

Filosofía y (Neuro)ciencia: Sobre la “naturalización” de la filosofía práctica.

Daniel González Lagier
Filosofía del Derecho
Universidad de Alicante

1. Cuestiones regionales y universales en la Filosofía.

Hay cuestiones filosóficas susceptibles de un enfoque “regional”¹, y hay también cuestiones filosóficas que son todo lo contrario. Algunos problemas de este segundo tipo también son de relevancia para los primeros, porque no pueden ser obviados por estos enfoques y se convierten muchas veces en su marco o su presupuesto. Uno de estos problemas filosóficos universales que, sin embargo, es necesario plantearse como presupuesto para el desarrollo de una filosofía del Derecho “regional” es el de la relación entre la filosofía (en especial nos interesa la filosofía práctica) y la ciencia. Es un viejo problema, pero el desarrollo actual de la neurociencia lo ha vuelto a poner en primer plano, a veces con el rótulo de “naturalización de la filosofía”. Puede afirmarse que la filosofía en torno a la responsabilidad, la filosofía práctica, está experimentando un proceso de “naturalización”, que es una consecuencia o manifestación de lo que se ha llamado “la naturalización de la filosofía” en general². En este trabajo trataré de dar cuenta de algunos intentos y algunos problemas de “naturalizar” la filosofía práctica y la noción de responsabilidad.

2. ¿Qué entienden los filósofos por “naturalizar”?

No hay un claro acuerdo acerca de qué entender por “naturalización” de la filosofía, pero se podría coincidir en que implica la reconstrucción de la filosofía a partir de conceptos admitidos -o, al menos, admisibles- por las ciencias de la naturaleza o ciencias empíricas (frecuentemente identificadas, en

¹ La expresión se usa en el sentido desarrollado por Manuel Atienza en “Una propuesta de filosofía del Derecho para el mundo latino”, en *Doxa*, núm. 30, 2007..

² Para un análisis de este fenómeno puede verse T. Grimaldos y J. Pacho (eds.), “La naturalización de la filosofía: problemas y límites”, en *Pre-Textos*, 2005.

un sentido estricto de “ciencia”, con la física y la química)³. Esta es una caracterización amplia (y mínima) que incluye tanto a las posturas más científicas, que pretenden la reducción de la filosofía (o alguna parte de la misma) a alguna ciencia empírica (como propuso Quine a propósito de la epistemología, a la que consideraba psicología cognitiva⁴) como las posturas menos radicales, que entienden que hay cierto continuo entre filosofía y ciencia, aunque ambas puedan tener dominios distinguibles. El enorme desarrollo que las investigaciones sobre el funcionamiento del cerebro y el sistema nervioso están teniendo desde mediados del siglo XX, junto con el papel central de los procesos cerebrales en el razonamiento y la acción, explican que la neurociencia se haya convertido en el paradigma conceptual y metodológico al que se está intentando reconducir la filosofía práctica.

Suele situarse el nacimiento de la neurociencia moderna en 1888, cuando Santiago Ramón y Cajal descubre que el cerebro es una red de células, las neuronas. Posteriormente, Charles Sherrington analizó las conexiones entre ellas y Edgard Adrian registró actividad bioeléctrica en todo el sistema nervioso. Los avances en psicofarmacología a mediados del siglo XX y, sobre todo, la aparición de las técnicas de neuroimagen en los noventa y principios del siglo XXI, han dado un impulso decisivo a la neurociencia. La gran mayoría de noticias espectaculares que trascienden a la opinión pública tienen que ver con estas técnicas de neuroimagen, que han dado lugar a lo que se ha llamado la neurociencia cognitiva, esto es, el estudio del funcionamiento del cerebro en los procesos de adquisición del conocimiento y de la formación de estados mentales⁵. Del tronco central de la neurociencia se han ramificado nuevas disciplinas que pueden verse como aplicaciones de la misma a distintos ámbitos. En 2002 se celebró en San Francisco el congreso *Neuroética: esbozando un mapa del terreno* que consagraba esta nueva disciplina,

³ Carlos Moya, “La naturalización de la responsabilidad moral”, en T. Grimaldos y J. Pacho (eds.), *La naturalización de la filosofía: problemas y límites*, pág. 59.

⁴ W.V. Quine, “Naturalización de la epistemología”, en *La relatividad ontológica y otros ensayos*, Ed. Tecnos, 2002.. Este artículo puede verse como uno de los impulsos más importantes de la naturalización de la mente desde el lado de la filosofía.

⁵ Una descripción detallada de la evolución del conocimiento del cerebro desde el Antiguo Egipto puede verse en J.P. Changeux, *El hombre neuronal*, Ed. Espasa-Calpe, 1985. Véase también, J. M. Giménez Anaya y J. I. Murillo, “Mente y cerebro en la neurociencia contemporánea. Una aproximación a su estudio interdisciplinar”, en *Scripta Theologica*, 39, 2007, págs. 607 y ss.

dedicada a la aplicación de la neurociencia a los temas tradicionales de la ética, y en 2007 la MacArthur Foundation creó el proyecto “Derecho y Neurociencia” para reunir a varios conocidos neurocientíficos, filósofos y juristas de diversos países con el fin de profundizar en la intersección entre las neurociencias y el Derecho.

En realidad, hay dos maneras de entender la “neuroética” y el “neuroderecho”: (1) como *ética (o Derecho) de la neurociencia*, una parte de la ética (o del Derecho) que trataría de establecer un marco ético (o jurídico) para las investigaciones neurocientíficas y sus aplicaciones; y (2) como *neurociencia de la ética (o del Derecho)*, el estudio de la conducta ética (o de problemas jurídicos) desde el punto de vista de las investigaciones sobre el cerebro. Algunos problemas de la ética y el Derecho de la neurociencia son los siguientes: Si está justificado o no el uso de los descubrimientos neurocientíficos para la mejora de las capacidades mentales o sensoriales de los humanos (el llamado “transhumanismo”), en qué condiciones es legítimo el uso en los tribunales de pruebas basadas en técnicas neurocientíficas (como la prueba P300 o brainfingerprinting, que permite determinar si el sujeto miente observando las variaciones en las ondas cerebrales ante ciertos estímulos), qué valor en relación con la atribución de responsabilidad hay que conceder a determinadas disfunciones cerebrales, o si es correcto –y en qué casos- usar técnicas de control de la conducta basadas en conocimientos neurocientíficos. Y algunos problemas de la neurociencia de la ética y del Derecho son los siguientes: La discusión general sobre el libre albedrío y su relación con la responsabilidad, la cuestión de la fundamentación de las normas penales en emociones como la repugnancia, el análisis del papel de la oxitocina o de las llamadas “neuronas espejo” en nuestra conducta ética, o si la neurociencia puede fundamentar conclusiones normativas acerca de la corrección de nuestros juicios morales (o nuestras decisiones jurídicas).

La “neuroética” y el “neuroderecho”, entendidos como neurociencia de la ética y neurociencia del Derecho, pueden verse como el principal intento actual de naturalizar la filosofía práctica y la responsabilidad⁶ (aunque el ámbito de la

⁶ Una reflexión acerca del impacto que la neurociencia puede tener en el Derecho puede verse en M. Narváez, “El impacto de la neurociencia en el Derecho: El caso de la responsabilidad subjetiva”, en Revista Telemática de Filosofía del Derecho, núm. 15, 2012. Una muestra de los

filosofía práctica puede considerarse más amplio que el de la “filosofía de la responsabilidad”, me centraré en este segundo).

3. Las dos vías de “naturalización” de la responsabilidad: La naturalización de las normas y la naturalización de la mente.

De acuerdo con la concepción clásica de la responsabilidad (estoy pensando en la responsabilidad moral y jurídico-penal, entendida como reproche), para adscribir responsabilidad a un agente es necesario –al menos– (1) que exista un sistema de reglas (con el que juzgamos la conducta del individuo) y (2) que el agente haya actuado libremente, en el doble sentido de que tenga *libertad de acción* (que nuestras acciones sean consecuencias de la combinación de nuestros deseos con las creencias acerca de cómo satisfacerlos) y *libertad de voluntad* (que esos deseos y creencias sean a su vez, al menos en cierto grado, libres, controlables por el agente)⁷. Consecuentemente, el proceso de fundamentar la filosofía práctica a partir de las ciencias empíricas se está llevando a cabo por dos vías (siendo necesario recorrer las dos para el éxito completo del proyecto): La primera, la “naturalización de la normatividad”; la segunda, la “naturalización de la mente” (y, con ella, de la acción).

La naturalización de las normas es el intento de dar cuenta de la normatividad a partir de las ciencias empíricas, en la línea del “darwinismo moral” de Spencer, que trataba de fundamentar las normas éticas en el proceso evolutivo, o la “sociobiología” de E.O. Wilson, que pretendía explicar el comportamiento de todos los animales extrapolando el egoísmo genético del mundo de los genes a los demás ámbitos de la vida. En la actualidad, desde la neuroética muchos autores (J. Haidt, M.Hauser, Patricia Churchland, M.S. Gazzaniga,...) proponen dar cuenta de las opiniones y creencias morales como un conjunto de intuiciones, emociones y capacidades en gran parte innato, inscrito en nuestro cerebro por las “fuerzas” de la evolución: “Hemos desarrollado –escribe Hauser– un instinto moral, una capacidad que surge en

temas tratados bajo el rótulo de neuroderecho puede verse en S. Zeki y O. Goodenough, *Law & the Brain*, Oxford University Press, 2006.

⁷ Asumo, de entrada, la idea tradicional de que la responsabilidad presupone la libertad, sin entrar de momento en las discusiones entre compatibilistas e incompatibilistas.

cada niño, diseñada para generar juicios inmediatos sobre lo que está moralmente bien o mal, sobre la base de una gramática inconsciente de la acción. Una parte de esta maquinaria fue diseñada por la mano ciega de la selección darviniana millones de años antes de que apareciera nuestra especie; otras partes se añadieron o perfeccionaron a lo largo de nuestra historia evolutiva y son exclusivas de los humanos y de nuestra psicología moral”⁸.

La naturalización de la mente, por su parte, es el intento de reconstruir los conceptos mentales -como creencia, decisión, intención, deseos, emociones, dolor, etc.- de manera que puedan ser aceptados por las ciencias de la naturaleza. La neurociencia ofrece técnicas –como, por ejemplo, la resonancia magnética funcional- que permiten detectar los cambios en el flujo sanguíneo en el cerebro en el momento en que el individuo realiza determinadas tareas motoras o está en ciertos estados cognitivos o emocionales, lo que parece permitir correlacionar estados mentales con estados cerebrales, de manera que ante la presencia de un estado cerebral determinado se podría suponer que el sujeto tiene el estado mental correspondiente. Esto sugiere una estrecha conexión entre estados mentales y estados cerebrales, conexión que los “neurofilósofos” interesados en la mente pueden tratar de usar para explicarla (veremos en el apartado 6 estos intentos de explicación de la mente).

4. Algunos problemas de la naturalización de las normas.

Aunque las aportaciones de los autores que han enfocado la normatividad desde la neurociencia son heterogéneas, se pueden señalar las siguientes tesis como una generalización aproximada de sus propuestas:

- a) Los juicios morales provienen más de corazonadas, intuiciones y emociones que de la razón. Se apoya, por tanto, una concepción de la ética más cercana a un emotivismo moral como el de Hume que al racionalismo de Kant.
- b) La razón tiene meramente un papel racionalizador “ex post” de nuestras opiniones morales, rara vez decisivo en la conformación del juicio moral. Y, cuando sí tiene carácter decisivo, no lo tiene directamente, sino por la vía de

⁸ M. Hauser, *La mente moral*, Paidós, 2008, pág. 17.

permitirnos ver otros puntos de vista que favorecen el surgimiento de nuevas intuiciones morales. Por ello, somos presa de dos ilusiones: “La primera puede ser denominada *ilusión de mover el perro*: creemos que nuestro propio juicio moral (el perro) es conducido por nuestro propio razonamiento (la cola). La segunda ilusión puede ser denominada *ilusión de mover la cola del otro perro*: En un argumento moral, esperamos que la réplica eficaz a los argumentos de un oponente cambie la mente de éste. Tal creencia es como pensar que forzar el movimiento de la cola de un perro, moviéndola con tu mano, hará feliz al perro”⁹.

c) Este conjunto de intuiciones y emociones (en las que la empatía y la capacidad de reciprocidad tienen un papel fundamental) es en gran medida innato y en buena parte resultado de la historia de la evolución humana, y sólo en una pequeña medida adquirido culturalmente o por medio de la educación (y ello en los estrechos márgenes que permiten nuestras capacidades, condicionadas evolutivamente). “La moral, como el lenguaje –escribe J. Haidt– es una gran adaptación evolutiva de una especie intensamente social, construida en múltiples regiones del cerebro y del cuerpo, y es mejor descrita como emergente que como aprendida”¹⁰.

d) Las capacidades morales innatas del hombre, y sus intuiciones y emociones, pueden servir de suelo para fundamentar una moral de carácter universal¹¹.

Los problemas a los que se enfrenta esta vía de naturalización de la normatividad son enormes. Me limitaré a apuntarlas:

a) Falta una adecuada distinción entre capacidades morales (esto es, capacidad para poder plantearse y tratar de resolver problemas morales) y propiamente normas o valores morales. Como ha señalado Francisco Ayala, es necesario distinguir dos problemas distintos: “(1) ¿está determinada por la naturaleza biológica del ser humano la capacidad para la ética?; y (2) están determinados biológicamente los sistemas o códigos de normas éticas aceptadas por los seres humanos?”¹² (su respuesta es afirmativa a la primera

⁹ J. Haidt, “El perro emocional y su cola racional”, en Adela Cortina (Ed.), “Neurofilosofía práctica”, Ed. Comares, 2012, pág. 184.

¹⁰ J. Haidt, “El perro emocional y su cola racional”, pág. 192.

¹¹ M.S. Gazzaniga, *El cerebro ético*, Ed. Paidós, 2006, pág.179.

¹² F. J. Ayala, *Evolución, ética y religión*, Universidad de Deusto, 2013, pág. 61.

pregunta y negativa a la segunda). La confusión entre ambas cuestiones es apreciable en algunos neuroéticos.

b) Hay una constante tendencia a pasar del “ser” al “deber ser”, en sustituir explicación de la normatividad por justificación de las normas. Esto ha sido denunciado por muchos, pero es especialmente relevante la denuncia de un neuroético como Joshua Greene, quien (refiriéndose a las tesis de Casebeer según las cuales, de acuerdo con la neurociencia, la toma de decisiones real se parece a la recomendada por Aristóteles), cuestiona: “¿Cómo, de la afirmación de que el pensamiento moral aristotélico es más natural que el de Kant y Mill, se sigue que el de Aristóteles es mejor?”¹³.

c) Los intentos de situar a la empatía y la capacidad de reciprocidad¹⁴ como fundamento de la moral plantean el problema del altruismo. Si realmente la moral nace a partir de la cooperación y benevolencia entre los integrantes de la comunidad, *porque de esa manera aumentamos nuestras posibilidades de supervivencia*, ¿cómo es posible derivar de ahí una benevolencia universal o la imparcialidad que parece exigir la ética? ¿Y cómo es posible explicar el altruismo desinteresado?

d) Y, relacionado con lo anterior: Supongamos que aceptamos la tesis fuerte de acuerdo con la cual la evolución ha inscrito en nuestro cerebro no sólo capacidades morales, sino algún tipo de código de normas o valores. ¿No es precisamente el hecho de que esas intuiciones y corazonadas en las que se manifiesta tal código sean un residuo biológico de la evolución en unas circunstancias ya superadas lo que las hace sospechosas o, en muchos casos, directamente inaceptables? Como señala Peter Singer, “la investigación reciente en neurociencia nos aporta nuevas y poderosas razones para tomar una postura crítica hacia las intuiciones comunes”¹⁵. Tenemos intuiciones de acuerdo con las cuales es peor matar a alguien que dejarlo morir (causar un

¹³ J. Greene, “Del ‘es’ neuronal al ‘debe’ moral”, en Adela Cortina (Ed.), “Neurofilosofía práctica”, Ed. Comares, 2012, pág. 151.

¹⁴ Cuyo origen se analiza a partir de datos neurofisiológicos, como las “neuronas espejo”. Estas son un tipo de neuronas, descubiertas por Giacomo Rizzolatti en los simios, incluyendo los humanos, que se activan tanto cuando realizamos una acción como cuando vemos que otros la realizan, y que se supone que posibilitan, sin necesidad de razonamientos o inferencias, comprender las acciones del otro, sus intenciones y sus emociones. Serían, por tanto, las neuronas de la “empatía”.

¹⁵ P. Singer, “Ética e intuiciones”, Adela Cortina (Ed.), “Neurofilosofía práctica”, Ed. Comares, 2012 pág. 297.

daño por acción es peor que permitir ese mismo daño por omisión), o que es prioritario ayudar a nuestra familia o a los miembros de nuestra comunidad, frente a las necesidades de personas de lugares lejanos a las que no conocemos. Podemos explicar esas intuiciones como convenientes para la adaptación a las circunstancias en la que ha tenido lugar nuestra evolución. ¿Quiere eso decir que tales intuiciones tienen hoy algún valor normativo? Lo que la neuroética parece mostrar es más bien “la necesidad de separar aquellos juicios morales que debemos a nuestra historia evolutiva y cultural de aquellos otros que tienen un fundamento racional”¹⁶.

Toda esta discusión es de gran relevancia para la comprensión de la naturaleza de la ética misma. No obstante, no me voy a ocupar aquí más de la “naturalización de la normatividad”, sino de la otra vía, que, en cierto sentido, es lógicamente previa.

5. Algunos problemas de la naturalización de la mente: Tres objeciones.

El intento de naturalizar la mente plantea tres objeciones filosóficas básicas, que podemos llamar a) la objeción del reduccionismo, b) la objeción del determinismo y c) la objeción de la falacia mereológica.

a) *La objeción del reduccionismo*: De acuerdo con esta objeción las propuestas de naturalización de la mente basadas en las aportaciones de la neurociencia identifican dos fenómenos que, en realidad, son distintos: la mente y el cerebro. Al hacerlo, incurren en un reduccionismo ilegítimo, dado que los estados mentales tienen ciertos rasgos (en lo que sigue me referiré a ellos como “las propiedades de lo mental”) que no pueden ser explicados en términos de actividad neurofisiológica y, en general, encajan mal con lo que las ciencias empíricas nos dicen del mundo: En primer lugar, nuestros estados mentales nos son accesibles a nosotros mismos, por introspección, de una manera directa, al margen de la evidencia empírica y de inferencias a partir de ella (es el rasgo de la *conciencia*). En segundo lugar, y relacionado con lo anterior, hay cierta diferencia cualitativa en cómo emergen a mi conciencia los dolores, deseos o preocupaciones, o en la manera en que experimentamos escuchar un

¹⁶ P. Singer, “Ética e intuiciones”, pág. 301.

concierto de Bach o el sabor de un trozo de chocolate. Puede decirse, entonces, que estos estados mentales se corresponden con diferentes sensaciones internas¹⁷, que los filósofos -en analogía con el término *quanto* usado en física- llaman *qualia*. En tercer lugar, muchos estados mentales (las creencias, las intenciones,...) poseen un contenido, un significado, versan sobre otros hechos (es el rasgo de la “intencionalidad” o “contenido mental”). En cuarto lugar, los estados mentales se relacionan con nuestra conducta externa, pareciendo tener eficacia causal sobre ella (es el fenómeno de la *causación mental*)¹⁸. Conciencia, carácter cualitativo, intencionalidad y eficacia causal son cuatro características de la mente que no es evidente que puedan explicarse por referencia exclusiva a procesos físico-químicos y leyes empíricas. Supongamos que la mente se identifica con el cerebro y que los estados mentales son conexiones eléctricas, físicas o químicas entre neuronas: ¿cómo es posible que algo material, físico, nos resulte accesible sin observación externa y posea un contenido semántico?¹⁹ Y si la mente no forma parte del mundo físico, ¿cómo puede interactuar causalmente con él? Las dificultades para dar cuenta de las propiedades de lo mental han llevado a muchos filósofos a ignorarlas o, incluso, a negarlas, presentándonos un mundo sin una verdadera “vida mental”.

b) La objeción del determinismo: De acuerdo con este argumento, identificar la mente con el cerebro lleva a aceptar los postulados del determinismo. Desde el ámbito del Derecho, Michele Taruffo ha expresado esa posible objeción en un artículo sobre “Proceso y neurociencia”²⁰. En palabras suyas: “Se trataría, de hecho, de adoptar una perspectiva teórica estrictamente materialista y determinista, según la cual toda la vida mental, espiritual, volitiva y cognitiva se reduciría al funcionamiento de las células cerebrales y sus conexiones. Siguiendo esta perspectiva, se acabaría descartando que exista la voluntad,

¹⁷ T. Nagel, “Qué se siente ser un murciélago”, Ed. Fondo de Cultura Económica, 2000.

¹⁸ Una explicación más detallada de estos rasgos puede verse en C. Moya, *Filosofía de la mente*, Capítulo I (“Algunos rasgos de nuestro concepto preteórico de lo mental”), Universidad de Valencia, 2006.

¹⁹ Una presentación de una estrategia para la naturalización de la intencionalidad puede verse en Carlos Moya, “Las emociones y la naturalización de la intencionalidad”, en *Anales del Seminario de Metafísica*, núm. 28, Ed. Complutense, 1994.

²⁰ M. Taruffo, “Proceso y neurociencia”, en M. Taruffo y J. Nieva (Dirs.), *Neurociencia y proceso judicial*, Ed. Marcial Pons, 2013.

que se pueda elegir y tomar decisiones, que se pueda hablar de responsabilidad por los propios comportamientos, que exista conciencia de estos comportamientos y que estos puedan referirse a conocimientos de cualquier naturaleza. En esencia, el hombre se vería reducido a una especie de máquina del cerebro, y toda su vida estaría determinada mecánicamente por el funcionamiento sin control de esta máquina”²¹. Las consecuencias son tan distorsionadoras de nuestra visión del mundo que podrían verse como un buen argumento para rechazar la identificación entre mente y cerebro, si realmente se siguieran de esta identificación. De hecho, algunos autores, como Peter Strawson, han tratado de mostrar la imposibilidad de aceptar el determinismo a partir de la constatación de que ciertas actitudes espontáneas y naturales hacia las acciones de los demás, como el resentimiento, el agradecimiento, el perdón, el enfado, etc., que no tenemos cuando aceptamos que el sujeto no podía haber actuado de otro modo, están en la base de nuestras relaciones interpersonales y, en definitiva, de la vida social²².

c) *La objeción de la falacia mereológica*: La tercera objeción afirma que al identificar mente y cerebro se estaría incurriendo en una falacia. Al identificar cierta actividad cerebral detectada por alguna técnica neurocientífica con una intención, una mentira, un recuerdo o una emoción, se están atribuyendo estos hechos al cerebro, o a una parte del mismo, cuando en realidad son estados sólo atribuibles al sujeto como un todo. Por ejemplo, señala también Taruffo que “aun admitiendo que una determinada actividad cerebral sea necesaria para mentir (al igual que para decir la verdad), sin embargo resulta conceptualmente imposible identificar la mentira con esa actividad (...) Las mentiras no suceden en el cerebro”²³; igualmente, “las intenciones no son procesos cerebrales, y del mismo modo la intencionalidad no se 'coloca' en una zona del cerebro y no se reduce a un estado cerebral”²⁴. Asimismo, “la conciencia no es algo que 'sucede' en el cerebro, como la digestión se produce en el estómago, sino que implica el contacto con el mundo exterior, en una compleja interacción de cerebro, cuerpo y mundo”²⁵. Toda estas

²¹ M. Taruffo, “Proceso y neurociencia”, pág. 23.

²² P. Strawson, “Libertad y resentimiento”, Ed. Paidós, 1995.

²³ M. Taruffo, “Proceso y neurociencia”, pág. 22.

²⁴ M. Taruffo, “Proceso y neurociencia”, pág. 20

²⁵ M. Taruffo, “Proceso y neurociencia”, pág. 21

identificaciones, en definitiva, "incurren en una falacia fundamental, que consiste en atribuir a una parte del sujeto humano, es decir, al cerebro, habilidades y funciones que son propias del sujeto en su totalidad, con toda la complejidad que se manifiesta en la vida"²⁶.

Taruffo toma esta objeción de los trabajos de Michael S.Pardo y Dennis Patterson, quienes, a su vez, como veremos más adelante, aplican al ámbito jurídico una crítica que Peter Hacker y Maxwell Bennet, un filósofo y un neurocientífico, dirigen a los fundamentos filosóficos de la neurociencia (y a destacados filósofos, como John Searle y Daniel Dennet, que siguen de cerca en sus explicaciones de la mente y la conciencia las aportaciones de la neurociencia), considerando que ésta incurre en lo que ellos llaman la "*falacia mereológica*", al atribuir a una parte (el cerebro) lo que debería ser atribuido al todo (la persona).

6. Un mapa de teorías de la mente.

Para valorar el alcance de las anteriores objeciones es necesario conocer con más detalle los distintos intentos de establecer una relación entre la mente y el cerebro. Para ello trataré de presentar un mapa de las teorías de la mente más extendidas y conocidas, clasificándolas a partir de sus respuestas a una serie de preguntas encadenadas:

1) ¿Está la realidad constituida por entidades físicas y mentales, entendidas como sustancias separadas, o sólo existe un tipo de entidades?

El *dualismo*, como el sostenido por Descartes, sostiene lo primero; el *monismo*, lo segundo. El dualismo ha sido la concepción de la mente dominante en la historia de occidente al menos desde Descartes y ha "empapado" el lenguaje común y lo que podríamos llamar la "psicología popular" (esto es, la teoría que comúnmente usamos en las explicaciones de la conducta y las atribuciones de estados mentales). De acuerdo con el dualismo, el pensamiento y el mundo físico son sustancias distintas, cada una de ellas con propiedades distintas (para Descartes, por ejemplo, la mente es consciente, se conoce directamente, es libre, indivisible e indestructible,

²⁶ M. Taruffo, "Proceso y neurociencia", pág. 23.

mientras que el cuerpo posee extensión, se conoce indirectamente, es determinado, infinitamente divisible y destructible). Hoy en día las propiedades que los dualistas asignarían a cada sustancia no son las mismas, pero la idea de dos tipos de sustancias con propiedades irreducibles permanece, por lo que la mente no puede ser estudiada por las ciencias de la naturaleza.

2) Si se asume el monismo, ¿qué tipo de sustancia compone la realidad?

El monismo puede ser *idealista* o *materialista*²⁷. El monismo idealista se inclina del lado de la mente: por ejemplo, al sostener, como Berkeley, que las cosas físicas son ideas en la mente de Dios, o al analizar la realidad, como el positivismo lógico, como una construcción a partir de datos de los sentidos. El monismo materialista es el intento de decantar la distinción hacia el lado de lo físico, entendiendo que lo físico es lo único que es irreduciblemente real y que, por lo tanto, la mente debe ser explicada a partir de lo material (o eliminada). Así como el dualismo es la concepción que está detrás de la psicología popular, el materialismo es la concepción dominante hoy entre científicos y filósofos de la mente.

3) Si se asume el monismo materialista, ¿tiene sentido hablar de estados mentales?

Para el dualismo, los *estados* mentales existen genuinamente en el mundo mental, por lo que los *conceptos* mentales forman legítimamente parte de la psicología, que es una ciencia distinta de las ciencias físicas. Pero si se acepta el monismo materialista (dejaré de lado al idealismo, que es una postura prácticamente abandonada hoy día), entonces se plantea la cuestión de si es posible acomodarlos en una concepción fisicalista del mundo. Surge, entonces, la alternativa entre el eliminacionismo, el reductivismo y el no reductivismo. El *materialismo eliminacionista* cree que la tarea de encontrar los correlatos neuronales de los estados mentales está condenada al fracaso, porque no hay estados mentales, tal como los considera la psicología y el sentido común.

²⁷ Dejo de lado el monismo neutral de autores como Bertrand Russell, para quien el componente último del mundo es una sustancia que no es ni material ni espiritual (éstas serían formas de observar esa sustancia neutral). Puede verse, por ejemplo, B. Russell, *Análisis de la materia*, 1969 o *El conocimiento humano*, 1993. Para una explicación actual (y defensa) del monismo neutral puede verse Rafael A. Alemañ Berenguer, "La relación mente-materia y el monismo neutral", en *Naturaleza y libertad. Revista de estudios interdisciplinarios*, núm. 1, 2012.

Presuponer que la neurociencia permitirá establecer una correlación entre estados cerebrales y mentales es dar carta de existencia a estos últimos. Pero la psicología está equivocada, trata de dar explicaciones de la conducta que la investigación científica mostrará que no logran decir nada verdadero sobre la realidad. Como la teoría del flogisto (una supuesta sustancia, en realidad inexistente, que se creía que se liberaba en los procesos de combustión u oxidación), las teorías y los conceptos mentales están condenados a desaparecer, simplemente porque hacen referencia a cosas que no existen. A medida que avance la neurociencia, podremos sustituir el lenguaje de los estados mentales por el lenguaje de los estados cerebrales, de igual manera que el descubrimiento de los virus y las bacterias permitió rechazar las explicaciones de las enfermedades en términos de demonios, maldiciones y espíritus. Éste sería, simplemente, un ámbito más donde la extensión de la ciencia permitiría la expulsión de la superstición. El *materialismo reduccionista* no impugna el sentido de los conceptos mentales (mantiene la psicología como una disciplina epistemológicamente adecuada o conveniente, al menos en cierto grado), pero cree que las teorías mentales ganarán en precisión y poder explicativo si son traducidas o reducidas a un lenguaje en última instancia no mental. Una teoría es reducida a otra, de acuerdo con Paul Churchland, cuando “una teoría nueva y eficaz logra abarcar un conjunto de proposiciones y principios que reflejan perfectamente (o casi perfectamente) las proposiciones y principios de una teoría o marco conceptual anteriores”²⁸. Si el ejemplo preferido de los eliminacionistas es el flogisto, el de los reduccionistas es la explicación del calor en términos de promedio de energía cinética de las moléculas (a diferencia de lo que ocurría con el flogisto, las teorías o leyes que usan el concepto de calor, como la que afirma que el calor dilata los metales, son útiles, pero su reducción a la nueva teoría proporciona una descripción más precisa del fenómeno). El *materialismo no reduccionista* sostiene que la traducción del lenguaje mental al lenguaje físico es imposible sin pérdidas significativas: aunque haya una sola sustancia, lo físico, podemos hablar de estados o propiedades cerebrales y de estados o propiedades mentales. El materialismo no reduccionista puede verse como un monismo respecto de las

²⁸ P. Churchland, *Materia y conciencia*, Ed. Gedisa, 1999, pág. 51.

sustancias y, simultáneamente, un dualismo respecto de las propiedades. Su desafío, por tanto, es explicar cómo de lo físico pueden surgir las propiedades de lo mental.

4) Si se asume que hablar de estados mentales no es un sinsentido, esto es, si se asume un materialismo no eliminacionista (reductivista o no) ¿qué son realmente los estados mentales?

Se han propuesto varias alternativas:

a) La primera opción consiste en identificar lo mental con la conducta o las disposiciones de comportamiento. A esta estrategia se le ha llamado *conductismo lógico o conceptual*. Lo que tienen en común las distintas concepciones conductistas (como las de Carnap, Ryle o Wittgenstein) es la idea de que existe algún tipo de relación constitutiva e interna entre la mente y el comportamiento²⁹. Esta conexión surge por el modo de construcción de los conceptos mentales y el aprendizaje de los términos correspondientes. De acuerdo con el conductismo este proceso de aprendizaje no va de lo “interno” a lo “externo”: no reconocemos un dolor en nosotros mismos y luego aplicamos el concepto a otros (si así fuera, como no es posible saber que lo que sienten los otros es exactamente lo mismo, no habría criterios de corrección en el uso de las expresiones mentales, esto es, no tendrían significado), sino que aprendemos a usar el concepto a partir de las manifestaciones externas del dolor que observamos en los demás y en nosotros mismos. El conductismo es una concepción reductivista: el lenguaje mentalista debe ser traducido a la descripción del comportamiento y a la expresión de disposiciones conductuales. “En su forma más fuerte y más directa -escribe Paul Churchland- el conductismo filosófico postula que toda oración acerca de un estado mental se puede parafrasear, sin pérdida de significado, por una oración larga y compleja acerca de cuál sería la conducta observable que se produciría si una determinada persona se encontrara en esta, o aquella o cualquier otra circunstancia observable”³⁰. Volveremos más adelante sobre el conductismo filosófico, puesto que la objeción según la cual la identificación neurociencia

²⁹ C. Moya, *Filosofía de la mente*, pág. 43.

³⁰ P. Churchland, *Materia y conciencia*, pág. 46.

incurrir en lo que hemos llamado la falacia mereológica parece conducir a esta postura o a una similar.

b) La segunda estrategia consiste en identificar los estados mentales de un sujeto con sus estados cerebrales. Esto es lo que hace el *materialismo de la identidad de tipos*, un materialismo reduccionista que sostiene que cada *tipo* de estado mental es una y la misma cosa que un *tipo* de estado cerebral, o que cada propiedad mental es idéntica a una propiedad cerebral. Esta teoría no niega que existan estados mentales: lo que niega es que sean algo distinto de estados del cerebro. De la misma manera que la luz son ondas electromagnéticas, el agua es H₂O y la temperatura es promedio de energía cinética molecular, las creencias son un determinado estado cerebral, los deseos otro, etc. Se trata de “identidades interteóricas”³¹. Para el materialismo de la identidad esta identificación entre estados mentales y estados cerebrales tiene el carácter de una hipótesis científica, que cuenta con algunos indicios a su favor y que se espera que llegue a ser confirmada en el futuro permitiendo la reducción de los conceptos mentales a la neurociencia. Un problema de esta teoría es que si los estados cerebrales y los estados mentales son la misma cosa, entonces todas las propiedades de los primeros han de ser también propiedades de los segundos, y viceversa. Esto es, los estados mentales deben tener propiedades espaciales (deben estar en algún sitio) y los estados cerebrales deben tener propiedades semánticas³². Las dificultades para entender esto suelen llevar a negar las propiedades de lo mental, o alguna de ellas.

c) Existe una versión distinta del materialismo de la identidad. Ante la desconfianza de que sea posible encontrar los correlatos neuronales de los estados mentales, algunos autores han defendido un tipo de materialismo que postula no una identidad entre tipos de estados mentales y tipos de estados cerebrales, sino sólo entre *casos* o *instancias* de estados mentales y *casos* o *instancias* de estados cerebrales. Son las *teorías de la identidad de instancias*. De esta forma, cada estado mental se corresponde con un evento físico, pero dos sujetos pueden tener el mismo estado mental y estar en estados cerebrales distintos, o tener estados mentales diferentes pese a estar en

³¹ P. Churchland, *Materia y conciencia*, pág. 52.

³² P. Churchland, *Materia y conciencia*, pág. 55.

estados cerebrales idénticos. Esto implica que no puede hacerse una reducción teórica -esto es, general- de los estados mentales a los estados cerebrales (aunque puede hacerse, como sugiere la siguiente estrategia, el funcionalismo, a otro tipo de entidades). Una de estas teorías es el *monismo anómalo* de Donald Davidson³³. Se trata de una teoría relevante, porque pretende compaginar el determinismo del mundo físico con el indeterminismo (al menos, epistemológico) de la conducta humana. Davidson parte de la asunción de la tesis de la identidad de instancias y de tres principios básicos: (a) El principio de interacción causal, de acuerdo con el cual al menos algunos eventos mentales interactúan causalmente con eventos físicos, (b) el principio del carácter nomológico de la causalidad, de acuerdo con el cual donde hay causalidad, debe haber una ley que correlacione tipos de eventos y (c) el principio del carácter anómalo de lo mental, según el cual no hay leyes estrictamente causales a partir de las cuales puedan predecirse los eventos mentales, dado que las interpretaciones de los estados mentales de otros sujetos -a diferencia de lo que ocurre con las relaciones causales entre eventos físicos- están sujetas al principio rector de la racionalidad.

d) La estrategia que tiene mayor difusión actualmente es el *funcionalismo*, que en realidad puede verse como una concepción que funde cierta herencia del conductismo con la tesis de la identidad de instancias. Para los funcionalistas los estados mentales se identifican con roles, papeles o funciones causales. De la misma manera que un reloj se caracteriza por su función para medir el tiempo con exactitud, un estado mental se caracteriza por la función que cumple. Dicha función se describe a partir de ciertas relaciones con el contexto (*inputs*) y con la conducta (*outputs*). Para el funcionalismo “lo esencial para que un determinado sujeto tenga la creencia de que *p* no es que en su cerebro haya una configuración neuronal de un tipo determinado, sino que se halle en un estado que guarde las relaciones adecuadas con otros estados y procesos y cumpla el papel causal adecuado en la organización de dicho sujeto”³⁴. El funcionalismo es compatible con el materialismo, porque tener una propiedad funcional requiere una estructura subyacente física, pero difiere del materialismo de la identidad de tipos porque admite que las mismas funciones

³³ D. Davidson, *Sucesos mentales*, en *Ensayos sobre acciones y sucesos*, Crítica, 1995.

³⁴ C. Moya, *Filosofía de la mente*, pág. 101.

podrían ser realizadas por estructuras físicas organizadas de manera distinta. Así, un ser (o una máquina) que tuviera una organización funcional suficientemente semejante a la de un hombre podría tener estados mentales. Por tanto, aunque todo estado mental se corresponde con un estado físico, no hay una correspondencia entre tipos de estados mentales y tipos de estados físicos. El funcionalismo puede verse, además, como una concepción reductivista, en el sentido de que asume la posibilidad de una descripción de las propiedades mentales que no haga referencia a otras propiedades mentales, pero a diferencia del conductismo y del materialismo de la identidad no reduce los estados mentales a estados cerebrales, sino a propiedades funcionales³⁵.

e) La última alternativa que analizaremos es la del materialismo emergentista. El *emergentismo* es una postura materialista que identifica los estados mentales con propiedades emergentes o sistémicas, es decir, propiedades que surgen a partir de cierto grado de complejidad de un sistema. Son propiedades del conjunto del sistema, pero no de cada una de sus partes. Para explicar qué es una propiedad emergente John Searle -cuya concepción, que él mismo llama *naturalismo biológico*, es una teoría emergentista- recurre al siguiente ejemplo: "Supongamos que tenemos un sistema S, compuesto de los elementos *a*, *b*, *c*... Por ejemplo, S podría ser una piedra y los elementos podrían ser moléculas. En general, habrá rasgos de S que no son, o no necesariamente, rasgos de *a*, *b*, *c*... Por ejemplo, S podría pesar 20 kilogramos, sin que las moléculas individualmente pesen 20 kilogramos. Denominemos a estos rasgos 'rasgos del sistema'. Algunos rasgos del sistema pueden ser deducidos, o determinados, o calculados a partir de los rasgos *a*, *b*, *c*..., simplemente por la forma en que se componen y ordenan (y, a veces, por sus relaciones con el entorno). Ejemplos de ellos serían la forma, el peso y la velocidad. Pero algunos otros rasgos del sistema no pueden ser determinados sólo a partir de los elementos que los componen y de las relaciones con el entorno: han de ser explicados a partir de las relaciones entre los elementos. Llamémosles 'rasgos del sistema causalmente emergentes'. La solidez, la liquidez y la transparencia son ejemplos de rasgos del sistema causalmente

³⁵ C. Moya, *Filosofía de la mente*, pág. 104. En contra, P. Churchland, *Materia y conciencia*, pág. 65.

emergentes”³⁶. De manera que, para Searle, una propiedad emergente es una propiedad que surge a partir de los elementos que componen el sistema, de las relaciones del sistema con el entorno y -esto parece lo decisivo- de las relaciones de esos elementos entre sí. Los estados mentales se identifican con propiedades emergentes de los cerebros. Los cerebros son sistemas físicos que han ido adquiriendo complejidad a lo largo de su historia evolutiva. El naturalismo biológico es descrito por Searle a partir de cuatro tesis³⁷:

- 1) Los estados mentales, con su ontología subjetiva de primera persona, son fenómenos reales del mundo real.
- 2) Los estados mentales son causados en su totalidad por procesos neurobiológicos de nivel inferior localizados en el cerebro.
- 3) Los estados mentales son rasgos del sistema cerebral en su conjunto y existen, por tanto, en un nivel superior al de las neuronas y sus sinapsis (“Por sí misma -dice Searle-, una neurona no es consciente, pero las partes del sistema cerebral compuestas por ella sí lo son”).
- 4) Los estados mentales, en la medida en que son reales, tienen eficacia causal (“Mi sed consciente, por ejemplo, me lleva a tomar agua”).

También ha sostenido una postura emergentista Mario Bunge (presentándola como una prolongación del programa de Santiago Ramón y Cajal de estudiar el espíritu en la materia), para quien (1) todos los estados, sucesos y procesos mentales son estados, sucesos y procesos en los cerebros de vertebrados superiores, (2) estos estados, sucesos y procesos emergen de los estados, sucesos y procesos de los componentes celulares del cerebro y (3) las relaciones entre la mente y el cerebro son relaciones entre subsistemas diferentes del cerebro o entre algunos de ellos y otros componentes del organismo³⁸.

El materialismo emergentista no es reductivista. De acuerdo con Searle, aunque los estados mentales pueden ser reducidos causalmente a -en el sentido de explicados por- los componentes del cerebro y sus relaciones, esta

³⁶ J. Searle, *El redescubrimiento de la mente*, Ed. Crítica, 1996, pág. 121.

³⁷ J. Searle, *La mente. Una breve introducción*, Ed. Norma, 2006, págs. 147 y 148.

³⁸ M. Bunge, *El problema mente-cerebro. Un enfoque psicobiológico*, Ed. Tecnos, 2002, pág. 42.

explicación no es suficiente, porque lo característico de lo mental es que es un fenómeno en primera persona, requiere un lenguaje subjetivo que no puede ser descrito con el lenguaje objetivo de la neurociencia. En el mismo sentido, para Bunge la neurofisiología es necesaria, pero no suficiente para dar cuenta de los estados mentales, dado que olvida conceptos como pensamiento, intención, etc.; y lo mismo puede decirse de la psicología, puesto que no presta atención al cerebro o Sistema Nervioso Central.

Teorías de la mente

1. Dualismo (dos sustancias)
2. Monismo (una sustancia).
 - 2.1. Monismo idealista (todo es espíritu).
 - 2.2. Monismo materialista (todo es materia).
 - 2.2.1. Materialismo eliminacionista (no hay estados mentales, es una descripción falsa).
 - 2.2.2. Materialismo reductivista (los estados mentales son estados de un tipo distinto).
 - 2.2.2.1. Tesis de la identidad de tipos (reducción a eventos neurofisiológicos)
 - 2.2.2.2. Conductismo (reducción a conducta).
 - 2.2.2.3. Funcionalismo (reducción a roles causales).
 - 2.2.3. Materialismo no reductivista.
 - 2.2.3.1. Tesis de la identidad de instancias (los estados mentales son estados físicos, pero no hay correlación entre tipos).
 - 2.2.3.2. Emergentismo (los estados mentales son propiedades emergentes. Monismo de sustancias, dualismo de propiedades).

7. Naturalización sin reduccionismo.

No puedo señalar aquí los problemas y las ventajas de cada una de estos enfoques distintos sobre la mente. Sin embargo, su exposición es suficiente para darse cuenta de que no todos los intentos materialistas de explicar la mente son eliminacionistas o reductivistas, ni lo son en el mismo grado. La objeción de eliminacionismo o reduccionismo (según la cual la neurociencia lleva a dejar de lado lo más peculiar de los estados mentales) no puede dirigirse, por ejemplo, contra el emergentismo, porque precisamente éste es un intento de mantener las propiedades de lo mental tal como se entienden en la psicología de sentido común. Precisamente, en mi opinión, sólo algún tipo de emergentismo causal puede satisfacer tres presupuestos que me

parecen básicos y que espero que se acepten, a efectos de este trabajo, sin necesidad de mayor discusión:

a) El primer presupuesto sería el "materialismo ontológico", esto es, la idea de que el mundo en el que vivimos es en última instancia el mundo físico o material. Es posible que tenga sentido hablar de un mundo de instituciones y de un mundo de ideas y conceptos (y es probable que en el análisis de estos mundos no haya que aceptar un "materialismo metodológico", esto es, la idea de que todo debe ser explicado de acuerdo con el modelo de las ciencias de la naturaleza), pero estos presuponen el mundo material (donde "presuponer" no significa que sean sin más "reducibles" a él). Este presupuesto nos compromete con el monismo materialista³⁹.

b) El segundo presupuesto sería el de la aceptación de la existencia de los estados mentales. Eso es algo de lo que todos tenemos experiencia directa. Es un dato (obtenido por introspección) innegable. Como ha insistido una y otra vez John Searle, el materialismo eliminativista "parece una locura. Parece una locura afirmar que yo nunca he sentido sed o deseos, que nunca he tenido dolor, que, en realidad, nunca he creído nada, o que mis creencias y mis deseos no juegan ningún papel en mi conducta"⁴⁰. Este presupuesto nos obliga a rechazar el materialismo eliminacionista.

c) El tercer presupuesto es la aceptación de que la mente (los estados mentales) tienen cuatro propiedades básicas: Son conscientes, poseen (al menos algunos estados mentales) *qualias*, pueden tener contenido semántico y pueden causar la conducta. Este presupuesto nos obliga a rechazar el reductivismo, pues las distintas formas de reductivismo plantean el mismo problema: no pueden dar cuenta del aspecto subjetivo o de primera persona de los estados mentales, al sustituirlo por un discurso en tercera persona. Las concepciones reductivistas, en palabras de Thomas Nagel, "omiten algo

³⁹ Las propias limitaciones del dualismo de sustancias nos conducen al materialismo, pues, en palabras de Searle: "1) Nadie ha logrado proporcionar jamás una descripción inteligible de las relaciones entre esos dos reinos. 2) la postulación es innecesaria. Es posible explicar todos los hechos de primera y todos los hechos de tercera persona sin postular reinos separados. 3) La postulación genera dificultades intolerables. De acuerdo con esta concepción, se hace imposible explicar de qué manera los estados y sucesos mentales pueden causar estados y sucesos físicos. En síntesis, es imposible evitar el epifenomenalismo" (J. Searle, *La mente. Una breve introducción*, pág. 168. También puede encontrarse una clara crítica al dualismo en P. Churchland, *Materia y conciencia*, págs. 40 y ss.

⁴⁰ J. Searle, *El redescubrimiento de la mente*, pág. 62.

esencial que descansa más allá de los fundamentos externamente observables para atribuir estados mentales a otros, a saber, el aspecto del fenómeno mental que es evidente desde la primera persona, el punto de vista interior del sujeto consciente: por ejemplo, la forma en que el azúcar te sabe o la forma en que ves el rojo o sientes la ira"⁴¹. Este rechazo del reduccionismo nos lleva a buscar una teoría que mantenga las propiedades de lo mental.

El materialismo emergentista trata de conciliar estos tres presupuestos, por lo que creo que debe ser asumido como un programa de investigación que permite evitar un reduccionismo eliminacionista, aunque ciertamente todavía no como una teoría plenamente satisfactoria (la clarificación de la noción de “emergencia” sigue siendo un tema pendiente de esta teoría).

8. Naturalización sin determinismo.

De acuerdo con el segundo argumento contra la neurociencia y la aplicación de sus técnicas al proceso, si se acepta la identificación de las actividades mentales con las actividades cerebrales, se acabaría descartando que exista la voluntad y toda explicación de la conducta debería hacerse en términos causales y mecanicistas, sin lugar para el libre albedrío. ¿Es esto necesariamente así? Es cierto que el determinismo se encuentra más cómodo entre las posturas materialistas y que la tesis del libre albedrío parece encontrar mejor acomodo en el dualismo, pero antes de afirmar que una cosa lleva a la otra es necesario tener en cuenta tres matizaciones: La primera es la discusión entre compatibilismo e incompatibilismo; la segunda es la distinción entre determinismo ontológico y determinismo epistemológico, y la tercera es el carácter gradual de la identificación entre mente y cerebro que podemos apreciar en las distintas teorías materialistas de la mente.

a) La compatibilidad o no entre el determinismo y el libre albedrío (y la responsabilidad moral) constituye ya una venerable cuestión filosófica. Quienes defienden la objeción del determinismo asumen una postura *incompatibilista*: si el determinismo fuera cierto, no podrían sostenerse ni el libre albedrío ni la responsabilidad. Para los *compatibilistas*, en cambio, sería posible sostener que nuestras acciones están determinadas sin que, sin embargo, se siguieran

⁴¹ T. Nagel, *Mente y cosmos. Por qué la concepción neo-darwinista materialista de la naturaleza es, casi con certeza, falsa*, Capítulo II (*La conciencia*), Ed. Biblioteca Nueva, 2004.

las consecuencias para la responsabilidad o para la disolución de la vida social que señala Taruffo. Si el compatibilismo tuviera razón, la aceptación de una concepción materialista de la mente no llevaría necesariamente al determinismo⁴².

b) En segundo lugar, debe tenerse en cuenta la distinción entre el determinismo como tesis ontológica (todo está causalmente determinado) y el determinismo como tesis epistemológica (conociendo el estado del mundo en un momento determinado y las leyes causales adecuadas, todo suceso es predecible). El primero requiere que las causas sean siempre condiciones suficientes de los efectos. Pero precisamente esto es algo que se niega en ocasiones. Kathinka Evers, en su libro *Neuroética. Cuando la materia se despierta*, distingue entre una causalidad necesaria, que produciría el mismo efecto en todos los mundos posibles, y una causalidad contingente, para la que habría al menos un mundo posible en el que esa misma causa ha producido un efecto distinto⁴³. Desde el lado de la neurociencia se han presentado algunas investigaciones que sugieren que en el cerebro (al menos, para algunas de sus funciones) las relaciones causales son de esta segunda manera (Evers se centra sobre todo en los trabajos del neurocientífico J.P. Changeux tratando de

⁴² En mi opinión, no obstante, siempre hay algo insatisfactorio en las propuestas compatibilistas, porque su objetivo lo alcanzan, en el mejor de los casos, redefiniendo explícitamente lo que se entiende por determinismo o por libre albedrío, y otras veces simplemente deslizándose inadvertidamente de un sentido a otro de estos términos. Hobbes, por ejemplo (un compatibilista clásico) sostenía que la libertad consiste en la ausencia de obstáculos para actuar conforme a nuestros deseos. Esto es compatible con que nuestros deseos estén determinados. Pero si aceptamos que un sujeto es realmente libre cuando hubiera podido actuar de una manera distinta a como lo hizo, entonces el libre albedrío requiere –como hemos visto al comienzo del artículo- no sólo que la acción sea libre (*libertad de acción*), en el sentido de que no haya obstáculos que impidan al agente actuar de acuerdo con sus planes y objetivos, sino también que la voluntad sea libre (*libertad de voluntad*), en el sentido de la elección por esos planes y objetivos no esté determinada por condiciones suficientes previas. Muchos compatibilistas suelen identificar la libertad con algo (la libertad de acción sin libertad de la voluntad, libertad como nómeno pero no fenomenológica, deseos de segundo orden, etc.) que no satisface plenamente nuestra percepción de ser libres, por lo que la conciliación entre libertad y determinismo se logra en la mayor parte de concepciones compatibilistas asumiendo un sentido fuerte de determinismo, pero débil de libertad. Creo que tiene razón Searle cuando afirma que "el compatibilismo hace un planteamiento lógico sobre los conceptos de 'libre' y 'determinado' y señala, con acierto, que hay un uso de estos según el cual decir que una acción es libre no significa, hasta aquí, formular interrogante alguno acerca de si está determinada o no, en el sentido de la existencia de condiciones causales previas suficientes. Pero una vez aceptado este planteamiento lógico, todavía queda abierta una cuestión empírica fáctica. ¿Es verdad que toda acción humana ocurrida en el pasado, que ocurre ahora o que ocurrirá alguna vez fue, es y será causada por condiciones previamente suficientes?" (J. Searle, *La mente. Una breve introducción*, pág. 278.)

⁴³ K. Evers, *Neuroética. Cuando la materia se despierta*, Ed. Katz, 2010, pág. 86 y ss.

explicar la creatividad a partir de la plasticidad neuronal, esto es, la capacidad del cerebro de modificarse físicamente en función del aprendizaje, incluso del proveniente del medio social). Si esto es así, "es teóricamente posible *influir* deliberadamente en el resultado para orientarlo en cierta dirección"⁴⁴. Creo que, respecto del determinismo ontológico, hay suficiente grado de discusión entre neurocientíficos como para poder dejar todavía la cuestión abierta, sin identificar necesariamente reducción de la mente al cerebro con aceptación del determinismo⁴⁵.

El determinismo epistemológico, por su parte, añade al requisito de la existencia de condiciones causalmente suficientes previas de la conducta la idea de una correlación entre tipos de estados cerebrales y tipos de estados mentales. Si esta correlación no existe o no puede establecerse, entonces la conducta humana, aunque pueda estar causalmente determinada, es impredecible. "Leyendo" los estados cerebrales no podríamos saber qué estados mentales tendría el sujeto y, por tanto, cómo va a actuar. El materialismo de la identidad de instancias, por tanto, no puede incurrir en un determinismo neurobiológico epistemológico.

c) En tercer lugar, como hemos visto al exponer las distintas teorías materialistas, cabe una graduación en el grado de identificación entre mente y cerebro, que tiene consecuencias para las implicaciones deterministas. De mayor a menor identificación: 1) El materialismo eliminacionista y el materialismo reduccionista proponen una identificación fuerte entre cerebro y mente: las explicaciones de la conducta basadas en creencias, deseos, emociones, etc. podrán ser eliminadas o al menos traducidas a explicaciones neurofisiológicas una vez que se hayan encontrado los correlatos neuronales

⁴⁴ K. Evers, *Neuroética. Cuando la materia se despierta*, pág. 96.

⁴⁵ Otro intento de introducir la indeterminación en el ámbito de la conducta ha tratado de apoyarse en la mecánica cuántica. Como es sabido, ésta parece introducir un elemento de azar e indeterminación en el nivel sub-atómico del mundo físico, de manera que sus leyes sólo pueden ser estadísticas. Esta indeterminación no afecta en condiciones normales a niveles superiores. ¿Es posible que en el cerebro se filtre algún grado de indeterminación por influencia de la mecánica cuántica? No han faltado autores que lo han sugerido así (Eddington, R. Penrose, J. Eccles, Robert Kane, etc.). Esta estrategia tiene dos enormes problemas (por no decir, como señala Searle, que trata de resolver un misterio con otro): ¿cómo es posible que la indeterminación de la mecánica cuántica afecte al nivel neuronal (superior al sub-atómico), a diferencia de lo que ocurre en el resto de contextos físicos? Y, más grave aún: ¿No es el libre albedrío también incompatible con el ingrediente azaroso que la indeterminación introduciría en nuestra conducta?⁴⁵ Tan desastroso es para el libre albedrío el determinismo rígido como la ausencia de toda determinación: El libre albedrío requiere del tipo adecuado de determinación.

de los estados mentales. Sin embargo, no llevarían necesariamente a sostener que todas las conductas están determinadas si se encuentra alguna explicación neurofisiológica para ello (como las que destacan la plasticidad del cerebro y la flexibilidad de algunas relaciones causales en su funcionamiento). 2) Un grado menor de identificación se da en el materialismo de la identidad de instancias o casos: es una identificación en el nivel de los casos, pero sin una correlación de tipos. De esta manera, como hemos visto, se abre la puerta a aceptar una noción de libertad como imposibilidad de predecir la conducta, esto es, como indeterminismo epistémico. 3) Menor identificación aún se da en las posturas que sostienen un monismo de sustancias pero un dualismo de propiedades. Es el caso del materialismo emergentista. Si la relación de emergencia es algún tipo de relación causal entre el sustrato físico y lo mental, el libre albedrío podría explicarse desde posturas emergentistas si pudiera haber influencia causal de las propiedades sistémicas sobre las propiedades de nivel inferior. Mario Bunge, por ejemplo, trata de ofrecer una explicación del libre albedrío, entendido como la capacidad del pensamiento de influir en la conducta, desde los postulados del materialismo emergentista. 4) El funcionalismo, por último, establece una conexión muy débil entre el cerebro y las propiedades mentales (que, recordémoslo, se identifican con propiedades funcionales). Puede ser compatible con algún tipo de determinismo, pero en la medida en que define la mente al margen de su sustrato neuronal (hasta el punto de que las mismas funciones mentales pueden realizarse en soportes físicos distintos), se aleja de un determinismo neurobiológico.

En definitiva, si se tienen en cuenta los distintos sentidos de determinismo y los distintos sentidos de materialismo, no parece que haya necesariamente una contradicción entre ser materialista y aceptar el libre albedrío. Es decir, *siempre hay un argumento disponible para cada tipo de teoría materialista que permite rechazar el determinismo.*

En todo caso, como he sostenido en otro lugar, hay argumentos conceptuales que permiten sostener que no cabe probar empíricamente la tesis determinista. Precisamente en esta idea cabe sustentar la responsabilidad: no en el hecho de que el determinismo sea una hipótesis que no ha sido todavía probada, sino en el hecho de que, por razones conceptuales, no puede serlo. No puede serlo porque las categorías conceptuales implicadas, como la de

relación causal, están construidas a partir del concepto de acción, que a su vez presupone fragmentos del mundo no determinados⁴⁶. Probablemente, éste y otros argumentos conceptuales semejantes descansan en la imposibilidad de construir un lenguaje lo suficientemente objetivizante como para demostrar que carecemos de algo de lo que tenemos una convicción subjetiva tan intensa. Como ha escrito Adela Cortina: "Si quiero comprobar experimentalmente si tenía posibilidad de actuar de otro modo o si tengo posibilidad de decidir, entonces he de asumir una perspectiva impersonal u objetiva para poder diseñar experimentos que me permitan comprobar la verdad o falsedad de las afirmaciones. En este caso puede muy bien ocurrir que no podamos traducir el lenguaje de primera persona al de tercera persona"⁴⁷

9. Naturalización sin confusión.

La tercera de las objeciones hacia las pruebas neurofisiológicas es la de que sus defensores incurren en una confusión grave al atribuir al cerebro, o a alguna de sus partes, estados mentales que sólo son atribuibles al sujeto en su conjunto. Esta objeción ha sido sostenida por Pardo y Patterson contra la neuroética y el neuroderecho. Estos, a su vez, la toman de la objeción que Hacker y Bennet dirigen a la neurociencia -al menos a la neurociencia cognitiva-.

Los argumentos centrales en Pardo y Patterson son, por un lado, la distinción entre criterios e indicios y, conectada con esta distinción, la identificación de los estados mentales con "habilidades" o con la conducta del sujeto. Esta definición de estado mental les lleva a negar que la mente (y los estados mentales) pueda ser localizada en algún lugar del cuerpo. El punto de partida de su argumentación es la distinción wittgensteniana entre pruebas que son criterios y pruebas empíricas o inductivas para la adscripción de predicados psicológicos, como creer, percibir, conocer, etc. Las primeras tienen una relación constitutiva o conceptual con dicho predicado psicológico, mientras que las segundas sólo prestan apoyo inductivo, en la medida en que

⁴⁶ D. González Lagier, *¿La tercera humillación? Sobre neurociencia, filosofía y libre albedrío*, en M. Taruffo y J. Nieva, *Neurociencia y proceso judicial*.

⁴⁷ A. Cortina, *Neuroética y neuropolítica*, Ed. Tecnos, 2012, pág. 157.

pueden estar empíricamente correlacionadas con el estado mental⁴⁸. De acuerdo con los autores, las pruebas "criteriales" de los atributos psicológicos son "varios tipos de comportamiento": "Comportarse de un modo determinado es lógicamente una buena prueba y, por eso, parte constitutiva de estos conceptos. Para la percepción visual esto incluye, por ejemplo, que uno de los ojos siga el fenómeno que se persigue, que uno reporte ajustadamente lo que se observa, y así respectivamente. El creer incluye, por ejemplo, que uno asevere o refrende lo que uno cree, que uno actúe en direcciones consistentes con lo que se cree, que uno no crea proposiciones directamente contradictorias y así sucesivamente"⁴⁹. La conexión entre el comportamiento y los estados psicológicos, según los autores, es conceptual, de manera que "el comportamiento no sólo aporta pruebas de que alguien en una ocasión dada está percibiendo o tiene una creencia, sino que también determina parcialmente lo que significa percibir o creer"⁵⁰. Por el contrario, la actividad neuronal, aunque pudiera establecerse una correlación entre ella y los estados mentales, sólo sería prueba inductiva o empírica de los mismos. Lo neuronal no podría ser criterio de los estados mentales porque para saber que cierto tipo de evento neuronal se corresponde con un determinado tipo de estado mental, previamente hemos de haber identificado el estado mental⁵¹ (a través del comportamiento).

La conexión interna o constitutiva entre la conducta y los estados mentales les lleva a definir la mente como "un conjunto diverso de habilidades ejercidas por una persona"⁵². "La mente -afirman- no es en absoluto una entidad o sustancia (ni física ni no física). Tener una mente implica poseer una serie de ciertos poderes racionales que se manifiestan con el pensamiento, con los sentimientos y las acciones"⁵³. Pardo y Patterson contraponen esta concepción al materialismo (por el que entienden la identificación de mente y cerebro) y al dualismo, presentándola como una tercera vía.

⁴⁸ M. S. Pardo y D. Patterson, "Fundamentos filosóficos del Derecho y la Neurociencia", en *In Dret*, 2/2011 pág. 13.

⁴⁹ M. S. Pardo y D. Patterson, "Fundamentos...", pág. 14.

⁵⁰ M. S. Pardo y D. Patterson, "Fundamentos...", pág. 14.

⁵¹ M. S. Pardo y D. Patterson, "Fundamentos...", pág. 15.

⁵² M. S. Pardo y D. Patterson, "Fundamentos...", pág. 7.

⁵³ M. S. Pardo y D. Patterson, "Fundamentos...", pág. 43.

El último paso en su argumentación es el siguiente: si la mente no es el cerebro, ni ningún otra parte del sujeto, no tiene sentido atribuir a un órgano los estados mentales. Hacerlo es incurrir en la falacia que -siguiendo a Hacker y Bennet- llaman falacia mereológica.

Esta argumentación sigue punto por punto la de Peter Hacker y Maxwell Bennet en sus *Fundamentos filosóficos de la neurociencia* y en *La naturaleza de la conciencia* (que reproduce parcialmente el primer libro): De acuerdo con estos autores, los neurocientíficos han heredado en cierta medida los errores del dualismo cartesiano: han pasado de adscribir atributos psicológicos a la mente a adscribir los mismos atributos al cerebro, pero “suponer que el cerebro es el portador de los atributos psicológicos no supone avance alguno sobre esa concepción falsa [el dualismo]”⁵⁴. Esta confusión les lleva a decir que el cerebro *sabe cosas, razona, construye hipótesis, decide*, o que las neuronas son *inteligentes, transmiten mensajes, calculan probabilidades*, etc.⁵⁵. Se trata en todos los casos, opinan, de un grave error conceptual que lleva a sinsentidos, y no una simple forma metafórica de hablar. Y es un grave error porque la atribución de estados mentales no es una cuestión empírica, sino conceptual. Para poder adscribir estos atributos al cerebro necesitaríamos algún criterio conceptual. En el caso del ser humano, el criterio de atribución es su comportamiento; pero en el caso del cerebro “ni siquiera sabemos qué es lo que demostraría que (...) posee tales atributos”⁵⁶. El comportamiento está, por tanto, constitutiva y conceptualmente vinculado con los estados mentales: es parte de lo que significa tener esos estados mentales. Por tanto, los estados mentales sólo son atribuibles al agente que actúa como un todo.

¿Cómo debemos valorar estos argumentos? Me parece que John Searle, en su contestación a las críticas de Hacker y Bennet, muestra algunas debilidades de estas objeciones a la neurociencia. Como es evidente, la concepción de la mente de estos autores se basa en el conductismo lógico de Wittgenstein⁵⁷. Uno de los rasgos distintivos del conductismo lógico es la

⁵⁴ P. Hacker y M. Bennet, “Fundamentos filosóficos de la neurociencia”, en M. Bennet, P. Hacker, D. Dennet y J.Searle, *La naturaleza de la conciencia. Cerebro, mente y lenguaje*, Ed. Paidós, 2008, pág. 19.

⁵⁵ P. Hacker y M. Bennet, “Fundamentos filosóficos de la neurociencia”, pág. 31.

⁵⁶ P. Hacker y M. Bennet, “Fundamentos filosóficos de la neurociencia”, pág. 35.

⁵⁷ El conductismo lógico no niega necesariamente la existencia de un elemento subjetivo en los estados mentales (al menos, al parecer, Wittgenstein no lo hacía), pero en la medida en

existencia de esta relación lógica o conceptual entre la conducta (o las disposiciones conductuales, o los poderes causales) y la mente que señalan tanto Pardo y Patterson como Hacker y Bennet. El argumento para establecer esta conexión, como hemos visto, parte de las consideraciones de Wittgenstein sobre la imposibilidad de un lenguaje privado: Si no hay criterios públicos de corrección del uso de un lenguaje, éste carece de sentido. Atribuir a un sujeto un estado mental -dolor, por ejemplo- es algo que hacemos porque le vemos actuar de determinada manera. Luego determinada conducta, o cierta disposición a ella, es parte de lo que significa tener un dolor. Frente a este argumento, Searle señala que Hacker y Bennet confunden los criterios de atribución de estados mentales con los estados mentales atribuidos: “La falacia -escribe Searle-, en pocas palabras, reside en confundir las reglas de uso de las palabras con la ontología. Del mismo modo que el antiguo conductismo confundía las pruebas de estados mentales con la ontología de los estados mentales, este conductismo lógico wittgensteniano confunde las bases para hacer la atribución con el hecho que se atribuye. Es una falacia decir que las condiciones para el correcto funcionamiento del juego del lenguaje son condiciones para la existencia del fenómeno en cuestión”⁵⁸. Thomas Nagel hace una objeción similar al conductismo lógico cuando señala que “todas estas estrategias son verificacionistas. Suponen que todo lo que se necesita decir sobre el contenido de un estado mental es lo que verificaría o confirmaría o justificaría su aserción desde el punto de vista de un observador. De una forma u otra reducen los atributos mentales a las condiciones de lo externamente observable con el criterio de que atribuimos estados mentales a los otros”⁵⁹.

que no lo introduce en la definición de los mismos se le presentan algunas paradojas. Así, por ejemplo, tendría que aceptar que un sujeto superespartano que nunca expresara su dolor, ni siquiera en situaciones extremas –por ejemplo, al extraerle una muela sin anestesia– no tendría realmente dolor, y que un actor tan magnífico que pudiera fingir expresiones externas de dolor que engañaran a todos tendría, sin embargo, dolor. Otra objeción importante es que analizar determinados estados mentales nos lleva necesariamente a otros estados mentales, lo que convierte a las definiciones de los conductistas en circulares. Así, por ejemplo, mi creencia de que llueve se manifestará en que yo me llevo un paraguas sólo si tengo también el deseo de no mojarme. Y mi deseo de no mojarme se manifestará en esta conducta de coger el paraguas sólo si tengo la creencia de que el paraguas me mantendrá seco. Véase W. Betchel, *Filosofía de la mente: Una panorámica para la ciencia cognitiva*, Ed. Tecnos, 1991.

⁵⁸ J. Searle, “Réplica a Philosophical Foundations of Neuroscience”, en *La naturaleza de la conciencia*, pág. 130.

⁵⁹ T. Nagel, *Mente y cosmos*, Capítulo II.

Con esta reducción de la mente a la conducta, a habilidades o a roles causales, el conductismo (y su heredero sofisticado, el funcionalismo) deja fuera la perspectiva subjetiva o de primera persona, que en opinión de autores como Searle o Nagel es precisamente lo característico de lo mental. El conductismo asume respecto de los estados mentales un punto de vista externo, desplazando el punto de vista interno característico de los estados mentales. Es concebible un estado mental que no se exterioriza en una conducta, pero no lo es un estado mental que no se perciba de una manera característica por el sujeto que lo tiene (esa percepción no tiene por qué ser una sensación; las creencias, por ejemplo, no tienen por qué llevar aparejada una sensación. Pero hay algo que nos hace saber que tenemos una creencia, sin necesidad de observarnos “desde fuera”). Los *qualia*, en definitiva, son negados explícitamente por Hacker y Bennet.

Tampoco otorgan la debida importancia estos autores al hecho de que la neurociencia ha establecido que existe alguna relación causal entre los estados cerebrales y los estados mentales. Por supuesto, no es algo que nieguen expresamente. Pero si se toma en serio que el funcionamiento del cerebro es causa de la existencia de las experiencias mentales, entonces la atribución de estados mentales al cerebro o componentes del mismo puede verse como una manera aceptable de hablar, no muy distinta de cuando decimos que el estómago hace la digestión. Se trata sólo de localizar la causa de un proceso, el encargado de una función. No parece haber aquí una falacia grave, ni por ello se está identificando al cerebro con el sujeto, ni se está reduciendo el concepto de estado mental al concepto de estado cerebral (pues un concepto completo de estado mental ha de incluir las “propiedades de lo mental”): No se está diciendo –por ejemplo- que las intenciones son estados cerebrales, sino que son causadas por estados cerebrales. El argumento de Hacker, Bennet, Pardo y Patterson para considerar esta atribución como una falacia es que los estados mentales son conductas o disposiciones a actuar de una u otra manera, por lo que sólo se pueden atribuir a quien puede actuar. Este es un concepto que tampoco da cuenta de las propiedades de lo mental: Ni da cuenta de los *qualia* ni de la eficacia causal de lo mental (pues lo mental se confunde con la acción causada por lo mental). Si se rechaza esta definición, que también es reductivista, la acusación de falacia a estas atribuciones cae por sí

sola. Pueden a veces resultar extrañas desde el punto de vista del lenguaje común, pero probablemente ello se explique por la carga de dualismo cartesiano que persiste en ese lenguaje.

10. Algunas reflexiones finales.

1. En las anteriores páginas se han revisado algunos de los problemas que plantea la naturalización de la filosofía (práctica). Podría distinguirse entre dos maneras de entender este proceso de naturalización: Una manera radical (el proyecto “fuerte” de naturalización), que trata de disolver la filosofía en la ciencia empírica (actualmente, en la neurociencia); y otra, más moderada (el proyecto moderado), que recuerda la necesidad de que la filosofía no contradiga la visión del mundo (y de la naturaleza del hombre) que está surgiendo de las ciencias, pero que no niega que la filosofía tiene un espacio donde la ciencia no puede llegar. Incluso, desde este segundo punto de vista, puede afirmarse que la filosofía tiene una tarea (el esclarecimiento conceptual) sin la cual la ciencia no puede proporcionar conclusiones relevantes acerca de las normas, la mente, la acción, etc. Las aportaciones de las ciencias han de ser interpretadas, y en esa interpretación la filosofía ejerce una labor de control.

2. En el actual estado de desarrollo de la neurociencia, el programa de naturalización “fuerte” de la filosofía práctica no puede ser realizado: nuestro conocimiento del funcionamiento del cerebro no permite naturalizar la normatividad sin incurrir en la falacia naturalista o confundir normas y regularidades, ni naturalizar la mente sin eliminar las propiedades de lo mental (e incurrir, por ello, en un reduccionismo que impide comprender bien la conducta). Pero, al mismo tiempo, la filosofía debe tener en cuenta las aportaciones de la ciencia. Podría distinguirse también entre un “monismo ontológico” (la aceptación de que sólo existe un tipo de entidades: las físicas) y un “monismo metodológico” (la tesis de que sólo es legítimo el método de las ciencias empíricas). El estado actual de la neurociencia parece avalar el primero, pero no es suficiente para aceptar el segundo. Y si llega a ser suficiente debería permitir explicar cómo surgen los estados mentales a partir de los estados cerebrales sin negar las propiedades de lo mental (significado, causalidad mental, consciencia, etc.), con lo que la reducción metodológica dejaría de ser necesaria. La teoría de la mente que hemos llamado

“emergentismo” parece la mejor situada para alcanzar este objetivo, aunque la relación de “emergencia” todavía es poco más que una metáfora.

3. En todo caso, la discusión sobre la naturalización de la filosofía práctica no es sólo una discusión acerca del fundamento y la naturaleza de la filosofía, sino que recorre también un buen número de cuestiones que son directamente relevantes para el desarrollo de una filosofía del Derecho para el mundo latino. La discusión de temas como el del peso de las emociones en los procesos de justicia transicional, el de las causas de la anomia, la función de las intuiciones en el razonamiento moral, los medios de reducción de la violencia, el papel de las imágenes sobre el mundo en la argumentación en contextos multiculturales, la indignación y la repugnancia como explicaciones (y, para algunos, justificaciones) de las normas básicas del Derecho penal, la posibilidad de un objetivismo ético, la posibilidad del control de las emociones y las vías de la educación emocional, el racismo, etc., etc. pueden verse beneficiados por un mejor conocimiento del funcionamiento del cerebro y las aportaciones de la neuroética y la neurociencia cognitiva.