivas que genera, entre otros). Por otro lado, está el concepto “fenoménico” de dolor, es decir, aquel que refiere al aspecto subjetivo y privado del dolor. Por hipótesis, este último concepto está constituido únicamente por cómo se siente estar en ese estado, y está completamente desconectado de cualquier determinación corporal.

En definitiva, puede apreciarse que autores como Chalmers creen básicamente que los predicados psicológicos son nombres de entidades internas, y que su significado puede ser captado independientemente de cualquier conexión conceptual con criterios conductuales³⁶. Sin embargo, en el capítulo anterior bosquejé una manera alternativa de pensar cómo se adquieren algunos conceptos psicológicos que no arrastraba las dificultades que conlleva postular la presencia de un acceso privado y privilegiado a los propios “estados” mentales. Básicamente, la idea era que las auto-atribuciones psicológicas elementales se aprendían como una extensión o sustitución lingüística de la *conducta* expresiva natural, y no a partir de algún logro cognitivo.

Con todo, debe recordarse que el aprendizaje en cuestión *parte* del hecho de que el individuo está teniendo la experiencia en cuestión (es decir, no estoy proponiendo un conductismo enmascarado). En todo caso, lo que quedaría descartado si este enfoque es correcto, es que los predicados o conceptos en cuestión funcionen bajo la lógica “nombre-objeto” (por lo menos en el sentido en que, por ejemplo, “rojo” es el nombre de cierta propiedad visible), y que apunten a algo que sólo cada uno de nosotros conoce “dentro” suyo, algo acerca de lo cual no podemos dudar y que tiene una conexión meramente contingente con la conducta.

Ahora bien, en virtud de lo dicho en el capítulo anterior y en los últimos párrafos del presente capítulo con respecto a la viabilidad del lenguaje privado, podríamos decir ahora que autores como Chalmers malinterpretan cómo funcionan

³⁶ Para una muestra del compromiso explícito de Chalmers con el lenguaje privado ver Chalmers (2010: 273) y Hutto (2006: pp. 78-79).

las atribuciones psicológicas y el modo en que ciertas expresiones adquieren significado, y por eso les parece que son concebibles los escenarios zombis, porque conciben a algunos “estados” mentales como “escarabajos en la caja”, es decir, como objetos internos análogos en gran medida a los objetos externos. Sin embargo, acabo de dar razones para abandonar esta concepción e incluso para dudar de la posibilidad de que conceptos como esos puedan adquirirse de la manera en que lo propone el lingüista privado o autores como Chalmers. Si esto es así, ni siquiera es claro que puedan *poseerse* los conceptos que Chalmers dice que poseemos y que se necesitarían para concebir el escenario zombi.

Llegados a este punto podríamos preguntarnos si es razonable sostener que existe algún sentido en el cual pueda decirse que la experiencia es inefable. Por mi parte, sostendré que el grano de verdad detrás de la tesis de la inefabilidad es el siguiente: para poseer una maestría completa de ciertos conceptos acerca de ciertas experiencias o estados perceptivos, no basta con tener ciertas *creencias* acerca de dichos estados mentales; para contar con una maestría completa es necesario, además, haber atravesado las experiencias relevantes y contar con las capacidades perceptivas funcionando de manera típica.³⁷ Así, por ejemplo, para poseer una maestría en el concepto de EXPERIENCIA DE VER ROJO, es necesario tener una maestría en el concepto de ROJO, y para cumplir con esta condición, se debe ser capaz de discernir mediante la vista el color rojo de otros colores.

Desde luego, ninguna descripción verbal de estas experiencias podrá dotar a alguien de esta capacidad de discriminación. Por este motivo, se sigue que quien no cuente con las facultades perceptivas relevantes funcionando de manera “normal” o “típica”, no podrá tener una maestría de los conceptos en cuestión. Así, una persona ciega, por ejemplo, no será capaz de comprender los conceptos de color o, más precisamente, no los comprenderá de modo cabal; de la misma manera, alguien que carece de la capacidad de audición, no podrá tener una comprensión completa de los conceptos de sonidos.

³⁷ Con todo, no debe creerse que esta restricción rige para la comprensión de todos los conceptos psicológicos ni para toda atribución psicológica. En efecto, de acuerdo con Bennett y Hacker 2003, es obviamente falso que para comprender el concepto de sentirse celoso o envidioso necesitemos haber sentido celos o envidia. “Cualquiera que haya leído a Proust encontrará los celos descritos con detalles perfectos, y no tendrá ninguna dificultad en entender qué son o qué se siente estar celoso. [...] Uno no necesita haber experimentado esas pasiones para captar los conceptos relevantes, y si uno lee [...] [ciertas novelas], será capaz de describir cómo se siente estar movido por esas pasiones (no a qué se parece, sino cuáles son sus características salientes). (Bennett y Hacker 2003: 290).

Con todo, esto no implica aceptar que quienes sí tienen estas capacidades funcionando adecuadamente, captan los conceptos mencionados gracias al conocimiento privado de ciertos objetos internos como el que fue descartado en el capítulo anterior. Tampoco se sigue de estas consideraciones que estas capacidades discriminativas no puedan ser explicadas en términos compatibles con el fisicalismo, de la misma manera en que el hecho de que no resulte posible adquirir o “derivar lógicamente” la habilidad de andar en bicicleta meramente a partir de una explicación teórica acerca de la dinámica de dicha práctica, no significa que esta explicación sea de algún modo deficiente o que haya que invocar misteriosas entidades adicionales e inefables para dar cuenta de ella. Simplemente, este tipo de “destrezas” o capacidades no son el tipo de habilidades que puedan obtenerse solamente a partir de la comunicación de cierta información teórica. Así, no debe confundirse (o al menos no es necesario hacerlo) el hecho de que sólo quien puede ver colores será capaz de tener una maestría más o menos completa de los conceptos de color (y, también, del de las experiencias de color), con la idea más controvertida de que los conceptos y términos de colores están definidos por referencia a entidades privadas e incommunicables.

En la sección siguiente emplearé esta idea para responder a uno de los argumentos más discutidos en el ámbito de la filosofía de la mente dedicada al problema de la consciencia: me refiero al argumento del conocimiento.

3. El argumento del conocimiento.

El llamado “argumento del conocimiento” fue desarrollado originalmente por Frank Jackson en su artículo de 1982. Su objetivo principal es negar la verdad del fisicalismo demostrando que dado que es posible poseer el conocimiento físico completo y aun así ignorar algunas verdades acerca de la experiencia consciente, deberíamos concluir que dicha experiencia involucra propiedades no físicas. Más específicamente, el argumento del conocimiento nos pide que imaginemos la historia de una súper-neurocientífica, llamada (solamente) María, que conoce todos los hechos físicos (es decir, todos los hechos que la Física, la Biología, la neurofisiología y la Química ideales o acabadas nos dicen que hay) involucrados en la percepción humana del color. Sin embargo, por haber sido confinada desde su

nacimiento a una prisión monocromática, María no ha tenido nunca experiencias de color.

Según Jackson, a pesar de toda su erudición, “parece perfectamente obvio” (Jackson 2003: 100) que María aprenderá algo nuevo cuando sea liberada de su encierro: aprenderá, por ejemplo, cómo es ver algo rojo cuando se enfrente con algún objeto que tenga ese color. Ahora bien, si el fisicalismo es verdadero, entonces si alguien sabe todo lo que se puede saber acerca de la naturaleza física del mundo, sabe todo lo que hay para saber *sin más*.

Sin embargo, dado que, como ilustraría el caso de María, es posible poseer todo el conocimiento físico pero de todos modos aprender algo nuevo, debería concluirse que eso nuevo que se conoce no es físico sino de otra naturaleza. De esto se seguiría que el fisicalismo es falso (Jackson 2003: 111-112). Y lo sería en un doble sentido: ni todo el conocimiento acerca de lo real es conocimiento físico (tesis epistemológica), ni todo lo real es físico o reducible a lo físico (tesis ontológica).³⁸ En resumidas cuentas, podríamos reconstruir el argumento del conocimiento (en su versión “fuerte”³⁹) de la siguiente manera:

Premisa 1: María tiene un conocimiento físico completo acerca de la visión humana del color.

Consecuencia 1: María conoce todos los *hechos* físicos acerca de la visión humana del color.

Premisa 2: Hay un tipo de conocimiento concerniente a los hechos acerca de la visión humana de color que María no posee antes de ser liberada. Por lo tanto, se seguiría que:

³⁸ En palabras originales de Jackson: “María es una brillante científica que, por alguna razón, ha sido forzada a investigar el mundo desde una habitación en blanco y negro, mediante un monitor de televisión en blanco y negro. Se especializa en neurofisiología de la visión y adquiere, supongámoslo, toda la información física que hay que obtener acerca de lo que pasa cuando vemos tomates maduros, o el cielo, y usamos términos como ‘rojo’, ‘azul’, y demás. Descubre, por ejemplo, precisamente qué combinaciones de longitudes de onda del cielo estimulan la retina, y exactamente cómo esto produce, *mediante* el sistema nervioso central, la contracción de las cuerdas vocales y a expulsión de aire de los pulmones que dan por resultado que se pronuncie la oración ‘El cielo es azul’. [...].

¿Qué pasará cuando se libere a María de su habitación en blanco y negro o cuando se le dé un monitor de televisión a color? ¿Aprenderá algo o no? Parece perfectamente obvio que aprenderá algo acerca del mundo y de la experiencia visual que tenemos de él. Pero entonces es ineludible que su conocimiento previo era incompleto. Sin embargo, tenía *toda* la información física. *Ergo*, hay más conocimiento por obtener que eso, y el fisicalismo es falso. (Jackson 1982/2003: 99-100. *La cursiva es del autor*).

³⁹ De acuerdo con la distinción de Martine Nida-Rümelin 2015.

Consecuencia 2: Hay algunos hechos acerca de la visión humana del color que María no conoce antes de ser liberada. Por lo tanto, de C1 y C2 se seguiría que:

Consecuencia 3: Hay hechos no físicos acerca de la visión humana del color.⁴⁰

En principio, podría creerse que parte de lo razón por la cual María aprende algo al ser liberada, residiría en el hecho de que “saber lo que se siente” al experimentar la sensación de rojo es un conocimiento inefable, que obliga a recurrir a frustrantes deícticos: “Se siente...así”. Por este motivo, resultaría imposible saber lo que se siente a partir de algún tipo de deducción, inducción o abducción basada en alguna porción del corpus científico impersonal. ¿Pero es esto necesariamente así?

Desde luego, frente a este argumento hay diversas respuestas por parte de los fisicalistas. Entre otras, se encuentra aquella que consiste en rechazar el supuesto de que María realiza efectivamente algún descubrimiento cognitivo luego de ser liberada.

Desde esta perspectiva, María podría saber perfectamente cómo se siente ver rojo dentro de su habitación monocromática. (Esta es la posición que defiende, casi en solitario, Dennett 1995). Una reacción más habitual frente al argumento de Jackson asume como un hecho que María realiza algún tipo de progreso cognitivo después de ser liberada, pero sostiene que este progreso no afecta las pretensiones del fisicalismo. Dentro de esta segunda línea de respuestas, los intentos preferidos por los filósofos buscan mostrar que existe una ambigüedad en el tipo de conocimiento involucrado en las premisas: así, algunos (como David Lewis 1988/2003 y más recientemente Laurence Nemirow 2007) han señalado que en un caso (esto es, en el momento anterior a la liberación) se hace alusión a un saber teórico o “saber que” (*know that*) y en el otro (es decir, en el momento posterior a la liberación), a la adquisición de distintas habilidades o de un “saber cómo” (*know how*). La imposibilidad de adquirir este último tipo de saber, que es una suerte de

⁴⁰ Adicionalmente, Jackson cree que su argumento se adecúa plenamente a nuestras intuiciones pre-teóricas según las cuales, nada de lo que pueda decirse acerca de lo que ocurre en un cerebro vivo desde el punto de vista físico (sus tipos de estados, sus relaciones con otros elementos, etc.) parece capaz de capturar, por muy inteligentes que seamos para unir toda esa información física, el carácter peculiar del perfume de una rosa, de lo doloroso de los dolores o de las punzadas de los celos (Jackson 2003:96)

destreza o habilidad, únicamente a través de una comunicación verbal, permitiría dar cuenta de nuestras intuiciones a favor de la idea de que María obtendría un nuevo conocimiento después de ser liberada sin necesidad de reconocer ningún tipo de dualismo ontológico.

Desde otra perspectiva, los defensores de la llamada “estrategia de los conceptos fenoménicos” (que es la estrategia más extendida en la actualidad para lidiar con el argumento del conocimiento), consideran que si bien en todas las premisas se está hablando de un conocimiento de tipo “saber qué”, hay sin embargo una variación en el *modo de presentación* del objeto conocido involucrado en cada una de ellas, sin que esto implique una variación del *objeto* al que estas premisas refieren. Básicamente, la idea es que el concepto que le falta a María antes de ser liberada es de tipo reconocitivo, es decir, un tipo peculiar de concepto cuya posesión requiere haber pasado por ciertas experiencias o ciertos estados físicos determinados a los que, en este caso, María no tenía acceso desde su prisión.

Como puede y podrá verse, mi propuesta en algún sentido toma algo de cada una de estas dos últimas respuestas.

3.1. El argumento del conocimiento y el externismo social.

Como ya mencioné, mi idea es que aquello de lo que carece alguien como María es de una maestría completa en los conceptos acerca de las experiencias de color. En este apartado voy a ofrecer algunas razones y argumentos para fundamentar la tesis de que el tipo de saber o de capacidad necesaria para alcanzar esa maestría no es de tipo proposicional o conceptual. Para eso voy a servirme, entre otras cosas, de la propuesta de quienes defienden lo que Torin Alter 2013 llama “el externismo social” en torno a los conceptos fenoménicos.

Con respecto a estos conceptos, señala Derek Ball, es posible comprenderlos o bien como conceptos que refieren a ciertas propiedades de las experiencias conscientes, o bien como conceptos que refieren a ciertas propiedades de los objetos de esas experiencias. Para la segunda opción, el concepto fenoménico ROJO refiere a cierta propiedad de los tomates maduros y de los camiones de bomberos, mientras que de acuerdo a la segunda opción el concepto relevante referiría a una propiedad de las experiencias que son típicamente causadas por ese tipo de objetos. Por mi

parte, si bien simpatizo con la primera opción, todo lo que digo en este apartado podrá aplicarse a ambas alternativas por igual.

Ahora bien, tal como señalan Gabriel Rabin (2011) y Daniel Stoljar (2005) entre otros, gran parte de los defensores de la llamada “estrategia de los conceptos fenoménicos”, consideran que haber atravesado la experiencia relevante es una condición necesaria para poseer el concepto fenoménico correspondiente. En otras palabras, de acuerdo con lo que podríamos llamar el “requerimiento experiencial”, la *posesión* de los conceptos fenoménicos o experienciales exige haber tenido la experiencia que cae bajo ese concepto. Junto con este requisito experiencial, algunos autores consideran otros dos requisitos adicionales que deberían darse para que estos conceptos puedan ser poseídos: por un lado, encontramos el requisito imaginativo, según el cual, la posesión de un concepto fenoménico requiere la habilidad de imaginar instancias de la cualidad fenoménica que cae bajo el concepto. Por otro lado, se menciona el requisito reconocitivo, según el cual, la posesión de un concepto fenoménico exige la capacidad de reconocer experiencias que caigan bajo el concepto en cuestión.

En contra de la necesidad de reconocer la satisfacción del requisito experiencial como condición para poseer los conceptos fenoménicos, filósofos como Michael Tye (2009) y Derek Ball (2009) han desarrollado, una serie de ejemplos y argumentos persuasivos para demostrar que es posible poseer al menos algunos conceptos meramente en virtud de las relaciones que se tengan con cierta comunidad lingüística, aun cuando las *concepciones* que tengamos de dichos conceptos (es decir, el conjunto de creencias asociadas a ellos), junto con la capacidad de reconocer las instancias u objetos que caen bajo ellos, sean profundamente inexactas y pobres.

Por ejemplo, de acuerdo con el externismo social uno podría adquirir el concepto de ARTRITIS sin saber mucho acerca de la artritis, simplemente por haber adquirido una palabra que expresa el concepto en su comunidad lingüística. En el célebre caso del filósofo Tyler Burge (1979), se solicita imaginar la historia de un hombre que cree falsamente que tiene artritis en su muslo. Cuando el doctor le explica que la artritis no es un tipo de dolencia que pueda acontecer en esa zona del cuerpo sino que se trata de una dolencia que afecta las articulaciones, es razonable

pensar que el hombre reconocería que tenía una creencia equivocada acerca de la artritis.

Ahora bien, la posibilidad de establecer una contraposición entre lo que el sujeto creía antes y después de ser instruido por su médico, supone que ambos, el lego y el experto, compartían el concepto de artritis. Presumiblemente, lo que ambos compartían es un término del lenguaje público que expresa el concepto en cuestión – un término como “artritis”. En otras palabras, aceptar la corrección del médico como una corrección, parece implicar aceptar que ambos comparten un mismo concepto (el de artritis) acerca de un mismo fenómeno.

Veamos otro ejemplo de Tye para sustentar el externismo social. Como resulta claro, las personas ciegas carecen de la habilidad de reconocer cosas rojas mediante la vista. Ahora bien, ¿alcanza esto para considerar que éstas personas carecen del concepto de rojo en absoluto?

Supongamos que una persona ciega llega a convertirse en una experta en olmos. Tal persona no sería capaz de reconocer olmos por medio de la vista. ¿Carecería entonces de un concepto que sí posee el experto con visión normal? Ambos pueden pensar que los olmos son árboles. Cuando la persona ciega afirma erróneamente ‘Hay un olmo aquí a la derecha’ (quizás recuerde erróneamente la ubicación del olmo) y el experto con visión normal dice ‘No, ese árbol no es un olmo’, ciertamente está corrigiendo a la persona ciega. ¿Cómo podría corregirlo, si los conceptos que expresan mediante ‘olmo’ fuesen diferentes? Sin duda, ambos utilizan el concepto *olmo* en sus pensamientos. Sin duda, ambos significan *olmo* mediante ‘olmo’ (Tye 2009: 147).⁴¹

Como puede verse, la intuición detrás del externismo social consiste en hacer notar que solemos atribuir la posesión de determinados conceptos a ciertos usuarios aun cuando el dominio o la maestría que tengan con respecto a su uso no sea perfecta, (por ejemplo porque tienen creencias equivocadas acerca de sus

⁴¹ Otro ejemplo de Tye. Un experto puede ser muy bueno a la hora de reconocer olmos a simple vista; yo no lo soy. De este modo, el experto tiene una habilidad reconocitiva, perceptiva de la que yo carezco. Sin embargo, de esto no parece seguirse necesariamente que el experto tiene dos conceptos de olmo, uno que comparte conmigo y otro que él tiene pero yo no. En principio, no resultaría intuitivo multiplicar de esta forma los conceptos. Por el contrario, lo más natural sería considerar que el experto tiene ciertas habilidades reconocitivas de las que yo carezco; conoce mucho más que yo acerca de los olmos –su concepción de los olmos (esto es, el conjunto de creencias que asocia con el concepto *olmo*) es mucho más rica que la mía. Pero intuitivamente ambos compartimos un pensamiento con el mismo contenido cuando decimos ‘Esto es un olmo’ con base en nuestras respectivas experiencias perceptivas.

referentes, o porque no son capaces de reconocer instancias de objetos que caen bajo esos conceptos). Desde esta perspectiva, lo que les permite a estos usuarios poseer los conceptos en cuestión es haber llevado a cabo cierto tipo de intercambios con una comunidad lingüística. Ahora bien, si se extiende el externismo social, que parece funcionar bastante bien en los casos mencionados, a los conceptos fenoménicos, se socavan los fundamentos para sostener el “requisito experiencial”, es decir, la idea de que sea necesario haber tenido ciertas experiencias relevantes para poseer los conceptos pertinentes. Sin necesidad de recurrir a experimentos extravagantes, esto quedaría ilustrado con el ejemplo de la persona ciega.

En este sentido, aun en su cuarto blanco y negro, María podría compartir muchas de las creencias acerca de la percepción del color con muchos de sus interlocutores que tienen experiencias cromáticas fuera de la habitación de María. Por ejemplo, ella podía acordar, basada tal vez en ciertos testimonios, que ver rojo es fenoménicamente más parecido a ver negro que a escuchar una trompeta, que los tomates maduros causan típicamente experiencias de rojo o que si algo es un número entonces no es cómo es ver rojo. Ahora bien, puesto que ella tiene esas creencias, debe ser capaz de considerar sus contenidos. Al mismo tiempo, eso implica que ella posee todos los conceptos que esos contenidos tienen a pesar de no haber experimentado jamás las experiencias de ver objetos de color rojo. Así, del mismo modo que en el caso del concepto ARTRITIS, no es ningún misterio el determinar de qué modo María podría alcanzar tales logros: adquiriendo las palabras que los expresan. Esas palabras podrían aparecer tanto en los libros de texto que lee como en las conversaciones que mantiene con sus colegas del mundo exterior.⁴²

En otras palabras, si los conceptos fenoménicos son públicamente expresables a través del lenguaje, *tal como aceptan los partidarios de la estrategia de los conceptos fenoménicos*, entonces no sería imposible que María pudiera aprender el lenguaje fenoménico ¡*aun dentro de su habitación!*

⁴² Existen otro tipo de experimentos más cercanos a la ciencia ficción que desafían el requerimiento en cuestión. Así, es posible imaginar un duplicado átomo por átomo de cualquiera de nosotros, generado por una fuerza errática y azarosa entre la niebla de un pantano, que tuviera los mismos conceptos en torno a la experiencia que nosotros pero que, en sentido estricto, nunca hubiera atravesado la experiencia relevante.

Si, en cambio, se insiste con la idea de que María de todos modos no podría poseer ciertos *conceptos* fenoménicos porque no atravesó las experiencias relevantes aunque sí maneje las palabras que suelen expresarlo, se tendrían que aceptar ciertas consecuencias que los partidarios del externismo social juzgan implausibles. En particular, se debería admitir que María no expresa los mismos conceptos en torno a la experiencia consciente antes que después de ser liberada, a pesar de utilizar en cada una de estas ocasiones las mismas palabras.

En otros términos, de acuerdo con el externismo social, una vez que se acepta (i) que hay conceptos cuya posesión requiere cumplir con el requisito experiencial y (ii) que, tal como parecen conceder la mayor parte de los propulsores de la estrategia de los conceptos fenoménicos, que dichos conceptos son expresables en el lenguaje público, se seguirían dos consecuencias:

(a) las mismas palabras o términos del lenguaje público podrían expresar *dos* conceptos diferentes pero co-referenciales: un concepto estrictamente fenoménico cuando se ha atravesado la experiencia relevante, y un concepto “no fenoménico” (por llamarlo de alguna manera) cuando, como en el caso de María antes de ser liberada, todavía no se ha tenido la experiencia en cuestión.

(b) que los perceptores normales con una historia de exposición al color poseerán al menos dos conjuntos de conceptos co-referenciales acerca de propiedades fenoménicas. (En este trabajo no me ocuparé de este aspecto de la cuestión).

De este modo, cuando María dice en su habitación cosas tales como “Si algo es un número, entonces no es cómo ver rojo (o cómo tener una experiencia de ver rojo)”, o bien “Objetos rojos como los camiones de bomberos causan experiencias de rojo”, no expresa necesariamente el mismo concepto que podría expresar pronunciando las mismas palabras luego de ser liberada, precisamente porque María habría adquirido, a partir de ese momento, nuevos conceptos acerca de la experiencia consciente. Más específicamente, cuando María, después de ser liberada, pronuncia la siguiente afirmación:

(1) Tuve una experiencia de ver rojo.

Expresa el siguiente pensamiento:

(1)' TUVE UNA EXPERIENCIA R* (donde R* es un concepto fenoménico que María adquirió después de ser liberada).

En contraposición, cuando María, antes de ser liberada, pronuncia una oración como la siguiente:

(2) Nunca tuve una experiencia de ver rojo.

Está en rigor expresando un pensamiento como el siguiente, distinto de (1)':

(2)' NUNCA TUVE UNA EXPERIENCIA DE VER ROJO, en donde "EXPERIENCIA DE VER ROJO" no constituye un concepto fenoménico.

En estos casos, a primera vista resulta natural e intuitivo considerar que (1) y (2) expresan pensamientos contrarios. Sin embargo, el partidario de los conceptos fenoménicos se ve obligado a negar esta afirmación, precisamente porque adhiere al "requisito experiencial" que sostiene que debido a la ausencia de las experiencias relevantes, María no podría tener los mismos pensamientos en su habitación.

Para bombear con mayor fuerza nuestras intuiciones en la misma dirección que el ejemplo anterior, Ball propone parafrasearlo de la siguiente manera. En este caso, consideremos que María, luego de ser liberada, piensa verdaderamente:

(3) Solía preguntarme cómo es ver rojo, pero ahora lo sé.

Si se acepta que María es capaz de afirmar con verdad una oración como (3) después de ser liberada, (lo cual, en principio, no parece objetable), resulta todavía menos claro que el concepto cuyo contenido hace referencia a la experiencia que María deseaba tener o conocer en su encierro, no sea el mismo contenido que refiere a la experiencia que atraviesa luego de ser liberada. Si María hubiera adquirido un nuevo concepto después de ser liberada, la primera parte de (3) resultaría falsa: María jamás hubiera podido ni siquiera considerar el pensamiento expresado por esa oración porque no dispondría del concepto necesario para hacerlo. Pero esto último, otra vez, parece poco intuitivo.

Ahora bien, si se rechaza el requisito experiencial, se pierde con él una manera de contestar al argumento del conocimiento. En efecto, si el requisito experiencial tuviera vigencia, podría sostenerse que puesto que María en su encierro no tuvo jamás las experiencias cromáticas relevantes, tampoco puede poseer ciertos conceptos acerca de la misma y, por este motivo, no puede ni siquiera *considerar* (al menos algunos de) los pensamientos acerca de la experiencia consciente, ni mucho menos conocer su verdad. Sin embargo, la constatación de esta imposibilidad no constituiría una razón para rechazar el fisicalismo. Veamos por qué.

El argumento del conocimiento sostiene que puesto que es posible poseer toda la información física acerca de la percepción del color y sin embargo desconocer o no ser capaz de inferir a partir de esa información ciertas verdades o proposiciones verdaderas en torno a la experiencia consciente, deberíamos concluir que los hechos mencionados por este último tipo de enunciados no son hechos físicos. Pero es un error arribar a esta conclusión metafísica a partir de un caso, como el de María, en el cual el agente, de acuerdo con el “requisito de la experiencia”, no posee todos los conceptos relevantes, porque podría aducirse que es precisamente este último aspecto de su situación lo que le impide realizar la derivación o inferencia pertinente, y no la ausencia de una implicación metafísica. Un ejemplo sencillo que proporciona Rabin puede ayudar a ilustrar esta cuestión.

Imaginemos que la hermana de María, Jane, conoce toda la información física, toda la información acerca de género, y toda la información acerca de quién está casado con quién. Pero supongamos también que Jane no posee, ni está en condiciones de poseer, el concepto de SOLTERO. Jane sabe que Úrsula es una mujer. Pero no sabe, ni está en condiciones de saber, que Úrsula no es un soltero. Si nosotros inferimos desde este hiato epistemológico un hiato metafísico, concluiremos que el hecho de que Úrsula es una mujer no implica, metafísicamente hablando, que ella no es un soltero. Pero claramente esto es falso. Que Úrsula sea una mujer *de hecho* implica que no es un soltero (Rabin 2011: 127-128).

En el caso de María, vemos que es posible tener toda la información física y sin embargo no ser capaz de derivar de ahí ciertas proposiciones verdaderas en torno a la experiencia consciente. Con todo, el caso de María no puede servir para mostrar la falsedad del fisicalismo porque María carece de algunos conceptos relevantes y esto podría explicar su incapacidad para realizar la inferencia solicitada. En otras palabras, aunque cierto conjunto de proposiciones P implique otro conjunto de proposiciones T uno podría conocer P sin conocer T porque no dispone de los conceptos relevantes de T . Sin embargo, esto no dice nada acerca de la implicación ontológica entre los hechos referidos por las distintas proposiciones. En todo caso, el argumento del conocimiento debería intentar probar que aun poseyendo todos los conceptos pertinentes es imposible derivar al menos algunas proposiciones acerca de la experiencia consciente de la totalidad de las

proposiciones físicas verdaderas. Pero esto supone operar una modificación en el argumento original.

Ahora bien, si como pretende el externismo social el cumplimiento del requisito experiencial no es una condición necesaria para la posesión de los conceptos fenoménicos, entonces se pierde una manera atractiva de afrontar el argumento de Jackson desde una posición fisicalista.

En mi opinión, la conclusión a la que habría que arribar si se aceptan las premisas del externismo social es ligeramente distinta: en todo caso, si se aceptan dichas premisas y, al mismo tiempo, se quiere conservar la idea de que María logra algún tipo de progreso epistémico después de ser liberada, entonces debería admitirse que dicho progreso no está vinculado con un conocimiento de tipo proposicional o conceptual.

En principio, podría pensarse que haber atravesado las experiencias, poder reconocerlas cuando se las tiene o imaginarlas, no es una condición necesaria para poseer los conceptos fenoménicos pero sí para tener una maestría en dichos conceptos. En efecto, de acuerdo con Rabin (2011), podemos distinguir dos niveles de comprensión de un concepto (en Alter 2013 también encontramos una distinción parecida). El primero es la *posesión del concepto*. Los requisitos para que pueda afirmarse que alguien posee un concepto no son muy exigentes. Si se es capaz de captar en cierta medida las proposiciones que contienen el concepto, o pensar contenidos donde alguno de sus componentes es el concepto en cuestión, entonces uno posee el concepto, aun cuando, como en el ejemplo del hombre de la artritis de Burge, se tenga una concepción muy pobre del mismo. De acuerdo con Rabin, el doctor tendría una relación más sofisticada y demandante con el concepto ARTRITIS, o en otras palabras, lo que tendría es una *maestría conceptual* de dicho concepto. Uno puede poseer un concepto y sin embargo estar terriblemente equivocado acerca de su extensión, e incluso acerca de verdades *a priori* constitutivas de ese concepto. La maestría conceptual sería menos tolerante con tales errores. (Rabin 2011: 130).

Si se acepta esta distinción, podríamos reacomodar las enseñanzas que deberían extraerse del externismo social. La moraleja que debería extraerse de los ejemplos vistos es que si bien no es posible tener una *maestría* o un dominio más o menos completo del concepto de rojo sin tener las experiencias relevantes, esto no

impide que, aun en esas circunstancias, no pueda poseerse, por ejemplo, el concepto de *rojo* en absoluto (o de EXPERIENCIA DE ROJO). Ahora bien, llegados a este punto contamos con otro argumento para afrontar el caso María. De acuerdo con esta nueva réplica, desarrollada por Rabin (2011), para poder inferir la falta de una implicación ontológica de una incapacidad para derivar ciertas proposiciones verdaderas de otras (como sucede en el caso de María), se debe exigir, no sólo que el agente *posea* los conceptos relevantes, sino que también tenga una *maestría* con respecto a dichos conceptos. Para ver la necesidad de esta exigencia, consideremos el siguiente ejemplo.

Imaginemos otra vez a la hermana de María, Jane. Por interactuar con sus colegas, Jane llega a poseer el concepto de SOLTERO. Ella sabe que Ulises, que está casado, no es un soltero, y que Aquiles es un soltero. Pero Jane no tiene una maestría en el concepto de SOLTERO; piensa que las mujeres no casadas pueden ser solteros. Jane tiene un conocimiento completo de P y un conocimiento del género y del estatus marital de cada persona. Jane sabe que Úrsula es una mujer, y que Úrsula no está casada. A diferencia del caso anterior, puesto que Jane posee el concepto SOLTERO, puede considerar la proposición de que Úrsula no es un soltero. Pero por su visión errónea de que una mujer puede ser un soltero, Jane no llegará a saber que Úrsula no es un soltero. Si no exigimos, en la inferencia desde un hiato epistémico o uno metafísico, que los agentes epistémicos relevantes tengan maestría conceptual (i.e. si la mera posesión de un concepto fuera suficiente), entonces estaremos forzados a concluir, sobre la base de la incapacidad de Jane para saber que Úrsula no es un soltero, que el estatus de Úrsula como una mujer no implica que no es un soltero. Pero esa conclusión es ridícula. Por lo tanto, debemos requerir, si vamos a inferir desde la falta de implicación [epistémica] una falta de implicación [ontológica], que los agentes epistémicos relevantes tengan maestría en todos los conceptos de la proposición(es) [...] [a derivar] (Rabin 2011: 131).

De este modo, a partir de la distinción entre posesión y maestría⁴³, contamos con otro recurso eficaz para contrarrestar al argumento del conocimiento. El hecho

⁴³ Alter (2013) da el siguiente ejemplo para ilustrar una distinción semejante a la consignada en el cuerpo del texto entre *poseer* un concepto y tener una *maestría* del mismo: “Considera a alguien que ha escuchado hablar de los números primos y entonces posee el concepto NÚMERO PRIMO sin maestría, en la manera en que alguien que escuchó hablar de la artritis podría poseer el concepto ARTRITIS sin maestría. Tal persona podría no estar en posición de deducir que hay infinitamente muchos números primos, incluso si su capacidad de razonamiento fuera en todos los otros aspectos ideal. No se sigue que no es *a priori* [cierto] que hay infinitamente muchos números primos. Ahora considera a alguien que tiene una maestría [del concepto de] NÚMERO PRIMO y asume que su capacidad de razonamiento es ideal. Si no pudiera deducir que hay infinitamente muchos números primos, esto podría ser una evidencia fuerte de que no es *a priori* [cierto que] hay infinitamente muchos números primos. Esto hace resaltar el punto general de que la maestría de un concepto está atada a la aprioricidad de un modo en que la posesión no lo está.” (Alter 2013: 489).

de que María no pueda derivar ciertas proposiciones verdaderas en torno a la experiencia visual del color desde la totalidad de la información física, puede explicarse, como en el caso de Jane, no por la falta de conceptos (que, después de todo, María podría haber adquirido a partir de cierta interacción con otros investigadores o de sus abundantes lecturas), sino por la falta de una *maestría* o de un dominio adecuado de los conceptos en cuestión. Pero, si esto es así, por las razones aducidas en la cita anterior, no es legítimo derivar una falta de implicación ontológica de una carencia epistémica como la de María. Para que pudiera contar con la maestría necesaria, podemos conjeturar que es necesario haber tenido la experiencia relevante, poder reconocerla cuando la tenemos, e imaginarla.

Por mi parte, acuerdo con la idea de que aquello de lo que carece María es de una maestría de los conceptos relevantes y que dicha maestría se logra, en circunstancias normales, luego de atravesar las experiencias pertinentes. Sin embargo, como ya lo dije, considero que *si* se aceptan las premisas del externismo social, junto con otros argumentos que presentaré enseguida, cobra fuerza la idea de que si efectivamente María carece de algún tipo de saber antes de ser liberada, se trata de alguna especie de saber cómo o de cierta habilidad no conceptual.

En efecto, en los ejemplos que motivan la aceptación del externismo social (como el de la artritis), y en aquellos mediante los cuales se explica la distinción entre la mera posesión de un concepto y la maestría en él (como el del concepto de SOLTERO y el de los números primos), nada impide en principio que el sujeto pueda adquirir la maestría de la que carece a través de algún testimonio o de una interacción adecuada con los miembros de la comunidad lingüística de modo tal que ninguna experiencia sea necesaria para lograrla. “Por ejemplo, si bien el hombre con artritis de Burge se equivoca en no creer que se trata de una dolencia que afecta a las articulaciones, no hay nada que le impida que llegue a conocer este hecho. Es perfectamente capaz de aprender que la artritis es una dolencia de las articulaciones mediante testimonio” (Ball 2009: 956-957).

Por otro lado, en los otros casos en los que sí podría ponerse en duda que el sujeto fuera capaz de alcanzar la maestría en cuestión únicamente a través de ciertos testimonios o, en última instancia, de algún tipo de interacción adecuada con su comunidad lingüística (como el de la persona inexperta en reconocer olmos mediante la percepción visual o el de la persona ciega) son precisamente casos en

los cuales, por diferentes motivos, el sujeto no es capaz de realizar ciertas discriminaciones perceptivas.

3.2. La réplica de Dennett al argumento del conocimiento.

Para reforzar la idea de que si María carece de algún conocimiento antes de ser liberada este conocimiento no es de tipo proposicional, podríamos retomar un aspecto de la réplica de Dennett al argumento del conocimiento que, hasta donde sé, no ha recibido demasiada atención en las discusiones filosóficas. De acuerdo con esta propuesta, María podría conocer, en virtud de su vasto saber, “exactamente qué efectos –descriptos en términos neurofisiológicos- tendría cada color sobre su sistema nervioso” (Dennett 1995: 411) y también, qué *pensamientos* tendría una vez que se enfrentara con las experiencias cromáticas relevantes, incluso *antes* de tenerlas efectivamente: “porque, después de todo, la <<mera disposición>> a pensar sobre esto o aquello no es uno de sus famosos *qualia*, ¿no?” (op. cit.).

En efecto, dado que en principio se supone que los pensamientos no son el tipo de estados mentales que deberían ser afectados por el argumento del conocimiento (en la medida en que sería posible dar cuenta de ellos en términos de estados físico-funcionales), María podría anticipar algunas reacciones que tendría al ver cosas coloreadas entre las cuales encontraríamos pensamientos como “Así es como es ver algo rojo” y cosas por el estilo.

Ahora bien, dado que esto parece suponer al mismo tiempo que María debería poder comprender los conceptos involucrados en dichos pensamientos, no resulta del todo claro que no pudiera saber entonces cómo se siente tener experiencias de color antes de ser liberada. Howard Robinson (1993) advierte la fuerza de la réplica de Dennett y considera que el partidario del argumento del conocimiento se enfrenta en este punto a una suerte de dilema según el cual, o bien reconoce que María podría saber qué se siente antes de ser liberada gracias a su conocimiento físico completo tal como muestra Dennett, o bien debería admitir que el fisicalismo fracasa, no sólo porque no puede dar cuenta de las experiencias conscientes, sino porque, además, tampoco es capaz de dar cuenta de algunos otros estados mentales como ciertos pensamientos en torno a la experiencia consciente. Pero, en principio, esto es más de lo que pretende probar el *qualófilo*. De ser así,

El argumento del conocimiento no debería ser moldeado en la forma de 'el fisicalismo puede funcionar para todos los estados mentales pero no para los *qualia*', sino en la forma de 'incluso si pudiera parecer como si el fisicalismo funcionara para estados [...] como los pensamientos, el caso de María muestra que no funcionará para los *qualia*, y podemos ver a partir de esto que no funcionará tampoco para los pensamientos –al menos, para cierta categoría de pensamientos [...]'(Robinson 1993: 177).

3.3. El argumento del conocimiento recargado.

Dentro de la literatura especializada, algunos autores consideran que es necesario efectuar modificaciones al argumento de Jackson para que pueda verse con claridad en qué sentido dicho argumento pone en duda la verdad del fisicalismo. Dado que esta modificación del escenario original podría considerarse como un argumento en contra de la idea sostenida a lo largo de este capítulo acerca del tipo de carencia epistémica de la que sufre María en su encierro, voy a ocuparme de contrarrestarlo.

De acuerdo con autores como Martine Nida-Rumenin (1998) y Daniel Stoljar (2005), en el argumento de Jackson aparecen confundidas o no del todo separadas dos cuestiones: (i) la adquisición de los conceptos en torno a la experiencia consciente que permiten tener ciertos pensamientos acerca de las mismas y (ii) la cuestión de si es posible derivar todas las proposiciones verdaderas acerca de la experiencia consciente de la totalidad de las proposiciones físicas. De acuerdo con los autores mencionados, es la segunda cuestión la que está verdaderamente en juego a la hora de evaluar al fisicalismo, dado que aun cuando María poseyera todos los conceptos relevantes, si no lograra derivar la totalidad de las proposiciones verdaderas en torno a la experiencia consciente de su conocimiento físico completo, debería concluirse que el fisicalismo es falso. Si esto es así, la cuestión acerca de las condiciones de posesión de los conceptos fenoménicos es secundaria con respecto a la cuestión de la derivación mencionada.

Ahora bien, desde el punto de vista que favorezco en este capítulo acerca de qué tipo de capacidades o saberes no proposicionales incluye el hecho de contar con una maestría completa de los conceptos fenoménicos, es esperable que María no pueda deducir estas capacidades de ningún tipo de saber proposicional, precisamente porque no constituyen un conocimiento de esa clase, de la misma

manera en que no es posible derivar la habilidad de andar en bicicleta de una serie de instrucciones. Por el contrario, quienes abogan por modificar la formulación del argumento del conocimiento de la manera indicada, suponen, al mismo tiempo, que es un saber de tipo proposicional el que no puede ser derivado de la totalidad del conocimiento físico. Con todo, no creo que logren asentar este punto de forma convincente. Para ver por qué, repasemos en primer lugar cómo se desarrolla esta nueva versión del argumento de Jackson.

En este nuevo experimento mental, se nos pide que consideremos el caso de Mariana, quien, como María, es una neurocientífica brillante que posee toda la información física completa en torno a la percepción humana del color y que también creció en una habitación donde sólo pudo percibir los colores blanco y negro. Sin embargo, a diferencia de María, Mariana acepta participar en un experimento psicológico (con una predisposición sorprendente para colaborar con la filosofía y la ciencia a pesar de haber tenido una vida realmente espantosa) que requiere que, por un tiempo, sin abandonar la habitación en donde vivió desde siempre, observe distintas manchas de colores en una pared sin que se indique el nombre del color de cada mancha. Acto seguido, se le solicita que indique cuál de esas manchas coincide con el color del cielo diurno y despejado. En principio, parece que Mariana no puede basarse en su vasto conocimiento en torno a la percepción del color para responder esta pregunta, como sí podría hacerlo si se le hubieran presentado objetos familiares como tomates maduros, fotografías de paisajes o cuadros realistas, a partir de los cuales podría inferir el color del cielo porque conoce de antemano los colores que habitualmente poseen dichos objetos.

En este punto, Mariana elige la mancha roja impresionada por su belleza y a sabiendas de cuán bello se supone que es el color del cielo. De este modo, si bien Mariana (quien se supone que posee una visión “normal” y que sabe que rojo y azul son colores diferentes que no pueden ocupar simultáneamente el mismo espacio), cree (desde antes de ser liberada) que (a) el cielo aparece azul a las personas con visión normal, ahora cree, además, que (b) el cielo aparece rojo a las personas con visión normal (es decir, del color de *esa* mancha en la pared). De acuerdo con Nida-Rümelin, no hay verdadera contradicción entre (a) y (b) porque la última es una creencia *fenoménica* acerca de la experiencia visual del color mientras que la segunda no lo es y, por este motivo, el contenido de ambas varía. Con todo, la

creencia fenoménica de Mariana es falsa, y cuando es definitivamente liberada y ve por sí misma el color del cielo, se da cuenta de que no aplicó correctamente la palabra “azul” a la mancha en cuestión cuyo color supuso que se parecía al del cielo. Para Nida-Rümelin, “este nuevo conocimiento acerca del lenguaje es acompañado por la adquisición de un rico cuerpo de nuevo conocimiento fenoménico” (Nida-Rümelin 1998: 56). Y puesto que dicho conocimiento, no pudo ser derivado de la totalidad de la información física aun cuando Mariana ya contaba con los conceptos fenoménicos relevantes tras haber visto las manchas de la pared, se seguiría que el conocimiento que ganó después de ser liberada definitivamente es un conocimiento *proposicional* sobre una propiedad no física de la experiencia.

Por mi parte, no creo que esta conclusión tenga que ser aceptada obligatoriamente. En efecto, si se acepta que el tipo de saber del que carecen María o Mariana una vez liberadas se asemeja a un saber cómo o a una habilidad no conceptual que incluye, esencialmente, la capacidad de distinguir mediante la vista distintos colores (cuyo ejercicio supone además un aparato perceptual funcionando de manera “normal” o “típica”), es esperable que no pueda derivarse dicha capacidad o habilidad de ningún tipo de saber proposicional, sencillamente porque no es el tipo de “saber” que pueda ser derivado lógicamente de otro saber ni que pueda ser obtenido meramente a partir de información teórica. De este modo, ni antes ni después de ser liberada, primero parcial y después completamente, Mariana será capaz de derivar esta capacidad de la información física completa.

Como ya se dijo, esto no constituye un problema para el fisicalismo porque, por un lado, si así fuera, debería admitirse absurdamente que toda habilidad o saber de estas características constituye una amenaza en el mismo sentido. Por otro lado, porque el hecho de que estas capacidades no pueden ser en sentido estricto derivadas de alguna porción de información teórica proposicional, no es un impedimento para que pueda ofrecerse una explicación del funcionamiento de las mismas en términos compatibles con el fisicalismo.

Ahora bien, por otra parte, si bien es cierto que existe un saber del que carece Mariana después de percibir las manchas de color sin que se le indique a qué colores corresponden cada una de ellas es un saber proposicional, no es el tipo de saber proposicional que conformaría una amenaza para el fisicalista. De hecho, si el experimento mental en cuestión no violara subrepticamente sus propias premisas,

se debería admitir que Mariana podría determinar, en virtud de su conocimiento científico completo, a qué color corresponde cada mancha antes de que le dijeran su nombre o de compararla con el color del cielo. En efecto, si Mariana se ve llevada a creer erróneamente que el cielo luce como esa mancha roja en la pared, es porque carece de la información física relevante que cuya presencia impediría que cometa esa equivocación. “En particular, ella no conoce ni la composición química ni la reflectancia de las manchas coloreadas. Si ella supiera estos hechos, podría inferir [de qué color es cada mancha].” (Ball 2009: 942). Para decirlo de otra manera: incluso María, desde su encierro en blanco y negro, podría saber que aquel tipo de experiencia que Mariana reconoce (imagina y recuerda) cuando es expuesta a las manchas de color rojo no es del tipo del que se etiqueta con el nombre de azul porque sabría, entre otras cosas, gracias a su conocimiento físico ideal, qué muestras de color Mariana elegiría si se le pidiera que elija aquellas coinciden con el color del cielo y, a partir de una comparación de la longitud de onda de estas muestras con las del color del cielo, concluiría sin dificultades que Mariana está equivocada.

Ahora bien, puesto que María y Mariana poseen el mismo grado y tipo de saber antes de ser liberadas, se sigue que Mariana también debería poder determinar de qué color es la mancha que es capaz de reconocer (imaginar y recordar) cuando es expuesta por primera vez a ciertas experiencias cromáticas (si efectivamente contara con toda la información física pertinente).

4. Conclusiones

Como se advirtió al comienzo de este capítulo, la infabilidad, junto con el carácter privado del acceso a los estados conscientes, son las dos propiedades frecuentemente atribuidas a la experiencia consciente que obstaculizan una aproximación científica a la misma, porque si la experiencia consciente es privada, entonces la perspectiva científica no puede alcanzarla, y si es infable, entonces incluso si la ciencia lograra alcanzarla no podría describirla en una teoría (Ross 1993: 443).

Por mi parte, me propuse mostrar que hay un sentido sensato en que puede atribuírsele esta propiedad a la experiencia pero que no conduce a estos resultados escépticos. Para encaminarme a esa conclusión, comencé haciendo notar (de la mano de las apreciaciones de Bennett y Hacker 2003 en la sección 2.1) que, de

hecho, describimos habitualmente, tanto nuestra experiencia como sus objetos, sin mayores dificultades. Por otra parte, si bien podría admitirse, como señala Dennett, que la experiencia es inefable en la *práctica* debido al alto grado de información que puede llegar a portar, o que es temporalmente inefable, o que puede parecer inefable por la dificultad psicológica que implicaría “ponerse (imaginativamente) en los zapatos” de personas muy alejadas de nuestra horizonte cultural y temporal, nada de esto convierte a la experiencia en un fenómeno desconcertante ubicado más allá de todo conocimiento científico. Por otra parte, contemplé también la posibilidad de que la tentación de considerar que la experiencia es inefable en algún sentido problemático se hallara en la adhesión, más o menos explícita, a la tesis del lenguaje privado. En efecto, un lenguaje de estas características explicaría por qué quien no tuvo las experiencias relevantes no comprendería ni los términos ni los conceptos acerca de las experiencias conscientes por mucho que otras personas se esfuercen intentando explicárselo mediante múltiples y detalladas descripciones: quien no tuvo las experiencias en cuestión, no puede tener acceso a las entidades privadas a las que dichos conceptos y términos refieren y que agotan su contenido y significado.

Sin embargo, retomé algunas de las consideraciones realizadas en el capítulo anterior para desestimar la necesidad de postular un lenguaje de estas características. Además, hice notar algunas diferencias cruciales entre el modo en que podrían determinarse, a través de una suerte de ostensión, los significados o contenidos de los términos y conceptos en un lenguaje privado y en el lenguaje público, con el objetivo de poner en cuestión precisamente la idea de que la postulación una ostensión privada no resulte problemática o que vaya de suyo.

A la luz de estas observaciones, cobraba fuerza mi propuesta de que el grano de verdad detrás de la tesis de la inefabilidad consiste simplemente en que para poseer una maestría completa de ciertos conceptos acerca de sensaciones o estados perceptivos, es necesario haber atravesado las experiencias relevantes y contar con las capacidades perceptivas funcionando de manera “normal” o típica. Así, por ejemplo, para poseer una maestría en el concepto de EXPERIENCIA DE VER ROJO, es necesario tener una maestría en el concepto de ROJO, y para cumplir con esta condición, se debe ser capaz de discernir mediante la vista el color rojo de otros colores. Desde luego, ninguna descripción verbal de estas experiencias podrá dotar

a alguien de esta capacidad de discriminación. Con todo, subrayé que no debe confundirse el hecho de que sólo quien puede ver colores será capaz de tener una maestría más o menos completa de los conceptos de color (y, también, del de las experiencias de color), con la idea más controvertida de que los conceptos y términos de experiencias cromáticas están definidos por referencia a entidades privadas e incommunicables.

Por otro lado, a partir de la sección 3, me propuse enfrentar el argumento de Jackson para reforzar mi propuesta. De acuerdo con este argumento, el caso de María muestra que es posible contar con *toda* la información física ideal acerca de la percepción consciente del color y sin embargo ignorar todavía ciertas verdades en torno a dichas experiencias. En principio, de esto se seguiría que existen verdades acerca de la percepción consciente del color que refieren a hechos que están más allá de los hechos físicos. Sin embargo, busqué mostrar que si se aceptan las premisas del llamado “externismo social” aplicado a los conceptos fenoménicos debe aceptarse que es posible poseer dichos conceptos en virtud de mantener distintas interacciones adecuadas con una comunidad lingüística.

De estas consideraciones se concluyó, en última instancia, que si María carece de algún tipo de conocimiento antes de ser liberada (es decir, si hay algo que no pueda derivar de toda la información física que posee), este saber no es de tipo proposicional o conceptual. Para alcanzar esta última conclusión por otra vía, y reforzar de este modo mi punto, utilicé un aspecto de la réplica de Dennett 1995 al argumento de Jackson, según el cual, María debería poder saber qué pensamientos tendría luego de ser liberada (dado que se supone que los pensamientos no son físicamente irreducibles) y, a partir de ahí, podría saber qué se siente tener experiencias conscientes. Otra vez, si se acepta esta propuesta, debe admitirse que, si María carece de algún tipo de conocimiento antes de la liberación, este conocimiento no puede ser de tipo proposicional o conceptual. Posteriormente, desestimé una reformulación del argumento del conocimiento a cargo de Nida-Rümelin 1998 que podría leerse como crítica a mi propuesta en torno al carácter no proposicional o no conceptual del conocimiento que le falta a María. En particular, mostré que el argumento de Nida-Rümenin sólo puede llegar a esa conclusión si viola sutilmente sus propias premisas.

Finalmente, consideré que si se acepta que la experiencia consciente es inefable en el único sentido en que parece sensato atribuirle esta propiedad, es posible explicar por qué no se puede derivar cierto conocimiento no proposicional o no conceptual acerca de los estados conscientes sin comprometerse con algún tipo de dualismo ontológico tal como pretende el argumento de Jackson.

CAPÍTULO 5

YO, CEREBRO

1. Introducción

En muchos de los capítulos anteriores se hizo notar, de manera incidental en diversos pasajes, que las discusiones que se estaban llevando adelante en ellos resultaban relevantes no sólo para la filosofía de la mente sino también para las ciencias cognitivas, ya que, en muchos casos, ambas comparten varios de los supuestos disputables señalados y atribuyen a la consciencia las propiedades controvertidas que fueron objeto de análisis pormenorizado en este trabajo (me refiero, obviamente, a la intrinsecidad, a la infabilidad y al presunto carácter privado y directo del conocimiento de nuestros propios estados mentales). En particular, muchos neurocientíficos, como veremos enseguida, adoptan acríticamente buena parte de la terminología filosófica para caracterizar a la consciencia como, por ejemplo, la noción de *qualia* y la expresión “cómo es estar” en determinado estado mental.

Así las cosas, en este capítulo trataré de exorcizar otra de las dificultades que ya habíamos visto aparecer con anterioridad pero desde otra perspectiva. En particular, voy a ocuparme de un fantasma que recorre las neurociencias contemporáneas: el fantasma de Descartes. Por lo menos, esto es lo que denuncian autores como Alva Noë (2010, 2015), Anthony Kenny (2009, 2010), Peter Hacker (Bennett y Hacker 2003), Harry Smit (2014, 2017) y (en cierto modo) Kevin O'Regan (2011) entre otros.

En efecto, de acuerdo con Noë (2010, 2015) el “postulado fundacional” de gran parte del trabajo en neurociencias consiste en partir de la idea de que, en definitiva, es *algo* dentro de nosotros lo que piensa, siente, quiere y decide: para Descartes, ese “algo” era la sustancia pensante o *res cogitans*, una entidad inmaterial cuya conexión con lo físico resultaba problemática. Así, la famosa frase de Descartes “Cogito, ergo sum” parece certificar la idea de que nuestra naturaleza fundamental es la consciencia o el pensamiento (en el sentido amplio en que el filósofo francés

entendía este término), y todo lo demás, nuestra situación, nuestro cuerpo, el lugar que ocupan los demás en nuestra vida, es puro accidente.

Ahora bien, de acuerdo con los autores mencionados, la única diferencia esencial entre esta postura y aquella que subyace en gran parte de los trabajos de la neurociencia actual es que ahora “es el cerebro, antes que algo no físico, lo que siente, piensa, decide” (Noë 2010: 22). Así las cosas, los problemas consabidos que surgen a la hora de dar cuenta de la interacción entre dos sustancias con propiedades completamente heterogéneas desaparece. Sin embargo, otros inconvenientes permanecen más o menos incommovibles.

En efecto, dado que desde esta perspectiva cada uno de nosotros no es otra cosa que un cerebro en una cubeta de carne y huesos: no sabemos nada acerca de lo que nos rodea excepto por lo que se proyecta en unas pantallas internas. En este modelo, el mundo, sus objetos y las demás personas son estímulos incognoscibles que afectan a nuestro sistema nervioso (Noë 2015: 120). Lo que realmente percibimos de manera directa son representaciones o imágenes, creadas por nuestro cerebro, acerca de ese mundo conjeturado. Ahora bien, si esto es así, reaparecen algunos de los espectros que creímos haber dejado atrás en los capítulos anteriores.

Por un lado, si aceptamos esta caracterización de la situación, deberíamos aceptar también que podríamos ser víctimas de una suerte de ilusión masiva a gran escala pergeñada por nuestro cerebro, porque cuando vemos, tocamos, oímos, creemos, erróneamente, estar en contacto directo con las cosas mismas cuando, al parecer, no es así. Además, si ver un objeto, por ejemplo, consiste en ver o construir una imagen o una representación de ese objeto, entonces parece que no es un error, después de todo, considerar que cada sujeto, cuando percibe, tiene un acceso privilegiado y privado a su propia experiencia. Al parecer, otra vez estaríamos frente al problema de los objetos internos.

En segundo lugar, veremos que la adscripción de predicados psicológicos por parte de los neurocientíficos a ciertos procesos cerebrales de manera más o menos literal, puede conducir, de acuerdo con Bennett y Hacker 2003, a caer en ciertas incoherencias conceptuales (en particular, a lo que los autores denominan, la “falacia mereológica”) y, sobre todo, a tornar misterioso el “surgimiento” de la experiencias perceptiva consciente.

Por mi parte, en este capítulo voy a explicitar cuáles son algunas de las consecuencias nocivas que se siguen de adoptar una perspectiva cartesiana como la mencionada. Luego, voy a mostrar cuáles serían las motivaciones centrales que podrían estar detrás de la adopción de dicha perspectiva. Frente a ellas, sostendré, apoyándome en el trabajo de algunos de los partidarios de lo que se conoce como el “enfoque sensoriomotor” de la percepción y del filósofo Peter Hacker, que la postulación de dichas entidades internas, además de comprometernos con una serie de consecuencias perniciosas, resulta innecesaria e insuficiente para dar cuenta de la naturaleza de nuestras experiencias perceptivas.

2. Los problemas de adoptar la tesis del cerebro creador

De acuerdo con Noë 2010 y muchos otros autores⁴⁴, el “postulado fundacional” de la mayor parte de la neurociencia contemporánea sostiene que lo que realmente percibimos de manera directa son representaciones o imágenes, creadas por nuestro cerebro, acerca de ese mundo conjeturado. Así, por ejemplo, el científico galardonado con el premio Nobel, Francis Crick, afirma que:

⁴⁴ En efecto, como dije, Noë no es el único que considera que gran parte de la neurociencia contemporánea adopta la tesis del “cerebro creador”. Así, por ejemplo, el filósofo Anthony Kenny (2010) declara que “en sí mismo no hay nada equivocado es hablar de imágenes en el cerebro, si uno quiere decir que hay patrones en el cerebro, observables para el neurofisiólogo, que pueden ser mapeados con rasgos del entorno visible. Lo que resulta equivocado es tomar estos mapeos como representaciones, considerarlos como visibles a la mente, y decir que ver consiste en la percepción [...] de estas imágenes” (Kenny 2010: 111). Al parecer, esto último es lo que harían muchos científicos especializados, al menos de manera implícita, ya que si bien “no todos los científicos cognitivos llegan tan lejos como para identificar explícitamente la mente con el cerebro, pero todos buscan un paralelismo entre lo mental y lo físico” (Kenny 2010: 112). Por su parte, otros autores contemporáneos que se adscriben dentro de la corriente fenomenológica sostienen que: “En términos generales, muchos científicos de la visión todavía conciben ésta como un proceso en el que, sobre la base de lo que está presente de más cerca, es decir, de imágenes en la cabeza o en el ojo, pretendemos descubrir (por medio de inferencias) lo que está presente ahí afuera, en el mundo”. (Gallagher y Zahavi 2014: 147). De hecho, los mismos protagonistas parecen confirmar esta interpretación acerca de su trabajo. En efecto, esta concepción se verifica en el planteo del clásico problema de las neurociencias llamado “problema de la unión” (*binding problem*). En este caso, el problema consiste en explicar cómo es que dado que diversas partes o células del cerebro en diferentes ubicaciones responden, respectivamente, al color, la forma, la ubicación y el movimiento de los objetos exteriores, logra el cerebro unir todos esos “datos dispersos” para posibilitar las experiencias perceptivas unificadas como las que de hecho tenemos. Según el neurocientífico Francis Crick (1995), “seguimos sin saber cómo el cerebro une todo” (Crick 1995: 22). Por su parte, otro científico de renombre como Antonio Damasio adopta una concepción similar al sostener que: “cuando tú y yo miramos un objeto fuera de nosotros, formamos imágenes comparables en nuestros cerebros respectivos. Lo sabemos porque podemos describir el objeto de formas muy similares, hasta los matices. Pero esto no significa que la imagen que vemos sea la copia de cómo es el objeto que sea que está ahí afuera. En términos absolutos, no sabemos a qué se parece. (Damasio 1999: 320).

Lo que ves no es lo está ahí realmente; es lo que tu cerebro cree que está ahí...Tu cerebro hace la mejor interpretación que puede de acuerdo con su experiencia previa la información limitada y ambigua provista por tus ojos...lo que el cerebro hace es preparar [...] interpretación de la escena visual (Crick, 1995: 30).

Por su parte, otro referente de la neurociencia cognitiva, Christopher Frith (2007), sostiene más recientemente que “nuestro cerebro crea la ilusión de que tenemos contacto directo con objetos en el mundo físico” (Frith, 2007: 17).

Ahora bien, si esto es así, deberíamos aceptar, como dije al inicio, que podríamos ser víctimas de una suerte de ilusión masiva a gran escala pergeñada por nuestro cerebro, porque cuando vemos, tocamos, oímos, creemos, erróneamente, estar en contacto directo con las cosas mismas cuando, al parecer, no es así. Al mismo tiempo, recaeríamos en un escepticismo también con respecto a las otras mentes, a las que, por hipótesis, nunca tendríamos acceso de “primera mano”. Por otro lado, si esta perspectiva fuera cierta, jamás podríamos confirmarla o refutarla porque, por hipótesis, nunca podremos correlacionar nuestras representaciones o imágenes internas con aquello que representan puesto que sólo tenemos acceso a las primeras.

Por otra parte, dentro de este marco cartesiano, es habitual, de acuerdo con Bennett y Hacker 2003, (y también Smit y Hacker 2013), que en las explicaciones de la neurociencia cognitiva contemporánea, se adscriban atributos psicológicos al cerebro y a sus partes para dar cuenta de la posesión y ejercicio (y de las deficiencias en el ejercicio) de los poderes cognitivos de los seres humanos u organismos (Bennett y Hacker 2003: 3). A juicio de los autores, este procedimiento comete lo que ellos denominan “falacia mereológica”, es decir, el error de atribuir propiedades psicológicas en un sentido más o menos literal a distintas partes de un animal que sólo pueden ser atribuidas inteligiblemente al animal como un todo (Bennett & Hacker 2003: 73; Smit y Hacker 2013: 1).

En efecto, mientras que en principio contamos con criterios más o menos claros para atribuir propiedades psicológicas a los seres humanos y animales que nos permiten determinar cuándo éstos tienen experiencias perceptivas, creen o saben cosas y toman decisiones, no resulta obvio qué querría decir, o cómo podríamos determinar, que el cerebro ve u oye, toma decisiones, presenta

argumentos o estima posibilidades. En última instancia, la idea es que esta práctica supone cometer la misma equivocación que se cometería si se dijera que porque un auto es rápido, entonces sus carburadores son rápidos, o que porque dado que es necesario, para que un avión vuele, que sus motores estén funcionando correctamente, que son los motores y no sólo el avión, los que vuelan.⁴⁵

Parafraseando a Wittgenstein, Bennett y Hacker adhieren a la idea de que “sólo de seres humanos vivos y de lo que se les asemeja (se comporta de modo semejante) podemos decir que tienen sensaciones, ven, están ciegos, oyen, están sordos son conscientes o inconscientes” (Wittgenstein 2010: 237). En este sentido, es importante aclarar que gran parte del esfuerzo argumentativo de Bennett y Hacker 2003 tiene como objetivo mostrar que en las neurociencias la atribución de propiedades psicológicas a ciertas partes o procesos del cerebro es una atribución *literal*, donde no se hace un uso “técnico”, metafórico, analógico o metonímico de términos homónimos. Por el contrario, los problemas surgen, a juicio de estos autores, precisamente porque quienes utilizan estos predicados psicológicos, suelen mezclar un significado “técnico” de los mismos, con el significado que dichos predicados tienen en el lenguaje ordinario. De hecho, no implicaría ninguna incoherencia conceptual usar homónimos de los predicados psicológicos, aun cuando pueda resultar confuso, siempre y cuando “*los científicos que usan estos*

⁴⁵ Alva Noë realiza una denuncia similar con respecto a los supuestos que subyacen en los trabajos clásicos de las ciencias cognitivas estándar y de las neurociencias. Como se dirá más adelante, de acuerdo con la teoría clásica de la visión, el sistema visual (ubicado en el cerebro) desarrolla un trabajo de procesamiento de información según el cual, extrae información sobre el entorno a partir de la imagen retiniana y, de ese modo, construye una representación interna de ese entorno. Sin embargo, Noë apunta que hay razones obvias para dudar de la idea de que el cerebro es un procesador de información. Y algunas de estas razones se vinculan con la falacia mereológica. Así, de acuerdo con Noë, si bien tenemos una idea clara de qué querría decir que un detective extrae información acerca de un ladrón a partir de una huella en la escena del robo, no es tan claro qué es lo que querría decir que el cerebro extrae información del entorno a partir de la imagen retiniana. Por empezar, “el cerebro no es científico, ni detective; no sabe nada y carece de ojos para examinar la imagen retiniana. No tiene capacidad de inferir nada y, mucho menos, de inferir remotas causas ambientales [i.e. los objetos del mundo y sus rasgos] del estado de observación de la retina. Entonces ¿cómo vamos a creer que el cerebro sea un artefacto para procesar información?” (Noë 2010: 198). Y más adelante afirma: “Nosotros, los humanos adultos y otros animales [es decir, los organismos como un todo], pensamos; vemos, sentimos, juzgamos, inferimos.” (op.cit.). Por otra parte, Noë considera que el hecho de que existan computadoras capaces de procesar información, realizar cálculos y reproducir imágenes tridimensionales a partir de dibujos lineales, no indica que las computadoras “piensen” o sigan reglas o “recetas” ciegamente. De acuerdo con Noë, “de la misma manera en que un reloj pulsera no sabe la hora, a pesar de que lo usamos para saberla, el ordenador tampoco entiende las operaciones que hacemos con él. Nosotros pensamos con los ordenadores, pero los ordenadores [...] son herramientas” (Noë 2010: 201). Así, para Noë, si queremos comprender la contribución que el cerebro hace a la cognición, se debería desechar la idea de que la mente se dé en nuestro interior gracias a unos mecanismos internos.

términos así no olviden que los términos no tienen su uso habitual. Desafortunadamente, es típico que lo olviden y *proceden a cruzar el nuevo uso con el viejo*, generando incoherencia” (Bennett y Hacker 2003:76. La cursiva es de los autores).

En principio, este tipo de falencias se evidenciarían en las obras de lo más destacados representantes de la ciencia cognitiva estándar actual.⁴⁶ Así, de acuerdo con Bennett y Hacker, lejos de ser nuevos homónimos, las expresiones psicológicas que usan están siendo invocadas en el sentido habitual, de otro modo los neurocientíficos no podrían sacar las inferencias de ellas que sacan. Cuando el neurocientífico británico, Francis Crick afirma que “lo que tú ves no es lo está *realmente* ahí; es lo que tu cerebro *cree* que está ahí” (Crick, 1995), es importante que él considere que “creer” tiene sus connotaciones habituales porque es parte de la historia de Crick que la creencia es el resultado de *una interpretación* basada en la *experiencia e información previa* (y no el resultado de una interpretación* basada en la *experiencia* e información* previa* [es decir, no como procesos entendidos en un sentido meramente “técnico” o no habitual]). Cuando el neurobiólogo británico Semir Zeki observa que la adquisición de conocimiento es una “función primordial del cerebro” (Zeki, 2000: 547), quiere decir conocimiento, y no conocimiento* -de otro modo no podría pensar que la tarea de la neurociencia futura es resolver los problemas de la epistemología (sino sólo, presumiblemente, los de la epistemología*) (Bennett y Hacker, 2003: 75).

Por otra parte, con respecto a la propuesta del neurocientífico David Marr, Bennett y Hacker sostienen que es evidente, por lo que él dice acerca de las representaciones y las inferencias que realiza a partir de la postulación de representaciones *en* el cerebro, que lo que entiende por ‘representación’ no es simplemente el hecho de que existe una correlación causal entre ciertos eventos del cerebro y ciertos rasgos o propiedades del entorno que permitirían, de algún modo, dar cuenta de la visión. En principio, esto constituiría una manera “técnica” y legítima de usar la noción de “representación” dentro de dicho contexto. Sin embargo, de acuerdo con Bennett y Hacker, Marr confunde el sentido semántico de

⁴⁶ Para una caracterización breve pero clara de los rasgos principales de lo que denominé “la ciencia cognitiva estándar”, así como también de sus limitaciones, junto con un panorama y evaluación de perspectivas alternativas corporeizadas, puede consultarse Shapiro (2011).

la noción de representación, propio de las descripciones (que es uno de los sentidos con los que utilizamos el término en el lenguaje ordinario) con el sentido causal. Así, Marr compara las representaciones del cerebro con las “representaciones que usamos todo el tiempo”, pero las representaciones que menciona como ejemplos son notaciones simbólicas (como las notaciones romanas, arábicas, y binarias en aritmética), regidas por reglas, es decir, son representaciones en un sentido semántico y no causal (Bennett y Hacker 2003: 76, 145).⁴⁷

En otras palabras, el problema que señalan Bennett y Hacker 2003 no constituye una crítica al intento de “naturalizar” las propiedades semánticas o representativas de los estados mentales (aunque seguramente tampoco considerarían viable ese proyecto)⁴⁸; en todo caso, su crítica pretende hacer notar que el uso que se hace de los términos “psicológicos” en el contexto de las explicaciones científicas es ambiguo, porque se realizan inferencias a partir del uso “natural” o causal de la noción de representación que sólo podrían realizarse a partir del uso semántico y/o ordinario del término. De este modo, aunque Bennett y Hacker no saquen esta consecuencia de su propio planteo, se sigue que una explicación formulada en estos términos corre el riesgo de presuponer aquello que debe explicar, en la medida en que estaría atribuyendo predicados psicológicos al cerebro para explicar la atribución de predicados psicológicos al organismo como un todo.

Sin embargo, aunque esta crítica advierte sobre el peligro de confundir los significados de términos que deberían ser entendidos meramente como homónimos, no resulta forzoso aceptar que esta confusión sea inevitable para el neurocientífico. En todo caso, no se ve por qué, en principio, no podría evitar cometer estas inconsistencias.

⁴⁷ Por otro lado podría decirse que la idea de atribuir propiedades representacionales al cerebro para dar cuenta de las propiedades representacionales de la cognición del animal como un todo, corre el riesgo de dar por sentado aquello que se supone que debe explicar (es decir, el carácter representacional).

⁴⁸ En líneas generales, podríamos caracterizar a este proyecto como un intento de explicar o reducir ciertas propiedades de los estados mentales (que, en principio, las ciencias naturales no parecen poder admitir en sus ontologías), en término de propiedades no-mentales (ya sean físicas, biológicas, funcionales o alguna combinación de estas). Para una evaluación de este proyecto así como también la formulación de una propuesta alternativa ver Skidelsky (2015).

De todos modos, existen *otras* dificultades de peso relacionadas con estas cuestiones que pueden contribuir a aumentar el misterio que rodea la experiencia consciente aun cuando se deje de lado lo dicho hasta acá. En efecto, de acuerdo con la teoría de Marr, por ejemplo, se supone que el resultado de los procesos de computación que realiza el cerebro para dar lugar, de algún modo, a la visión, es la producción de una *descripción*.⁴⁹ Sin embargo, lo que no se explica es cómo estas descripciones permiten *ver* al organismo que las posee, es decir, cómo se da la transición entre la formación de una descripción en el cerebro a la experiencia de ver lo que está delante de los ojos, dado que es posible argüir, como hacen Bennett y Hacker (2003), que *ver* consiste en el ejercicio de un poder o de una capacidad, en usar la facultad visual, y no en acceder a una representación o descripción aun cuando se la entienda en un sentido análogo al semántico.

En todo caso, aun cuando se aceptara que la percepción y la descripción que surge en el cerebro comparten el mismo contenido, el modo en que pueden llegar a representarlo parece ser tan diferente como el modo en que una fotografía y una descripción verbal pueden representar una bandada de aves cruzando el cielo. Pero entonces, el error de partida tal vez consista en pretender que una representación cerebral realice una función de estas características.

En otras ocasiones, el “misterio” de la consciencia se alimenta porque se considera que, para que ésta surja, el cerebro debe crear una *imagen* del mundo (en un sentido más menos literal del término). Así, por ejemplo, Damasio sostiene que “hay un misterio, sin embargo, con respecto a *cómo* las imágenes emergen de patrones neurales. Cómo un patrón neural *se convierte* en una imagen es un

⁴⁹ En líneas generales, podríamos decir que gran parte de la ciencia cognitiva estándar está comprometida con la idea de que la cognición en general (es decir, la percepción, la memoria, el lenguaje, la resolución de problemas, el aprendizaje, etc.) involucra procesos algorítmicos (alojados en el cerebro) sobre representaciones, que funcionan de un modo análogo a los procesos de procesamiento de información que llevan adelante las computadoras. Desde esta perspectiva, es posible que las neuronas o sus actividades tengan propiedades simbólicas, es decir, que puedan de algún modo representar cosas (como objetos o aspectos del entorno), que a su vez puedan combinarse de acuerdo con ciertas reglas. En el caso particular de Marr, su propuesta central consiste en sostener que la visión es esencialmente un proceso de procesamiento de información, cuyo objetivo es descubrir cómo son las cosas del entorno (qué hay en él y dónde) a partir de unos patrones de luz en la retina. Así, el cerebro extraería información sobre el entorno a partir de la imagen retiniana y, de este modo, construiría una representación interna de ese entorno. De esta manera, el cerebro, a través de una serie de operaciones computacionales sobre un sistema de símbolos que representan rasgos del entorno, produce, en el estadio final del proceso visual, una descripción de las formas, de los objetos del entorno, su distancia, orientación e identidad.

problema que la neurobiología todavía no ha resuelto” (Damasio 1999: 322). Así, se suscita la pregunta acerca de cómo tales eventos neurales pueden producir algo tan categorialmente distinto de la excitación nerviosa como una imagen. Como veremos en la sección siguiente, la necesidad de postular una imagen de estas características puede ser puesta en duda.⁵⁰

Con todo, Bennett, Hacker y Smit, advierten que estas observaciones no deberían llevarnos a pensar que el estudio y comprensión del funcionamiento del cerebro no sea relevante para la comprensión de los fenómenos mentales. Por el contrario:

Dado que el normal funcionamiento de las estructuras y procesos del cerebro son condiciones causales de los fenómenos mentales, es interesante estudiarlas. Por ejemplo, las alteraciones al nivel de las células y los circuitos neurales son pre requisitos para la retención de información, [...] Pero si bien estas alteraciones son pre requisitos para la retención de la información, esto no muestra que las células, los circuitos neurales, o la estructura del cerebro retenga información. [...] Es el animal como un todo que retiene información y ejercita esta habilidad de retener información, no sus partes (Smit y Hacker 2013: 2).

De todos modos, no queda del todo claro, cuál sería el rol exacto que debería cumplir la neurociencia y de qué modo podría contribuir al esclarecimiento de la realidad mental. En particular, no se vislumbra cuál es el aporte que el cerebro o, para el caso, el resto de las partes del organismo, que realizan en el desarrollo y ejercicio de sus capacidades cognitivas y afectivas. Sin embargo, esta clarificación se hace necesaria si se quiere evitar reinsertar otra forma de dualismo ahora entre las “partes” del organismo o animal y las capacidades cognitivas que posee como un todo.

Por lo demás, aun si se aceptara todo lo dicho hasta ahora, no se sigue que no resulte provechoso en algún sentido recurrir a la postulación de representaciones

⁵⁰ Por lo demás, algunos autores como Kevin O’Regan (2011, 2016) y Alva Noë (2004), que defienden lo que en ocasiones se etiqueta con el nombre de “enfoque sensoriomotor” de la percepción, consideran que el primer paso en falso que se da con respecto a la consciencia es cometer el error categorial de creer que es algo que puede ser “generado” por algo o que puede “emerger” de algo (el cerebro, en este caso). Por el contrario, la idea general de este enfoque, de acuerdo con una de sus formulaciones más recientes (O’Regan 2016), es que la percepción está constituida por una colección de capacidades o destrezas (fundamentalmente corporales) para actuar e interactuar con el entorno, junto con un conocimiento práctico, por parte del agente, de ciertas regularidades que rigen dichas interacciones con el entorno.

internas o cálculos computacionales para dar cuenta de la cognición. En todo caso, lo que parece necesario, a partir de estas observaciones, es replantear o esclarecer el tipo de relación que se establecería entre el cerebro y las capacidades cognitivas de los sujetos que los poseen.

3. Algunas de las motivaciones detrás del fantasma cartesiano

Luego de señalar algunas de las dificultades que arrastra la concepción cartesiana de la mente, quisiera indagar qué motivaciones podrían llevar a adoptarla. Después, intentaré mostrar que el compromiso con una perspectiva cartesiana es innecesaria e insuficiente para dar cuenta de aquello que pretende.

Como habíamos visto, existe una asimetría entre las atribuciones mentalistas en primera persona en tiempo presente y las atribuciones correspondientes en tercera persona, según la cual, mientras que las primeras no se basan en la observación previa de la conducta, las segundas sí. Además, las atribuciones en primera persona se consideran infalibles, o al menos mucho más seguras que el resto de las afirmaciones empíricas, mientras que las atribuciones en tercera persona no.

Desde la posición favorecida en este trabajo, fomenté la idea de que la perspectiva que denominé “cartesiana” malinterpreta esta asimetría en términos de una asimetría epistémica, de acuerdo con la cual, el sujeto tiene un acceso epistémico directo y privilegiado a sus propios estados mentales y un acceso siempre indirecto e inseguro a los estados mentales de las demás personas. Ahora bien, de acuerdo con Bennett y Hacker 2003, parte de las razones por las cuales adoptan la tesis del cerebro creador tienen que ver precisamente con que interpretan la asimetría en cuestión como una asimetría epistémica. La única diferencia es que conciben el presunto acceso privilegiado ya no como un acceso a las operaciones mentales de una sustancia inmaterial sino como un acceso parcial y limitado a ciertos procesos que ocurren en el cerebro (Bennett y Hacker 2003: 85).⁵¹

⁵¹ Algunas declaraciones de científicos reconocidos parecen abonar esta hipótesis de lectura: en efecto, de acuerdo con Francis Crick, “hay un acuerdo general de que las personas no son conscientes de todos los procesos que suceden en sus cabezas...Mientras que puedes percartarte de muchos de los resultados de los procesos perceptuales y de la memoria, sólo tienes un acceso limitado al proceso que produce tu percartación” (Crick 1995: p. 19f.)

Vinculado con esta cuestión, podría decirse que algunas de las dificultades con las que tropiezan los científicos a la hora de dar cuenta de la naturaleza de la consciencia y de su relación con el mundo físico, se deben a que adoptan acríticamente buena parte de la terminología controvertida propia de los filósofos, como la noción de *qualia*⁵², o la de “como es estar en determinado estado mental para un organismo”⁵³. Así, por ejemplo, Antonio Damasio afirma que “los qualia son las cualidades sensoriales simples encontradas en la azulidad del cielo o en el tono del sonido producido por un chelo [...]” los científicos Gerald Edelman y Giulio Tononi sostienen que “cada experiencia consciente diferenciable representa un quale diferente”, y agregan que el “problema” de los qualia es “tal vez el problema más desalentador de la consciencia” (Edelman y Tononi, 2000: 157). De este modo, si se piensa a la experiencia consciente a través de estos conceptos, es decir, como algo esencialmente privado y accesible sólo al sujeto, inefable, etc., la relación con el mundo físico, objetivo y público se convierte, por todo lo dicho en los capítulos anteriores, en una cuestión verdaderamente inabordable para la ciencia y para el pensamiento, porque lo que se debería explicar es la vinculación entre dos ámbitos que lucen completamente heterogéneos a primera vista, y entonces la pregunta acerca de cómo algo tan diferente a la “materia” podría “emerger” de ella, nos deja completamente perplejos. Ahora bien, si tuve y tengo éxito en este trabajo, se sigue que deberíamos abandonar estas caracterizaciones problemáticas de la consciencia ya que tenemos buenas razones para hacerlo. Y si esto es así, al menos algunos de los problemas que entraña la pregunta por la vinculación entre el mundo físico y el mundo mental, se disuelven.

Por otra parte, existen *otras* motivaciones que están a la base de la aceptación de la tesis del cerebro creador. De acuerdo con Noë 2010, esta tesis carece en realidad de sustento empírico y, en rigor, cuando se presiona a los científicos para que ofrezcan algún fundamento convincente en favor de dicha doctrina, suelen apelar a argumentos filosóficos tradicionales, como el de los sueños. Sin embargo, Noë considera que es posible responder a estos argumentos. Veamos.

⁵² Como apunté en la Introducción de este trabajo, dadas las notas que comprende, la noción de *qualia* constituye un modo problemático de caracterizar la experiencia consciente para cualquier enfoque fisicalista y funcionalista de la mente, y también para enfoques no funcionalistas pero naturalistas y no dualistas como el enfoque sensoriomotor de la percepción (para ver por qué esto último es así ver, entre otros, O’Regan y Noë (2001: pp. 959-963).

⁵³ De esta expresión en particular me ocuparé en el capítulo próximo.

Es cierto que cada uno de nosotros ha tenido la experiencia de despertar y de encontrar que lo que parecía tan vívido, inquietante y real, era sólo un sueño. Sin embargo, no puede decirse que en un sueño exploro activamente el mundo, ni que interactúo de forma dinámica con él. De esto modo, parece que habría que renunciar a la idea de que estos dos últimos requisitos sean necesarios para que haya experiencia perceptiva y aceptar que, en definitiva, es el cerebro el que hace todo el trabajo.

Con todo, Nöe presenta diversos argumentos para rechazar esta conclusión. Por un lado, sostiene que aun si se aceptara que soñar depende sólo de lo que pasa en el cerebro, “de esta premisa no se infiere que, en lo tocante a la consciencia, lo único determinante sea lo que ocurre en la cabeza. Lo que se infiere, como mucho, es que las experiencias oníricas dependen únicamente de lo que ocurre en la cabeza” (Nöe 2010: 216). Y esto es así, en primer lugar porque, según el autor, no “toda experiencia puede tener lugar en un sueño”.

En otras palabras, aun cuando se admitiera que los sueños son semejantes a algunas de las experiencias que tenemos en nuestra vigilia, de esto no se seguiría que *todas* nuestras experiencias son “creadas” del mismo modo por el cerebro. Sin embargo, un objetor podría responder que esta suposición resultaría más económica, desde el punto de teórico, que la postulación de dos mecanismos diferentes, aquellos que generan las experiencias oníricas y los que generan las experiencias que tenemos durante la vigilia.

De todos modos, aun aceptando que no podemos distinguir entre el sueño y la vigilia (porque, en línea con el viejo argumento cartesiano, cualquier intento de determinar si estamos o no despiertos -como pellizcarnos-podría también formar parte de un sueño), de esta comprobación no se sigue lógicamente que no haya una diferencia entre ambos. Lo único que se sigue, como mucho, es que *nosotros* no podemos distinguirlos. De la misma manera, “el hecho de que yo no pueda distinguir a simple vista si un edificio es de una época o de otra no implica que no exista una diferencia entre estilos según la época.” (Nöe 2010: 217).

Por su parte, Hacker utiliza esta misma estrategia ante una objeción análoga pero agrega que parte de lo que conduciría a confundir la experiencia perceptiva con una ilusión (o, para el caso, un sueño) es que ambas pueden tener el mismo *contenido*. Más específicamente, el “contenido” en cuestión sería aquello cuya

descripción permitiría responder a las preguntas “¿Qué te pareció percibir?” y “¿Qué percibiste?”. A su juicio, esto sería lo que lleva a creer que la única diferencia entre ambas es su origen causal (es decir, que lo único que distingue a una experiencia verídica de una que no lo es, es su causa; en sí mismas, las dos experiencias serían iguales). Sin embargo, si bien “es verdad que las respuestas a estas preguntas pueden muy bien ser las mismas”, el hecho de que lo sean “no implica que describan una experiencia perceptual común a la percepción y a lo que parece percibirse [por ejemplo, en los casos de las ilusiones]”. De ser así, “la pregunta ‘¿qué no percibiste cuando no percibiste que Jack subió a la montaña?’ tiene la misma respuesta que la pregunta ‘¿Qué percibiste cuando percibiste que Jack subió la montaña?’ a saber, ‘Jack subió la montaña’. Pero esta respuesta no describe algo en común a percibir y a no percibir algo -¡es una respuesta común a dos preguntas diferentes!” (Hacker 2013: 312).

Pero además, es dudoso que alguna experiencia onírica sea efectivamente semejante a las experiencias perceptivas. Y recordemos que de esta semejanza dependía fundamentalmente la argumentación que pretendía concluir que la percepción consiste en poseer una representación en el cerebro. En efecto, la experiencia perceptiva, a diferencia de la onírica, es estable. Por ejemplo, de acuerdo con el trabajo de Noë, un investigador en sueños, Stephen Laberge, sostiene que si leemos un cartel en un sueño y después, en el sueño, nos damos vuelta y volvemos a mirarlo, las palabras habrán cambiado. Según Noë, esto no debería sorprendernos porque, después de todo, en la percepción normal -a diferencia de los sueños-, no tenemos que tomarnos el trabajo de estabilizar el detalle porque el detalle está ahí en el mundo. Hagamos lo que hagamos -cerrar los ojos, darnos vuelta, distraernos-, las cosas que nos rodean no se alteran (Noë 2010: 218).

Por otra parte, con respecto a las experiencias con realidad virtual, es posible esgrimir argumentos semejantes. Por un lado, no podemos afirmar tajantemente que la experiencia virtual dependa, en sentido estricto, únicamente de lo que ocurre en el cerebro. En este sentido, se podría argumentar que depende de la interacción entre el cerebro, los movimientos corporales y el aparato con el cual se completa la experiencia en cuestión. Además, también puede ponerse en duda que dichas experiencias y las experiencias perceptivas en el mundo real sean efectivamente iguales y que la mera imposibilidad de distinguirlas (si es que existe una

imposibilidad tal) no implica que no exista entre ambas diferencias significativas. “En el escenario de la realidad virtual [...] no oímos, ni vemos, ni actuamos como solemos; es un escenario donde erróneamente, nos parece oír, ver y actuar. En el mejor de los casos, la realidad virtual nos proporciona una experiencia virtual y unas mentes virtuales” (Nöe 2010: 215). Desde luego, esta réplica de Nöe se vería fortalecida si se ocupara de señalar algunas diferencias importantes entre un caso y el otro en lugar de limitarse a afirmar que las existen.

En resumen, en todos estos casos, es precisamente el supuesto de la semejanza el que puede ser puesto en duda. De hecho, si es acertada la propuesta de los partidarios del enfoque sensoriomotor (entre los que se encuentra Noë), según la cual, la estabilidad del campo visual percibido está dada en función de la estabilidad del mundo visual, es decir, el mundo sobre el cual la actividad exploratoria perceptual está dirigida, es esperable que, sin un mundo estable, la experiencia debería perder el carácter consistente, coherente y estable de una percepción genuina. Y esto es justamente lo que podría alegarse que ocurre en el caso de los sueños y las alucinaciones. De este modo, se esboza una posibilidad para debilitar la idea de la semejanza entre experiencias verídicas y de las otras. Sin embargo, no debe sobredimensionarse el alcance de los argumentos que ofrece Noë en contra de la tesis del cerebro creador. Como puede verse rápidamente, muchos de ellos parecen dar por sentado que habría *experiencias* (como las oníricas) que podrían interpretarse como siendo el resultado exclusivo de la labor del cerebro, aun cuando sean muy diferentes de las experiencias que tienen lugar dentro de la vigilia. Así, en el mejor de los casos, los argumentos de Noë logran *restringir* el alcance de la tesis del cerebro creador.

Ahora bien, fuera de estas motivaciones más típicas para rechazar la idea de que percibir es, en definitiva, una relación epistémica entre un perceptor y un objeto percibido y que los objetos que percibimos no están en nuestra cabeza (excepto que cuando estamos observando algún objeto colocado en nuestra cabeza a través de un espejo o cuando observamos nuestro cerebro a través de algún tipo de aparato), algunos autores, como Hodgson 2005 sostienen, en el contexto de una discusión en torno al problema de la consciencia, que si bien “ver un objeto no es ver [literalmente] una imagen [*picture*] interna del objeto, [...] está más allá de toda duda que de hecho vemos objetos vía imágenes construidas por nuestro cerebro” (85).

Por ejemplo, cuando miro la puesta de sol justo antes de que se desaparezca detrás del horizonte, hay algo en mi campo visual que es naranja y de apariencia circular, pero que no es (no puede ser) el Sol directamente. ¿Por qué? En el mejor de los casos, se trataría del sol como era hace ocho minutos, porque es un hecho que toma más o menos ese tiempo que la luz viaje desde el objeto en cuestión en el que se refleja hasta los ojos del observador.

Ahora bien, supongamos que el sol se desintegra durante ese período. “Entonces, si la cosa naranja es literalmente el sol, en lugar de algún tipo de imagen o representación visual del sol, [entonces deberíamos admitir que] el sol existe (está justo aquí, justo ahora) y al mismo tiempo no existe (se desintegró) al momento en que me percato de la cosa naranja”. (Hodgson 2005: 85). En otras palabras, es un hecho que cuenta con aval científico que efectivamente toma tiempo que la luz viaje desde el objeto (percibido) en el que se refleja hasta los ojos del observador. Por lo tanto, cuando observamos la puesta de sol, lo que aprehendemos no puede ser al sol porque la luz que impacta en nuestra retina dejó el sol ocho minutos antes, y por todo lo que sabemos, el sol podría haber dejado de existir en ese intervalo. Así, a menos que queramos caer en la contradicción de admitir que el sol tiene que existir para que lo veamos ahora pero, al mismo tiempo, que dejó de existir de acuerdo con la hipótesis propuesta, deberíamos conceder que lo que vemos no es el sol en sí mismo sino una “imagen” de sol construida por nuestro cerebro. Pero una vez que se acepta que el cerebro construye tales imágenes, “se hace innegable que sólo el sujeto de la experiencia se percata de estas imágenes; y esto puede ser llamado razonablemente acceso privilegiado a esas imágenes” (Hodgson 2005: 87).

Ahora bien, ¿qué tipo de acceso tenemos a esta imagen y en qué sentido podemos considerarla como tal?

La cosa naranja debe seguramente ser una imagen del sol construida por mi cerebro [...], si bien no en el sentido de una imagen del sol que podría en sí misma ser vista. De hecho, no veo esta imagen, y no es un tipo de cosa que exista independientemente de ser parte de mi experiencia consciente. Es diferente a una imagen en que me es inmediatamente presente, y en que su existencia es completamente dependiente de mi existencia como sujeto inmediatamente familiarizado con ella. Pero es una imagen sin embargo, en el sentido de que no es el sol en sí mismo sino una representación visual del sol para mí (Hodgson 2005: 85-86).

Como se desprende de las palabras de Hodgson, la imagen del cerebro no es literalmente una imagen (como afirma el autor, no es una imagen en el sentido en que un cuadro o una foto podrían ser una imagen del sol, ni es un tipo de cosa que exista independientemente de ser parte de la experiencia consciente). Ahora bien, ¿por qué no podría decirse que es literalmente una imagen del sol?

En primer lugar, porque si se entiende esta idea de “imagen” en un sentido más o menos habitual del término, parece sencillamente falso que percibir sea como ver una imagen, un cuadro o una fotografía, a menos que efectivamente estemos viendo alguno de estos objetos. En este sentido, en la medida en que no vemos literalmente imágenes del mundo exterior, no resultaría necesario poseerlas en el cerebro para dar cuenta de lo que percibimos.⁵⁴ Pero dejando de lado esta cuestión en la que no voy a profundizar acá, otra de las razones por las cuales resultaría problemático plantear que la percepción supone, literalmente, la presencia de imágenes en el cerebro es que esto nos arrastraría a alguna variante del problema conocido como “problema del homúnculo”. Si la percepción de un objeto externo (el sol en este caso) se explicara postulando una percepción de un objeto interno (la imagen del sol), lo único que se habría logrado dando este paso es retrotraer la necesidad de dar una explicación ahora con respecto a la percepción interna.

Como puede verse, a menos que se llegue a un punto en la cadena de explicaciones en el cual ya no quede ningún tipo de percepción de la que haya que dar cuenta (porque, por ejemplo, se logró explicar mediante otro mecanismo) se corre el riesgo de caer en un regreso al infinito. Por otra parte, en la medida en que, otra vez, en nuestros actos perceptivos habituales no percibimos lo que ocurre dentro de nuestro cerebro, esta explicación representacionista o pictórica de la percepción debe explicar cómo las imágenes internas, logran dar lugar a nuestras percepciones. La mera postulación de imágenes o representaciones internas no alcanza para explicar el fenómeno de la percepción. En otras palabras, ¿por qué una representación del sol en mi cerebro explica mi percepción del sol y no una imagen del sol en mi bolsillo? Aun aceptando que hay imágenes internas o representaciones, el trabajo explicativo no puede detenerse ahí. Volveré sobre este punto.

⁵⁴ En contra de esta concepción fotográfica de la percepción, que no parece ser un mero “hombre de paja” sino que estaría presente de manera más o menos implícita en gran parte de los trabajos científicos acerca de la percepción visual, argumentó en extenso Noë 2004 en el capítulo 2.

Con todo, la réplica de Hacker (2005) al tipo de argumentos como el que plantea Hodgson es concisa y contundente:

El hecho de que a la luz desde el sol le tome ocho minutos alcanzar nuestros ojos no muestra que no veamos ahora al sol como era hace ocho minutos. [...]

Lo que muestra es simplemente que cuando hablamos con precisión de los objetos celestes a grandes distancias de nosotros, necesitamos especificar por separado el tiempo de ver y el tiempo de la existencia u ocurrencia del objeto de la visión (Hacker 2005: 62).

En otras palabras, otra vez se demuestra que no es necesario postular una imagen interna para dar cuenta de nuestra percepción, en este caso, de objetos alejados en el espacio. En definitiva, de estas observaciones se concluye que las imágenes en cuestión no constituyen un recurso ni necesario ni suficiente para dar cuenta de la percepción.

Por último, podemos distinguir otras de las motivaciones centrales que se formulan para sustentar la tesis del “cerebro creador”, esta vez provenientes específicamente del ámbito de la ciencia de la visión. De acuerdo con Noë 2010, desde hace siglos la ciencia de la visión parte de la idea de que lo que vemos representa un estímulo sensorial mucho mayor que el que recibimos, es decir, que existe una discrepancia entre lo que vemos y lo que realmente recibimos del mundo que nos rodea, en forma de estímulo sensorial. Por este motivo, la tarea del cerebro, proponen algunos, es compensar de algún modo la empobrecida “señal de origen”.

En otras palabras, el supuesto de gran parte de la teoría visual es que dada la existencia de ciertas discrepancias entre el carácter de nuestra visión y el carácter de la información que contiene la imagen en la retina, se infiere que es la labor del cerebro compensar la falta. Veamos algunos de los ejemplos que recoge Noë que se supone que avalarían esta inferencia.

(a) La imagen retiniana invertida y el carácter ciclópeo de la vista.

Después de mucho especular en la Edad Media, Kepler averiguó cómo funcionaba la óptica del ojo. En particular, demostró que la imagen retiniana está necesariamente invertida: que la escena visible, tienen que proyectar una imagen invertida en el interior del globo ocular. La cuestión que se plantea inmediatamente, y que sin duda confundió a Kepler, es la siguiente: ¿cómo es que conseguimos ver el mundo en la posición correcta cuando nuestra imagen ocular del mundo está invertida? El problema se agudiza cuando notamos además que hay dos imágenes retinianas, y

no sólo una. ¿Por qué, visualmente, no se muestran las cosas por duplicado y un poco borrosas?

Como puede verse, existiría acá una discrepancia entre lo que vemos (una impresión visual, singular bien enfocada) y lo que recibimos (dos imágenes invertidas, discrepantes). De alguna manera, se pensó, desde la época de Kepler, que el cerebro debía de algún modo compensar esta diferencia dando vuelta las imágenes y combinándolas para formar una única representación correcta del mundo, que se corresponda con lo que realmente experimentamos y vemos.

(b) La imagen retiniana es inestable.

Los ojos se mueven continuamente. Varias veces por segundo se agitan y rebotan; también experimentan pequeñas sacudidas o microsacudidas –es decir, bruscos movimientos balísticos-. De esto resulta que, en realidad, la proyección de un objeto que percibimos como estático, salte en el globo ocular, y que cuando seguimos un objeto en movimiento, su imagen se mantiene fija en la retina, mientras que la del trasfondo inmóvil corretea por nuestros ojos.

De nuevo, para explicar cómo logramos experimentar una visión estable del mundo, según parece, tenemos que suponer que la estabilidad se logra en una etapa posterior del procesamiento de la información retiniana inicial.⁵⁵

Como señala Noë, todos estos ejemplos exhiben una estructura común. Cada ejemplo ilustra que el carácter de nuestra experiencia parece “ir más allá” de lo que nos es dado en los estímulos y, por lo tanto, el cerebro tiene que estar “completando” o “corrigiendo” lo que el mundo deja afuera. Así, la tarea principal de la ciencia de la visión, consistiría en explicar cómo partiendo de dos imágenes retinianas, diminutas, discrepantes, distorsionadas, saltonas, invertidas, agujereadas, desiguales, sólo parcialmente sensibles al color, logramos ver una escena unificada

⁵⁵ Existen otros ejemplos como el fenómeno conocido como “punto ciego”. En cada ojo existe un punto ciego donde no hay fotorreceptores. Sin embargo, no experimentamos ningún agujero o ninguna discontinuidad en el campo visual. Cerremos un ojo y miremos una extensión uniforme de color, la pared, por ejemplo. ¿Notamos alguna discontinuidad? Sin embargo, existe una discontinuidad en la imagen retiniana de la pared. De modo que la presencia de un mundo continuo debe ser el resultado, o eso es lo que nos quieren hacer creer, del relleno activo del cerebro de la imagen defectuosa. Algo similar ocurre con el color. Existen pocos fotorreceptores (conos) sensibles al color en la periferia. No obstante, experimentamos la escena visual en pleno color y en toda su extensión. Además, vemos los objetos del mismo color, aunque varíen radicalmente las condiciones lumínicas. Por ejemplo, la mayor parte del tiempo el color de un libro no parece cambiar, aunque el carácter físico de la composición de la luz en la superficie del libro que nos llega a los ojos cambie de manera radical cuando entramos en un edificio con luz artificial, después de haber estado afuera, bajo un sol cenital.

y estable, de objetos y propiedades esparcidos a nuestro alrededor, a todo color y a todo volumen, en un espacio tridimensional. La tarea de la ciencia desde Kepler hasta acá es entender cómo logra el cerebro este aparente milagro.

Con todo, la postulación de mecanismos compensatorios y de representaciones cerebrales que rellenen los “huecos” de nuestra estimulación resulta innecesaria por diversos motivos. Por un lado, tal como señalan autores como Kevin O`Regan y Alva Noë 2001, porque el mundo es “su propio modelo”. Es decir, si tenemos acceso al fenómeno original, ¿por qué aceptar imitaciones? Es cierto que nuestra experiencia del mundo es más rica en detalles de todo lo que podamos imaginar o pensar. Sin embargo, no es necesario postular una imagen rica y detallada del mundo que de algún modo supla la pobreza del estímulo visual. En cambio, podría pensarse que es el mundo el que es detallado y que lo que tenemos, gracias a la percepción, es una cierta disposición a acceder a cada aspecto del mundo de manera inmediata a medida que resulta necesario. Así, si bien no me represento todos los detalles a la vez, tengo acceso potencial y rápido a estos detalles.

En otras palabras, tenemos una experiencia rica en detalles, pero para esto no es necesario que los detalles estén plasmados en una representación fotográfica del mundo que el cerebro construye a partir de unos elementos precarios. El detalle está en el mundo mismo; y nosotros tenemos un acceso rápido a estos detalles. Así, no es ninguna ilusión que el mundo aparezca ante nosotros con detalles sólidos y variados, aunque no sea cierto que vemos imágenes internas, detalladas y estables del mundo exterior, porque es el mundo el que es detallado.

Algo similar puede decirse con respecto al resto de los presuntos “defectos” mencionados acerca de la visión. Por una parte, no es cierto que la evidencia empírica muestre que el cerebro sea el creador de nuestro mundo y que, por este motivo, la tarea científica consista en explicar cómo se lleva adelante ese milagro. La necesidad de un milagro sólo se explica si se asume de antemano que de algún modo percibimos lo que aparece en la imagen retiniana, es decir, si se acepta de antemano la tesis del cerebro creador. Pero dado que ni nosotros, ni el cerebro, ni nadie (excepto quizás el oftalmólogo) ve la imagen retiniana, no hay razón para pensar que su orientación sea relevante para dar cuenta de la manera en que el mundo se nos muestra en la percepción visual. Qué aspecto visual tiene dicha imagen es irrelevante para llevar adelante su labor neurofisiológica.

Ahora bien, cuando nos damos cuenta que la imagen no es algo que vemos, ya no tiene sentido decir que está invertida, ya que no hay observador con respecto al cual trazar esa distinción. En otras palabras, el error central que parece subyacer a la formulación del problema que se plantea la ciencia de la visión consiste en creer que la imagen funciona como *imagen*, es decir, como algo percibido. Pero una vez que renunciamos a la idea de que la imagen retiniana posibilita la visión *en tanto imagen*, no hay más razones para creer que su orientación sea relevante para explicar las características de nuestra experiencia visual. Si no hay quien la observe, ¿con respecto a qué podría decirse que está invertida o doble? De nuevo, como no experimentamos la imagen retiniana, no es necesario operar sobre ella ninguna corrección. Por ejemplo, como no vemos mediante imágenes retinianas, que haya dos imágenes retinianas o cientos de ellas no cambia necesariamente lo que vemos. O, mejor dicho, su influencia no puede estar dada por su apariencia visual.

En la ciencia actual, si bien ya nadie sostiene explícitamente que alguien (un hombrecito interno o el mismo cerebro) *ve*, literalmente, la imagen retiniana como condición para que haya experiencia visual, el problema fundamental que se propone solucionar (i.e. el problema de explicar cómo es posible que veamos como vemos dadas las imperfecciones de la imagen retiniana), tiene básicamente la misma forma (Nöe 2004: 46). Al parecer, todavía continúa la tendencia de postular mecanismos compensadores, pero es difícil ver por qué habría que pensar que existe un problema o un enigma en lo relativo a la imagen retiniana, y mucho menos que el cerebro necesite “ajustar la inversión”, si realmente ya se hubiera desterrado definitivamente el supuesto disputable de que la imagen retiniana sea como la imagen de un cuadro.

Con respecto a otros de los defectos enumerados arriba, ocurre algo similar: ¿por qué habríamos de suponer que el movimiento de los ojos (o la imagen retiniana) equivale a un defecto del estímulo visual si no fuera porque damos por sentado que el cerebro representa el movimiento del mundo mediante un código según el cual los movimientos de la imagen retiniana corresponden a los movimientos de las cosas que nos rodean? Al parecer, la idea de que existe algún tipo de isomorfismo entre ambos sigue vigente en algunos casos. Sin embargo, por los mismos motivos que antes, es posible argumentar que este isomorfismo no es necesario. En otras palabras, como ya escribió Dennett (1995) hace tiempo, no debe

pensarse que es necesario que las propiedades de nuestra experiencia deban igualarse (de algún modo) con las propiedades físicas de los vehículos neurales que transportan la información acerca de ellas o que las posibilitan. Así, por ejemplo, no es necesario considerar que el tiempo debe ser representado temporalmente y el espacio espacialmente.⁵⁶ No debe confundirse la representación de la posición de un objeto en el mundo con la posición dentro del cerebro de la representación de ese objeto. Análogamente, no es necesario que el movimiento en el mundo sea representado por cierto movimiento en mapas topográficos en el cerebro. (Desde luego, afirmar que una semejanza o isomorfismo tal no es necesario no implica que, de hecho, estos no se den).

Ahora bien, si se aceptan estas observaciones, cobra fuerza la idea de que los problemas apuntados arriba en torno a la visión, no son otra cosa que pseudoproblemas que tienen su origen en el supuesto inicial y disputable de que el cerebro aprende sobre el mundo mediante el escrutinio de la imagen retiniana. Así, lejos de apoyar la idea del cerebro creador y del mundo percibido como ilusorio, estos varios rompecabezas son, en sí mismos, *el resultado de ese supuesto espurio inicial*. Pero hay algo más.

En efecto, una vez que se reconoce que no percibimos literalmente aquello que ocurre en el cerebro, debe reconocerse también que la mera postulación de una representación interna (más o menos semejante o isomorfa) no es *suficiente* para explicar los rasgos de nuestra experiencia perceptiva. ¿Cómo una representación de la Torre Eiffel en mi cabeza explica mi percepción de la presencia de la Torre Eiffel justo ahí al cruzar la calle? ¿Qué tiene de especial esa representación mental que no tenga la postal en mi bolsillo? Consideremos qué pasaría si encontráramos de hecho un conjunto de neuronas que se correlacionara perfectamente con nuestra percepción visual; un pequeño cuarto de proyección en el cual lo que es visto por nosotros es proyectado allí de manera idéntica. Si se reflexiona, es claro que dicho

⁵⁶ En general, debemos distinguir los rasgos de los representantes de los rasgos de los representados. Uno puede chillar <<¡Con cuidado, caminad de puntillas!>> a pleno pulmón, hay imágenes enormes de objetos microscópicos, y no hay nada imposible en una pintura al óleo de un artista haciendo un boceto al carboncillo. La primera frase de una descripción escrita de un hombre de pie no tiene por qué describir su cabeza, ni la última frase tiene por qué describir sus pies. Este principio también se aplica, con menos claridad, al tiempo. Considérese la siguiente frase proferida oralmente: <<un brillante y breve destello de luz roja>>. Empieza con <<un brillante>> y acaba con <<luz roja>>. Las porciones de dicha proferencia no son en sí mismas representaciones del principio o el final del breve destello rojo [...] (Dennett 1995:162).

descubrimiento, por sí solo, no nos acercará más a comprender cómo es que vemos. Porque todavía nos faltaría una explicación de cómo la imagen en esa “glándula pineal” genera la visión; esto es, como hace posible o controla o modula las formas de la actividad en que consiste ver (O’Regan y Noë 2001:966-967). Desde luego, estas observaciones no implican que esta explicación ausente no pueda darse. Tampoco permiten concluir que no haya representaciones en el cerebro. Lo único que pretendía apuntar es que la mera presencia de las mismas, no es suficiente para explicar las propiedades de la experiencia visual.

Sin embargo, si extendemos un poco más la moraleja que mencionamos anteriormente acerca de que no deben confundirse las propiedades de nuestra experiencia con las propiedades físicas de los vehículos neurales que transportan la información acerca de ellas o que, al menos, las posibilitan, debemos admitir que aun en el caso en que se conceda que hay representaciones en el cerebro (lo cual puede significar simplemente la presencia de patrones de activación neuronal), no es necesario consentir que la percepción misma tiene un carácter representacional, porque, otra vez, estaríamos identificando innecesariamente las propiedades de los diferentes niveles mencionados.

En otras palabras, una cosa es sostener que las representaciones sub-personales están entre las condiciones internas que posibilitan la percepción. Pero otra cosa muy diferente (y que, en todo caso, quedaría abierta aun aceptando la afirmación anterior), es que la percepción misma tiene una naturaleza representacional.

Consideraciones similares pueden aplicarse a otro de los problemas que deberían enfrentar, a juicio de Tye (1995, 2007), quienes se propongan explicar la consciencia: me refiero al llamado “problema de la unión” (*binding problem*). ¿De qué se trata este problema? Veamos. Los científicos suelen considerar que el sistema visual está compuesto de numerosos subsistemas (o módulos) más o menos independientes que extraen información acerca de una variedad de atributos de los objetos externos como el color, contraste, profundidad, orientación y textura a partir de los estímulos visuales. El hecho de que estos módulos operen de manera independiente y que estén en general ubicados en diferentes regiones del cerebro, hace surgir la pregunta acerca de cómo es posible que flujos separados de información se unifiquen para dar lugar a una percepción unificada de la realidad

tal como la que de hecho tenemos. Así, se postulan diferentes mecanismos a partir de los cuales “todo se reuniría”.

Sin embargo, la idea misma de que una unificación así resulte necesaria y suficiente para dar cuenta de los rasgos de nuestras experiencias visuales resulta impugnable. En efecto, suponer que una unificación temporal y espacial de las distintas vías de información presentes de forma paralela e independiente en el cerebro es necesaria para que haya una percepción unificada en el espacial y temporalmente, parece apoyarse en el prejuicio de que las propiedades de los vehículos de representación deben replicar a las propiedades de lo representado por dichos vehículos. Al mismo tiempo, la postulación de un mecanismo de unificación parece basarse en el prejuicio homuncular de que aquello que efectivamente percibimos no es el mundo externo sino una imagen o representación de él creada por el cerebro.

Pero entonces, una vez más, lejos de constituirse como una prueba empírica a favor de la tesis del cerebro creador, el problema de la unión se origina por la aceptación previa de dicha tesis. Por otra parte, y tal como observan Noë y Bennett y Hacker, precisamente porque no percibimos literalmente las representaciones que presuntamente fabricaría nuestro cerebro, su mera presencia y unificación (sea lo que sea que esto signifique y por mucha semejanza que exista entre la ficción y la realidad), no bastaría por sí sola para dar cuenta de los rasgos y características de nuestras experiencias perceptivas.

4. Conclusiones

En este capítulo me ocupé de mostrar cierta evidencia a favor de la idea de que el fantasma de Descartes sigue vivo no sólo en la filosofía, sino en buena parte de las neurociencias contemporáneas. De acuerdo con este “nuevo” dogma, somos, casi literalmente, nuestro cerebro, de forma tal que, en rigor, es el cerebro, antes que algo no físico, lo que siente, piensa y decide.

A partir de la constatación de esta situación, mostré las consecuencias nocivas que se siguen de adoptar esta doctrina a partir de los aportes de los trabajos de Noë y Bennett y Hacker: entre dichas consecuencias mencioné, por un lado, la caída en un escepticismo acerca de la existencia y naturaleza del mundo externo y de las otras mentes. En efecto, una vez que se asume esta perspectiva cartesiana, se

supone que lo que realmente percibimos de manera directa son representaciones o imágenes, creadas por nuestro cerebro, y no las cosas “en sí mismas”. Al mismo tiempo, de este modo llegamos por otra vía a la postulación de objetos privados como los objetos propios de la percepción consciente.

Por otro lado, se seguía también de la adopción de esta perspectiva cartesiana, la posibilidad de caer en lo que Bennett y Hacker llamaban “falacia mereológica”, es decir, el error de atribuir predicados psicológicos a las partes de un organismo que, en principio, sólo pueden ser aplicados de manera inteligible al organismo como un todo. Las confusiones conceptuales que se seguían de esta falacia ponían en duda la potencialidad explicativa de la neurociencia cartesiana porque, como observé, parecía presuponer aquello que debía explicar.

Por último, en la sección 2, consideré algunas de las razones que suelen citarse para adoptar la tesis del cerebro creador. Entre ellas, mencioné la presencia de ciertos fenómenos más o menos cotidianos que de algún modo parecen abonar a la idea de que percibimos representaciones creadas por el cerebro: se trata de experiencias (como los sueños, las alucinaciones, la realidad virtual o la percepción de objetos muy alejados en el espacio) que parecen idénticas o muy semejantes a distintas experiencias perceptivas verídicas pero en las cuales el objeto externo está ausente. En estos casos, la experiencia se genera, en principio, de una manera que no involucra ningún objeto externo, y lo más razonable parece ser adjudicar al cerebro esta tarea. Ahora bien, el hecho de que exista esta semejanza sugiere con plausibilidad que los mecanismos que subyacen en uno y otro caso deben ser también muy similares e incluso que *todas* nuestras experiencias (y no sólo los sueños, o las experiencias con realidad virtual) requieren de algún tipo de intermediario (una representación creada por el cerebro) para configurarse como tales. Otra de las motivaciones esgrimidas para postular este tipo de intermediarios tiene que ver con la necesidad de postular mecanismos compensadores que puedan dar cuenta de nuestras experiencias perceptivas dadas la “pobreza” e “incoherencia” de los estímulos que afectan nuestros sentidos. En este caso, el cerebro se encargaría de corregir los defectos de la señal de origen.

Frente a estas observaciones, mostré, siguiendo a Noë y Hacker, que esas experiencias que parecían semejantes a las experiencias perceptivas ordinarias pero en las cuales no está presente ningún objeto externo o bien no son, de hecho,

semejantes, o bien, cuando lo son (como en el caso de la percepción de objetos externos lejanos) el objeto externo está presente y entonces debe considerárselas como experiencias verídicas normales.

En segundo lugar, y siguiendo por esta última línea argumentativa, se verá que la presencia de estas representaciones cerebrales no resultaría *necesaria* para dar cuenta de las experiencias perceptivas. Así, por ejemplo, aquellos presuntos defectos de la percepción visual que el cerebro debería corregir no son tales o, al menos, no es necesario considerar que lo sean. Por el contrario, considerarlos como tales *presupone* la perspectiva cartesiana a la que, en teoría, estos hechos deberían brindarle apoyo empírico independiente.

En tercer lugar, mostré que aun aceptando que existan tales representaciones, lo cierto es que no resultarían *suficientes* para explicar aquello que se supone que vienen a explicar.

CONCLUSIÓN

La prestidigitación explicada: desde la (di)solución del “problema” de la consciencia hacia nuevos horizontes

1. Introducción

Como vimos, la consciencia ha sido fuente de numerosas y prolongadas discusiones entre filósofos de la mente, pero también entre científicos cognitivos y neurocientíficos. De hecho, en muchas ocasiones científicos y filósofos coinciden en señalar que persiste una profunda oscuridad en torno a la naturaleza de la consciencia y a su vinculación con el resto del mundo físico. Así, por ejemplo, científicos cognitivos como Phillip Johnson-Laird (1988), confiesan que “nadie sabe qué es la consciencia, o si sirve a algún propósito”. Y filósofos, como Chalmers (1999), consideran que la consciencia es el último obstáculo pendiente en nuestra búsqueda de una explicación científica del universo. En este contexto, se suele creer que las dificultades involucradas a la hora de dar cuenta de este fenómeno tienen su origen no en una falta de información empírica relevante, sino en la falta de una idea clara de cómo deberían responderse las preguntas que se plantean.

Por mi parte, en este trabajo sostuve que la existencia de estas dificultades mencionadas se explica, en buena medida, en virtud de que quienes se ocupan de estas cuestiones atribuyen a la consciencia, de manera más o menos implícita, ciertas propiedades que la convierten en un atolladero verdaderamente insuperable, *y por principio*, para una posición fisicalista de la mente. Sin embargo, el recorrido que se llevó adelante en este trabajo reveló, por un lado, que hay buenas razones para pensar que algunas de esas propiedades problemáticas, como la intrinsecalidad y el carácter presuntamente privado y directo del conocimiento del fenómeno en discusión, no forman parte de la naturaleza de la consciencia (aun cuando pueda parecerlo a primera vista). Por otro lado, argumenté también en favor de la idea de que otra de esas propiedades supuestamente problemáticas atribuidas a la consciencia, la infabilidad, si bien pertenece a la naturaleza del fenómeno en

cuestión, no es especialmente problemática cuando se la comprende de manera adecuada.

A modo de cierre, en este último capítulo voy a enfocar la discusión desde otro ángulo pero con el objetivo de arribar a conclusiones similares. En particular, me ocuparé de analizar una frase que la mayor parte de los filósofos involucrados en estas discusiones utilizan típicamente para hacer referencia al aspecto problemático de la consciencia, sin preocuparme ahora por determinar si esta caracterización asume implícitamente algunas o todas las propiedades mencionadas recién. Concretamente, estoy pensando en la célebre frase acuñada por el filósofo Thomas Nagel en su artículo del 74, según la cual, el aspecto problemático de la experiencia se describe como “lo que es como estar en un estado mental para el organismo”⁵⁷ (Nagel 2003: 46).

En este marco, argumentaré que es posible concluir, a partir de la discusión que se suscita entre los trabajos de P.M. S. Hacker (2002) y Greg Janzen (2005, 2008, 2011) que, en definitiva, quien utiliza el giro nageliano, se enfrenta a una suerte de dilema, según el cual, (a) o bien admite que utiliza la frase en cuestión en un sentido que no es claramente inteligible, y entonces tampoco es claro que haga referencia a un aspecto especialmente problemático de la mente, (b) o bien acepta que la usa en un sentido inteligible (ya sea técnico o cotidiano) pero que claramente no es problemático en absoluto. En cualquier caso, ninguna de las opciones mentadas ayuda a favorecer la idea de que hay algo inasible en torno a la consciencia para el que adopte un enfoque fisicalista / funcionalista de la mente.

2. ¿La consciencia? ¿Qué es eso? La crítica de Hacker

De acuerdo con Nagel, es la consciencia lo que vuelve al problema mente-cuerpo realmente inextricable (Nagel 2003: 45): “sin la conciencia, el problema mente-cuerpo sería mucho menos interesante; con la conciencia parece no tener solución.” (Nagel 2003: 46). Ahora bien, ¿cómo define Nagel específicamente este concepto y porque su referente constituye un problema sin solución visible?

⁵⁷ De hecho, es el trabajo de Nagel el que motiva gran parte de la discusión en torno a la consciencia dentro de la filosofía de la mente contemporánea.

En principio, podríamos comenzar por mencionar aquellos fenómenos concomitantes cuya explicación no resultaría verdaderamente inextricable y que no conformarían el eje de la discusión. En primer lugar, parece que las dificultades que menciona Nagel no se vinculan con la distinción entre lo que se llama, a partir del trabajo de Malcom (1984), consciencia *intransitiva* y consciencia *transitiva*. De acuerdo con esta clasificación, el primer tipo de consciencia se contrasta con varias formas en las cuales un organismo puede no estar consciente, como cuando está dormido, anestesiado o en estado de coma. Por otro lado, la consciencia transitiva es la consciencia *de algo*, ya sea algo percibido que captura nuestra atención, un estado mental ocurrente, o algún hecho o pensamiento que se pondera en una deliberación.

En principio, los estados de la consciencia transitiva son aquellos estados mentales que ocurren mientras se es consciente intransitivamente. Sin embargo, como dije, ninguno de estos dos sentidos del concepto de consciencia recoge el aspecto problemático del fenómeno.

En todo caso, para acercarnos a él, sería adecuado señalar que se trata de un aspecto de la mente que mostraría las limitaciones de cualquier perspectiva funcionalista acerca de ella. Como vimos, podría decirse que, en líneas generales, el funcionalismo sostiene que un estado mental se define en términos de las relaciones causales que mantiene con ciertos *inputs* (los estímulos sensoriales), los *outputs* del sistema u organismo (su conducta) y otros estados mentales internos. Por lo demás, habíamos visto que este enfoque resultaba compatible con el fisicalismo en la medida en que se supone que los estados y roles causales relevantes, se instancian o realizan en sistemas *físicos* de una manera semejante al modo en que el *software* “corre” en el *hardware* de una computadora. Sabemos, sin embargo, que se suele pensar que existe un aspecto de lo mental que se resistiría a ser abordado desde una perspectiva funcionalista: se trata del aspecto cualitativo de la mente que sería paradigmáticamente exhibido por las sensaciones corporales, las experiencias perceptivas y las emociones. En efecto, tal como señala Thomas Nagel, este aspecto:

No puede analizarse en términos de ningún sistema explicativo de estados funcionales o de estados intencionales, pues éstos podrían atribuirse a robots o a autómatas que se comportaran como personas, aunque no tuvieran ningún

tipo de experiencia [i.e. experiencia consciente]. Debido a razones similares, tampoco puede analizarse en términos del papel causal de las experiencias en relación con la conducta humana típica. No niego que los estados y sucesos mentales conscientes causen la conducta, ni que se les pueda dar una caracterización funcional. Sólo niego que ello agota su análisis (Nagel 2003: 47).

En tiempos más recientes, a partir de los trabajos de Chalmers, se popularizó una distinción, (que mencioné en la Introducción de este trabajo), entre los problemas fáciles en torno a la consciencia, y *el* problema difícil, que parece cuadrar bastante bien cuando se intenta delimitar el sentido del concepto de consciencia que está en juego para Nagel. De acuerdo con Chalmers, los llamados problemas fáciles son fáciles⁵⁸ porque todo lo que se necesita en estos casos es una explicación de ciertas *habilidades y funciones* cognitivas. Y, al parecer, los métodos de la ciencia cognitiva resultan perfectamente adecuados para afrontar este tipo de cuestiones. Entre estos problemas fáciles se encontrarían, entre otros, la explicación de la habilidad, por parte de un organismo o individuo de: discriminar, categorizar, y reaccionar ante los diversos estímulos ambientales, de integrar la información proveniente de diferentes vías en un todo coherente, de reportar los propios estados mentales, de acceder a sus propios estados mentales, de dirigir el foco de la atención a algún evento interno o externo, de controlar de manera racional la propia conducta en base a la información disponible, o la diferencia entre el estado de vigilia y el sueño, entre otros. (Como puede verse, parte de estos fenómenos coinciden que los que apuntamos a propósito de la distinción entre consciencia transitiva y consciencia intransitiva). Sin embargo, lo que hace verdaderamente difícil al problema mente-cuerpo es la dificultad que se presenta al intentar explicar, en términos compatibles con el conocimiento científico contemporáneo y con sus compromisos filosóficos y ontológicos, el aspecto subjetivo o cualitativo de la experiencia antes mencionados. Y resulta difícil precisamente porque, en este caso, tal como sostenía Nagel, no sería una *función o distintos roles causales* lo que resulta necesario explicar.

⁵⁸ La caracterización de estos problemas como “fáciles” es obviamente una caracterización relativa. Son fáciles con respecto al problema difícil, pero de ningún modo Chalmers piensa que no pueda tomar unos cuantos siglos poder resolverlos.

En definitiva, de acuerdo con la caracterización de Nagel, el aspecto problemático de la consciencia es el que se cifra en la siguiente frase: “un organismo tiene estados mentales conscientes si y sólo si hay algo que es cómo es *ser* ese organismo, algo que es cómo es *ser para* ese organismo” (Nagel 2003: 46). De acuerdo con la mayor parte de los intérpretes (incluyendo a aquellos que serán objeto particular de análisis en este capítulo), la verdad de la frase citada dependería, presumiblemente, del presupuesto de que una criatura tiene experiencias conscientes si hay algo que es como es *ser* ese organismo para ese organismo *en virtud de estar o tener* esas experiencias. De este modo, la afirmación básica de Nagel parece ser un intento de identificar la propiedad (o propiedades) que convierten a un estado mental en un estado mental *consciente*.

En otras palabras, de acuerdo con esta propuesta, la frase en cuestión podría ser parafraseada de la siguiente manera: un sujeto u organismo S está en un estado mental consciente M en el tiempo t, si y sólo si hay algo que es como estar en M en t para ese sujeto u organismo (Janzen 2011: 273; Hacker 2013: 50).

Dicho esto, habría que apresurarse a aclarar, en primer lugar, que la frase nageliana no pretende hacer alusión a algún tipo de comparación entre lo que es tener un estado mental consciente y otra cosa u otro estado, cómo cuando se pregunta a qué se parece caminar rápido y se responde: “Bueno, es como correr sólo que más despacio”: en efecto, en palabras del propio Nagel, “la forma analógica de la expresión ‘cómo es ser’ [un murciélago, por ejemplo] provoca confusiones. No significa ‘a qué se *parece* (en nuestra propia experiencia)’, sino más bien a ‘cómo es para el sujeto mismo’ (Nagel 2003: 50). Al mismo tiempo, se suele considerar que cada experiencia o estado mental consciente tiene su propio carácter cualitativo distintivo, su propia “sensación fenoménica” o *quale*. Así, se suele decir también que el problema de explicar estas cualidades es el problema de explicar la consciencia (Hacker 2013: 51; Chalmers 1999: 26).

Por su parte, Hacker va a rechazar precisamente (i) que haya algo que sea como es tener una experiencia o estado mental consciente y (ii) que haya algo que sea como es ser un murciélago para un murciélago o, para el caso, algo que es como es ser un ser humano para un ser humano. De todo esto se seguiría, según el autor, (iii) que no hay algo (i.e. el carácter subjetivo de la experiencia) que sólo pueda ser aprehendido desde el punto de vista del sujeto de la experiencia pero nunca desde

el punto de vista objetivo de la ciencia. Veamos cómo Hacker lleva adelante su estrategia argumentativa para concluir (i) y (ii).

Como dije recién, en ocasiones se define este concepto o aspecto de ciertas experiencias o estados mentales en términos de los modos distintivos en que se *sienten* las experiencias: “un estado mental es consciente si existe algo que es como estar en ese estado mental. Para plantearlo de otra manera, podemos decir que un estado mental es consciente si está ligado a una *sensación cualitativa*, una cualidad asociada a la experiencia” (Chalmers 1999: 26). Ahora bien, en este punto cabe preguntarse, ¿hay realmente un modo específico en que se *siente* ver, escuchar u oler?

La respuesta de Hacker es negativa. En primer lugar, porque no es cierto que cada experiencia tenga de alguna manera asociada una sensación (*feeling*) particular, si por esta sensación se entiende algún tipo de reacción emocional o actitudinal frente a ella o ante ella, como por ejemplo cuando se dice que cierta experiencia es agradable, placentera o maravillosa. En efecto, mientras que en circunstancias habituales podría responderse, frente a preguntas como “¿cómo se siente probar el primer vaso de cerveza?” o “¿cómo se siente escuchar la novena sinfonía de Beethoven?” cosas tales como, “Se siente bastante desagradable” o “Es muy placentero” respectivamente, no es cierto que atribuyamos este carácter a *todo* tipo de experiencias. Por ejemplo, ver un buzón es diferente de ver una lámpara o una mesa. Sin embargo, ninguna de estas experiencias perceptivas evoca algún tipo de reacción emocional como las mencionadas recién. Desde luego que se trata de experiencias diferentes. Pero, en principio, la diferencia entre ellas no parece estar dada por el modo distintivo en que cada una de ellas se *siente*.

En todo caso, una diferencia notable que podría señalarse es que todas difieren en aquello de lo que son experiencias (un buzón, una lámpara, una mesa). De la misma manera, podríamos decir que ciertas experiencias que típicamente se señalan como portadoras de este misterioso carácter fenoménico, como por ejemplo, la experiencia de oler “el aroma del pan recién horneado” o de ver “el matiz particularmente intenso de púrpura proveniente de un libro colocado en un estante”, son experiencias verdaderamente diferentes pero que se distinguen por sus diferentes objetos, que en este caso son propiedades que se perciben como propiedades cualitativas de los *objetos* de las experiencias, y no de las experiencias

mismas. Pero una cualidad percibida no es una cualidad de la percepción. De esta manera, también deberían ser descartadas como aquellas propiedades cuya naturaleza estamos tratando de captar.

En segundo lugar, aun aquellas experiencias que admiten ser consideradas bajo estos parámetros, no pueden ser distinguidas por el modo en que se sienten, porque experiencias de diversos tipos comparten una misma actitud o reacción emocional. En efecto, distintas experiencias pueden *sentirse* desagradables, interesantes o aburridas, de modo tal que no resulta correcto afirmar que cada una de ellas posee una sensación cualitativa propia o específica que la distingue del resto. Es cierto que no es posible derivar el placer que produce la cerveza de ver el amanecer, pero esto podría explicarse, según Hacker, porque “la identidad del placer o del disfrute depende del objeto del placer” (Hacker 2002: 163).⁵⁹

Ahora bien, la frase misma “hay algo que es como estar en M para S”, es una expresión gramaticalmente incorrecta de acuerdo con Hacker que surge de cruzar la forma gramatical que tendría una respuesta a la pregunta que inquiriera a qué se parece tener una experiencia, con la forma gramatical de una respuesta a la pregunta que inquiriera qué se siente tener una experiencia (es decir, si resulta agradable, terrorífica, etc.). Sin embargo, en este último sentido, resultaría gramaticalmente incorrecto decir que hay algo que “es como” tener una experiencia, porque no decimos, por ejemplo, “es *como* desagradable escuchar Reggaetón”. Simplemente decimos (o podríamos decir) que *es* horrible. Por este motivo, sostiene Hacker, dejando los juicios comparativos de lado, no sería gramaticalmente correcto decir que “hay algo que es como tener la experiencia auditiva de escuchar reggaetón”. En todo caso, podríamos reconstruir la estrategia de Hacker de la siguiente manera: hay dos modos gramaticalmente legítimos (dentro del lenguaje ordinario) en los cuales podría interpretarse y utilizarse la expresión de Nagel (o mejor dicho, una similar): (a) cuando se pretende responder a qué se parece tener una experiencia

⁵⁹ Por otra parte, Hacker apunta que, en estos casos, decir que una experiencia (como por ejemplo ver, observar, entrever, escuchar, probar esto o aquello, pero también caminar, hablar, bailar, jugar un partido de fútbol, trepar una montaña o pintar un cuadro) tiene un aspecto cualitativo (es decir, que es, por ejemplo, agradable, encantadora, desagradable, etc.), es una manera de decir que el sujeto de la experiencia, es decir, la persona que ve, escucha o prueba, las encuentra o considera agradables, maravillosas, etc. En otras palabras, en estos casos, el carácter cualitativo de una experiencia E (es decir, como se siente tener esa experiencia), es la actitud “afectiva” del sujeto que tiene la experiencia (Bennett y Hacker 2003: 275) con respecto a ella.

determinada (y así señalar, por ejemplo, que tener una experiencia de tipo E es semejante a tener una experiencia de tipo E2), o bien (b) cuando se pretende responder cómo se siente tener una experiencia determinada (y, se afirma, por ejemplo, que es desagradable, maravillosa, etc.).

Sin embargo, el primer sentido queda explícitamente descartado por Nagel como una interpretación adecuada de la pregunta que le interesa tratar, y el segundo sentido no serviría para caracterizar ni distinguir a los estados conscientes, ya que, como vimos antes, no todos los estados mentales poseen este “aspecto cualitativo” y distintos estados pueden poseer el mismo tipo. Así las cosas, (b) también debería descartarse en principio como una interpretación adecuada de la frase (Hacker 2002: 166-167). Pero entonces, ¿cómo deberíamos interpretar esta frase exactamente? De estas observaciones Hacker pretende extraer la primera de las conclusiones apuntadas, a saber: (i) que no es cierto que haya algo que es como tener una experiencia o estado mental consciente. Veamos ahora qué caminos lo conducen a justificar la segunda, es decir, (ii) que haya algo que sea como es ser un murciélago para un murciélago o, para el caso, algo que es como es ser un ser humano para un ser humano.

En primer lugar, apunta Hacker, en ocasiones hacemos preguntas de la forma “cómo es ser X”, como cuando preguntamos “¿cómo es ser un soldado (o un marinero o un ganador)? y entonces se pretende que se responda describiendo algunas características típicas de esos “roles” (o “tipos” de vida): qué es lo que se debe hacer y atravesar, sus pro y sus contra, etc. En otras palabras, el tipo de respuesta solicitada en estos casos consiste en efectuar algún tipo de descripción de las actividades y vicisitudes características de la vida de, pongamos por caso, un soldado o un marinero. Ahora bien, Hacker señala que en estos casos también resultaría gramaticalmente incorrecto decir que “hay algo que es como ser X”. En efecto, supongamos que se preguntara a alguien cómo es ser un soldado. Ante esta pregunta, la persona interrogada podría decir cosas tales como: “¡Es maravilloso! Me llena de orgullo servir a mi país”, o bien, “Es horrible, hay que arriesgar la vida para defender los intereses de una minoría acomodada que pretende hacernos creer que sus intereses son los intereses de la nación”.

Ahora bien, en este caso, no resultaría gramaticalmente correcto afirmar que “hay algo que es como ser soldado” en la medida en que no resultaría correcto

afirmar que es *como* maravilloso (u horrible) serlo. Otra vez, sencillamente diríamos que *es* maravilloso (u horrible) serlo.

Por otro lado, continua Hacker, generalmente no resulta necesario especificar a qué clase de individuos está dirigida la pregunta, porque el contexto lo hace evidente. Así, por ejemplo, si preguntamos cómo es ser un doctor, es claro que la pregunta queda restringida a los seres humanos que cuentan con ciertas credenciales académicas. Y cuando preguntamos cómo es estar embarazada, resulta obvio que la pregunta está dirigida únicamente a las mujeres. Sin embargo, en ocasiones podríamos estar interesados en realizar una pregunta dirigida a una clase particular de individuos, como cuando preguntamos “¿cómo es para una *mujer* ser un soldado (en oposición a cómo es para un hombre ser un soldado)?” o “¿cómo es para un adolescente debutar en la primera división del fútbol argentino (en contraposición a cómo es para alguien mayor de edad jugar en primera)?”. Otra vez, este tipo de preguntas demanda una declaración acerca de las impresiones, dificultades, experiencias vividas, satisfacciones y sinsabores de quienes ocupan esos roles (el de soldado o el de jugador de fútbol) en tanto miembros del “subconjunto” al que pertenecen (es decir, el de las mujeres o los adolescentes).

Como puede verse, se trata de preguntas y respuestas perfectamente admisibles pero que, en principio, no dejan ver de qué modo podrían hacer foco en algún aspecto problemático de la experiencia consciente.

Con todo, Hacker admite que hay tres razones por las cuales no puede ser éste el sentido en el que debe interpretarse la pregunta de Nagel. En primer lugar, porque la pregunta de Nagel tiene la forma de “cómo es para X (un murciélago, por ejemplo) ser un X”, mientras que las preguntas del lenguaje cotidiano consideradas en el párrafo anterior, tienen la forma “cómo es para X ser Y”, donde X es distinto de Y. Así, se pregunta, por ejemplo, cómo es para una *mujer* ser un soldado, pero *no* cómo es para un soldado ser un soldado o para una mujer ser una mujer.

Ahora bien, en caso de que la pregunta se realizara de este último modo, estaríamos dispuestos a considerarla, de acuerdo con Hacker, como una forma un tanto extravagante de preguntar simplemente, por ejemplo, cómo es ser un soldado (es decir, como si fuera equivalente a una pregunta de la forma “cómo es ser X”).

En segundo lugar, la pregunta filosófica es diferente a las preguntas del lenguaje cotidiano que tienen una forma similar porque, en este último caso, cuando

se pregunta cómo es para X ser Y, suele haber un contraste entre la clase de los Xs posibles a los que está concretamente dirigida la pregunta y otros posibles Xs que también podrían ser Y. En otras palabras, cuando se pregunta cómo es para una mujer ser un soldado, lo que se quiere saber es cómo es para una mujer ser un soldado en contraposición, por ejemplo, a cómo es *para un hombre* ser un soldado. Por el contrario, cuando se interroga, “cómo es ser un murciélago para un murciélago”, no podría darse un contraste similar porque, dejando de lado la mitología y sus transformaciones extravagantes, ninguna criatura que no sea un murciélago puede ser un murciélago. O, con otro ejemplo, no parece correcto preguntar cómo es para un doctor ser un doctor como si alguien que no fuera ya un doctor pudiera serlo. Esto es, desde luego, absurdo, porque alguien que no fuera un doctor no puede ser, al mismo tiempo, un doctor (aun cuando pudiera llegar a serlo más adelante).

En tercer lugar, cuando preguntamos, en el lenguaje ordinario, como es para X ser Y, existe otro tipo de contraste, esta vez entre cómo es para X ser Y en oposición a cómo sería para X ser otra cosa distinta de Y, pero que, en principio, también podría haber sido. Por el contrario, tampoco parece correcto preguntar cómo es para un murciélago ser un murciélago por oposición a otra criatura que también podría haber sido, porque el murciélago no podría ser otra criatura que la que es (a menos que dejara de ser un murciélago). En todo caso, la única interpretación adecuada posible (i.e. gramaticalmente correcta) que podría hacerse de la pregunta nageliana consistiría en considerarla como una manera un tanto rebuscada de interrogar, indirectamente, cómo es ser Y (cómo es ser un murciélago o un ser humano), es decir, como una manera extraña de solicitar alguna caracterización de la vida de los murciélagos o algún tipo de cualificación sobre dicha existencia. En este punto, los murciélagos tal vez quisieran respondernos, hobbesianamente: “La vida de un murciélago es desagradable, brutal y corta”. Sin embargo, este tipo de preguntas pueden responderse perfectamente sin necesidad de ser un murciélago. Bastaría con contar con buena información acerca de la vida de estas criaturas (Hacker 2002: 170). De hecho, creo que ya contamos con una información tal.⁶⁰

⁶⁰ En efecto, sabemos que parte de la vida de los murciélagos consiste en vivir en una cueva o una gruta; y en las ciudades, en los campanarios, los edificios abandonados o los más estrechos taparrollos de las persianas. Sabemos que duermen durante el día colgando cabeza abajo, y que salen por la noche volando en búsqueda de alimento, cuyo menú, que varía en función del tipo de

En definitiva, de estas consideraciones se seguiría que “es equivocado [...] sugerir que ‘sabemos cómo que es [para nosotros] ser nosotros’, que hay algo ‘preciso que es [para nosotros] ser nosotros’” (Hacker 2002: 170). Por este motivo, concluye Hacker la afirmación (iii): no hay nada acerca de la experiencia consciente que sólo pueda ser aprehendido desde el punto de vista de la criatura en cuestión.

En resumen: la estrategia de Hacker podría ser interpretada de la siguiente manera: según el autor, quien pretenda defender la legitimidad del uso de la frase nageliana o bien debe aceptar que dicha frase expresa (de manera gramaticalmente inapropiada) algunas de las expresiones admitidas por el lenguaje ordinario (y entonces no es claro que haga referencia a un aspecto iluminador, característico o problemático acerca de la experiencia consciente), o bien, en la medida en que no coincide con ninguno de los sentidos habituales considerados, debería reconocer que carece de significado y que, por lo tanto, no refiere a ninguna propiedad *real* de la experiencia consciente.

3. La respuesta de Janzen a Hacker

Por su parte, el filósofo Greg Janzen (2011), sostiene, contra Hacker, que el hecho de que la frase nageliana no tenga un uso en el lenguaje cotidiano es completamente irrelevante para desacreditarla, precisamente porque, en el contexto de las discusiones en torno a la experiencia consciente, lo que se pretende recoger con ella es un concepto filosófico, no un concepto propio del lenguaje cotidiano.

Insistir con la idea de que un término o un concepto no debería ser utilizado de cierta manera porque no coincide con el significado o uso que tiene en el lenguaje ordinario, equivaldría a sugerir que puesto que muchos no-expertos en física no usan el término “agujero negro” del mismo modo en que lo hacen los expertos en física, entonces los últimos deberían abandonar el término en cuestión. Por lo pronto, esta última observación de Janzen me parece acertada. No parece una movida legítima prohibir el uso de una expresión por el sólo hecho de que no

murciélago del que se trate, puede incluir desde frutas hasta insectos. Al mismo tiempo, diversas investigaciones empíricas nos otorgan información acerca de lo que podría considerarse la estructura del mundo perceptivo y comportamental de los murciélagos. Así, sabemos, por ejemplo, que perciben el mundo externo mediante un sonar, o mediante la ecolocalización, detectando los reflejos de sus propios chillidos rápidos, sutilmente modulados y de alta frecuencia, reflejos que provienen de los objetos a su alcance. Sus cerebros están diseñados para correlacionar los impulsos emitidos con los ecos subsecuentes, y la información así adquirida les permite hacer determinaciones precisas de distancia, tamaño, forma, movimiento y textura.

coincide con ningún uso cotidiano de los términos en juego. De todos modos, el análisis de Hacker subraya algunas cuestiones que no siempre son puestas sobre la mesa en las discusiones filosóficas al respecto.

Por un lado, hace notar que la frase de Nagel no está siendo utilizada en un sentido corriente u ordinario, porque, repito, si así fuera, no haría referencia a un aspecto distintivo o problemático tal como supone quien la utiliza. Ahora bien, esta última constatación obliga, a quien pretenda utilizar la frase, a ofrecer algún tipo de elucidación acerca de su significado, porque éste no va de suyo. La necesidad de una clarificación conceptual se hace todavía más evidente cuando nos percatamos de que no todos los filósofos usan la frase en el mismo sentido. En este contexto, podría decirse que Janzen acepta el desafío de elucidar el concepto en cuestión y ofrece una caracterización “técnica” del mismo.

Sin embargo, mostraré que si se acepta esta definición, ya no resulta claro que el aspecto de nuestra vida mental al que se alude con ella, represente algún tipo de problema para una teoría fisicalista de la mente.⁶¹

Como dije al comienzo, de acuerdo con Janzen, al formular su célebre frase, Nagel estaría intentando, en última instancia, “identificar la propiedad (o las propiedades) que hacen a un estado mental consciente”, es decir, aquellas propiedades cuya posesión por parte de un estado mental resultan necesaria (y tal vez suficientes) para que dicho estado pueda ser considerado verdaderamente *consciente*. En otras palabras, de acuerdo con Janzen, la frase de Nagel podría ser parafraseada de la siguiente manera: necesariamente, un sujeto S está en un estado mental consciente M en t si y sólo si hay algo para S que es cómo estar en M en t (Janzen 2011: 273-274). Luego, a partir de una exégesis un tanto polémica de distintos textos de Nagel⁶², Janzen va a concluir que aquello que éste intenta decirnos con esta frase en última instancia es que “cuando un sujeto está en un

⁶¹ En este punto, cabe aclarar que Janzen (2008, 2011) no se propone explícitamente mostrar que la consciencia constituye un problema insalvable para una perspectiva fisicalista y funcionalista de la mente; su objetivo, en cambio, es ofrecer una elucidación del fenómeno en cuestión. Sin embargo, sus observaciones sí resultan relevantes para los objetivos que yo persigo en este capítulo, a saber: mostrar que no es claro que haya un aspecto problemático de la consciencia que no pueda ser abordado por una posición como la mencionada.

⁶² En esta discusión no me interesa evaluar la adecuación o inadecuación de la interpretación de Janzen de las ideas de Nagel. Mis objetivos están consignados en detalle al comienzo y al final de este capítulo. Con todo, Janzen no está sólo en esta interpretación. Véase, por ejemplo, Gallagher y Zahavi (2014: 89).

estado mental consciente, el sujeto, de algún modo y en algún grado, se *percata* [*aware*] de estar en él” (Janzen 2011: 281. La cursiva es del autor.).

En otras palabras, la idea sería que “necesariamente, un sujeto S está en un estado mental consciente M en t si y sólo si, en algún modo apropiado, se percata de estar en M en t” (Janzen 2011: 282). A juicio de Janzen, esta definición es una versión de un principio aceptado por muchos autores y al que el filósofo David Rosenthal llama “principio de transitividad” (a partir de ahora PDT). De acuerdo con este principio, “un estado mental consciente involucra que uno sea consciente [*aware*] de ese estado del modo apropiado” (Rosenthal 2000: 265). Para justificar esta interpretación, Janzen se apoya en parte en una suerte de analogía entre esta especie de “auto-consciencia” enunciada por el principio y la percepción de objetos externos.

Observa que los objetos, eventos, estados, etc., extra-mentales ordinarios, pueden ser como algo para nosotros también, en el sentido de que pueden verse, sentirse, etc., como algo. Una rosa, por ejemplo, podría verse roja, sentirse aterciopelada, etc. Incuestionablemente, sin embargo, un objeto (evento, etc.) ordinario no puede ser como algo para un sujeto –esto es, no puede verse, sentirse, etc., como algo para el sujeto- a menos que el sujeto se percate (o se haya percatado), de un modo u otro, de ese objeto. Para que una rosa se vea roja o se sienta aterciopelada, por ejemplo, uno tiene que percibir la rosa o percatarse de ella de algún modo. Por lo tanto, el ser como algo de un objeto ordinario para un sujeto presupone que el sujeto se percata (o se percató) de ese objeto. [Ahora bien, según Janzen, este principio puede extenderse a la experiencia consciente misma]: así como el ser como algo de un objeto ordinario para un sujeto presupone que el sujeto se percata de ese objeto, el ser como algo de un estado mental para un sujeto presupone que el sujeto se percata de ese estado mental (Janzen 2011: 283).

Con todo, de acuerdo con Janzen, el PDT se mantiene neutral en cuanto a la caracterización precisa del modo en que el sujeto se percata de sus propios estados mentales. En este sentido, el principio en cuestión no exige que todos los estados mentales conscientes involucren necesariamente una auto-percatación *reflexiva* (o *introspectiva*); ni tampoco que esta percatación sea ella misma auto-consciente. Lo único que dice es que una experiencia consciente es como algo para el sujeto de dicha experiencia sólo en caso de que el sujeto se percate, en algún sentido

apropiado a definir, de que está atravesando dicha experiencia. De esta manera, el único tipo de teorías que quedaría descartado si se acepta el PDT son las llamadas teorías de la consciencia de “primer orden”, como la que sostiene, por ejemplo, Michael Tye (1995), según las cuales, la caracterización de los estados mentales conscientes no se construye en términos de algún tipo de percatación de los mismos por parte del sujeto de la experiencia sino en términos de estados mentales que hacen conscientes al sujeto de algún objeto.⁶³

Con todo, en sus diferentes trabajos Janzen (2005, 2008, 2011) considera razonable exigir que el tipo de auto-percatación en juego cumpla con ciertos requisitos para resultar verdaderamente adecuada.

Es razonablemente claro que tendría que ser inmediato, i.e., no estar basado ni en la inferencia ni en la observación [...], puesto que hay casos en los cuales nos volvemos conscientes de un estado mental particular por medio de inferencias u observación, incluso cuando no consideramos al estado mental en sí mismo como consciente. Por ejemplo, S podría inferir que está enojado en base al testimonio de otra persona. Pero incluso si S está verdaderamente enojado por algo, su percatación [por esta medio] de que está enojado no convierte a su enojo consciente. Es también razonable claro que debería ser no-reflexivo y no-introspectivo puesto que, en general, [nuestras experiencias conscientes no son] postuladas [*posit*] como objetos [...]. Si huelo una rosa, por ejemplo, puedo no reflexionar acerca de mi experiencia olfatoria presente, pero mi experiencia de todos modos es como algo para mí. Es considerablemente menos claro, sin embargo, si la auto-consciencia debería ser construida como ‘extrínseca’ o ‘intrínseca’ a los estados mentales conscientes (Janzen 2005: 730).⁶⁴

En resumen, de acuerdo con Janzen el tipo de acceso adecuado a los propios estados mentales debe ser, por un lado, (i) no inferencial o directo (esto es, al parecer, no debe estar basado en algún tipo de inferencia a partir de la observación

⁶³ Ahora bien, salvo esta restricción, el PDT resulta perfectamente compatible con las llamadas teorías de la consciencia de “segundo orden”, en donde un estado se hace consciente en virtud mantener una cierta relación apropiada con otro estado mental que, en principio, podría no ser él mismo objeto de algún tipo de percatación consciente. Al mismo tiempo, también resultaría compatible con aquellas posturas que postulan algún tipo de percatación “tácita” o “implícita”, más afín a lo que algunos denominan “teorías auto-representacionales de la consciencia”. Volveré sobre estas cuestiones en el cuerpo del texto.

⁶⁴ Unas exigencias similares en torno al tipo de auto-percatación adecuada aparecen en Janzen 2011: 282 y Janzen 2008, Cap. 4.

directa de la conducta propia –como cuando me miro al espejo y digo “¡Qué cansado estoy!”-, o como cuando, como sugiere Janzen, otra persona me hace notar que estoy celoso. Por otro lado, el tipo de auto-percatación adecuado debe ser (ii) no-reflexivo ni introspectivo, es decir, no se trataría de algún tipo de consideración deliberada en torno a los propios estados mentales.⁶⁵ En este sentido, quedarían excluidos como ejemplos pertinentes del tipo de auto-percatación en cuestión casos como aquellos en los que reflexiono acerca de cierto estado o proceso mental propio con posterioridad a su ocurrencia, o aquellos en los cuales una persona reflexiona acerca de sus emociones, de los motivos profundos que guían su accionar, de sus compromisos, sus rasgos del carácter, destrezas, disposiciones, fortalezas y debilidades, y obtiene esta información ya sea por intermedio del testimonio de otras personas (incluyendo analistas), o bien apelando a su memoria en donde busca, por ejemplo, algún tipo de patrón al que obedezca su comportamiento.

En tercer lugar, Jansen reclama, en 2005 y 2008, además, que (iii) una teoría apropiada de la consciencia debe hacer lugar a una distinción entre el conocimiento en primera persona (no inferencial) y en tercera persona (mediado) de los estados mentales (Janzen 2005: 715-716). En este sentido, si bien el autor reconoce que hay muchas cuestiones complicadas en torno a la noción de “conocimiento de primera persona” (como por ejemplo, acerca de cómo se adquiere y en qué se diferencia del conocimiento en tercera persona), considera que no es necesario responder estos interrogantes para sostener que, intuitivamente, un estado es consciente sólo en caso de que sea cognoscible en primera persona. Todo lo que se requiere es que *haya* una distinción entre el conocimiento en primera y en tercera persona, dado que el hecho de que la diferencia existe es “muy indiscutible” (Janzen 2008: 42-43; 2005: 715-716. Las cursivas son del autor).

Si pienso acerca del cielo azul, el modo en que *yo* sé que estoy pensando acerca del cielo azul es muy diferente del modo en que *vos* sabés que estoy pensando

⁶⁵ En efecto, de acuerdo con Janzen, “los actos de introspección [que habría que rechazar como característicos del tipo de auto-percatación en cuestión] son deliberados, moderadamente esforzados, y generalmente no demasiado prolongados, mientras que la auto-percatación implícita es un elemento constante, [que se da] sin esfuerzo, y [de manera] no deliberada en todos los estados mentales conscientes. Uno no puede interrumpir o detener su percatación implícita del propio acto de ver, escuchar, oler. Además, [...] la auto-percatación implícita es una condición necesaria para los estados conscientes, mientras que la introspección no” (Janzen 2008: 106).

acerca del cielo azul. A tu conocimiento de que estoy pensando acerca del cielo azul se arriba *inferencialmente*, mientras que a mi conocimiento de que estoy pensando acerca del cielo azul no. Dicho de otra forma, tu conocimiento de que estoy pensando acerca del cielo azul, a diferencia de mi conocimiento de que estoy pensando en el cielo azul, es *mediado*" (Janzen 2008: 43. Las cursivas son del autor).

Ahora bien, teniendo en cuenta estas demandas a la hora de caracterizar el tipo de auto-percatación adecuada, mi punto consistirá en señalar, sin cuestionar la legitimidad de las mismas, que es posible satisfacerlas sin comprometerse con una postura incompatible con el fisicalismo / funcionalismo. Si esto es correcto, o bien debería concluirse que el aspecto cualitativo de la experiencia no resulta problemático después de todo, (a pesar de lo que Nagel declara explícitamente en su artículo de 1974) o bien que todavía no encontramos (ni Hacker ni tampoco Janzen) una definición apropiada del mismo. Pero optar por este último cuerno del dilema abona la sospecha de Dennett (1988/2003) de que no está para nada claro a qué se refieren los filósofos cuando hablan acerca del "aspecto cualitativo" o "subjetivo" de la experiencia.

4. La propuesta de Janzen: ¿la consciencia explicada?

Una de las teorías estándar dentro de la tradición de la filosofía analítica que puede considerarse que se ocupa de la "consciencia fenoménica" son las teorías de "segundo orden". De acuerdo con dichas teorías, lo que hace que un estado mental sea consciente es el hecho de que el sujeto se percata de él en un sentido apropiado. Y es solamente cuando esto ocurre que el estado en cuestión se vuelve consciente.

En otras palabras, de acuerdo con este enfoque un estado mental M de un sujeto S es consciente cuando, y sólo cuando, S tiene otro estado mental M* tal que M* es una representación apropiada de M. El hecho de que M* represente M asegura que "haya algo que es como tener M para S" (Kriegel 2007: 46). Además, de acuerdo con esta perspectiva, M es numéricamente distinto de M*. Las distintas versiones de este tipo de teorías difieren en el modo en que caracterizan el proceso de monitoreo que da lugar a estas representaciones y en la manera en que conciben las representaciones mismas.

En el caso de la teoría de Rosenthal en particular, el estado de segundo orden es una especie de *pensamiento* acerca de otro estado mental que tiene por “objeto”. Por mi parte, entre algunos de los aspectos centrales de esta teoría quisiera destacar aquellos que muestran que cumple con las exigencias del PDT. En primer lugar, como acabo de indicar, el tipo de auto-percatación presente en esta perspectiva *no* es de tipo perceptual ni inferencial: no se basa ni en la observación, por parte del agente, de su conducta, ni en la información que otras personas pudieran brindarle acerca de sus estados mentales. En otras palabras, el pensamiento de segundo orden no es el producto de algún tipo de inferencia a partir del estado de primer orden. En particular, se supone que el pensamiento en cuestión se forma a través de algún mecanismo de procesamiento de información, pero este proceso es automático e inconsciente. Así, a partir de este rasgo se pretende reflejar “la *inmediatez*, o al menos la *inmediatez sentida*, de nuestra percatación de nuestros estados mentales” (Kriegel 2007: 47. La cursiva es del autor).

Por último, es claro que este tipo de “auto-consciencia” propuesto no es un tipo de percatación reflexiva. En efecto, de acuerdo con esta teoría es cierto que en ocasiones podemos reflexionar o introspectar deliberadamente nuestros propios estados mentales. En estos episodios, se distinguen, de acuerdo con el presente enfoque, tres estados separados: la experiencia de primer orden (mi percepción del cielo por ejemplo), mi percatación de segundo orden de dicha experiencia, y una representación de tercer orden de la percatación de segundo orden. Con todo, desde esta perspectiva no es necesario que se den estos tres estados para que el estado de primer orden se vuelva consciente. En otras palabras, no se pretende sugerir que el tipo de percatación relevante sea una percatación de tipo reflexiva. Así, cuando dejo de introspectar y vuelvo mi atención hacia el objeto de la experiencia de primer orden (el cielo azul), el estado de tercer orden se evapora y el de segundo orden se torna inconsciente. En este sentido, la teoría sostiene que el estado que vuelve consciente al estado de primer orden es en sí mismo un estado que puede permanecer no consciente. De hecho, esto es lo que sucedería normalmente. El único requisito que no cumple este tipo de teorías es aquel que Janzen señalaba, en cierto modo, como optativo, a saber, el del carácter intrínseco de la auto-consciencia con respecto al estado mental en cuestión.

En efecto, a diferencia de las llamadas teorías auto-representacionales de la consciencia, de acuerdo con las cuales un estado mental es consciente si y sólo se *representa a sí mismo* de modo apropiado, en las teorías de la consciencia de segundo orden que estuvimos analizando, el estado mental de segundo orden por medio del cual el estado de primer orden se vuelve consciente es externo o “extrínseco” a este último. Con todo, mostraré que es posible cumplir también con este último requisito sin salirse de los límites que impone una perspectiva fisicalista / funcionalista de la mente.

Para eso, puede resultar provechoso retomar el planteo del filósofo Andy Clark (2000). En ese artículo, nos pide considerar un sistema u organismo con la capacidad de detectar, no sólo una variedad de diferencias entre distintos estímulos (por ejemplo, una diferencia de color entre dos estímulos presentados visualmente), sino también con la capacidad de detectar, o de acceder a, ciertos aspectos de la actividad sensorial o perceptiva misma, es decir, con la capacidad de percatarse no inferencialmente de qué modalidad perceptiva está usando cuando percibe un estímulo dado.

En otras palabras, Clark nos pide concebir un organismo o sistema que no sólo se percata de la diferencia entre los estímulos sino que también se percata de que *ve*, en lugar de, *escuchar* u *oler*, la diferencia entre ellos. Por otro lado, y este punto es importante, Clark sostiene que este acceso forma parte de lo que Ned Block categorizó como “consciencia de acceso”, que es un tipo de acceso que puede ser perfectamente explicado dentro de un modelo funcionalista de la mente.⁶⁶ Así, desde esta perspectiva,

⁶⁶ El filósofo Ned Block (1997) popularizó una distinción entre dos tipos de consciencias cualitativamente diversas: la *consciencia de acceso* y la *consciencia fenoménica*. De acuerdo con esta distinción, se dice que un estado es consciente en el primer sentido si su contenido está dispuesto o listo (*poised*) para su uso libre en el control del pensamiento y la acción. Más específicamente, un estado con cierto contenido es consciente en este sentido si, en virtud de tenerlo, una representación con ese contenido está (1) dispuesta para ser usada libremente como premisa en un razonamiento, (2) dispuesta para ser usada libremente para el control de la acción y, en el caso de los organismos usuarios del lenguaje, (3) para dar origen a un reporte verbal sobre dicho contenido. Con respecto al segundo concepto de consciencia, la consciencia fenoménica, Block sostiene que es imposible definirlo de una manera no circular. Por esta razón, una posibilidad de acercarnos al significado de este concepto es “señalando” los estados mentales que se consideran casos paradigmáticos de estados fenoménicamente conscientes: básicamente, se recurre siempre en este punto a las sensaciones corporales y las experiencias perceptivas. En última instancia, el concepto hace referencia al rasgo problemático de la consciencia que viene siendo objeto de crítica en este trabajo: su aspecto “cualitativo”. Nuevamente, la parte más controversial de este último concepto de consciencia es que referiría a un fenómeno que se distingue de toda capacidad cognitiva que

en lugar de enfocarse [...] en la noción de disponibilidad generalizada y sostenida para el reporte y el control, [se] hace hincapié en el [...] rol de un tipo específico de reporte y de acceso, es decir, acceso a los *actos de detección* en lugar de [el acceso a] la información acerca de los *estados de cosas detectados*. [...] El conocimiento acerca de los actos de detección es un fenómeno de ‘primer orden’, a la par con el conocimiento acerca del mundo. No es estrictamente un conocimiento acerca del conocimiento, o pensamiento acerca del pensamiento. La reportabilidad verbal del acto de detección tampoco es necesaria para, ni constituye parte de, la presencia de tal acceso. Lo que importa es solamente que el acto por medio del cual la diferencia es detectada esté marcado no inferencialmente, para la criatura en cuestión, como involucrando (por ejemplo) una modalidad visual” (Clark 2000: 31).

Más específicamente, Clark considera que un modo adecuado de caracterizar este acceso podría hacer hincapié en los diferentes modos en que los contenidos de las experiencias pueden estar dispuestos [*poised*] para el control de acciones, de manera tal que aquello que “marca” a una información como perteneciendo a una modalidad en lugar de a otra es precisamente el modo en que está dispuesta o lista para guiar la acción. Así, Clark sostiene que, “aquello a lo que tenemos acceso cuando tenemos acceso a la modalidad involucrada en un acto de detección es a la batería específica de destrezas que podríamos haber desplegado. En la medida en que los conjuntos de destrezas difieren de acuerdo a la modalidad involucrada [...], el acceso a los conjuntos de destrezas que podrían haber sido desplegadas constituiría el acceso directo no-inferencial a la modalidad en uso [...]” (Clark: 2000: 35).

Como puede verse, la posición que esboza Clark da cuenta de un tipo de auto-consciencia o de acceso a los propios estados mentales (en este caso, a los estados mentales perceptivos), que cumple con los requisitos de la no inferencialidad y no reflexibilidad que exigía Janzen y, al mismo tiempo, y a diferencia de lo que sucedía con las teorías de “segundo orden”, cumple también con el requisito de la “intrinsicidad”. En efecto, con respecto a la no inferencialidad, Clark sostiene que en este tipo de auto-percatación el acceso por parte del agente a su propia actividad

involucre el pensamiento, de todo rasgo intencional (es decir, de la propiedad de un estado de ser *acerca* de algo), y de toda propiedad funcional.

perceptiva no es generado por ningún tipo de inferencia consciente o de nivel personal, como sucedería si razonara de la siguiente manera: “si cierro los ojos, dejo de percibir el objeto, por lo tanto, lo debo estar percibiendo visualmente”. Al mismo tiempo, en este modelo el mismo acto por el cual el sujeto percibe un objeto externo es aquel por el cual se percata, más o menos implícitamente, de su propio estado perceptivo, en la medida en que puede detectar qué modalidad perceptiva se está poniendo en juego en cada caso.

En un artículo posterior escrito en colaboración con otros autores, Clark retoma algunas de las ideas expuestas hasta acá. En esta nueva intervención, los autores desarrollan un modelo de la percepción visual consciente al que llaman “el modelo del espacio de acción” (*action-space model*). De acuerdo con este modelo, la experiencia perceptiva visual está constituida por el conocimiento directo, inmediato, implícito y práctico (es decir, no conceptual ni proposicional), por parte del agente, de los modos en los cuales está actualmente dispuesto (o, más precisamente, los modos en que implícitamente considera estar dispuesto) sobre un “espacio de acción”. Un espacio de acción, en este contexto, debe ser entendido como una matriz de posibilidades para la persecución y cumplimiento de las acciones intencionales, objetivos y proyectos (Ward, Roberts, Clark 2011: 383). Así, desde esta perspectiva, se establece una conexión estrecha entre la experiencia perceptiva y la conducta, en la medida en que la primera está constituida, al menos en parte, por la percatación de las oportunidades para la acción, de la “apreciación perceptual de qué puede ser hecho dados los constituyentes de la escena presente” (Ward, Roberts, Clark 2011: 383).⁶⁷

Ahora bien, como puede verse, otra vez estaríamos en presencia de un modelo según el cual la percepción supone (o mejor, está *constituida* por) un acceso no inferencial, por parte del sujeto de la experiencia, a un rango de destrezas (*skills*) o acciones habilitadas por el entorno presente. De este modo, podría decirse, otra vez, que en la medida en que los conjuntos de destrezas o acciones difieren de acuerdo a la modalidad involucrada, el acceso a los mismos constituiría el acceso directo y no-inferencial a la modalidad en uso en dicho acto perceptivo. Ahora bien,

⁶⁷ Por lo demás, de acuerdo con los autores, este conocimiento no necesita ser verbalizado ni de algún modo explícito: “Simplemente sabemos, con precisión razonable, qué clases de objetivos o proyectos nuestro contacto visual actual con el mundo nos permite llevar a cabo” (Ward, Roberts, Clark 2011: 383-384).

el lugar dentro de un “espacio de acción” que de algún modo capta el sujeto de la experiencia es parte del *contenido intencional* de la experiencia perceptiva, que presenta al mundo como una “arena” para la acción intencional (Ward, Roberts y Clark 2011:388). De hecho, Clark y compañía afirman, explícitamente, que podemos ver este estado como si el sujeto estuviera ocupando un estado *representacional* orientado a la acción (Ward, Roberts y Clark 2011: 388).⁶⁸ Pero el aspecto representacional o intencional de los estados mentales es precisamente aquel que se supone que puede ser incorporado dentro de un proyecto de “naturalización” de lo mental, es decir, dentro de un proyecto que podría dar cuenta de él en términos compatibles con los principios y entidades aceptadas por (alguna de) las ciencias naturales contemporáneas.⁶⁹ Si esto es así, habría que aceptar, o bien que no hay un problema específico con respecto a la consciencia, o bien que si existe un problema

⁶⁸ Al mismo tiempo, los autores reconocen la afinidad de su enfoque con un análisis gibsoniano de la percepción en términos de *affordances*. De acuerdo con esta perspectiva, sólo puedo percibir una silla como una silla en virtud de comprender (de alguna manera) qué acciones la silla permite (*afford*) (sentarse en ella; ser colocada debajo de escritorios y mesas, etc.).

⁶⁹ De hecho, esta es la apuesta de las teorías representacionistas de la consciencia, según las cuales, “todos los hechos mentales son hechos representacionales” (Dretske 1995: xiii). En principio, si bien la noción de “representación” en juego no resulta siempre del todo clara, parece legítimo suponer que los representacionistas identifican “representación” con *intencionalidad* (Janzen 2008: 138). Es decir, la idea es que en la medida en que nuestras experiencias conscientes son experiencias *de* algo o *acerca* de algo (de un objeto, de una propiedad o de un evento), *representan* esos objetos, propiedades o eventos. Por otra parte, se supone que dichas experiencias conscientes son estados físicos (presumiblemente neurológicos) que portan información acerca de varias propiedades perceptibles externas como el tamaño, la forma, el color o la textura de un objeto. Más específicamente, la idea es que en nuestro cerebro tenemos *vehículos* representacionales que “representan” o “indican” ciertos contenidos, es decir, cosas tales como la presencia del color azul o de algo redondo, por ejemplo. Sin embargo, esto no significa que haya necesariamente algo azul o redondo en el cerebro, es decir, algo que deba poseer estas propiedades representadas. Por otra parte, el modo específico en que se interprete la idea de que un vehículo “indica” o “representa” algo, dependerá de la formulación particular de la teoría representacional que estemos considerando. Dretske, por ejemplo, opta por una semántica “teleo-evolutiva”, que sostiene que el concepto relevante para explicar la representación es el de “función propia”, mientras que Tye opta por las nociones de “correlación” o “covariación causal” *bajo ciertas circunstancias óptimas*. En el caso de Dretske, la noción de información en juego no supone ningún proceso de interpretación por parte de un agente cognitivo, sino que se define en términos de relaciones nómicas entre eventos, por lo tanto, es una noción que puede ser utilizada para explicar la representación mental dentro de un proyecto naturalista y científicista. Ahora bien, para que ciertos estados cerebrales, además de portar información acerca de ciertos hechos del mundo, los representen, se requiere, según Dretske, que ésa sea su *función*. En efecto, así como sería legítimo decir que un velocímetro S representa (y no sólo porta información acerca de) la velocidad de un auto A porque esa es la función o el propósito con el que se lo diseñó, para Dretske también resultaría legítimo decir que cierto estado cerebral representa ciertos hechos o propiedades del mundo si esa es la función que dicho estado adquirió mediante cierta historia evolutiva. Dicho todo esto, debe aclararse que el tipo de representación o contenido representacional que debería considerarse como apropiado para satisfacer las demandas de Janzen, no es el que proponen ni Dretske ni Tye (por las razones apuntadas en el cuerpo del texto en torno a las limitaciones que Janzen observa en los enfoques de “primer orden” de la consciencia), sino uno que esté en sintonía con las propuestas consideradas de Clark o Noë.

en torno a ella, éste se identifica con el problema de explicar el aspecto intencional o representacional de los estados mentales.

Por último, podemos mencionar algunas observaciones afines a las de Clark hechas por el filósofo Alva Noë (2002, 2003) para reforzar mi punto. En el contexto de una revisión de la teoría causal de la percepción, Noë señala que cómo las cosas (los objetos del mundo externo) se nos “aparecen” o cómo se ven (en el caso de la percepción visual), depende no sólo de cómo las cosas sean verdaderamente sino también de la relación del perceptor con cómo las cosas son. “Hay aspectos del contenido perceptual –aspectos de cómo la experiencia presenta cómo son las cosas- que no están determinados únicamente por cómo son las cosas, sino que dependen, además, de la relación del perceptor con cómo las cosas son” (Noë 2003: 94).

En otras palabras, de acuerdo con esta idea, no sólo experimentamos cómo son las cosas, sino también como se ven “desde acá”, desde cierta perspectiva o punto de vista. Así, experimentamos, por ejemplo, que el plato es redondo y que luce elíptico “desde acá”. Su aspecto (o “aparecer”) elíptico “desde acá” es una propiedad genuina del plato según Noë, pero es también una propiedad relacional, una que depende de dónde se sitúe el “acá”. Ahora bien, para Noë, las propiedades perspectivas (como el aspecto elíptico del plato) forman parte de lo que muchos filósofos considerarían como parte del contenido *representacional* de la experiencia.

Por otra parte, la forma perspectiva captada del plato (o, para el caso, de cualquier objeto dentro del mundo tridimensional en el que habitamos), cambia en función del movimiento del perceptor con respecto al plato y de la posición del plato con respecto al perceptor. De acuerdo con Noë, resulta plausible pensar que nuestra captación⁷⁰ tácita del modo en que cambian las propiedades perspectivas cuando el

⁷⁰ La caracterización de esta “captación” o “saber” es motivo de controversia dentro del enfoque de Noë. En principio, podríamos decir que el conocimiento en cuestión no debe confundirse con un juicio explícito deliberado. Sin embargo, como señala Silverman (2017), permanece abierto a dos interpretaciones distintas. De acuerdo con una de ellas, el conocimiento de las leyes o “contingencias sensoriomotoras” que posee el perceptor para constituirse como tal es una suerte de capacidad o destreza corporal. De esta manera, la percepción no sería otra cosa que la posesión y el ejercicio de estas capacidades corporales. En otras palabras, de acuerdo con esta interpretación, la percepción involucraría un tipo de saber cómo. De acuerdo con la segunda interpretación, en cambio, el conocimiento involucrado en la percepción supondría o estaría constituido por la presencia de una *representación* interna (ubicada en el cerebro) de las contingencias sensoriomotoras (Silverman 2017:2). En función de los fines que persigo en este trabajo, cualquiera de las dos interpretaciones que se considere resultará adecuada. Si se interpreta como un estado representacional, lo será por el tipo de consideraciones mencionadas en el cuerpo del texto con respecto a la viabilidad del proyecto

organismo se mueve (cuando mueve, por ejemplo, sus ojos o su cabeza), es precisamente el modo en que logramos experimentar las propiedades no perspectivas de los objetos (en el caso del plato, su redondez) (Noë 2003: 94-95). De este modo, podría decirse que lo que algunos filósofos llaman contenido representacional de una experiencia, es decir, el modo cómo la experiencia presenta al mundo, tiene dos dimensiones: una *factual*, con respecto a cómo las cosas son, y otra *perspectiva*, con respecto a cómo las cosas se ven o aparecen desde el punto de vista del perceptor. Según Noë, la experiencia visual siempre presenta ambas dimensiones del contenido (Noë 2003: 95).

Ahora bien, si esto es así, podríamos decir nuevamente que, desde este tipo de enfoque, la auto-percatación por parte del agente de sus estados mentales forma parte del mismo acto por el cual se perciben los objetos del mundo externo. En efecto, dado que desde esta perspectiva la percepción está constituida, al menos en parte, de un cierto conocimiento tácito, por parte del sujeto de la percepción, de las regularidades que rigen los cambios en las propiedades perspectivas y que estas regularidades varían en función de las características y ubicación de los objetos externos, pero también de *la modalidad perceptiva que esté en juego*⁷¹, se sigue que percibir un objeto supone comprender tácitamente la modalidad perceptiva con la

de “naturalizar” el aspecto representacional o intencional de lo mental. Si se opta por la segunda interpretación, tampoco nos enfrentamos con una amenaza para una perspectiva fisicalista de la mente. En efecto, para decirlo brevemente, en este caso no se estaría adoptando ningún tipo de dualismo ontológico, ni se le estaría atribuyendo a la percepción consciente alguna propiedad que pueda resultar problemática, como la intrinsecalidad o la privacidad. Por último, si bien O’Regan y Noë 2001 consideran que la percepción es inefable, (y la inefabilidad, como vimos, podría considerarse en principio como un rasgo problemático de la experiencia perceptiva), el tipo específico de inefabilidad en juego, tampoco representa un problema: en principio, de acuerdo con los autores, no hay nada misterioso acerca de la incapacidad de poder dar una expresión verbal completa de los rasgos de la percepción. “Es muy difícil describir todo lo que hacemos cuando vemos, del mismo modo en que es difícil describir todo lo que hacemos cuando estamos involucrados en actividades [...] como [la realización de] algún esfuerzo atlético, tocando un instrumento o hablando un lenguaje. La mayor parte de nuestro dominio de las contingencias sensoriomotoras [dominio que forma parte del acto perceptivo] toma la forma de un saber cómo [*know how*]. [...] En general, la capacidad de saber cómo hacer algo no viene acompañada con la capacidad de reflexionar acerca de lo que uno está haciendo cuando ejerce la capacidad en cuestión” (O’Regan y Noë 2001: 962), ni tampoco, podríamos agregar, con la capacidad de expresar verbalmente todo lo que sabemos hacer. Sin embargo, este hecho por sí solo no constituye una prueba de que la experiencia posea un carácter particularmente especial dentro del orden natural del mundo.

⁷¹ Como ejemplos de las distintas regularidades propias de cada modalidad perceptiva podemos señalar la “ley” que determina que un objeto se vuelve más grande en el campo visual a medida que nos acercamos a él, la que determina que un sonido aumenta cuando nos aproximamos a la fuente que lo produce o que la suavidad de una esponja está constituida, entre otros, por el hecho de que cuando la aprieto, cede ante la presión.

cual se lo percibe, precisamente porque supone comprender las regularidades que rigen específicamente cada modalidad en cuestión.

Por último, con respecto al tercer requisito que habíamos mencionado que exigía Janzen para cualquier propuesta aceptable, según el cual, una teoría apropiada de la consciencia debe hacer lugar a una distinción entre el conocimiento en primera persona (no inferencial) y en tercera persona (mediado) de los estados mentales (Janzen 2005: 715-716), creo que existen distintas maneras de satisfacerlo. En el capítulo 3 y 4 di cuenta de estos diferentes “accesos” a partir de cierta caracterización particular de los conceptos que forman parte de las atribuciones psicológicas en primera y en tercera persona. Ahora bien, de acuerdo con lo dicho en este capítulo, podría decirse que, en las propuestas consideradas de Clark y Noë, toda persona tiene un acceso directo y no inferencial (en los sentidos apuntados) a sus propios estados mentales, entendidos ahora como estados representacionales. En cambio, el resto de las personas no cuenta con ese acceso y debe inferirlos a partir de la evidencia que pueda brindarle la observación de la conducta ajena (sea o no verbal).

Sin embargo, siempre cabe la posibilidad de que una persona no exteriorice sus estados mentales a través de su conducta o que no sea sincero en sus manifestaciones. En este sentido, el acceso en tercera persona siempre tiene el estatus de una conjetura que se realiza a partir de la observación directa de un *efecto* conductual de los estados mentales ajenos. Por esta razón, dicho acceso puede ser calificado como inferencial y mediado. En principio, puesto que Janzen no ofrece una caracterización demasiado detallada del modo en que el tercer requisito que postula debería ser satisfecho, estos pocos elementos mencionados bastarán para plantear el esbozo de una respuesta a su demanda.

5. ¡No hay nada como ser un murciélago!

A lo largo de este último capítulo intenté mostrar, hasta ahora, que existen serias dificultades a la hora de intentar caracterizar e identificar el fenómeno al que se alude con la célebre frase “hay algo como es estar en cierto estado mental” y que generaría el problema difícil de la consciencia.

En particular, Hacker nos muestra que la frase en cuestión no debería interpretarse en el sentido con el que podría utilizarse dentro del lenguaje ordinario

porque, de ser así, no se estaría señalando un aspecto característico, distintivo o especialmente problemático de los diferentes tipos de experiencias conscientes tal como pretenden quienes la utilizan dentro de las discusiones filosóficas relevantes.

En efecto, si con esta frase se quiere hacer referencia a la “sensación cualitativa” de la experiencia, y se entiende por ella una cierta apreciación o valoración de las mismas por parte del agente que las atraviesa (como cuando se dice que tal o cual experiencia es agradable o espeluznante), debemos aceptar que no todas las experiencias poseen este “aspecto cualitativo” (después de todo ¿cómo se siente ver de una lámpara o un buzón?) y que diferentes experiencias generan la misma sensación cualitativa (después de todo, puede ser agradable tener una cierta experiencia auditiva como una gustativa). De este modo, este rasgo no nos permitiría identificar un aspecto distintivo ni característico de las experiencias conscientes.

Por otra parte, si por “aspecto cualitativo” de la experiencia se pretende hacer referencia a cosas tales como “el aroma del pan recién horneado” o “el intenso color púrpura de la tapa de un libro en un estante”, podría argumentarse que estas son propiedades cualitativas de los *objetos* que se experimentan (el pan o el libro), pero no de las experiencias en sí mismas. Por lo tanto, tampoco podrían servir para caracterizarlas.

Por otra parte, si a propósito de las observaciones de Hacker, se acepta, siguiendo a Janzen, que la frase nageliana no debe ser interpretada en ninguno de estos sentidos cotidianos apuntados, sino como una expresión que haría referencia a un modo peculiar en que el agente se percata de sus propios estados mentales, puede probarse, como hice en la última sección, que es factible ofrecer una caracterización de esta auto-percatación que satisfaga los requisitos que Janzen sostiene que debería satisfacer (es decir, el carácter no inferencial, no reflexivo y no introspectivo), sin desafiar los compromisos metodológicos y metafísicos de una perspectiva fisicalista de la mente.

En última instancia, todo este análisis apoya la conclusión de que, o bien no resulta del todo inteligible de qué estamos hablando cuando hablamos del “aspecto cualitativo” de la experiencia, o bien que, si resulta inteligible, no aparece como un aspecto especialmente problemático de nuestra vida mental.

Desde luego, en este punto el partidario de los *qualia* y los misterios podría replicar que ninguna de las interpretaciones mencionadas en este capítulo (ni la de Hacker, ni la de Janzen), responde adecuadamente al significado de la expresión de Nagel. Si opta por esta opción, sin embargo, no hace otra cosa que confirmar las sospechas que mencioné con anterioridad: que no es del todo claro de qué hablamos cuando hablamos de “cómo es estar en determinado estado mental”. Por lo demás, hasta que no se ofrezca una caracterización más o menos definida del concepto en disputa, o se señalen las deficiencias de las definiciones propuestas, podremos mantenernos firmes en nuestra sospecha de que no existe una idea clara de qué es aquello que una teoría fisicalista de la mente dejaría de lado necesariamente.

6. La prestidigitación explicada: desde la (di) solución del “problema” de la consciencia hacia la apertura de nuevos horizontes

¿Cómo es posible que el agregado de miles y miles de millones de células de tipos diferentes que nos conforman y cuyo grado de consciencia individual no supera al de una levadura, genere o de lugar la experiencia subjetiva? Durante este trabajo sostuve que la dificultad a la que se pretende hacer alusión con preguntas como esta radica, en gran medida, en que su respuesta parece requerir una elucidación de la relación entre dos ámbitos que, a primera vista, lucen como ámbitos completamente heterogéneos.

En efecto, se hace referencia a la consciencia inocentemente como “lo mejor conocido”, como si fuera un objeto o un proceso interno y privado, o como si realmente supiéramos con claridad a qué nos referimos cuando la consideramos como un residuo que permanece inexplicado una vez despejados otros fenómenos concomitantes. Pero entonces el truco del prestidigitador ya fue llevado a cabo. Y la explicación de la consciencia se convierte, no meramente en un “problema difícil”, sino en un arcano impenetrable por principio. En este contexto, sostuve, (siguiendo en buena medida los aportes de Dennett 1988, Pérez 2002 y Ward 2006, entre otros), que gran parte de la “magia” y el “misterio” alrededor de la consciencia radica en que se le atribuyen a este fenómeno, de forma más o menos implícita, ciertas propiedades que resultan incompatibles con las propiedades admitidas por las posiciones fisicalistas ortodoxas actuales acerca de lo mental y las ciencias naturales

contemporáneas⁷²: entre esas propiedades identifiqué a la intrinsecalidad, la infabilidad, y el carácter privado y directo (y tal vez infalible) del conocimiento acerca de los propios estados mentales conscientes.

Frente a este panorama, busqué probar precisamente que hay buenas razones para creer que algunos de estos rasgos (como la intrinsecalidad o la privacidad y el acceso privado y directo), a pesar de lo que puede parecer a primera vista, no son propiedades reales de la consciencia ni de nuestro acceso epistémico a la misma. Con respecto a la infabilidad, si bien reconocí que hay un sentido en el que las experiencias conscientes son infalibles, argumenté que este hecho no representa un auténtico problema para las teorías fisicalistas mencionadas.

En este sentido, mis esfuerzos estuvieron destinados a promover una apuesta arriesgada y polémica: en efecto, mi tesis central puede sintetizarse diciendo que, planteado en estos términos, el problema de la consciencia es, en realidad, *un falso problema*, porque no hay ningún aspecto de nuestra vida mental que posea rasgos verdaderamente problemáticos en principio. En este marco, mi intención consistió, por un lado, en promover, mediante diversas estrategias argumentativas, el rechazo de cierta *caracterización* inadecuada pero bastante extendida de la consciencia. Al mismo tiempo, me propuse esclarecer las razones que podrían llevar a adoptarla con la esperanza de que, una vez descubiertos los “trucos” involucrados, buena parte del misterio se disipe. En estas páginas finales voy a recapitular los resultados obtenidos en los capítulos precedentes.

*

En el **capítulo 1** me ocupé de rechazar las razones por las cuales se podría considerar que la experiencia consciente tiene propiedades intrínsecas. De acuerdo con esta idea, si se considera que la intrinsecalidad es una de las propiedades definitorias de la consciencia, este fenómeno se convierte en un verdadero problema para las teorías fisicalistas ortodoxas acerca de la mente. En efecto, si la consciencia cuenta con propiedades intrínsecas, entonces parece imposible que estas propiedades puedan ser explicadas en términos aceptables para el funcionalismo, es decir, en términos de las interacciones causales entre *inputs* (estímulos

⁷² En realidad, tal como señalé oportunamente, estas propiedades también la convertirían en un obstáculo insalvable para posiciones no ortodoxas como el enfoque sensoriomotor.

sensoriales), *outputs* (respuestas conductuales) y otros estados mentales, y en términos de las relaciones causales de éstos últimos con objetos del entorno, porque si hay aspectos intrínsecos en los estados conscientes, no pueden ser reducidos a relaciones de ningún tipo.

Ahora bien, si la experiencia consciente es intrínseca y no física, debería ser metafísicamente posible y, en principio, concebible, que haya dos individuos idénticos en cuanto a sus propiedades físicas y sus relaciones con un entorno también idéntico, pero que difirieran en sus propiedades cualitativas e intrínsecas. En otras palabras, si la consciencia contara con propiedades intrínsecas y no físicas, entonces debería ser concebible y metafísicamente posible un “escenario zombi”, es decir, un escenario física y funcionalmente idéntico a nuestro mundo pero en el cual no hay experiencia consciente. Ahora bien, si se acepta este planteo, debería admitirse que si un escenario como el descrito no fuera concebible, entonces se pone en duda que se trate efectivamente de una posibilidad metafísica y, en última instancia, que la experiencia consciente cuente verdaderamente con propiedades intrínsecas y no físicas.

Precisamente, en el capítulo 1 me ocupé de mostrar que, a pesar de lo que puede parecer a primera vista, los zombis no son claramente concebibles. Más específicamente, a partir de un examen pormenorizado de la conexión entre las experiencias conscientes y ciertos estados mentales en principio no problemáticos para una perspectiva funcionalista de la mente, esto es, los juicios y creencias acerca de dichas experiencias, mostré que el argumento de los zombis se ve obligado a atribuir a estas criaturas, subrepticia e ilegítimamente, algunas incapacidades cognitivas a la hora de evaluar su propia situación epistémica que no están presentes en sus réplicas conscientes y que conforman la base para marcar una diferencia entre ellos que va más allá de una diferencia con relación a las propiedades fenoménicas de la experiencia consciente. Por lo tanto, debería admitirse que los zombis no son (no pueden ser) réplicas exactas de sus “gemelos” conscientes tampoco desde el punto de vista físico / funcional.

Adicionalmente, dentro del contexto de una crítica a otro de los argumentos típicos que suelen esgrimirse en contra de la viabilidad de un enfoque fisicalista y funcionalista de la consciencia, el argumento del hiato explicativo, en el **capítulo 2** reconstruí las objeciones a dicho argumento que muestran plausiblemente que aun

cuando se acepte que ciertos *conceptos* legítimos que poseemos acerca de la experiencia consciente no la caracterizan en términos de ciertos roles causales o funcionales, esto no constituye, por sí solo, un impedimento para identificar a la propiedad referida por dichos conceptos con una propiedad físico / funcional. Así, vimos que el argumento del hiato explicativo supone en su formulación que es una condición necesaria para la aceptación justificada de un enunciado de identidad (como “agua = H₂O” o “dolor = disparo de fibras C”), que resulte posible realizar cierto análisis particular de los conceptos a reducir / explicar.

En particular, de acuerdo con dicho argumento se requeriría un análisis del concepto del fenómeno a explicar / reducir en términos de la explicitación de los distintos roles causales que agotarían la noción pre - teórica u ordinaria que tenemos de dicho concepto. Sin embargo, es precisamente esta exigencia la que puede ponerse en cuestión. En efecto, de acuerdo con distintos autores, muchos enunciados de identidad, en diferentes disciplinas científicas, se aceptan por el poder explicativo que brindan o como el resultado de una inferencia a la mejor explicación. Así, por ejemplo, asumir que el calor *es* *mke*, o que la presión *es* transferencia de energía cinética, nos permite explicar hechos que no podríamos explicar de otra manera. Por lo tanto, estamos justificados por el principio de inferencia a la mejor explicación a inferir que estas identidades son verdaderas. Por otra parte, autores como Papineau 1998, sostienen que identidades que involucran descripciones (o conceptos que se agotan en la descripción de ciertos roles casales) como lo serían “agua = H₂O”, no son comparables con enunciados como “dolor = disparo de fibras C”, en las cuales se incluye algún concepto fenoménico. En particular, el autor considera que este último tipo de enunciado de identidad es más semejante a aquellos que involucran dos nombres propios, como “Mark Twain = Samuel Clemens”. Pero, si esto es así, el hecho de que no podamos, en el caso de las identidades psicofísicas relevantes, hacer el tipo de análisis funcional que Levine pretende, se debe simplemente a que los conceptos involucrados acerca de la experiencia consciente, al igual que los nombres propios, no tienen *a priori* asociadas descripciones de roles causales.

Ahora bien, si se aceptan las críticas mencionadas, entonces el defensor del argumento del hiato explicativo ya no puede esgrimir como una razón para rechazar las identidades psicofísicas o para señalar una limitación propia de los enfoques

funcionalistas de la mente, la imposibilidad de realizar cierto análisis de los conceptos fenoménicos involucrados porque existirían otro tipo de identidades (consideradas legítimas) que tampoco se aceptan mediante este procedimiento. Por lo tanto, si verdaderamente hay un hiato explicativo aquí, la causa no parece deberse a la imposibilidad de realizar dicho análisis conceptual.

En el **capítulo 3** me ocupé de otro de los aspectos que torna problemática a la experiencia consciente: el presunto carácter privado y directo de nuestro conocimiento acerca de ella, esto es, la idea de que, en sentido estricto, sólo cada uno de nosotros puede *saber* qué pasa por su propia mente, (al menos en lo que respecta a gran parte de sus estados mentales). En contraposición, se supone que el resto de las personas sólo puede *conjeturar* o realizar *hipótesis* acerca de la presencia y los rasgos de esos estados basándose en algún tipo de evidencia indirecta, como el comportamiento observable, o algún indicio del entorno.

Por mi parte, me propuse desestimar que exista un conocimiento de estas características. Con este objetivo, desarrollé una argumentación en dos etapas. La primera de ellas consistió, por un lado, en mostrar las consecuencias inadmisibles que podrían seguirse después de adoptar la tesis del conocimiento privado, entre las cuales se destacan: la imposibilidad de tratar los estados mentales conscientes desde una perspectiva científica, el compromiso con el “problema de las otras mentes”, el compromiso con la aceptación escenarios de dudosa coherencia interna (como el escenario zombi) y con una concepción del lenguaje que, en la medida en que supone que los significados de los términos mentalistas están constituidos por entidades privadas, hace difícil comprender cómo es que lograríamos comunicarnos unos con otros, tal como parece que de hecho hacemos, a través de él. En segundo lugar, mostré que, si se adopta el enfoque expresivista de filósofos como Dorit Bar-On (2004, 2010), es posible dar cuenta de algunos de los hechos que podrían motivar la aceptación de la tesis del conocimiento privado, sin verse obligados a postular un conocimiento de esas características ni asumir sus consecuencias perniciosas. En este sentido, consideré que una de estas motivaciones era la constatación de cierta asimetría entre las atribuciones mentalistas en primera persona de tiempo presente y las atribuciones mentalistas en tercera persona correspondientes según la cual, las primeras contarían con un grado considerable de certeza o de “presunción de verdad” que se no se le asigna a las segundas.

La segunda motivación para postular la existencia de un conocimiento privado la adjudiqué a otras peculiaridades de las atribuciones psicológicas, a saber: que mientras que aquellas que se realizan en tercera persona se basan fundamentalmente en la observación previa de la conducta de la persona u organismo, para realizar las auto-atribuciones mentalistas en primera persona, no nos basamos en este tipo de evidencia. Así, de acuerdo con una perspectiva que llamé “cartesiana”, estas asimetrías y peculiaridades de las atribuciones psicológicas se explicarían en última instancia en virtud de una asimetría epistémica, según la cual, contamos con un conocimiento directo, seguro (tal vez infalible) y privado con respecto a nuestros estados mentales conscientes que se contrapone con el conocimiento indirecto e irremediabilmente expuesto al error que tenemos con respecto a los estados mentales de las demás personas.

En relación con este punto, el cartesianismo acepta un tanto acríticamente la idea de que la auto-atribución de estados mentales conscientes se asemeja a la descripción de fenómenos físicos, en donde primero se daría un acto de observación “interna” de dichos fenómenos y luego, con base en esa evidencia, se daría la posibilidad de llevar adelante la mencionada descripción, de la misma manera en que una afirmación acerca del desorden que domina en mi habitación se basa o puede basarse en una observación previa de la misma y su estado.

Por mi parte, indiqué que el expresivismo (o neo-expresivismo) de Dorit Bar-On (2004, 2010) ofrece el inicio de una alternativa al paradigma cartesiano, por lo menos en lo que respecta a las atribuciones psicológicas. De acuerdo con esta posición, la seguridad especial que identifica a las mencionadas atribuciones se podría justificar por el carácter expresivo de los *actos* que las producen. Al mismo tiempo, esta concepción expresivista explica que por el mismo motivo por el cual no necesitamos ningún tipo de fundamento epistémico cuando expresamos, por ejemplo, nuestro dolor gritando o llorando, es por el cual tampoco necesitamos -ni es legítimo reclamar- una evidencia conductual o de algún tipo para realizar determinadas auto-atribuciones mentalistas en primera persona. Sin embargo, como vimos, esto no impide en principio que dichas atribuciones mentalistas digan, además, algo acerca de ciertos hechos del mundo y que, por lo tanto, tengan valor de verdad. En este contexto, el enfoque expresivista se destaca como una opción que, aunque su presencia no resulta nada frecuente en las discusiones en torno a la

consciencia, merece ser explorada en la medida en que representa una ruptura con ciertas ideas de toda una larga tradición en la filosofía y la psicología en torno al tipo de acceso que tenemos a nuestras mentes y a las mentes de los demás.

Con respecto a este último punto, el expresivismo considera que así como aquellos que están adecuadamente situados y poseen el conocimiento necesario, *ven* el significado de las palabras en una página, si tenemos las sensibilidades adecuadas y el conocimiento relevante, *vemos* (determinados) fenómenos psicológicos en el comportamiento de las personas. Por este motivo, la idea de que la evidencia conductual constituye siempre un signo indirecto y sistemáticamente poco confiable de los estados mentales de las demás personas pierda gran parte de su fuerza porque, si el expresivismo representa una opción aceptable, no hay un conocimiento directo que sirva como base a las auto-adcripciones mentalistas en primera persona con el cual contraponerlo.

En el **capítulo 4** señalé que la inefabilidad, junto con el carácter privado del acceso a los estados conscientes, son las dos propiedades frecuentemente atribuidas a la experiencia consciente que obstaculizan una aproximación científica a la misma, porque si la experiencia consciente es privada, entonces la perspectiva científica no puede alcanzarla, y si es inefable, entonces incluso si la ciencia lograra alcanzarla no podría describirla en una teoría (Ross 1993: 443). Por mi parte, me propuse mostrar que hay un sentido sensato en que puede atribuírsele esta propiedad a la experiencia pero que no conduce a estos resultados escépticos. Para encaminarme a esa conclusión, comencé haciendo notar que, de hecho, describimos habitualmente, tanto nuestra experiencia como sus objetos, sin mayores dificultades. Por otra parte, si bien podría admitirse, como señala Dennett, que la experiencia es inefable en la *práctica* debido al alto grado de información que puede llegar a portar, o que es temporalmente inefable, o que puede parecer inefable por la dificultad psicológica que implicaría “ponerse (imaginativamente) en los zapatos” de personas muy alejadas de nuestro horizonte cultural y temporal, nada de esto convierte a la experiencia en un fenómeno desconcertante ubicado más allá de todo conocimiento científico. Por otra parte, contemplé también la posibilidad de que la tentación de considerar que la experiencia es inefable en algún sentido problemático se hallara en la adhesión, más o menos explícita, a la tesis del lenguaje privado.

En efecto, un lenguaje de estas características explicaría por qué quien no tuvo las experiencias relevantes no comprendería ni los términos ni los conceptos acerca de las experiencias conscientes por mucho que otras personas se esfuercen intentando explicárselo mediante múltiples y detalladas descripciones: quien no tuvo las experiencias en cuestión, no puede tener acceso a las entidades privadas a las que dichos conceptos y términos refieren y que agotan su contenido y significado. Sin embargo, retomé algunas de las consideraciones realizadas en el capítulo anterior para desestimar la necesidad de postular un lenguaje de estas características. Además, hice notar algunas diferencias cruciales entre el modo en que podrían determinarse, a través de una suerte de ostensión, los significados o contenidos de los términos y conceptos en un lenguaje privado y en el lenguaje público, con el objetivo de poner en cuestión precisamente la idea de que la postulación de una ostensión privada resulte plausible o que vaya de suyo.

En este contexto, aproveché también para volver al argumento de los zombis y sostuve que si la viabilidad de un lenguaje privado es sumamente dudosa como parece serlo, la concebibilidad del escenario zombi también tambalea, porque dicho escenario parece presuponer esta tesis. Además, si se interpreta, junto con el neo-expresivismo, que nuestras atribuciones psicológicas en primera persona en tiempo presente son en realidad *actos expresivos* que no involucran conceptos formados a partir de nuestra familiaridad con ciertas entidades internas conocidas de forma privada y separadas de toda determinación corporal, se suman nuevos motivos para dudar de la concebibilidad del escenario zombi, porque es precisamente una caracterización de este tipo de los estados mentales la que vuelve plausible dicha concebibilidad.

En suma, a la luz de estas observaciones cobraba fuerza mi propuesta de que el grano de verdad detrás de la tesis de la infabilidad consiste simplemente en que para poseer una maestría completa de ciertos conceptos acerca de sensaciones o estados perceptivos, es necesario haber atravesado las experiencias relevantes y contar con las capacidades perceptivas funcionando de manera “normal” o típica. Así, por ejemplo, para poseer una maestría en el concepto de EXPERIENCIA DE VER ROJO, es necesario tener una maestría en el concepto de ROJO, y para cumplir con esta condición, se debe ser capaz de discernir mediante la vista el color rojo de otros colores. Desde luego, ninguna descripción verbal de estas experiencias podrá dotar

a alguien de esta capacidad de discriminación. Con todo, subrayé que no debe confundirse el hecho de que sólo quien puede ver colores será capaz de tener una maestría más o menos completa de los conceptos de color (y, también, del de las experiencias de color), con la idea más controvertida de que los conceptos y términos de experiencias cromáticas están definidos por referencia a entidades privadas e incommunicables o que las capacidades en juego no podrán ser explicadas en términos compatibles con una perspectiva funcionalista y fisicalista acerca de lo mental.

Por otra parte, a lo largo de este escrito señalé en varias oportunidades que las discusiones que se llevaban adelante en él resultaban relevantes no sólo para la filosofía de la mente sino también para las ciencias cognitivas ya que, en muchos casos, ambas comparten varios supuestos disputables y atribuyen a la consciencia las propiedades controvertidas que fueron objeto de análisis y crítica (intrinsecidad, inefabilidad y acceso privado y directo).

En particular, muchos neurocientíficos adoptan acríticamente buena parte de la terminología filosófica para caracterizar a la consciencia como, por ejemplo, la noción de *qualia* y la expresión “como es estar” en determinado estado mental. Por este motivo, en el **capítulo 5** me propuse encarar más directamente algunas de estas cuestiones. En particular, me ocupé de mostrar cierta evidencia a favor de la idea de que el fantasma de Descartes sigue vivo y no sólo en la filosofía, sino en buena parte de las neurociencias contemporáneas.

De acuerdo con este “nuevo” dogma, somos, casi literalmente, nuestro cerebro, de forma tal que, en rigor, es el cerebro, antes que algo no físico, lo que siente, piensa y decide. A partir de la constatación de esta situación, mostré las consecuencias nocivas que se siguen de adoptar esta doctrina a partir de los aportes de los trabajos de Noë y Bennett y Hacker: entre dichas consecuencias mencioné, por un lado, la caída en un escepticismo acerca de la existencia y naturaleza del mundo externo y de las otras mentes. En efecto, una vez que se asume esta perspectiva cartesiana, se supone que lo que realmente percibimos de manera directa son representaciones o imágenes, creadas por nuestro cerebro, y no las cosas “en sí mismas”. Al mismo tiempo, y esto es lo que fundamentalmente me interesaba discutir en el contexto del presente trabajo, se observa cómo se llega, por otra vía, a

la postulación de objetos privados como los objetos propios de la percepción consciente.

Por otro lado, se seguía también de la adopción de esta perspectiva cartesiana, la posibilidad de caer en lo que Bennett y Hacker llamaban “falacia mereológica”, es decir, el error de atribuir predicados psicológicos a las partes de un organismo que, en principio, sólo pueden ser aplicados de manera inteligible al organismo como un todo. Las confusiones conceptuales que se seguían de esta falacia ponían en duda la potencialidad explicativa de la neurociencia cartesiana porque, como observé, parecía presuponer aquello que debía explicar.

Todavía dentro de ese capítulo, consideré algunas de las razones que suelen citarse para adoptar la tesis del “cerebro creador”. Entre ellas, mencioné la presencia de ciertos fenómenos más o menos cotidianos que de algún modo parecen abonar la idea de que percibimos representaciones creadas por el cerebro: se trata de experiencias (como los sueños, las alucinaciones, la realidad virtual o la percepción de objetos muy alejados en el espacio) que parecen idénticas o muy semejantes a distintas experiencias perceptivas verídicas pero en las cuales el objeto externo está ausente. En estos casos, la experiencia se genera, en principio, de una manera que no involucra ningún objeto externo, y lo más razonable parece ser adjudicar al cerebro esta tarea.

Ahora bien, el hecho de que exista esta semejanza sugiere con plausibilidad que los mecanismos que subyacen en uno y otro caso deben ser también muy similares e incluso que todas nuestras experiencias (y no sólo los sueños, o las experiencias con realidad virtual) requieren de algún tipo de intermediario (una representación creada por el cerebro) para configurarse como tales. Otra de las motivaciones esgrimidas para postular este tipo de intermediarios tiene que ver con la necesidad de postular mecanismos compensadores que puedan dar cuenta de nuestras experiencias perceptivas dadas la “pobreza” e “incoherencia” de los estímulos que afectan nuestros sentidos. En este caso, el cerebro se encargaría de corregir los defectos de la señal de origen.

Frente a estas observaciones, mostré, siguiendo, otra vez, a Noë y Hacker, que esas experiencias que parecían semejantes a las experiencias perceptivas ordinarias pero en las cuales no está presente ningún objeto externo o bien no son, de hecho, semejantes, o bien, cuando lo son (como en el caso de la percepción de objetos

externos lejanos) el objeto externo está presente y entonces debe considerárselas como experiencias verídicas normales.

En segundo lugar, y siguiendo por esta última línea argumentativa, mostré que la presencia de estas representaciones cerebrales no resultaría necesaria para dar cuenta de las experiencias perceptivas. Así, por ejemplo, aquellos presuntos defectos de la percepción visual que el cerebro debería corregir no son tales o, al menos, no es necesario considerar que lo sean. Por el contrario, considerarlos como tales *presupone* la perspectiva cartesiana a la que, en teoría, estos hechos deberían brindarle apoyo empírico independiente.

En tercer lugar, mostré que aun aceptando que existan tales representaciones, lo cierto es que su postulación no resultaría suficiente para explicar aquello que se supone que vienen a explicar: la percepción.

En resumen: como vimos, la consciencia ha sido fuente de numerosas y prolongadas discusiones entre filósofos de la mente, pero también entre científicos cognitivos y neurocientíficos. De hecho, en muchas ocasiones científicos y filósofos coinciden en señalar que persiste una profunda oscuridad en torno a la naturaleza de la consciencia y a su vinculación con el resto del mundo físico. En este contexto, se suele creer que las dificultades involucradas a la hora de dar cuenta de este fenómeno tienen su origen no en una falta de información empírica relevante, sino en la falta de una idea clara de cómo deberían responderse las preguntas que se plantean.

Por mi parte, en este trabajo sostuve que la existencia de estos inconvenientes mencionados se explica, en buena medida, en virtud de que quienes se ocupan de estas cuestiones atribuyen a la consciencia, de manera más o menos implícita, ciertas propiedades que la convierten en un atolladero verdaderamente insuperable, *y por principio*, para una posición fisicalista de la mente.

Sin embargo, el recorrido que se llevó adelante en este trabajo reveló, por un lado, que hay buenas razones para pensar que algunas de esas propiedades problemáticas, como la intrinsecalidad y el carácter presuntamente privado y directo del conocimiento del fenómeno en discusión, no forman parte de la naturaleza de la consciencia (aun cuando pueda parecerlo a primera vista).

Por otro lado, argumenté también en favor de la idea de que otra de esas propiedades supuestamente problemáticas atribuidas a la consciencia, la

inefabilidad, si bien pertenece a la naturaleza del fenómeno en cuestión, no es especialmente problemática cuando se la comprende de manera adecuada.

A modo de cierre, enfoqué finalmente la discusión desde otro ángulo, pero con el objetivo de arribar a conclusiones similares. En particular, analicé en detalle una frase que la mayor parte de los filósofos involucrados en estas discusiones utilizan típicamente para hacer referencia al aspecto problemático de la consciencia, sin preocuparme por determinar si esta caracterización asume implícitamente algunas o todas las propiedades mencionadas recién. Concretamente, me ocupé de la frase acuñada por el filósofo Thomas Nagel, según la cual, el aspecto problemático de la experiencia se describe como “lo que es como estar en un estado mental para el organismo”.

En este marco, argumenté que es posible concluir, a partir de la discusión que se suscita entre los trabajos de P.M. S. Hacker (2002) y Greg Janzen (2005, 2008, 2011) que, en definitiva, quien utiliza el giro nageliano, se enfrenta a una suerte de dilema, según el cual, (a) o bien admite que utiliza la frase en cuestión en un sentido que no es claramente inteligible, y entonces tampoco es claro que haga referencia a un aspecto especialmente problemático de la mente, o bien (b) acepta que la usa en un sentido inteligible (ya sea técnico o cotidiano) pero que claramente no es problemático en absoluto.

En cualquier caso, ninguna de las opciones mentadas ayuda a favorecer la idea de que hay algo inasible en torno a la consciencia para el que adopte un enfoque fisicalista / funcionalista de la mente.

Ahora bien, una vez excluido este aspecto “problemático” de la experiencia consciente, se sigue que los aspectos restantes vinculados con el concepto de consciencia que pueden reclamar legítimamente una explicación, son precisamente aquellos que Chalmers ubica entre los “problemas fáciles” de la consciencia, como la habilidad, por parte de un organismo o individuo de discriminar, categorizar, y reaccionar ante los diversos estímulos ambientales mediante el ejercicio de sus facultades perceptivas, de reportar los propios estados mentales, de dirigir el foco de su atención hacia algún objeto o hecho, de controlar de manera racional la propia conducta en base a la información disponible y de reflexionar acerca de sus propios estados mentales, de sus motivaciones, de su carácter o sus acciones.

Pero entonces, el “misterio” de la consciencia, “el último obstáculo pendiente en nuestra búsqueda de una comprensión científica del universo” (Chalmers 1999), no es en realidad un misterio, sino un acto de magia, pero de magia de verdad, es decir, aquella que se logra con un conjunto de trucos y pases de manos.

Este es el resultado al que arribé una vez examinados los bastidores (conceptuales y filosóficos) de la consciencia. A la vera de la (di) solución que propuse del “problema” de la consciencia quedan abiertos, pues, nuevos horizontes para posibilitar nuestra búsqueda de una comprensión filosófica y científica del universo y de nuestro lugar en él.

BIBLIOGRAFÍA

- Alter, T. (2013), "Social Externalism and the Knowledge Argument", en *Mind*, Vol. 122, 486, pp. 481-496.
- Armstrong, D. (1985), *A Materialist Theory of The Mind*. London, Routledge.
- Ball, D. (2009), "There are No Phenomenal Concepts", en *Mind*, 118, pp. 935-962.
- Bar-On, D. (2004). *Speaking my Mind. Expression and Self-Knowledge*, Clarendon Press Oxford.
- Bar-On, D. (2007), "Expressing as 'Showing What's Within': On Mitchell Green's, Self-Expression, en *Philosophical Books*, Vol. 51, 4, pp. 212-227.
- Bar-On, D. (2010), "Avowals: Expression, Security, and Knowledge: Reply to Matthew Boyle, David Rosenthal and Maura Tumulty", en *Acta Anal* 25: pp. 47-63.
- Bayne, (2001), "Chalmers on the Justification of Phenomenal Judgments", *Philosophy and Phenomenological Research*, 62, 2, pp. 407-419.
- Beaton M. (2016) "Sensorimotor direct realism: How we enact our world". *Constructivist Foundations* 11(2): 265-276. <http://constructivist.info/11/2/265>.
- Beckermann, A. (2012), "Property Identity and Reductive Explanation", en Gozzano, S. y Hill, C. (eds.) (2012), *New Perspectives on Type Identity. The Mental and the Physical*, Cambridge University Press, New York., pp. 66-87.
- Bennett, M. R. y Hacker, P.M.S. (2003), *Philosophical Foundations of Neuroscience*, Blackwell Publishing.
- Block y Stalnaker (1999), "Conceptual Analysis, Dualism and the Explanatory Gap", *The Philosophical Review*, Vol 108, 1, pp-1-46.
- Block, N, Flanagan, O. y Güzeldere, G. (eds) (1997), *The Nature of Consciousness: Philosophical Debates*, Cambridge, MIT Press.
- Block, N. (1997), "On a Confusion about a Function of Consciousness", en Block, N, Flanagan, O. y Güzeldere, G. (eds) (1997), pp. 375-416.

- Campbell, M. y O'Sullivan, Michael. (comp.), (2015), *Wittgenstein and Perception*, Routledge.
- Carruthers, P. (2004), "Reductive Explanation and the 'Explanatory Gap'", en *Canadian Journal of Philosophy*, vol. 34, núm. 2, pp. 153-174.
- Carruthers., G. y Schier, E. (2014), "Why are We Still Being Hornswoggled? Dissolving the Hard Problem of Consciousness, *Topoi. An International Review of Philosophy*, pp. 1-13.
- Chalmers, David (1995), "Facing Up to the Problem of Consciousness", *Journal of Consciousness Studies*, 2 (3):200-19.
- Chalmers, D. (1999). *La mente consciente. En busca de una teoría fundamental*, Gedisa Editorial, Barcelona.
- Chalmers, D. (2010). *The Character of Consciousness*, Oxford University Press.
- Chalmers, D. (2010a), "The Content of Phenomenal Concepts", en Chalmers, D. (2010), *The Character of Consciousness*, Oxford University Press, New York, pp. 251- 275.
- Chalmers, D. (2010b), "The Epistemology of the Phenomenal Belief", en Chalmers, D. (op. cit.), pp. 277-304.
- Chalmers, D. y Jackson, F. (2001), "Conceptual Analysis and Reductive Explanation", en *The Philosophical Review*, vol. 110, pp. 315-360.
- Child, William, (2015), "Wittgenstein and phenomenal concepts", en Campbell y O'Sullivan (comps.), *Wittgenstein and Perception*, Routledge, pp. 84-103.
- Churchland, P. (1997), "The Hornswoggle Problem", en Shear, J. (ed.) *Explaining Consciousness: The Hard Problem*. Cambridge, Mass, MIT Press.
- Clark, A. (2000), "A case where access implies qualia?", en *Analysis*, Vol. 60.1, pp. 30-38.
- Crane, T. (2005), "Papineau on Phenomenal Concepts", *Philosophy and Phenomenological Research*, vol. LXXI, 1, pp. 155-162.
- Crane, T. (2012), "Tye on Acquaintance and The Problem of Consciousness", *Philosophy and Phenomenological Research*, Vol. LXXXIV, 1, pp. 190-198.
- Crick, F. (1995), *The Astonishing Hypothesis*, Touchstone, Londres.
- Damasio, A. (1999), *The Feeling of What Happens*, Heinmann, Londres.

- Dennett, D. (1995) *La conciencia explicada. Una teoría interdisciplinar*, Ediciones Paidós, Barcelona.
- Dennett, D. (1995), "The Unimagined Preposterousness of Zombies", *Journal of Consciousness Studies*, 2: 322-6
- Dennett, D. (2006), *Dulces sueños. Obstáculos filosóficos para una ciencia de la conciencia*, Katz, Buenos Aires.
- Dennett, D.C. (2003/1988) "Quinear los *qualia*" en Ezcurdia, Ezcurdia, en M. y Hansberg, O. (2003) *La naturaleza de la experiencia. Volumen I: Sensaciones* UNAM, México.
- Descartes, R. (1982), *Meditaciones Metafísicas*, trad. al español M. García Morente, Espasa-Calpe, Madrid.
- Díaz-León, E. (2011), "Reductive explanation, concepts, and a priori entailment", *Philosophical Studies*, 155, 1, pp. 99-116.
- Dretske, F. (1995), *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Dupré, J. (2009), "Hard and Easy Questions about Consciousness", en Glock, H. y Hyman, J, (eds.) *Wittgenstein and Analytic Philosophy. Essays for P.M.S. Hacker*, pp. 228-249.
- Edelman, G.M. y Tononi, G. (2000), *Consciousness: How Matter Becomes Imagination*, Allen Lane, The Penguin Press, Londres.
- Elpidorou, A. (2014), "Blocking the *A Priori* Passage", *Acta Analytica*, Vol.29, 3, pp. 285-307.
- Ezcurdia, M. y Hansberg, O. (2003) *La naturaleza de la experiencia. Volumen I: Sensaciones*, UNAM, México.
- Finkelstein, D. (2010), *La expresión y lo interno*, KRK ediciones.
- Frith, C. (2007), *Making up the Mind: How the Body Creates our Mental World*, Blackwell, Oxford.
- Gallagher, S. y Zahavi, D. (2014), *La mente fenomenológica*, Alianza Editorial, Madrid (traducción Marta Jorba).
- Glynn, I. (1999), *Anatomy of Thought*, Weidenfield an Nicolson, Londres.
- Hacker, P. (2005), "Goodbye To Qualia And All What?", en *Journal of Consciousness Studies*, 12, No 11, pp. 61-66.

- Hacker, P.M.S. (1990), *Wittgenstein. Meaning and Mind. An analytical commentary on the Philosophical Investigations*, Vol 3, Basil Blackwell, Oxford.
- Hacker, P.M.S. (2002), "Is There Anything it is Like to be a Bat?", en *Philosophy*, 77, pp. 157-174.
- Hacker, P.M.S. (2005), "Of knowledge and of knowing that someone is in pain", en Alois Pichler & Simo Saatela (eds.), *Wittgenstein: The Philosopher and His Works*, The Wittgenstein Archives at the University of Bergen, pp. 203-235.
- Hacker, P.M.S. (2013), *The Intellectual Powers: a Study of Human Nature*, Wiley-Blackwell, Oxford.
- Hodgson, D. (2005), "Goodbye To Qualia And All That?", *Journal of Consciousness Studies*, 12, No 2, pp. 84-88.
- Hüttemann, A. (2004), *What's Wrong With Microphysicalism?*, Londres, Routledge.
- Hutto, D. (2006), "Turning hard problems on their heads", en *Phenomenology and the Cognitive Science*, 5, pp. 75-88.
- Hymers, M. (2017), *Wittgenstein on Sensation and Perception*, Routledge, New York y Londres.
- Jackson, F. (1982), "Epiphenomenal Qualia", *Philosophical Quarterly* 32: 127-36.
- Jackson, F. (1982/2003) "Qualia epifenoménicos", en Ezcurdia, M. y Hansberg, 2003.
- Janzen, G. (2005), "Self-Consciousness and Phenomenal Character", en *Dialogue*, Vol. 44, 04, pp. 707-733.
- Janzen, G. (2008), *The Reflexive Nature of Consciousness*, John Benjamin Publishing Company.
- Janzen, G. (2011), "In Defense of The What-It-Is-Likeness of Experience", en *The Southern Journal of Philosophy*, Vol. 19, 3, pp. 271-293.
- Johnson-Lair, P. (1988), *The Computer and the Mind*, Fontana Press, p. 353.
- Kenny, A. (2000), *La metafísica de la mente*, trad. Rodríguez Consuegra, F., Editorial Paidós, España.

- Kenny, A. (2009), "Cognitive Scientism", en Glock, H. y Hyman, J., (eds.) *Wittgenstein and Analytic Philosophy. Essays for P.M.S. Hacker*, Oxford, Oxford University Press, pp. 250-262.
- Kenny, A. (2010), "Concepts, Brains, and Behaviour", en *Grazer Philosophische Studien*, 81, 105-113.
- Kirk, R. (1974) "Zombies vs. Materialists", en *Proceedings of the Aristotelian Society* (suppl. vol.). 48: 135-52.
- Kirk, R. (2008), "The Inconceivability of Zombies", *Philosophical Studies*, 139: 73-89.
- Kirk, R. (2015), "Zombies", en *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2015/entries/zombies/>>.
- Kriegel, U. (2007), "Philosophical Theories of Consciousness: Contemporary Western Perspectives", en Zelazo, D., Moscovitch, M., Thompson, E. (2007), *The Cambridge Handbook of Consciousness*, Cambridge University Press, New York, pp. 35-66.
- Levin, J. (2007), "What Is a Phenomenal Concept?", en T. Alter and S. Walter, eds., *Phenomenal Concepts and Phenomenal Knowledge: Essays on Consciousness and Physicalism*. New York:Oxford University Press.
- Levin, J. (2008), "Taking Type-B Materialism Seriously", en *Mind & Language*, Vol. 23, 4, pp. 402-425.
- Levine, J (2007), "Phenomenal Concepts and the Materialist Constraint", en en Alter, T. y Walter, S. (eds) (2007), *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press, pp. 145-166.
- Levine, J. (1983), "Materialism and Qualia: the Explanatory Gap", *Pacific Philosophical Quarterly*, 64, 354-361.
- Levine, J. (1997/1993), "On Leaving Out What It's Like", en Block, N. , Owen, F. y Güzeldere, G. (1997), *The Nature of Consciousness. Philosophical Debates*, pp. 543-555.
- Levine, J. (2001), *Purple Haze: The Puzzle of Consciousness*, Oxford, Oxford University Press.
- Lewis, D. (2003) "Lo que enseña la experiencia", en Ezcurdia, M. y Hansberg, (eds.), 2003.

- Loar, B. (1997), "Phenomenal States", en N. Block, O. Flanagan, and G. Güzeldere, eds., *The Nature of Consciousness*, Cambridge, Mass.: MIT Press.
- Lynch, (2006), "Zombies and the Case of the Phenomenal Pickpocket". *Synthese* 149: 37-58.
- Malcom, N. (1984), "Consciousness and Causality", en Armstrong y Malcom (1984) *Consciousness and Causality*, Blackwell, Oxford.
- Marcus, E., 2004, "Why Zombies are Inconceivable", *Australasian Journal of Philosophy*, 82: 477-90.
- Mc Ginn, C. (2003/1989), "¿Podemos resolver el problema mente-cuerpo?", en Ezcurdia, M. y Hansberg, 2003.
- Nagel, T. (2003/1974). "¿Qué se siente ser murciélago?", en Ezcurdia, M. y Hansberg, 2003.
- Nemirow, L. (2007) "So this is what it's like: a defense of the ability hypothesis," en Alter, T y S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, Oxford University Press, pp. 32-51.
- Nida-Rümelin, M. (1998), "On Belief about Experiences. An Epistemological Distinction Applied to the Knowledge Argument", en *Philosophy and Phenomenological Research*, Vol 58, 1, pp. 51-73.
- Nida-Rümelin, M. (2015), "Qualia: The Knowledge Argument", *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2015/entries/qualia-knowledge/>>.
- Noë, A. (2002), "Is Perspectival Self-Consciousness Non-Conceptual?", *The Philosophical Quarterly*, Vol. 52, 207, pp. 185-194.
- Noë, A. (2003), "Causation and perception: the puzzle unravelled", *Analysis*, 63.2, pp. 93-100.
- Noë, A. (2004), *Action in perception*, MIT Press.
- Noë, A. (2010), *Fuera de la cabeza. Por qué no somos el cerebro y otras lecciones de la biología de la consciencia*, Editorial Kairós, Barcelona, España.
- Noë, A. (2015), *Strange Tools: Art and Human Nature*, Hill and Wang, New York.

- O'Regan, K. (2010), "Why Red Doesn't Sound Like a Bell. Understanding the Feel of Consciousness", *International Journal of Philosophical Studies Humana.Mente*, pp. 5-24.
- O'Regan, K. (2011), *Why Red Doesn't Sound Like a Bell. Understanding the Feel of Consciousness*, Oxford: Oxford University Press.
- O'Regan, K (2014), "The Explanatory Status of the Sensorimotor Approach to Phenomenal Consciousness, and Its Appeal to Cognition", en Bishop, M. y Martin, A. (eds) (2014), *Contemporary Sensorimotor Theory*, Springer, pp. 23-35.
- O'Regan, K. (2016), "How the Sensorimotor Approach to Consciousness Bridges Both Comparative and Absolute Explanatory Gaps. And Some Refinements of the Theory", *Journal of Consciousness Studies*, 23, 5-6, pp. 39-65.
- O'Regan, K. y Noë, A. (2001), "A sensorimotor account of vision and visual consciousness", en *Behavioral and Brain Sciences*, 24, 939-1031.
- Orlando, E. (1997), "Contenido y conciencia: el debate en torno a los qualia", *Dianoia*, XLIII, pp. 1-29.
- Papineau, D. (1998), "Mind the Gap", *Philosophical Perspectives*, Vol. 12, Language, Mind, and Ontology (1998), pp. 373-388.
- Papineau, D. (2002), en *Thinking about Consciousness*. New York: Oxford University Press
- Papineau, D. (2007), "Phenomenal and Perceptual Concepts", en Alter, T. y Walter, S. (eds) (2007), *Phenomenal Concepts and Phenomenal Knowledge*, Oxford, Oxford University Press, pp. 111-144.
- Papineau, D. (2011), "What Exactly is the Explanatory Gap?", en *Philosophia*, vol. 39, pp.5-19.
- Pérez, D. (2002), "Los qualia desde un punto de vista naturalista", *Azafea*, 4, pp. 65-83.
- Pérez, D. (2007), "¿La conciencia?, ¿Qué es eso?", en *Estudios de Psicología: Studies in Psychology*, 28:2, 127-140.
- Pérez, D. (2009), "Cómo pensamos acerca de nuestra experiencia consciente. Acerca de la idea misma de 'concepto fenoménico', en Pérez, D., Skidelsky, L.

- y Minervino, R. (2009), *Conceptos. Debates Contemporáneos en filosofía y psicología*, Buenos Aires, Catálogos, pp. 155-180.
- Pérez, D. (2011), "Phenomenal Concepts, Color Experience, and Mary's Puzzle", en *Teorema*, Vol. XXX/3, pp. 113-133.
 - Pérez, D. (2013), "Cómo pensamos acerca de nuestra experiencia consciente", en *Sentir, desear, creer. Una aproximación filosófica a los conceptos psicológicos*, Prometeo Libros, Buenos Aires.
 - Pérez, D. (2014), "Asimetría y conceptos psicológicos", en *Páginas de Filosofía*, año XV, n^o18, Departamento de Filosofía, Universidad Nacional del Comahue, pp. 5-26.
 - Rabin, G. (2011), "Conceptual Mastery and the Knowledge Argument", en *Philosophical Studies*, 154, pp. 125-147.
 - Robinson, H. (1993), "Dennett on Knowledge Argument", en *Analysis*, Vol. 53, 3, pp. 174-177.
 - Robjant, David y Nether Stowey. "Learning of Pains. Wittgenstein's own Cartesian Mistake at Investigations 246" *Wittgenstein-Studien*, 3.1 (2012): 261-286.
 - Rosenthal, D. (2000), "Two concepts of consciousness". *Philosophical Studies* 49: 329-59.
 - Ross, D. (1993), "Quining Qualia Quine's Way", en *Dialogue*, 32, 3, pp. 439-459.
 - Russell, B. (1995), *Los problemas de la filosofía*, trad. Joaquín Xirau, Editorial Labor, Colombia.
 - Schroeder, S. (2013), "Can I have your pain?", *Philosophical Investigations*, 36:3, pp. 201-209).
 - Schroer, R. (2010), "Where's the Beef? Phenomenal Concepts as Both Demonstrative and Substantial", *Australasian Journal of Philosophy*, Vol 88-3, pp. 505-522.
 - Searle, John (1984), *Mind, Brain and Science- the 1984 Reith Lectures*, BBC, Londres.
 - Shapiro, L. (2011), *Embodied Cognition*, Routledge, Taylor and Francis Group, New York.

- Silverman, D. (2017), "Bodily Skill and internal representation in sensorimotor perception", *Phenomenology and the Cognitive Sciences*, <https://doi.org/10.1007/s11097-017-9503-5>.
- Skidelsky, L. (2015), *Representaciones mentales: donde la filosofía de la mente y la filosofía de la ciencia cognitiva se equivocaron*, EUDEBA, Ciudad Autónoma de Buenos Aires.
- Smit, H. (2014), *The Social Evolution of Human Nature: From Biology to Language*, Cambridge University Press, New York.
- Smit, H. (2017), "Inclusive Fitness Theory and the Evolution of Mind and Language", *Erkenntnis. An International Journal of Scientific Philosophy*, pp. 1-28.
- Smit, H. y Hacker (2013), "Seven Misconceptions About the Mereological Fallacy: A Compilation for the Perplexed", *Erkenntnis*, Vol. 79, 5, pp 1077-1097.
- Stoljar, D. (2005), "Physicalism and Phenomenal Concepts", en *Mind & Language*, Vol. 20, 5, pp. 469-494.
- Taylor, E (2015), "Explanation and the Explanatory Gap", *Acta Analytica*, vol. 31, 1, pp. 77-88.
- Thomas, N. J. T., (1998), "Zombie Killer", in S. R. Hameroff, A. W. Kaszniak, and A. C. Scott (eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*. Cambridge, MA: MIT Press, pp. 171-177
- Tye, M (1995), *Ten Problems of Consciousness: A Representational Theory of Phenomenal Mind*, Cambridge, MA: MIT Press.
- Tye, M. (2006), "Absent Qualia and the Mind-Body Problem", *Philosophical Review*, 115: 139-68.
- Tye, M. (2007), "Philosophical problems of consciousness", en Velmans, M. y Schneider, S. (eds.), *The Blackwell Companion to Consciousness*, Blackwell Publishing.
- Tye, M (2009), *Consciousness Revisited: Materialism without Phenomenal Concepts*, Cambridge, MA: MIT Press.
- Tye, M (2010), "El fisicalismo y el recurso a los conceptos fenoménicos", en Pérez, D., Español, S., Skidelsky, L., Minervino, R. (comps.) (2010), *Conceptos*.

Debates contemporáneos en filosofía y psicología, Buenos Aires, Catálogos, pp. 105-153.

- Vaaja, T. (2013), "Wittgenstein's 'Inner and Outer': Overcoming Epistemic Asymmetry", en *Nordic Wittgenstein Review*, 2, pp. 107-129.
- Veillet, B. (2012), "In Defense of Phenomenal Concepts", *Philosophical Papers*, Vol. 41:1, pp. 97-127.
- Ward, D. (2006), "Deflationism about Qualia" en Di Nucci, E. y McHugh, C. (eds) *Content, Consciousness, and Perception: Essays in Contemporary Philosophy of Mind*, Cambridge Scholars Publishing, pp. 104-124.
- Ward, D., Roberts, T., Clark, A. (2011), "Knowing what we can do: actions, intentions, and the construction of phenomenal experience", en *Synthese*, Vol. 81, No 3, pp. 375-394.
- Wittgenstein, L. (2010/1953), *Investigaciones filosóficas*, Instituto de Investigaciones Filosóficas, U.N.A.M., Editorial Crítica, Barcelona.
- Wrisley, G. (2011), "Wherefore the Failure of Private Ostensive Definition?", *Australian Journal of Philosophy*, 89:3, pp. 483-498.
- Zeki, S. (2000), "Abstraction and idealism", *Nature*.