

El problema de la conciencia en filosofía de la mente

*Felipe De Brigard
Universidad de Carolina del Norte, Chapel Hill
brigard@email.unc.edu*

1. Introducción

Tanto psiquiatras como neuropsicólogos se encuentran a diario con pacientes cuyos síntomas incluyen trastornos o alteraciones de la conciencia. El espectro de manifestaciones conductuales que esta noción cubre es, no obstante, muy amplio, y en ocasiones tiende a aplicársele a casos que parecen tener poco en común. Estados vegetativos y de coma profundo son ejemplos neuropsicológicos paradigmáticos. Poco controversiales son también las amnesias prolongadas o temporales, así como los casos de agnosia, asomatognosia y muchos otros síndromes relacionados. Considérese el caso de Mirna, reportado por el doctor Todd E. Feinberg en 1990. Después de haber cumplido setenta años, y como consecuencia de una embolia, Mirna sufrió considerable daño en la corteza frontotemporal derecha, lo que conllevó a su vez a una profunda hemiatención de su hemicuerpo izquierdo. Es decir, Mirna no sólo parece desconocer la existencia del lado izquierdo de su cuerpo: tampoco reconoce como suyo su propio brazo cuando, arrastrado por la mano del doctor Feinberg, es forzado a entrar en su campo visual. Aunque la presencia del brazo se hace visualmente consciente para Mirna, ella sigue siendo incapaz de reconocerlo como suyo, como lo atestigua el siguiente fragmento de una de sus muchas conversaciones (Feinberg, 2001, 9-10):

Feinberg: Quiero preguntarle de nuevo: ¿qué hay aquí? Déle un vistazo a esta parte.

¿Qué es esto?

Mirna: Sus dedos.

Feinberg: ¿Mis dedos?

Mirna: Sí.

Feinberg: Mírelos de nuevo; déles un buen vistazo. Listo... dígame lo que son.

Mirna: Dedos... veo dos dedos y un bolsillo.

Feinberg: Déle un buen vistazo. ¿Qué es esto? Esto [le toco el respaldo de la mano].

Mirna: El respaldo de su mano.

Feinberg: ¿El respaldo de mi mano?

Mirna: Sí.

Feinberg: Supongamos que yo le dijera que ésta es su mano.

Mirna: No le creería.

Feinberg: ¿No me creería?

Mirna: No, no.

Feinberg: Esta es su mano.

Mirna: No.

Feinberg: Mire, he aquí su mano derecha, y he aquí su mano izquierda.

Mirna: OK.

Feinberg: Ahora, ¿qué es esto [pregunto mostrándole su mano izquierda]?

Mirna: ¡El respaldo de su mano!

No es que Mirna sea incapaz de ver su mano izquierda—después de todo no tiene problema al reportar su presencia cuando se le presenta en su campo visual derecho. Su dificultad es distinta: Mirna no tiene consciencia de su propia mano en tanto que suya, y ha llegado incluso a adscribírsele a su difunto marido. Al contrario que las personas normales, el lado izquierdo de su cuerpo no parece suyo; es como si el modo en que Mirna fuera consciente de sí misma hubiese variado con respecto al modo en que solía hacerlo antes del accidente.

Trastornos similares son usualmente descritos en pacientes con desórdenes mentales de origen psiquiátrico. No es raro encontrar descripciones de ciertos delirios, como el de Caprás o el de Frégoli, en términos de alteraciones de la conciencia o trastornos del yo. Incluso las ya clásicas descripciones de casos de esquizofrenia que hiciera Kraepelin a finales del siglo XIX estarían formuladas en términos de “la pérdida de la unidad de la conciencia”, como si el paciente fuera una suerte de “orquesta sin director” (Kraepelin, 1896). Para muchos, dicha metáfora sigue siendo adecuada para describir el modo en que los pacientes caracterizan sus sensaciones subjetivas. Piénsese, por ejemplo, en los llamados “síntomas de primer rango” de la esquizofrenia (Schneider, 1959; Mellor, 1970)¹, que incluyen una marcado deterioro en la percepción de uno mismo. Los pacientes esquizofrénicos suelen reportar la ocurrencia de sentimientos, sensaciones, impulsos, percepciones, e incluso pensamientos que no se encuentran bajo su control, como si no fueran suyos, como si la orquesta tocara sola, o como si ellos fuesen los actores de unas líneas escritas por otro autor (Gerrans, 2002; Sass y Parnas, 2003).

La noción de conciencia suele asociarse también con la capacidad del paciente para reconocer su enfermedad. Al igual que en neuropsicología, en la literatura psiquiátrica abundan las descripciones de pacientes que son incapaces de reconocer sus dificultades. Flashman, por ejemplo, reporta el caso de EF, un hombre de 27 años, divorciado y con 12 años de educación, que presenta recurrentes alucinaciones y delirios:

Su principal delirio involucraba la creencia de que se encontraba bajo investigación estatal y federal, sin razones claras. “Ellos” habían implantado un dispositivo en su oído para monitorear su comportamiento e interferir con su vida. [EF] no tenía historia psiquiátrica antes del episodio, que supuestamente empezó luego de que sufriera un accidente automovilístico (que, según reportó, no fue su culpa) y al cual llegó la policía estatal a investigar. [...] Él estaba convencido de que no tenía ninguna enfermedad mental, a pesar del hecho de que había perdido su trabajo, su esposa lo había dejado y había sido hospitalizado varias veces. Él negaba el que una enfermedad mental hubiera tenido impacto alguno sobre su vida. EF manifestaba deterioro en su introspección [*insight*] y auto-conciencia. Negaba cualquier enfermedad mental, limitaciones, o síntomas relacionados con una

¹ De acuerdo con el DSM IV, la esquizofrenia es un desorden psiquiátrico caracterizado por síntomas psicóticos que incluyen delirios, alucinaciones, desórdenes de lenguaje y catatonía (American Psychiatric Association, 1994). Sus síntomas se dividen en dos grupos: negativo y positivo. Los síntomas negativos son considerados parte de la presentación clínica y parecen reflejar disminución o deterioro de las capacidades cognitivas normales (v.gr. alergia, disminución afectiva, anhedonia y comportamientos evasivos o asociales). Por el contrario, los síntomas positivos reflejan exceso o distorsión de dichas capacidades (Flashman, 2004).

enfermedad mental, y no mostraba comprensión alguna de cómo sus limitaciones afectaban su nivel actual de funcionamiento. [Flashman, 2004, 149]

Aunque en la práctica médica se está haciendo más frecuente el uso del término ‘introspección’ [*insight*] para referirse a la incapacidad que EF presenta (Navarro et al, 2006), muchos autores siguen utilizando la expresión ‘falta (o pérdida) de conciencia de enfermedad’ para hablar del mismo fenómeno².

De manera que vale la pena preguntarse en qué consiste esta conciencia cuya alteración permea tantos y tan variados desórdenes mentales. Por décadas, muchos filósofos han intentado dar cuenta de la naturaleza de tan extraordinaria facultad. Como es de esperarse, las opiniones se encuentran enormemente divididas cuando se trata de explicar en qué consisten nuestras sensaciones conscientes, subjetivas; es decir, esas experiencias que están esencialmente atadas a nuestra perspectiva de primera persona, a nuestro punto de vista de sujetos experimentadores. No sólo es difícil encontrar unanimidad en el modo en que, se supone, debemos hablar de “conciencia” tanto en neuropsicología como en psiquiatría, sino que tampoco resulta fácil entender exactamente la manera en que la conciencia figura en los modelos y explicaciones científicas de dichas disciplinas. Porque, como nos lo advierte William Lycan, “parece haber una infranqueable tensión entre una visión científica, materialista del mundo y el carácter subjetivo, fenoménico de la experiencia” (1996, 2). ¿En qué consiste esa tensión?, ¿de qué modo la filosofía puede arrojar luces sobre el problema? y, finalmente, ¿cuáles son las teorías filosóficas y científicas sobre la conciencia más prominentes del momento?, son las preguntas que este capítulo introductorio busca ayudar a responder.

2. El problema de la formulación del problema

El primer problema con que nos topamos cuando queremos analizar el llamado “problema de la conciencia” es que no existe una formulación unánime del acertijo que, se supone, hay que resolver. Esto ocurre por dos razones. En primer lugar, parte del problema de la formulación radica en los múltiples significados asociados al término “conciencia”. William Lycan (1996), por ejemplo, encuentra al menos ocho significados distintos de la palabra “conciencia” o “consciente” (dejando de lado otros sentidos, menos “técnicos”—como cuando equivalemos “conciencia” a “sentido moral”—que son, en definitiva, aplicaciones metafóricas de usos psicológicos de los mismos términos); veamos:

1. Un organismo es consciente si y solo si tiene la capacidad de percibir, pensar, sentir, etc.
2. Un organismo es consciente y ejerce un control consciente sobre su cuerpo si y solo si está despierto, tiene estados mentales, gobierna sus acciones, etc.
3. El estado mental P, expresado como “A es consciente de X” es un estado mental consciente si y solo si P es un estado intencional (está dirigido hacia un objeto).

² Según Flashman (2004), al menos siete términos se emplean con frecuencia en la literatura psiquiátrica para referirse a la incapacidad que presentan los pacientes esquizofrénicos para reconocer su enfermedad. Estos son: ‘agnosia’, ‘anosognosia’, ‘negación de enfermedad’, ‘introspección’ [*insight*], ‘anosodiaforia’, ‘pérdida de auto-conciencia’ [*self-awareness*] y ‘conciencia’.

4. Un estado o evento mental es consciente si y solo si el sujeto advierte dicho evento.
5. Un estado o evento es consciente si y solo si es reportable (i.e. uno es consciente de las cosas que puede reportar).
6. Un organismo es consciente de un evento o suceso mental si y solo si tiene de él una conciencia introspectiva, una percepción del sentido interno.
7. La conciencia es una experiencia subjetiva, es decir, accesible únicamente desde una perspectiva de primera persona.
8. Se es auto-conciente si y solo si uno se advierte como siendo distinto de los demás.

Las definiciones 1 y 2 parecen muy relacionadas, y llaman a gritos explicaciones ulteriores, de forma que no pueden darse por concluyentes. Qué significa organismo, o sujeto, o cómo se ejerce control, o cómo se unifican las percepciones en un solo sujeto, etc., son todas preguntas que surgen de dichas definiciones y, por tanto, no pueden llamarse a ser definitivas. La definición 3 es, también, cuestionable, en el sentido de que no es claro que todos los estados intencionales y sólo los estados intencionales sean conscientes. Freud, por ejemplo, hablaba de representaciones mentales inconscientes, que, al ser representaciones, son intencionales también (ver capítulo XX en este volumen). Por el contrario, otros filósofos, como John Searle, sostienen que sólo los estados conscientes son intrínseca y esencialmente intencionales. De hecho, para él, el único modo en que un estado inconsciente puede ser a la vez intencional es si, al mismo tiempo, es potencialmente consciente (Searle, 1995).

Las definiciones 4 y 5 suenan circulares. “Advertir” no parece ser más que un sinónimo de “ser consciente de”, y “reportable” no más que un efecto del hecho de que uno puede hablar acerca de lo que advierte. Entre las definiciones 6, 7 y 8 parece haber un rasgo de familia—así como también varios inconvenientes. La definición 6, por ejemplo, nos dice que hay eventos que ocurren al interior del sujeto que son considerados objetos del sentido interno. Es difícil saber qué significa exactamente “objeto del sentido interno” en este contexto, porque claramente hay eventos que ocurren ahora mismo en mi cuerpo (v. gr. reduplicaciones genéticas, movimientos peristálticos) de los que no soy consciente aun cuando son internos. El filósofo John Locke, cuyo nombre se asocia con frecuencia a esta definición, termina correlacionando ese sentido interno con un tipo de experiencia subjetiva que es privada, propia del organismo y, por tanto, equivalente a la formulación 7 (Locke, 1688/1959). Así, por ejemplo, aunque yo no tenga una experiencia del crecimiento de mi propio pelo distinta de la que pueda tener un tercero que ve cómo mi propio pelo crece, yo sí puedo tener una experiencia distintiva del sabor del jugo de papaya, diferente de la experiencia que un tercero puede tener al tomarse el mismo jugo de papaya. De modo que la definición 6 converge con la 7. Finalmente, la definición 8, al introducir el prefijo “auto”, parece más bien describir una característica de la conciencia en lugar de definirla. Si bien dicha característica—como veremos—es necesaria para su definición, en ningún caso resulta suficiente. En consecuencia, parece que el problema de la conciencia tiene que ver con la definición 7: ¿cómo podemos dar cuenta de experiencias subjetivas, de primera persona, cuando nuestra ciencia es objetiva, de tercera persona?

La otra razón por la cual el problema de la conciencia es de difícil formulación radica en su novedad. La noción de conciencia, aunque puede decirse que en un sentido fue problemática desde tiempos de Aristóteles (v. gr. muchos historiadores de la filosofía consideran que hay una relación entre la noción de “intelecto agente” en Aristóteles (1994) y nuestra noción de conciencia relativa a la unidad de la subjetividad), sólo consiguió arrinconarse como concepto problemático hasta hace unos pocos años. Porque, a decir verdad, para los filósofos modernos el problema de la conciencia no era distinto del problema de la mente en general. Para Descartes, por ejemplo, la mente y la conciencia eran básicamente equivalentes. La mente, en tanto que *res cogitans*, no consistía más que en el acto de pensar, y para Descartes era obvio que no había pensamiento sin la conciencia o “advertencia de segundo orden” que nos dice que estamos pensando (ver, por ejemplo, en Descartes, 1985, *Meditaciones II, Objeciones IV, y Principios I, 9* en adelante.) . Esta misma idea vuelve a aparecer en Locke, para quien “pensar consiste en ser conciente de que uno piensa [...] La idea del pensamiento en ausencia de la conciencia es tan ininteligible como la idea de un cuerpo extendido sin partes” (Locke, 1688/1959, Libro II, 1 §19)). Y aunque Hume habría de poner en duda la existencia de un yo indivisible y permanente encargado de orquestar nuestra vida mental, su escepticismo, lejos de motivar una investigación empírica por la naturaleza de tan inquietante fenómeno, permitió que Kant le ajustara la estocada final al problema de la unidad de la conciencia al sugerir que el “yo aperceptivo”—la unidad final de todas las representaciones mentales (ver más adelante)—no podría ser más que un sujeto trascendental, por lo cual su estudio empírico es imposible (Kant, 1787/1929)

Así, la conciencia pasó bastante desapercibida por el ámbito filosófico y científico durante el siglo XVIII y parte del XIX, tal vez sólo hasta Freud. Con Freud ella no sólo alcanzó nuevamente un estatuto científico inimaginado, sino una complejidad tan amedrentante que las escuelas psicológicas post freudianas empezaron a temerle al concepto. Esto ocurrió, por ejemplo, con William James, quien durante la década de 1880, cuando se preparaba la publicación de la primera parte de los *Principios de Psicología* (1890), aseguraba que la conciencia era una entidad real, accesible por introspección, de la que sólo podía darse una definición ostensible o, a lo más, metafórica, pero de cuya existencia no podía dudarse. No obstante, en su artículo “*Does consciousness exist?*” [*¿Existe acaso la conciencia?*], publicado 25 años después, James aboga por una idea totalmente contraria a la suya original: la conciencia, decía, “es el nombre de una no-entidad, y no tiene ningún lugar al interior de los primeros principios [de psicología]” (James, 1904).

En un intento por rescatar la científicidad de la noción de conciencia, tan maltratada por James y tan conceptualmente poluta por el psicoanálisis, la escuela introspeccionista buscó una estrategia alternativa: la de sistematizar nuestras impresiones subjetivas. Sin embargo, el asombroso fracaso a la hora de correlacionar resultados (v. gr. mientras el laboratorio de Titchener, en Cornell, publicó en la década de los 20s, 44.435 sensaciones discriminables diferentes, en Alemania Külpe publicó sólo 12.000) no sólo le dio a la investigación sobre la conciencia una segunda muerte, sino que le proveyó al conductismo un nuevo empujón. Con el conductismo, claro, la conciencia ni se mencionaba. Es interesante, por ejemplo, advertir que las entradas “conciencia”, “introspección” y “subjetivo” se eliminaron de los diccionarios de psicología hasta hace cerca de unos 35 años, cuando tras la caída del conductismo se impone la ciencia

cognitiva. Ahora la conciencia empieza a ser problema otra vez; de hecho, el único problema que—según opinan algunos—no podrá ser resuelto por el enfoque del procesamiento de información. Ese es el paradigma de explicación dominante hoy día. Y por carecer de un lenguaje teórico para formular exactamente el problema de la conciencia, nos va a tocar hacer toda clase de piruetas para poder acorralar la duda.

3. El problema de la conciencia

Algunas exposiciones del problema de la conciencia son más claras que otras, y sus autores se han vuelto famosos, más que por dar repuesta al problema, por formularlo. Una de las más afamadas fue propuesta hace unos años por el filósofo David Chalmers (1995, 1996). Él considera que en la investigación científica de la mente nos encontramos con varios problemas fáciles y con un problema difícil. Preguntas como ¿de qué modo el ser humano discrimina percepciones?, o ¿por qué reacciona apropiadamente frente a uno u otro estímulo? o, incluso, ¿cómo forma el cerebro una representación somato-sensorial? son, de acuerdo con su formulación, problemas fáciles. El problema difícil, en contraposición, es el de averiguar cómo un sistema físico como el cerebro, cuya naturaleza es independiente de la existencia de las experiencias conscientes, y cuyas operaciones son analizadas desde el punto de vista objetivo y de tercera persona empleado por la neurociencia, cómo un sistema así—repito—puede dar lugar a experiencias subjetivas, privadas, cuyo acceso está limitado al sujeto que las experimenta y que, por tanto, son analizables sólo desde la perspectiva de primera persona. La explicación de la conciencia —sostiene él—

no es sólo un asunto de explicar la estructura y la función [como es el caso de los problemas fáciles]. Una vez hayamos explicado toda la estructura física en la vecindad del cerebro, y hayamos explicado cómo son efectuadas todas las variadas funciones del cerebro, habrá aún un tipo de *explanandum*: la conciencia en sí misma. ¿Por qué deberían todas estas estructuras y funciones dar lugar a la experiencia? La historia acerca de los procesos físicos no lo dice” (Chalmers, 1996, 107).

Eso es lo que hace tan difícil y tan único el problema de la conciencia: éste “va más allá de los problemas acerca del desempeño de las funciones”. (Chalmers, 1995, 12).

Antonio Damasio, de otro lado, ha abogado por una descripción del asunto un tanto divergente. Él considera que la investigación científica de la conciencia debe dar cuenta, en primer lugar, de cómo creamos con (y en) el cerebro una película integrada de todas nuestras diversas imágenes sensoriales. Y, en segundo lugar, debe dar cuenta de cómo, junto con la película, el cerebro construye un sentido de yo en el acto de percibir y conocer los objetos, un yo que se constituye como el actor y portador de sus experiencias, el sujeto de lo que él llama “el sentimiento de lo que pasa” (Damasio, 1999 y 2000). Formulaciones como ésta se enmarcan, en todo caso, en la misma distinción entre fenómenos físicos, objetivos y de tercera persona, por un lado, y fenómenos conscientes, subjetivos, de primera persona, por el otro. El propio Damasio acepta dicha dualidad:

“Tomo los procesos de la mente como hechos por el continuo flujo de patrones mentales o imágenes, accesibles sólo en una perspectiva de primera persona. Por

esta razón los patrones o imágenes mentales son distinguibles de los patrones neuronales, que son accesibles a observaciones hechas en una perspectiva de tercera persona” (Damasio, 2000, 111).

A pesar de la novedad de las expresiones, ambas maneras de abordar el problema nos recuerdan una vez más a Kant. Como ya lo mencioné, Kant consideraba que debía asumirse a priori la existencia de un sujeto trascendental que acompañara todas nuestras representaciones en lo que denominó “la unidad de la apercepción”. Tal sujeto sería el encargado, por decirlo así, de unificar todos sus conceptos e intuiciones y otorgarles el carácter de experiencias propias. Expliquemos esto en términos más mundanos. Mientras escribo este documento tengo frente a mí varios libros abiertos, veo también las letras que al presionar en el teclado aparecen en la pantalla, estoy escuchando una famosa canción de Pink Floyd y procuro desatender el teléfono que ha sonado toda la mañana. Además, soy consciente del peso de mi cuerpo sobre la silla, paladeo un sorbo de café y percibo un ligero dolor en el dedo que me corté anoche. Soy, en una palabra, el dueño de una cantidad de experiencias. ¿Cómo hace el cerebro para ligar en una sola experiencia de un único objeto múltiples estímulos? Y, más aún, ¿cómo genera el cerebro la unidad de todas esas experiencias en un único flujo de conciencia que tiene sólo un dueño: yo? En neurociencia, el primer problema se denomina “problema de la ligadura” (*binding problem* [Triesman y Gedale, 1980]) y el segundo sería, propiamente, “el problema de la conciencia”. Dicho sea de paso, Kant consideraba que el yo aperceptivo resulta, por la misma estructura de nuestra razón, empíricamente incognoscible. De esta forma, si adoptásemos los términos kantianos, el problema de la conciencia se convierte en la indagación por la naturaleza del yo trascendental, en un intento por demostrar que, contrario a lo que Kant pensaba, sí podemos saber cuál es su naturaleza. Y al igual que Kant, hoy día muchos se resisten a creer que la conciencia pueda jamás encajar dentro de nuestros marcos científicos objetivos.

Veamos, pues, en qué consiste el problema de la investigación científica de la conciencia hoy día. Empecemos por aceptar que la conciencia y la experiencia consciente son una y la misma cosa. Porque experimentar conscientemente algo es, precisamente, ser consciente o, si se quiere, tener conciencia. Ahora bien, hemos visto que la propiedad esencial de la conciencia es su subjetividad. Dicha subjetividad no debe entenderse en el mismo sentido en que decimos que son subjetivos ciertos juicios de valor, como por ejemplo “el 11 Caldas es mejor equipo de fútbol que el América de Cali” o “Rembrandt es peor pintor que Kandinsky”, como opuestos a juicios objetivos del estilo de “El sol es el centro del sistema solar” o “el agua hierve a 100° al nivel del mar”. La conciencia es subjetiva en el sentido de que esencialmente, y no de modo derivado, constituye un fenómeno de primera persona. ¿Pero en qué sentido debemos entender acá la noción de “subjetivo”? Según Norton Nelkin, en filosofía de la mente pueden encontrarse al menos tres definiciones de subjetividad (Nelkin, 1996):

1. La subjetividad como *punto de vista*: los sujetos experimentan el mundo desde diversas perspectivas.
2. La subjetividad como *origen de la voluntad*: el sujeto es dueño de sí; es el protagonista de su auto-control.

3. La subjetividad como *cualidad fenomenológica*: la privacidad ontológica ligada a la privacidad epistémica. En consecuencia, es un fenómeno dependiente del sujeto ontológicamente y, por tanto, cognoscible sólo por el sujeto.

Es en este último sentido como ha de entenderse la noción de “subjetivo” vinculada al problema de la conciencia. Por eso la conciencia parece oponerse, esencialmente, a los fenómenos de tercera persona, es decir, a los fenómenos que no dependen de la perspectiva de ningún sujeto en particular, que son objetivos, públicos, compartibles y quizás incluso verificables. En una palabra, los fenómenos de que se ocupa y de que puede ocuparse la ciencia. Y así las cosas, parece que nos enfrentamos a un dilema bastante complicado: o bien aceptamos que la investigación de la conciencia debe darse en términos objetivos y de tercera persona, como los utilizados actualmente por la física o las neurociencias, renunciando así a dar cuenta del aspecto esencial de la conciencia—es decir, su subjetividad, su esencia característica de ser un fenómeno de primera persona—o bien procuramos mantener la esencia subjetiva del fenómeno de la conciencia a sabiendas de que con esto renunciamos a la posibilidad de encuadrarla al interior de un marco científico objetivo, de tercera persona (Botero, 2003). ¿Existe alguna salida? ¿Será posible que la ciencia llegue algún día a sortear esta difícil “brecha explicativa” [*explanatory gap*] (Levine, 1983)?

4. Qualia

Antes de analizar algunas de las diversas teorías filosóficas y científicas que han buscado darle solución al problema de la conciencia, vale la pena detenerse a examinar otro asunto que si bien parece ser una dificultad subsidiaria, tras un breve vistazo resulta no ser más que una reformulación de la cuestión central por la relación entre los fenómenos subjetivos y objetivos. Me refiero, por supuesto, al problema de los qualia (problema que, para filósofos como John Searle, resulta literalmente equivalente al problema de la conciencia [Searle, 1997, capítulo 1]). El término qualia fue acuñado por C. I. Lewis en 1929 y se refiere a las propiedades cualitativas o fenoménicas de ciertos estados mentales, de los cuales los estados conscientes son paradigmáticos. Al estado mental en que me encuentro cuando paladeo un buen sorbo de café le pertenecen ciertas cualidades: sabe ligeramente amargo, calienta mi garganta, huele de un modo particular, etc. El modo en que mi experiencia de ese sorbo de café se presenta en mi conciencia puede variar significativamente con respecto al modo en el que se presenta en la conciencia de alguien más. Lo que determina que el sabor del café tenga el peculiar gusto que siento yo al tomarlo son las particularidades fenomenológicas, cualitativas, que distinguen mi experiencia consciente o quale (singular de qualia) de las demás. Con frecuencia los qualia son caracterizados como estados mentales que tienen las siguientes propiedades: son estados reales internos, privados y a los que solamente uno tiene acceso. Además, son inefables, directa e inmediatamente aprehensibles en la conciencia, y por tanto están intrínsecamente ligados a la perspectiva de primera persona. Finalmente, muchos sostienen que son inaccesibles, inanalizables e inexplicables por la ciencia.

Para efectos del presente capítulo, no importa si la anterior definición goza o no de unánime aceptación. Basta con aceptar que dichos estados son, esencialmente, subjetivos (en el sentido anteriormente mencionado). Valga anotar que, si bien la idea de los qualia sólo ha encontrado un lugar destacado en la filosofía y en neurociencia más o

menos recientemente, es tan antigua como cualquier doctrina que se ocupara de la percepción. Locke, por poner sólo un ejemplo, sostenía que las propiedades como los colores, olores y sabores, eran poderes que las cosas del mundo ejercían sobre nosotros en virtud de las cualidades primarias. De ahí que las denominara “cualidades secundarias”. Como nos lo relata Daniel Dennett, para Locke el rojo era “la propiedad disposicional o poder que poseen ciertas superficies de los objetos físicos, gracias a los rasgos microscópicos de su textura, para producir en nosotros la *idea de rojo* cada vez que la luz se reflejaba desde esas superficies hacia nuestros ojos” (Dennett, 1991, 371). Pero cabe preguntarse ¿qué es lo que es rojo en mi idea de rojo? Evidentemente no está coloreada así, ni tiene un rótulo que diga “rojo” debajo de ella, ni nada por el estilo. Parece no ser otra cosa que una cualidad netamente subjetiva, es decir, inextricablemente vinculada a mi propia experiencia de rojo lo que la hace ser roja para mí. Esto es mi *quale* (singular de *qualia*) particular de rojo, y averiguar cuál es la naturaleza de este *quale*, entendido como propiedad o cualidad de la experiencia perceptible del sujeto, es justamente de lo que se trata el problema de los *qualia*.

Ahora bien, ¿en qué sentido es un problema científico y en qué sentido es un problema metafísico? Hay un modo de responder ambas preguntas con la misma respuesta. Dado que la ciencia se enmarca dentro de una perspectiva metafísica materialista, cuyos términos se refieren a fenómenos objetivos y de tercera persona, la existencia de los *qualia* se convierte en premisa fundamental del que podríamos llamar “argumento de la subjetividad”. Éste corre así (Lycan, 1996, 45):

1. Si el materialismo es cierto entonces los seres humanos son sólo una colección de diminutos objetos físicos.
2. Si los seres humanos son sólo una colección de diminutos objetos físicos, entonces deben poder describirse con los términos objetivos de la física (de tercera persona).
3. Pero hay fenómenos humanos (*qualia*) que no pueden describirse desde una perspectiva de tercera persona.

Luego,

4. El materialismo es falso. (QED. MMT, 1,2,3)

Para defender la premisa 3 de este argumento, los filósofos acuden con frecuencia a experimentos mentales, algunos de los cuales se han vuelto muy famosos. El primero de ellos se encuentra subsumido bajo el epítome que Thomas Nagel popularizó con la pregunta “¿qué es ser como un murciélago?” (Nagel, 1974. Sobre el epítome ver: Dennett, 1981), dado que, en general, tendemos a pensar que tener una experiencia consciente significa, “básicamente, algo que es ser como ese organismo.” (Nagel, 1974). En su ejemplo, Nagel nos dice que la ciencia ha demostrado que los murciélagos carecen de visión pero que, sin embargo, poseen una suerte de sistema de sonar tan impresionantemente preciso que les permite percibir la distancia y configuración de los objetos de su entorno en virtud de la eco-localización. La idea, en breve, consiste en que, aunque sepamos objetivamente cómo funciona esa eco-localización a tal punto que podamos imaginarnos cómo sería, para nosotros, percibir el mundo de tal modo, no podríamos nunca saber exactamente cómo un murciélago lo percibe puesto que, para saberlo, deberíamos *ser* murciélagos. Del mismo modo, no sabemos cómo puede ser la

percepción de algo verde por otra persona, porque para hacerlo deberíamos *ser* esa persona, y no lo somos. Nuestros reportes al respecto de cómo funcionan las percepciones sonoras de los murciélagos, e incluso visuales de las demás personas, podrán ser todo lo objetivas que se quiera, pero el rasgo cualitativo—el carácter qualia—de las mismas, en tanto vinculado íntimamente con el punto de vista del sujeto que percibe, es inexpresable por cualquier visión externa (De Brigard, 2003b). De ahí que tengamos que aceptar que la pregunta por el “¿qué es ser como?...” está desafortunadamente cargada de una teoría subjetivista difícil de evadir para los reduccionistas, y que encuentra en el mismo artículo de Nagel su más explícita expresión:

Es imposible excluir las características fenomenológicas de la experiencia [con] una reducción del mismo modo en que uno excluye las características fenoménicas de una sustancia ordinaria [con] una reducción física o química de ellas—a saber, explicándolas como efectos sobre las mentes de los observadores humanos. Si el fisicalismo ha de ser defendido, las características fenomenológicas deben en sí mismas ser dadas en términos físicos. Pero cuando examinamos su carácter subjetivo parece que tal resultado es imposible. La razón es que cada fenómeno subjetivo está esencialmente conectado con un punto de vista particular, y parece inevitable que una teoría objetiva, física vaya a abandonar ese punto de vista. (Nagel, 1974, 520).

Otro famoso experimento mental que nos puede ayudar a entender el problema al que se enfrenta el materialismo cuando intenta explicar los qualia, fue originalmente ideado por Frank Jackson en 1982. Jackson nos pide que imaginemos a una súper neurofisióloga llamada Mary que conoce perfectamente toda la física de la percepción de los colores. Ella es capaz de dar un reporte absolutamente preciso de cómo las ondas lumínicas inciden en los conos de la retina, cómo dichos conos afectan células anacrinas, como se transmiten los impulsos nerviosos por el tracto óptico, etc. Luego se nos pide que imaginemos que Mary ha vivido toda su vida entre un salón en la que sólo hay cosas blancas y negras. Y se nos pide, finalmente, que imaginemos que Mary es liberada y que lo primero que ve al salir de la caja es una manzana madura. Ahora bien, la pregunta es obvia: ¿acaso Mary aprende algo nuevo acerca del color cuando por vez primera ve dicha manzana? La idea del experimento mental es llevarnos a concluir que, en realidad, Mary sí aprende algo nuevo al ver la manzana, a saber: cómo se ve el rojo. El problema es que, si es cierto que ella adquiere un nuevo conocimiento respecto de la percepción del color rojo, entonces el fisicalismo falso, dado que, *ex hipotesi*, Mary sabe todo al respecto de los procesos físicos que ocurren durante la percepción del color. Si el fisicalismo fuera cierto—sugiere Jackson—Mary no aprendería nada nuevo. No obstante, el experimento mental parece llevarnos a creer lo contrario.

5. Algunas enfoques filosóficos y científicos

Lo que estos y otros ejemplos—como el famoso argumento de los qualia invertidos (Shoemaker, 1981)—buscan generar en nosotros es la idea, aparentemente clara y distinta, de que los qualia no pueden ser reducidos a la física ni, *a fortiori*, a la neurofisiología, porque ningún reporte físico o neurofisiológico podrá dar cuenta de ellos. Y en virtud de esta dificultad han venido apareciendo toda suerte de respuestas. Hacerle

justicia a la plétora de hipótesis que se han tejido alrededor del problema de la conciencia es muy difícil. Por ello, el mapa conceptual que a continuación presento no busca, en modo alguno, ser ni exhaustivo ni crítico; mi objetivo es meramente el de presentarle al lector las ideas principales de algunas reconocidas perspectivas, así como ciertas dificultades a las que se enfrentan.

Un primer grupo de teorías—que siguiendo a Francisco Varela (1996) podemos denominar “misterianistas”—sostiene que la ciencia no puede dar cuenta del fenómeno de la conciencia. Existen varias versiones de esta tesis. Están, por ejemplo, los que se inclinan a pensar que el problema de los qualia está más allá de la capacidad explicativa de la ciencia *actual* (Nagel, 1998). De acuerdo con esta perspectiva, aunque en principio es posible que los seres humanos lleguemos a resolver el problema de la conciencia, ello no ocurrirá sino hasta que nuestra ciencia contemporánea sufra un cambio revolucionario que nos permita incluir fenómenos subjetivos dentro de sus marcos explicativos. Otra perspectiva sostiene que la conciencia está vinculada a procesos mecánico-cuánticos que tienen lugar en los microtúbulos de las neuronas y que, siendo esa su naturaleza, no puede ser explicada en términos de la física contemporánea. (Penrose, 1989). Por ello, la física tendrá primero que resolver el asunto de las explicaciones causales en mecánica cuántica antes de poder dar cuenta del problema de la conciencia. Y están, finalmente, los que optan por una salida casi kantiana, es decir, aquellos que, tras sostener que el problema de los qualia y de la conciencia supera la capacidad cognitiva del hombre (McGinn, 1989), creen que cualquier investigación empírica de la conciencia estará siempre destinada a fracasar (ver, también, White, 1986).

Pero la mayoría de los filósofos y los neurocientíficos son un poco más optimistas que los misterianistas. En términos generales, los intentos por resolver el problema de la conciencia podrían dividirse—con mucha dificultad—según la clase de pregunta que, primordialmente, buscan responder. Hay teorías metafísicas que se concentran, por ejemplo, en la naturaleza de la conciencia, y que nos dicen muy poco al respecto de los mecanismos en virtud de los cuales surge la experiencia consciente. Otras, no obstante, se enfocan más bien en darle respuesta a la pregunta por el cómo surge la conciencia, mientras que otras buscan averiguar por qué razón los seres humanos—y posiblemente algunos animales—son conscientes. Finalmente, están aquellas que se concentran en indagar acerca de cuál pueda ser el mejor modo para estudiar dicho fenómeno. Obviamente, todos estos conjuntos comparten muchos de sus componentes. Como veremos, hay enfoques cognitivos, por ejemplo, que son netamente materialistas, así como aproximaciones fenomenológicas dualistas y hasta eliminativistas. Veamos, a grandes rasgos, cuales son los lineamientos fundamentales de algunas de estas teorías.

5.1 Teorías dualistas

Tradicionalmente, los enfoques metafísicos pueden dividirse en dos grandes grupos. Al primero se le conoce como *dualismo*, y su tesis principal es que la mente y el cuerpo (o, en general, la materia) son ontológicamente distintos. Existen diversas versiones de dualismo, según el modo en que se traza dicha distinción. La versión más fuerte de dualismo, y que comúnmente se asocia con Descartes, se conoce como *dualismo de sustancias*. De acuerdo con esta versión, la mente y el cuerpo son dos sustancias enteramente distintas. Como la conciencia es función de la mente y no del cuerpo, es perfectamente plausible que haya cuerpos que se comporten cual si fueren

conscientes sin serlo. A estas entidades filosóficas se les conoce como zombis, y para muchos, el que puedan ser lógicamente concebibles habla a favor del dualismo (Chalmers, 1996). Si bien el dualismo de sustancias no goza actualmente de gran aceptación, debido en gran medida a su incapacidad por explicar satisfactoriamente el modo en que una mente inmaterial puede ejercer poder causal sobre el cuerpo, su valor histórico es incuestionable, así como su relación con el modo en que la mayor parte de las personas conceptualiza a la mente humana (ver el capítulo XX de este volumen, De Brigard, 2003a y Bloom, 2004).

Una segunda versión de dualismo conocida como *dualismo de propiedades*, sostiene que la distinción ontológica entre mente y materia no se da al nivel de las entidades, sino al nivel de las propiedades que dichas entidades instancian. Así, aunque es posible que sólo haya un tipo de sustancia en el mundo, las propiedades mentales y las propiedades físicas no sólo no son idénticas entre sí, sino que tampoco son reductibles las unas a las otras. Actualmente nos encontramos con al menos dos versiones relativamente distintas de dualismo de propiedades. En primer lugar, está la versión de dualismo que el filósofo David Chalmers (1996) ha propuesto. Según esta perspectiva, las propiedades mentales conscientes existen a la par con las propiedades físicas de la materia, de modo que ambas han de contarse como componentes fundamentales de la realidad. Y en segundo lugar, está el llamado *dualismo de propiedades emergentes*, según el cual las propiedades mentales conscientes, aunque distintas de las propiedades físicas de la materia, emergen de las complejas configuraciones de estas últimas (Hasker, 1999). En cierta medida, esta suerte de dualismo emergentista nos recuerda al viejo *monismo neutral* de William James, según el cual tanto la materia como la mente eran resultado de diversas configuraciones de una misma realidad que, en últimas, no era ni mental ni física (James, 1904; ver también Russell 1985, para una exposición más condensada).

Finalmente, hay una última versión de dualismo a la que se le conoce como *epifenomenalismo*. De acuerdo con esta teoría, aunque los eventos mentales conscientes son producto de causas físicas, ellos en sí mismos no son causa de ningún efecto físico. Si bien el epifenomenalismo tiene la ventaja de permitir las innegables conexiones entre las causas físicas y los efectos mentales—entre, por ejemplo, sufrir una herida y sentir dolor por ella, o recibir una cierta configuración lumínica en la retina y percibir un cierto objeto en la lejanía—se enfrenta sin embargo a problemas a la hora de intentar explicar como podemos tener conocimiento de nuestros propios estados mentales (para una defensa del epifenomenalismo en contra de esta crítica, ver Robinson 1982).

Al dualismo se le ha criticado desde muchos frentes. Hay quienes sostienen, por ejemplo, que tanto Descartes como Chalmers cometen el error de extraer una conclusión metafísica a partir de premisas epistemológicas: que podamos concebir a la mente sin el cuerpo, o que podamos imaginar un cuerpo que se comporte como si tuviera conciencia sin tenerla (i.e. que podamos concebir a un zombi), no significa que estemos hablando de posibilidades metafísicas genuinas (ver, por ejemplo, Dennett, 2005). En otras palabras: que algo sea concebible no significa que sea ni física ni metafísicamente posible. Otras críticas, motivadas por consideraciones metafísicas metodológicas, apuntan al hecho de que, puesto que el dualismo no parece representar un avance explicativo significativo, mejor haríamos quedándonos con una teoría más parsimoniosa que apelase a una en lugar de dos sustancias o propiedades irreducibles. Finalmente, otras críticas de índole empírica, sostienen que el dualismo es simplemente incapaz de acomodar una gran

cantidad de evidencia científica, por lo cual lo mejor es buscar alternativas monistas más acordes con nuestros marcos científicos.

5.2. *Teorías monistas*

Las teorías monistas, en consecuencia, constituyen el segundo grupo de enfoques metafísicos de la conciencia. Ellas comparten la idea de que sólo existe un único tipo de sustancia en el universo, a pesar de que difieren grandemente cuando se trata de describir, no sólo la naturaleza de esta sustancia, sino también el modo en que da lugar a la mente y a la experiencia conciente. Tal vez la más innovadora postura monista es el *panpsiquismo*. De acuerdo con una de sus versiones más contemporáneas, la de Galen Strawson (2006), aunque la materia prima del universo es únicamente física, nos hemos equivocado al definir “físico” en contraposición a “mental”. Para Strawson, absolutamente todo lo que existe exhibe—en mayor o menor grado—conciencia, subjetividad. A partir del hecho, evidente para él, de que los fenómenos conscientes son indudablemente parte fundamental de la realidad, y del hecho, evidente también para él, de que no parece posible explicar la emergencia de la conciencia a partir de componentes no-concientes, Strawson sugiere que la conclusión correcta es la de que toda la naturaleza, todos los componentes que conforman nuestro cerebro, nuestro cuerpo y nuestro entorno, son conscientes también. Así, el problema de la interacción mente-cuerpo desaparece, y la conciencia, al ser parte fundamental de la realidad, deja de ser un misterio. O se convierte en otro, quizás aun más difícil—puesto que, por una parte, dicha tesis resulta difícil de compaginar con la evidencia científica actual y, por otra, aunque dice resolver el problema mente-cuerpo, lo hace a expensas de crear el problema “mente-mente”: es decir, el problema de explicar cómo es posible que micro-conciencias den lugar a la macro-conciencia de *mi* propia subjetividad (Lycan, 2006).

Una segunda postura monista, radicalmente opuesta al panpsiquismo, se conoce como *eliminativismo*. De acuerdo con esta perspectiva, no sólo la física tradicional está en lo cierto al sostener que la materia es inconsciente, sino que los propios términos “conciencia” y “qualia” no se refieren a ningún tipo de realidad. Aunque no deja de ser arbitrario, es posible dividir las perspectivas eliminativistas en dos grupos: aquellas que vienen de la neurociencia y aquellas propuestas por filósofos. En realidad, ambos grupos le guardan igual pleitesía al enfoque científico, y aunque con frecuencia sus conclusiones se asemejan, los argumentos tienden a diferir. De las teorías eliminativistas propuestas por la neurociencia hablaré más adelante, cuando me refiera a los enfoques funcionalistas y neurocognitivos de la conciencia. Ahora, dentro de los filósofos eliminativistas, quizás los más famosos sean Paul y Patricia Churchland, por una parte, y Daniel Dennett, por la otra. Tanto los Churchland como Dennett consideran que la difícil tarea de explicar la naturaleza de la conciencia y de los qualia puede ser llevada a cabo por una buena teoría que ponga el significado de dichos términos en el lugar adecuado de una ciencia objetiva, es decir, en tercera persona. Sin embargo, ambos difieren a la hora de explicar el modo en que la ciencia—se supone—ha de proporcionarnos una explicación. Mientras que para los Churchland la neurociencia será capaz de explicar la conciencia a nivel meramente “subpersonal” (1995)—es decir, sin incluir datos sobre el sujeto de la experiencia, sino únicamente sobre partes de la experiencia de ese sujeto—Dennett cree que el nivel “personal” de explicación es indispensable (1981). Esto no implica, en todo caso, que la teoría no pueda ser todo lo objetiva que se quiera—a fin de cuentas, ella pondrá al

descubierto que “*qualia* es un término de los filósofos que sólo produce confusión y no refiere al final a propiedades o características en lo absoluto” (Dennett, 1988).

Un tercer enfoque monista, que gozó de gran popularidad entre 1950 y 1970, se conoce como “teoría de identidad” (Place 1956; Smart 1959). De acuerdo con esta perspectiva, tanto los estados conscientes como los estados cerebrales existen. Sin embargo, contrario a los dualistas, los teóricos de la identidad sostienen que ambos estados son idénticos, i.e. que no son más que dos maneras de referirse a la misma realidad. Mas si es verdad que son idénticos, se podría preguntar uno, ¿cómo puede ser posible que no nos hayamos dado cuenta? La respuesta es sencilla: no todas las identidades son evidentes. Hace un par de siglos la gente no sabía que el contenido de un vaso de agua era idéntico a un conglomerado de moléculas con estructura H₂O. La ciencia, no obstante, reveló que “agua” y “H₂O” son dos modos distintos de referirse a una misma entidad. De modo similar—sostienen los teóricos de la identidad—la ciencia demostrará que el término “conciencia” comparte el mismo referente con alguna expresión, existente o no, del lenguaje de la neurofisiología. Infortunadamente, la teoría de identidad pronto se encontró con devastadoras críticas. Por ejemplo, Saúl Kripke (1980) presentó argumentos muy sugestivos para demostrar la imposibilidad de que la relación entre los estados mentales y los estados cerebrales pueda tratarse de una identidad. Igualmente, Jerry Fodor (1974) nos proporcionó un buen número de argumentos para convencernos de que las propiedades mentales, en realidad, pueden ser realizadas por diversos tipos de sustratos, de modo que hablar de una identidad entre los dos resulta, a lo menos, chauvinista (ver también Block, 1980). No obstante las críticas, la teoría de identidad ha visto una suerte de renacimiento en los últimos años, propiciado primordialmente por el filósofo inglés David Papineau (1994, 2002), para quien los *qualia*—aunque no muchos otros estados intencionales—sí tienen las características necesarias para soportar una identidad con los estados cerebrales. La teoría de identidad, en consecuencia, vuelve a estar sobre la mesa.

El filósofo John Searle ha abogado por una suerte de *naturalismo biológico* según en cual la conciencia es una propiedad biológica emergente. Aunque muchos consideran que su teoría es sospechosamente cercana a un dualismo de propiedades, él insiste en que es una subespecie de monismo. Para Searle, los fenómenos mentales son causados por procesos neurofisiológicos y, por tanto, deben considerarse como características del cerebro (Searle, 1992). Ahora, la razón por la cual tan obvia conclusión nos parece sorprendente, es porque nos hace falta reconocer que en la naturaleza se pueden dar dos clases de fenómenos biológicos ontológicamente distintos. Las montañas, los glaciares, las lágrimas y los átomos, por ejemplo, son fenómenos ontológicamente objetivos, en tanto que no dependen de un sujeto para existir. Los dolores, las penas y las cosquillas, por el contrario, son ontológicamente subjetivos: ellos no pueden existir independientemente del sujeto (humano o animal) que los experimenta (Searle, 1998). La conciencia, de acuerdo con esta perspectiva, es un fenómeno ontológicamente subjetivo que, no obstante, puede ser estudiado a partir de un punto epistemológicamente objetivo, como el de la ciencia. ¿Cómo puede la ciencia, desde su perspectiva de tercera persona, abordar un fenómeno tan disímil ontológicamente a aquellos con los que normalmente trata? Y, más aún, ¿cómo puede dar cuenta de una cadena causal (supuestamente emergente) entre fenómenos ontológicamente objetivos y rasgos ontológicamente subjetivos? Su teoría, infortunadamente, no nos lo dice.

5.2.1. Teorías filosóficas (neuro)funcionalistas

Finalmente nos encontramos con la teoría materialista que, posiblemente, goza de más popularidad en la actualidad: el *funcionalismo*. De acuerdo con esta perspectiva, los estados mentales se definen según los roles funcionales que juegan al interior de un determinado sistema organizado (del cual nosotros, los seres humanos, somos solamente un caso). En esta medida, el funcionalismo es “ontológicamente neutro”, i.e. es independiente del sustrato en el cual la función se realiza. Mas como hoy día la mayor parte de filósofos funcionalistas son, a la vez, monistas y materialistas, no es extraño que se considere al funcionalismo como otra teoría materialista más. Las teorías varían, no obstante, en función del nivel de abstracción con que se describen los procesos mentales que dan lugar a la conciencia. Existe, pues, un continuo entre las teorías funcionalistas de la conciencia, empezando por aquellas que condicionan las descripciones de los roles funcionales a nivel neuronal y molecular (Bickle, 2003), hasta aquellas cuyas descripciones de los procesos mentales son mucho más abstractas, al punto en que casi ni mencionan las estructuras cerebrales en que se realizan. Ambos extremos tienen sus ventajas y sus desventajas. Los modelos de “bajo nivel”—por llamar de algún modo a los primeros—tienen la gran ventaja de ser neuroanatómicamente más realistas, lo cual les permite nutrirse de evidencia empírica. Infortunadamente, el excesivo énfasis en el funcionamiento de las estructuras cerebrales puede ocasionar un alejamiento del nivel psicológico: el nivel de los contenidos de las experiencias conscientes. Si la brecha entre los dos persiste, se corre el riesgo de que se forme una “capa limítrofe” entre los dos niveles que, en ocasiones, resulta simplemente inmanejable (De Brigard, 2006). De otro lado, los modelos de “alto nivel” tienen la gran ventaja de que pueden incorporar el lenguaje psicológico de los contenidos de la experiencia consciente sin mayor problema. Sin embargo, al encontrarse tan alejados de las restricciones impuestas por nuestra neuroanatomía, modelos funcionalistas de la conciencia que sean demasiado abstractos pueden resultar “intraducibles” al lenguaje de la implementación física, por lo cual su estatus de hipótesis científica se pone en entredicho—y, por tanto, su candidatura como posibles respuestas al problema de la conciencia.

A pesar de ello, los modelos de nivel alto son bastante populares en filosofía de la mente. Casi todos ellos defienden una u otra perspectiva *representacionalista*. De acuerdo con el representacionismo, los estados conscientes son representaciones mentales, y sus propiedades *qua* estados conscientes se agotan en sus propiedades *qua* representaciones (Harman, 1990; Dretske, 1995). Sin embargo, como el representacionismo, por sí mismo, no es una teoría de la conciencia sino únicamente una teoría de la naturaleza de los estados conscientes, algunos filósofos han utilizado esta perspectiva para soportar tesis más robustas sobre la naturaleza de la conciencia. Probablemente las dos teorías representacionistas más importantes del momento son la *teoría del pensamiento de alto nivel* [*Higher-order thought*] (o HOT, por su sigla en inglés) y la *teoría de la percepción de alto nivel* [*Higher-order perception*] (o HOP). La primera, propuesta inicialmente por David Rosenthal, sostiene que lo que hace que un estado mental sea consciente es el estar acompañado de otro estado mental, asertórico y no-inferencial, cuyo contenido es, precisamente, aquél estado mental (Rosenthal, 1986). Consecuentemente, los estados mentales inconscientes serán aquellas representaciones mentales que carecen de un estado mental de alto nivel para el cual son sus contenidos.

Aunque HOT parece poder contestar unas cuantas preguntas al respecto de la naturaleza de la conciencia, se le critica el que los estados conscientes se parecen mucho más a las percepciones que a los pensamientos. La inmediatez de la experiencia consciente, por ejemplo, no parece explicable si los estados conscientes tienen la naturaleza de un juicio; así tampoco el hecho de que algunas características de dichos estados mentales sean—aparentemente—inarticulables y, tal vez, no conceptuales. Las teorías HOP (Lycan 1987; 1996) logran sortear estos inconvenientes al identificar los estados mentales de alto nivel con estados representacionales perceptuales, cual si estos fueren estados funcionales de un sistema de monitoreo del sentido interno. Infortunadamente, al decirnos muy poco al respecto de cual pudiera ser la estructura cerebral encargada de dicho monitoreo, es difícil acercar la teoría HOP al examen empírico de la neurociencia.

Quizás la mejor alternativa para el funcionalismo sea la de adoptar un punto de vista de “nivel intermedio”: una perspectiva que atienda tanto a restricciones neuroanatómicas como a las descripciones psicológicas de los contenidos de la experiencia consciente. Algunas de las primeras propuestas funcionalistas de este estilo se encuentran en obras como *Consciousness and the Computational Mind*, de Ray Jackendoff (1987) y, en cierta medida, en *Consciousness Explained*, de Daniel Dennett (1991). Para este último, por ejemplo, una de las lecciones más importantes que debemos aprenderle a la neurociencia es que no parece haber un lugar especial en el cerebro donde nuestro flujo de conciencia “se reúne”. Aunque nuestro cerebro está constantemente operando contenidos mentales, estos no tienen una trayectoria determinada hacia un centro de la conciencia, o un “teatro cartesiano”, como lo llama él. La conciencia parece tener un modo de ser más *abstracto*. En consecuencia—nos sugiere Dennett—lo que la ciencia de la conciencia debe hacer es explicar por qué reportamos ciertos contenidos de nuestros estados mentales en lugar de otros; por qué—por ponerlo en sus más recientes términos—ciertos contenidos mentales alcanzan la “fama cerebral” (*cerebral celebrity*), puesto que “la conciencia se parece más a la fama que a la televisión; *no* es un “medio de representación” especial localizado en el cerebro al que deben transducirse los eventos con contenido para volverse conscientes.” (Dennett, 2005, 136. Ver también De Brigard, 2007). Finalmente, vale la pena mencionar una de las más recientes y representativas teorías funcionalistas de nivel intermedio: la teoría AIR, sugerida inicialmente por Jesse Prinz (2000; en preparación). De acuerdo con esta teoría, la conciencia se origina cuando, y sólo cuando, la atención hace que las representaciones perceptuales de nivel intermedio (i.e. información que ya ha sido sensitivamente procesada y está en camino de ser conceptualmente procesada) estén disponibles para que puedan entrar a la memoria de trabajo. De ahí el nombre AIR: representaciones atendidas de nivel intermedio (*Attended Intermediate-level Representations*).

5.2.2. Teorías neurocognitivas

Cuando los filósofos proponen teorías de la conciencia que no sólo atienden al nivel psicológico y fenomenológico de descripción de los contenidos conscientes, sino también a la evidencia neurológica que constriñe las posibilidades de implementación de sus modelos, resulta muy difícil dejar de agrupar sus teorías junto con los enfoques neurocientíficos reduccionistas. Porque algunos neurocientíficos—como Gerald Edelman, Rodolfo Llinás, Francis Crick y Christof Koch, entre muchísimos otros—creen que es la neurociencia la que podrá proporcionarnos una respuesta para el problema de la

conciencia. Crick y Koch, por ejemplo, sostienen que el mejor supuesto con que contamos para hablar de la conciencia y, *a fortiori*, para hablar de los qualia, es que dichos procesos se dan en el cerebro—hipótesis ésta que sólo podremos revalidar cuando “haya evidencia clara, decisiva para lo contrario (como opuesta a argumentos por ignorancia)” (Crick y Koch, 2000). Ambos soportan una idea metodológica reduccionista heredada de Attneave (1961) y de su metáfora del homúnculo inconsciente. En 1961, Attneave propuso la idea de que el yo consciente era una ilusión producida por el funcionamiento de la formación reticular que, a manera de un homúnculo inconsciente, hace conscientes representaciones de su propia actuación “en la forma de imaginiería y discurso hablado o no hablado” (Crick y Koch, 2000)—una conclusión parecida a la del hemisferio izquierdo como cerebro intérprete, que fuera extraída por el Dr. Gazzaniga de sus experimentos con pacientes comisurotomizados (por ejemplo, Gazzaniga, 1988). Crick y Koch complementan esta postura al sugerir que la base neuronal de la conciencia—su correlato neurológico, como a veces se le llama (Metzinger, 2000)—es producto de oscilaciones neuronales que se ponen en fase a una frecuencia de 40 Hz. No obstante, cautelosamente Crick y Koch renuncian a la idea de que ese reporte neurológico sea respuesta definitiva al problema de los qualia y de la subjetividad. Para ellos, los qualia “son posiblemente causados por la actividad de una pequeña fracción de todas las neuronas del cerebro, localizadas estratégicamente entre los mundos externos e interno” (Crick y Koch, 2000), es decir, el mundo del homúnculo inconsciente, pero—concluyen un poco más adelante—que cómo pueda esto producir un mundo subjetivo, sigue siendo un misterio.

Otra famosa teoría científica de la conciencia, proveniente esta vez de la ciencia cognitiva, fue propuesta inicialmente por Bernard Baars (1988). De acuerdo con esta perspectiva, conocida como “teoría del espacio de trabajo global” (*Global Workspace Theory*), la conciencia corresponde a una suerte de módulo de capacidad informativa limitada por los recursos cognitivos de la atención, de forma que los contenidos que se hacen conscientes son aquellos que “entran” al foco de atención en el espacio de trabajo. Mas esta, de nuevo, es únicamente una de las muchas teorías que encontramos en el mercado (Baars, 2003). Listarlas todas no sólo resultaría muy difícil sino, posiblemente, inútil. Baste con mencionar que la mayor parte de ellas busca darle respuesta al problema de la conciencia a partir del hallazgo de un correlato neurológico necesario y suficiente para que se de la experiencia conciente. Pero a pesar de la gran cantidad de hipótesis, la naturaleza de este correlato sigue siendo un misterio.

Así termina este breve, somero y nada exhaustivo recuento de algunas de las perspectivas filosóficas y científicas más importantes sobre la naturaleza de la conciencia. Cada una de ellas se aproxima al problema a partir de un enfoque ontológico y/o metodológico distinto, de forma que al sugerir respuestas relativamente distantes, cada perspectiva abre toda suerte de prometedoras líneas de investigación. Y así como en la ciencia no existe un consenso unificado respecto a la respuesta del problema de la conciencia—por no mencionar respecto a la manera en que debe siquiera abordarse—la filosofía tampoco ofrece explicaciones definitivas. A pesar de ello, el análisis concienzudo del concepto que subyace a nuestro uso del término conciencia tarde o temprano habrá de producir frutos en la práctica y, en particular, en la práctica médica. ¿Cuál puede ser la relevancia del estudio filosófico de la conciencia para la psiquiatría y

la neuropsicología? Con un intento por responder esta pregunta concluyo, a continuación, el presente capítulo.

6. El problema de la conciencia y su relevancia para la psiquiatría y la neuropsicología.

Hace unos años el filósofo Thomas Metzinger se preguntó cuál es la relevancia, para la filosofía, del estudio de los trastornos mentales (Metzinger, 2003). Yo quiero concluir formulando exactamente la pregunta opuesta: ¿cuál es la relevancia del análisis filosófico de la conciencia para el estudio de los trastornos mentales? Quizás la primera y más obvia consecuencia del análisis conceptual de la noción de conciencia tiene que ver con su clarificación lingüística. Desambiguar el término ‘conciencia’ no sólo lograría unificar su uso en la nosología psiquiátrica y neuropsicológica, sino que, además, contribuiría a darle un sentido mucho más unificado a las sintomatologías (ver, por ejemplo, Beitman y Nair, 2004; Kircher y David, 2003). Descripciones de sintomatologías más claras y definidas acarrearían, a su vez, criterios de diagnóstico precisos, y probablemente universales.

Y por si esto fuera poco, una segunda consecuencia práctica podría también seguirse. Supongamos que, luego de un completo—o, al menos, suficientemente avanzado—estudio filosófico y científico del problema de la conciencia, una determinada teoría funcionalista de corte cognitivo llegara a ser capaz de acomodar toda la evidencia disponible, proporcionándonos al mismo tiempo predicciones acertadas. Así las cosas, la aplicación de dicha teoría a la hora de determinar posibles tratamientos para los trastornos mentales de la conciencia se haría inminente, quizás incluso necesaria. De un lado, los presupuestos ontológicos subyacentes determinarían los posibles mecanismos de intervención, de tal modo que, si la teoría llegase a explicitar la manera en que la conciencia es realizada a nivel neurobiológico en el cerebro, ciertos tratamientos psicofarmacológicos se harían más o menos sugestivos. De otro lado, el propio lenguaje de la teoría ayudaría a demarcar los caminos que subsiguientes investigaciones habrían de recorrer, garantizando así—tanto como sea posible—el que los estudios llevados a cabo bajo tal marco teórico conduzcan a la consecución de resultados fácilmente interpretables.

Finalmente, es plausible pensar en una última consecuencia teórica que, no por ello, resulta menos llamativa ni importante. Me refiero a la curiosa posibilidad de llegar a entender, de una vez por todas, cuál es la relación de la conciencia con las demás facultades cognitivas del ser humano. Muchas y muy variadas hipótesis han sido propuestas para intentar explicar la manera en la que la conciencia cuadra al interior del mapa cognitivo. Daniel Schacter (1989), por ejemplo, propone entender a la conciencia fenoménica—es decir, aquella facultad cognitiva con que experimentamos la particular sensación subjetiva producida por nuestro estar en un determinado estado de conciencia—como si fuese un módulo de acceso entre los sistemas operativos que generan “conocimiento” de un objeto o evento tanto externo como interno, y un sistema ejecutivo central que, a su vez, estaría encargado del control directo de nuestra capacidad racional, y de nuestra habilidad para reportar percepciones introspectivas y generar acciones guiadas. Así, de acuerdo con este modelo, la función de la conciencia es la de integrar y transmitir información proveniente de nuestros mecanismos preceptuales y

conceptuales a los mecanismos con que llevamos a cabo razonamientos, reportes verbales y acciones.

Ahora, ¿es ésta la función de la conciencia? En un sentido, el modelo de Schacter parece cuadrar bastante bien con la evidencia empírica y experimental. Por ejemplo, el fenómeno de la ceguera cortical [*blindsight*] podría explicarse con tan sólo incluir un paso directo entre la información sensitiva y la respuesta conductual, una vía de acceso que se “salte”, por decirlo así, al módulo de la conciencia fenoménica. De ahí que los pacientes con ceguera cortical puedan discriminar acertadamente varios estímulos sin ser conscientes de que los están viendo. Sin embargo, esta no parece ser toda la historia. Algunos filósofos, como Ned Block (1995), han argumentado que el modelo explicativo de Schacter confunde dos nociones independientes y distintas de ‘conciencia’, y que en ausencia de una noción clara del fenómeno que se intenta entender, todo experimento o estudio empírico que se realice conforme a dicho modelo correrá el riesgo de generar que resultados ambiguos e inconclusos. Pero mi interés no es dimitir esta contienda a favor de uno u otro combatiente. Por el contrario, lo que intento es resaltar la estrecha relación mutualista que se establece entre el filósofo de la mente y el psiquiatra o el neuropsicólogo. El primero se encarga de proporcionar la estructura conceptual con que han de entenderse los datos que proporciona el segundo. Sin el riguroso estudio científico de los psiquiatras, psicólogos y neuropsicólogos, la filosofía de la mente y, en particular, el estudio de la conciencia, estaría andando absolutamente a tientas. No obstante, sin el trabajo conceptual y de clarificación que ofrecen la filosofía y, en cierta medida, las ciencias cognitivas, la evidencia empírica proporcionada por la ciencia correría también el riesgo vagar sin rumbo. Cuando se trata de entender la naturaleza de alteraciones de la conciencia, ambas podrán avanzar sólo si trabajan juntas.

Referencias

- American Psychiatric Association, 1994. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Press.
- Aristóteles. 1994. *Acerca del Alma*. Madrid: Gredos.
- Attneave, F. 1961. “In defense of homunculi” En: *Sensory Communication*. W. A. Rosenblith (ed.). Cambridge, MA: MIT Press.
- Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baars, B.J.; Banks, W.P.; y Newman, J.B. 2003. *Essential Sources in the Scientific Study of Consciousness*. Cambridge, Ma: MIT Press.
- Beitman, B. y Nair, J. 2004. *Self-Awareness Deficits in Psychiatric Patients*. Norton: N.Y.
- Bickle, J. 2003. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Kluwer.
- Block, N. 1980. “Troubles with Functionalism” En: *Readings in the Philosophy of Psychology*, Volume 1, Ned Block, ed., Cambridge, MA : Harvard University Press.
- Block, N. 1995. “On a confusion about the function of consciousness.” *Behavioral and Brain Sciences*, 18: 227-47.

- Block, N., Flanagan, O., y Güzeldere, G. 1997. *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Bloom, P. 2004. *Descartes' Baby: How the science of child development explains what makes us human*. New York: Basic Books.
- Botero, J. J. 2003. "The Intractable First-Person" (ms)
- Chalmers, D. 1995. "Facing up to the problem of consciousness". *Journal of Consciousness Studies*, 2: 200-19.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Churchland, P. M. 1995. *The Engine of Reason and Seat of the Soul*. Cambridge, MA: MIT Press.
- Crick, F. and Koch, C. 1990. "Toward a neurobiological theory of consciousness". *Seminars in Neuroscience*, 2: 263-75.
- Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt.
- Damasio, A. 2000. "A Neurobiology for Consciousness". En: T. Metzinger (ed.). *Neural Correlates of Consciousness*. Cambridge, Ma: MIT Press.
- De Brigard, F. 2003a. "Eliminando el fantasma de la máquina. Del alma al software 1". *Revista Colombiana de Psiquiatría*. 32 (2): 175-192.
- De Brigard, F. 2003b. "En busca de la mente cerebral. Del alma al software 2". *Revista Colombiana de Psiquiatría*. 32 (4): 373-390.
- De Brigard, F. 2006. "Capas limítrofes y dominios de evidencia en ciencia cognitiva". *Universitas Philosophica*. 46: 53-77.
- De Brigard, F. 2007. "Reseña de "Sweet Dreams: Philosophical Obstacles to a Science of Consciousness". Daniel C. Dennett. (MIT Press: Cambridge, MA. 2005)". *Ideas y Valores: Revista Colombiana de Filosofía*. 56 (134): 129-136.
- Dennett, D. C. 1981. "Three Kinds of Intentional Psychology". En: R. Healey (ed). *Reduction, Time and Reality*. Cambridge: Cambridge University Press. Reimpreso en Dennett, 1987.
- Dennett, D. C. 1987. *The Intentional Stance*. Cambridge, Ma: MIT Press.
- Dennett, 1988. "Quining Qualia". En: A. Marcel y E. Bisiach (eds). *Consciousness in Modern Science*. Oxford: Oxford University Press. Reimpreso en Block et al. (eds), 1997.
- Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown and Company.
- Dennett, D. C. 2005. *Sweet Dreams*. Cambridge, MA:MIT Press.
- Descartes, R. 1985. *The Philosophical Writings of Descartes*. J. Cottingham, R. Stoothoff, y D. Murdoch (trd.). Volúmenes I, II y III. Cambridge: Cambridge University Press.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, Mass: The MIT Press, Bradford Books

- Feinberg, T. E. 2001. *Altered Egos*. Oxford: Oxford University Press.
- Flashman, L.A. 2004. “Disorders of Insight, Self-Awareness, and Attribution in Schizophrenia”. En: Beitman y Nair, 2004.
- Jackendoff, R. 1987. *Consciousness and the Computational Mind*. Cambridge, Mass.: MIT Press,
- Fodor, J. 1974. “Special sciences”. *Synthese*, 28: 77-115.
- Gazzaniga, M. 1988. *Mind Matters: How Mind and Brain Interact to Create our Conscious Lives*. Boston: Houghton Mifflin.
- Gerrans, P. 2002. “Multiple Paths to Delusion”. *Philosophy, Psychiatry, & Psychology*. 9(1): 65-72.
- Harman, G. 1990. “The intrinsic quality of experience”. In J. Tomberlin, ed. *Philosophical Perspectives*, 4. Atascadero, CA: Ridgeview Publishing .
- Hasker, W. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Jackson, F. 1982. “Epiphenomenal qualia”. *Philosophical Quarterly*, 32: 127-136.
- James, W. 1890. *The Principles of Psychology*. New York: Henry Holt and Company.
- James, W. 1904. “Does Consciousness Exist?” *Journal of Philosophy, Psychology, and Scientific Methods*, 1, 477-491.
- Kant, I. 1787/1929. *Critique of Pure Reason*. Translated by N. Kemp Smith. New York: MacMillan.
- Kircher, T. y David, A. 2003. *The Self in Neuroscience and Psychiatry*. Cambridge: Cambridge University Press.
- Kraepelin, E. 1896. *Psychiatrie. Ein kurzes Lehrbuch für Studierende und Aerzte*. 5th ed. Leipzig: Barth.
- Kripke, S. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Levine, J. 1983. “Materialism and Qualia: The Explanatory Gap”. *Pacific Philosophical Quarterly*. 64: 354-61.
- Lewis, C.I., 1929. *Mind and The World Order: an Outline of a Theory of Knowledge* . Charles Scribner's Sons: New York.
- Locke, J. 1688/1959. *An Essay on Human Understanding*. New York: Dover.
- Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Lycan, W. 2006. “Resisting ?-ism”. *Journal of Consciousness Studies*. 13(10-11): 65-71.
- Metzinger, T. ed. 2000. *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Metzinger, T. 2003. *Why are identity-disorders interesting for philosophers?* En: T. Schramme y J. Thome (eds.). *Philosophy and Psychiatry*. Berlin: de Gruyter.

- McGinn, C. 1989. “Can we solve the mind-body problem?” *Mind*, 98: 349-66
- Mellor, C. S. 1970. “First Rank Symptoms of Schizophrenia” *The British Journal of Psychiatry*. 117: 15-23.
- Nagel, T. 1974. “What is it like to be a bat?” *Philosophical Review*, 83: 435-456.
- Nagel, 1998. “Conceiving the Impossible and the Mind-Body Problem”, *Philosophy*, 73 (285): 337-352
- Navarro, F.; Cardeño, C.A.; Cano, J.F.; Gómez, J.M.; Jiménez, K.; Palacio, C.A.; y García, J. 2006. “Introspección en pacientes con psicosis”. *Revista Colombiana de Psiquiatría*. 35(1): 61-70.
- Nelkin, N. 1996. “Subjectivity”. En: Guttenplan, S. *A Companion to the Philosophy of mind*. London: Blackwell.
- Papineau, D. 1994. *Philosophical Naturalism*. Oxford: Blackwell
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.
- Penrose, R. 1989. *The Emperor's New Mind: Computers, Minds and the Laws of Physics*. Oxford: Oxford University Press.
- Place, U. T. 1956. “Is consciousness a brain process?” *British Journal of Psychology*, 44-50.
- Prinz, J. 2000. “A Neurofunctional Theory of Visual Consciousness”. *Consciousness and Cognition*. 9: 243-259.
- Prinz, J. (En preparación). *The Conscious Brain*. Oxford: Oxford University Press.
- Robinson, W. S. 1982. “Causation, Sensations and Knowledge”. *Mind* 91: 524-540
- Rosenthal, D. 1986. “Two concepts of consciousness.” *Philosophical Studies*, 49: 329-59.
- Russell, 1985. *The philosophy of Logical Atomism*. Open Court: Illinois.
- Sass, L.A. y Parnas, J. 2003. “Schizophrenia, Consciousness, and the Self”. *Schizophrenia Bulletin*. 29(3):427-444
- Schacter, D. 1989. “On the relation between memory and consciousness: dissociable interactions and consciousness.” In H. Roediger and F. Craik eds. *Varieties of Memory and Consciousness*. Hillsdale, NJ: Erlbaum.
- Schneider, K. 1959. Schneider, K. *Clinical Psychopathology*. New York: Grune and Stratton.
- Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press
- Searle, J. 1997. *The Mystery of Consciousness*. New York: A New York Review Book.
- Searle, J. 1998. *Mind, Language and Society, Philosophy in the Real World*. New York: Basic Books, 1998
- Shoemaker, S. 1981. “Absent qualia are impossible.” *Philosophical Review*, 90: 581-99.

- Smart, J. 1959. “Sensations and brain processes.” *Philosophical Review*, 68: 141-56.
- Strawson, G. 2006. “Realistic Monism: Why Physicalism Entails Panpsychism”. *Journal of Consciousness Studies*. 13 (10-11): 3-31.
- Triesman, A. y Gelade, G. 1980. “A feature integration theory of attention.” *Cognitive Psychology*, 12: 97-136.
- Varela, F. J. (1996). “Neurophenomenology: A Methodological Remedy for the Hard Problem”. *Journal of Consciousness Studies*, 4(4), 344-346.
- White, S. 1986. “Curse of the Qualia”. *Synthese*, 68: 333-368