



Universidad
Zaragoza

Trabajo Fin de Grado

Título del trabajo: El problema mente - cerebro

Autor

José M^a Marín Herrero

Director

Juan Vicente Mayoral de Lucas

Facultad de Filosofía y letras
Julio de 2013

Resumen

Este trabajo fin de grado se enmarca dentro del campo de la filosofía de la mente, su objetivo general es establecer, a un nivel apropiado para un graduado de filosofía, el panorama actual de los diversos puntos de vista sobre la relación entre las operaciones mentales y los procesos cerebrales y se propone como objetivos particulares los dos siguientes

- i. Conocer la terminología y los principales conceptos utilizados por los especialistas en ciencias cognitivas, intentando precisar su significado.
- ii. Discriminar y relacionar las principales concepciones que se han dado a lo largo de la historia, y especialmente en la actualidad, del problema filosófico llamado *problema mente – cerebro*.

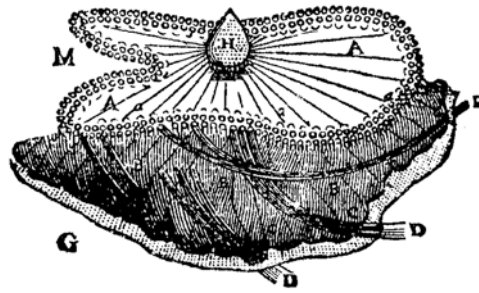
Tomando como punto de partida el planteamiento dualista de Descartes, se da un salto de más de doscientos años hasta finales del s. XIX, momento en el que se replantea el estatuto de las operaciones mentales desde múltiples frentes: filosofía, psicología y neurofisiología principalmente; cincuenta años después se suman la informática, la cibernética y la inteligencia artificial, suministrando una capacidad de análisis muy potente. Una descripción lo más actualizada posible de este panorama es el objetivo que se persigue con este trabajo fin de grado.

Introducción y objetivos

A pesar de que el terreno cultural humano venía siendo abonado durante cientos e incluso miles de años de su historia para que se asentara con fuerza, en la vida social y especialmente en la religiosa, la creencia en entidades distintas a la materiales, un punto que los expertos coinciden en señalar como el origen de la instalación de esta escisión ontológica en el seno de la propia filosofía son las *Meditaciones metafísicas* de René Descartes [1]. Ese punto de vista se empieza a cuestionar seriamente a partir del siglo XIX y en la actualidad el panorama que se respira, en términos generales, es una reacción a su planteamiento, aunque todavía sigan existiendo posturas de tipo dualista.

Descartes concibe la relación entre mente y cuerpo como la de un piloto que guía desde la glándula pineal la nave del cuerpo que queda reducido a un puro mecanismo. Como prueba clara de este planteamiento, lo más ilustrativo son unos breves fragmentos seleccionados de su *Tratado del hombre* [2] que explican mejor que ningún comentario su pensamiento respecto a la existencia en el hombre de dos principios distintos, uno material y otro espiritual:

“Supongo que el cuerpo no es otra cosa que una estatua o máquina de tierra a la que Dios forma con el propósito de hacerla tan semejante a nosotros como sea posible” (pg. 22)...”no debemos concebir en esta máquina alma vegetativa o sensitiva alguna, ni otro principio de movimiento y de vida” (pg. 109).



“Puede apreciarse en la zona de la figura señalada con M que los espíritus que salen de la glándula H, habiendo dilatado la parte del cerebro A y entreabierto todos sus poros, se dirigen desde tal punto hasta B, después hasta C y, finalmente, hacia D; alcanzando este punto, se propagan a través del interior de todos sus nervios. De este modo, los espíritus mantienen tan tensos todos los pequeños filamentos de los que se componen estos nervios y el cerebro que, por reducida que fuera la fuerza que para moverlos tengan las diversas acciones, ésta se comunica fácilmente” (pg. 84).

Así pues para Descartes los seres humanos están hechos de dos componentes: su cuerpo, que es una máquina carente de cualquier principio de movimiento autónomo, y otro muy distinto, contenido en la glándula pineal (H en la figura), que genera todos los movimientos, internos y externos, del cuerpo, -razones, sentimientos, memoria, imaginación y los movimientos propiamente corporales-, y que es la *res cogitans*. Este planteamiento es totalmente inadmisibles para la física actual porque viola uno de sus principios básicos: el *cierre causal* que significa que ningún ente externo al mundo físico es capaz de provocar ninguna modificación del mismo, es decir, puede tener capacidad causal. No obstante, la filosofía de Descartes tuvo consecuencias radicales y de largo alcance tanto en el campo de la ontología como en el de la epistemología y las preguntas fundamentales que se derivan de su planteamiento siguen presentes en el debate filosófico

1. ¿Los procesos mentales son distintos o idénticos a los procesos cerebrales?
2. Si son idénticos, ¿cómo se explican los procesos mentales a partir de los procesos cerebrales?
3. Si son distintos, ¿cómo interactúan entre sí?

El objetivo del presente trabajo fin de grado es dar cuenta del estado del llamado problema mente-cuerpo o, si se prefiere, mente-cerebro y para ello el procedimiento seguido es, en primer lugar, situar los orígenes actuales de este campo filosófico que se conoce principalmente como *filosofía de la mente* o también *ciencias cognitivas* y señalar qué campos del conocimiento se han congregado en torno a él. Para este primer apartado ha sido especialmente útil la referencia [3], que ha permitido trazar un hilo conductor y ha sido la fuente de muchas de las referencias que se han consultado y que, cuando ha procedido, se han utilizado y han quedado recogidas en el apartado de la

bibliografía. Posteriormente se han analizado más en detalle las propuestas más influyentes dentro de esta disciplina en la actualidad, todas ellas monistas, algunas reduccionistas, como las que representa el filósofo Jaegwon Kim, y otras no reduccionistas representadas por el filósofo Hilary Putnam y el neurobiólogo y premio nobel de Medicina R. W. Sperry. El trabajo se concluye con el punto de vista, compartido por el autor de esta memoria, del neurofisiólogo Antonio Damasio acerca de cómo se puede entender la *conciencia* y el *yo* desde un punto de vista científico que sea aceptable para la filosofía y con un resumen de las distintas posturas que se encuentran en la filosofía de la mente actual.

Inicios y bases de las ciencias cognitivas

En el terreno de la filosofía de la mente, una de las reacciones más contundentes al método introspectivo de Descartes fue el conductismo. Se considera fundador de esta corriente al psicólogo estadounidense J. Watson (1878 – 1958) y sus representantes más conocidos son el fisiólogo y psicólogo ruso I. Pavlov (1848 – 1936) y el filósofo y psicólogo experimental norteamericano B. F. Skinner (muerto en 1990). Los conductistas consideraban anticientífico el uso de fenómenos no observables como las representaciones mentales (símbolos, imágenes o conceptos) y querían restringir la atención a los fenómenos observables, que comprendían la acción del entorno sobre el cuerpo humano y la reacción de éste sobre aquél, es decir, la conducta. Su modelo científico era el de la caja negra y pretendían buscar correlaciones entre estímulos y reacciones. Este planteamiento encajaba muy bien en las corrientes positivistas pero, en el fondo, significaba una renuncia a investigar la mente humana.

En septiembre de 1948 un grupo de eminentes intelectuales, interesados y formados tanto en humanidades como en ciencias, se juntaron en el Instituto de Tecnología de California (Pasadena) auspiciados por la fundación Hixon [3] para estudiar la forma en que el sistema nervioso gobierna la conducta y se dieron cuenta de que el paradigma conductista era un obstáculo para el conocimiento de la relación mente-cerebro. Uno de los participantes más celebrados fue el psicólogo Karl Lashley quien advirtió que, para alcanzar nuevos conocimientos sobre el cerebro, se debía dar cuenta de conductas organizadas complejas, muy especialmente del uso del lenguaje, y que el marco teórico explicativo predominante en aquel momento -el conductista- no era capaz de dar cuenta de ninguna conducta de este tipo.

En aquella época empezaban a coger fuerza, impulsadas entre otros motivos por las circunstancias bélicas de la 2ª Guerra Mundial, la informática y las ciencias de la computación. Turing [4] propuso la idea de una máquina abstracta capaz de realizar, en principio, cualquier cálculo concebible. Esta máquina constaba de una cinta de longitud infinita que pasaba por la máquina, donde un dispositivo de lectura detectaría su contenido. La cinta se dividía en registros idénticos, cada uno de los cuales contenía un espacio en blanco o una marca. La máquina podía desarrollar cuatro

clases de movimiento con la cinta: desplazarla a la derecha, desplazarla a la izquierda, borrar la marca o imprimirla. Con estas operaciones sencillas solamente, era capaz de ejecutar cualquier clase de programa o de plan que se expresara en un código binario. Cualquier tarea que pudiera explicitar con claridad los pasos necesarios para llevarla a cabo podría programarse y ejecutarse en la máquina de Turing; los ordenadores actuales están concebidos según su esquema.

Los científicos interesados en el pensamiento humano advirtieron las implicaciones de estas ideas, al darse cuenta de que si eran capaces de describir con precisión los procesos de pensamiento o de conducta de un organismo, podrían diseñar un ordenador que operara en forma idéntica a él. Este modelo computacional ha sido uno de los más fértiles en la investigación del cerebro humano y en aquel momento parecía despejado el camino para construir máquinas que se comportaran como seres humanos. En paralelo, otra investigación estaba avanzando poderosamente en aquel momento: la neurofisiología del cerebro humano y de los animales. El esquema que se estaba elaborando concebía al sistema nervioso como una vasta red de neuronas (se considera que en el cerebro humano existen del orden de 10^{10} neuronas), distribuidas en estructuras muy elaboradas con interconexiones extremadamente complejas.

En 1943 McCulloch y Pitts elaboran el primer modelo matemático de neurona y de cerebro concebido como una red modular de neuronas [6], demostrando que era capaz de memorizar información y de realizar cálculos, es decir, que se comportaba como una

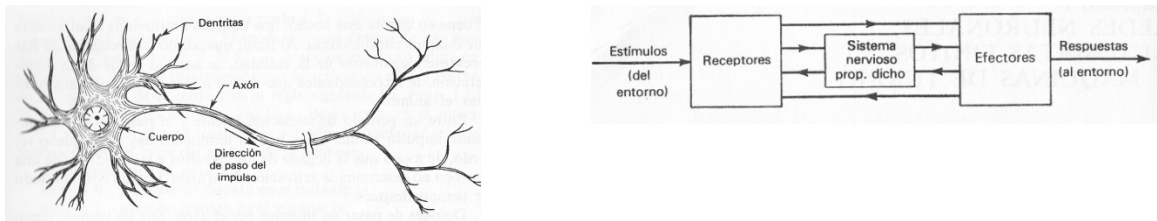


Figura 1.- Modelo de neurona y de red neuronal (Tomado de [5])

máquina de Turing¹; se tenía un modelo de cerebro que se comportaba como un ordenador. Sobra mencionar que se está considerando solamente una parcela del funcionamiento cerebral, importante pero no exclusiva, la de del almacenamiento y procesado de la información; otras actividades mentales mucho más difíciles de describir, como las emocionales o las intencionales, se quedaban fuera de esta modelización.

Cinco años después, en 1948, N. Wiener escribió uno de los libros más influyentes sobre los aspectos comunes del comportamiento de animales y máquinas, *Cybernetics*.

¹ La idea de su máquina la elaboró Turing en 1936, aunque su publicación definitiva fuera muy posterior

En él, el autor aborda no solo los aspectos concernientes al manejo de la información, sino que amplía sus elaboraciones teóricas a otros aspectos esenciales, para la vida en general y la vida humana en particular, como la autorregulación, la autoconservación, el lenguaje y la sociedad. Al igual que MacCulloch, Wiener pertenece al M. I. T. y desde esa época este instituto de investigación se convierte en referencia mundial en temas relacionados con la inteligencia artificial y la filosofía de la mente. De hecho, el nacimiento formal de lo que hoy se conoce como ciencias cognitivas tiene lugar en su sede, en un Simposio sobre *Teoría de la Información* realizado en 1956 en el que cabe destacar la presencia y las ponencias presentadas, entre otros, por los especialistas en ordenadores Newell y Simon, el lingüista N. Chomsky, el especialista en telecomunicaciones C. Shannon o el psicólogo G. Miller.

En 1959 aparece un artículo en la revista *Scientific American* titulado *La fisiología del imaginación* [7] en el cual su autor, J. Eccles, sostiene que no solo el puro procesamiento de información sino también las actividades correspondientes a la imaginación creativa surgen “de alguna manera” de la actividad del cerebro y que el objetivo de su artículo era ver como se podía relacionar el trabajo del cerebro con las experiencias de la mente. Indica que las líneas que conectan los sentidos a ciertas zonas del cerebro están tendidas pero que, una vez que llega la información al cerebro, debido al umbral que necesita cada neurona para entrar en actividad, no se propagaba a lo largo de cadenas lineales de neuronas sino como un frente de onda que avanza por muchos carriles paralelos y que puede barrer hasta 10^5 neuronas por segundo.

En la Figura 2 se muestra un gráfico que aparece en la pg. 140 del artículo [7] y donde se representan estas ondas cerebrales. Las señales procedentes de distintos frentes de onda integra la información que procede de los distintos sentidos de la percepción. Para Eccles el funcionamiento del cerebro es como el de un telar mágico en el que millones de lanzaderas tejen una estructura significativa pero temporal, y solo después de un largo periodo de educación el sujeto es capaz de interpretar los sucesos cerebrales como procedentes de un mundo exterior. La memoria depende de cambios producidos en la corteza cerebral por una activación previa que permanece; experimentos en la sinapsis neuronales habrían demostrado que su activación repetida provoca un aumento de su duración. En este contexto, ¿qué se puede decir de la imaginación creativa? La red neuronal está compuesta por un entramado altamente complejo de engramas² cuya permanencia proviene del incremento de la eficacia sináptica derivada del uso repetido. La operación subconsciente de la mente involucra una interrelación de engramas que generan nuevos engramas que se presentan como nuevas imágenes o ideas. En cuanto a la conciencia, Eccles reconoce que esta actividad mental fundamental se escapa de sus posibilidades explicativas.

Pocos años después empiezan a aparecer libros sobre psicología cognitiva, de los cuales el más influyente fue el de U. Neisser [9], psicólogo experimental versado en

² Un engrama es una estructura neuronal estable [8]

computadoras quien, reconociendo la aportación de las ciencias de la computación, señalaba la enorme distancia que existía entre una máquina de cálculo, limitada al procesamiento y almacenamiento de la información, y un ser humano capaz de actividades fuera del alcance de las máquinas como la intencionalidad o la afectividad. Neisser rebajaba el tono triunfalista procedente del mundo de la informática, diferenciando entre los procesos perceptivos, como la visión o la audición, -en los que los modelos propuestos eran específicos, se podían hacer muchos experimentos pertinentes y las hipótesis comprobables eran fáciles de formular-, de lo que denominaba procesos mentales superiores, en los que interviene la intencionalidad del sujeto; de éstos se ocupó del pensamiento y de la memoria y dejó fuera la imaginación y la consciencia.



Figura 2.- Trayectoria de las ondas cerebrales

Neisser fue preclaro al darse cuenta de las enormes limitaciones de la psicología experimental de su época, en la que se obligaba al sujeto a centrarse en un solo motivo y a un conjunto fijo de respuestas alternativas cuando la vida diaria era mucho más compleja, presentando simultáneamente una multiplicidad de motivaciones y de respuestas, ausentes de la mayoría de los experimentos sobre los procesos mentales superiores. Si la motivación y la cognición son inseparables, una teoría satisfactoria de los procesos mentales superiores sólo podría cristalizar cuando se tuvieran teorías de la motivación, de la personalidad y de la interacción social. De manera que el problema de la relación entre los procesos mentales y los procesos cerebrales volvía a quedar abierto a la reflexión filosófica.

Funcionalismo

Entre los filósofos de la época destacaba Hilary Putnam que trabajaba en Princeton y que en su libro *Minds and machines* de 1960 [10] propuso una aproximación al problema mente-cerebro denominada *funcionalismo* donde sostenía que, a semejanza de un ordenador, la mente corresponde a la actividad funcional del cerebro que es el soporte material para aquella. Las operaciones lógicas (o *software*) podían describirse en forma independiente del soporte material (*hardware*) en que fueran implementadas.

En palabras posteriores de Putnam [10], la concepción computacional de la mente fue una reacción contra la idea de que la materia era más importante que la función que desempeñaba. Como los fenómenos mentales son estados funcionales del organismo, no es posible conocerlos estudiando exclusivamente los procesos en los que están implicados sus sustratos materiales como el cerebro; el funcionalismo presupone que una misma función puede ser desempeñada por sistemas muy distintos, ya que la naturaleza de sus componentes no es esencial para el correcto desarrollo de su función. Del mismo modo, las creencias y deseos son estados funcionales de sistemas físicos que pueden estar hechos de diferentes tipos de materiales. Por tanto, para comprender un proceso mental lo que debe analizarse es la función y no el sistema y, por ello, se puede atribuir estados mentales a seres extraterrestres con una estructura fisicoquímica muy diferente a la nuestra o a un soporte no orgánico, como un ordenador digital. Cualquier sistema puede tener mente a condición de que sea capaz de realizar la función adecuada.

En su libro *Representación y realidad* Putnam hace una revisión a fondo de los planteamientos del funcionalismo y se retracta de ellos, generalizando su afirmación anterior de que es incorrecto identificar los estados mentales con estados físico-químicos para pasar a decir que es asimismo incorrecto identificar directamente los estados mentales con los estados funcionales (caracterizados computacionalmente). Se apoya especialmente en los planteamientos sobre el holismo del significado de Quine [12] y afirma que no hay forma de identificar un estado computacional idéntico en todas las personas que tienen una creencia determinada y, de hecho, los espacios de los

estados computacionales de diferentes especies con una conducta verbal suficientemente próxima, -para poder ser traducidas entre ellas-, en una misma circunstancia y con una misma creencia, pueden ser incluso incomparables [10] (pg. 134).

El requisito básico del funcionalismo es que existe alguna clase de ordenadores, como la máquina de Turing, que puede servir de modelo para los seres humanos. Esto significa que se puede encontrar una relación de equivalencia, -en el sentido de la teoría de conjuntos-, entre los estados de esa clase de ordenadores y los humanos que sea computable, es decir, describable en un lenguaje reglamentado (como los de la informática o la física). Pero para interpretar un lenguaje es necesario tener alguna idea de las teorías y los métodos de inferencia que comparten los miembros de la comunidad que habla ese lenguaje. Nadie podría determinar a qué se refiere "spin" en la mecánica cuántica sin conocer la mecánica cuántica o a que se refiere "carga negativa" sin saber algo sobre la teoría de la electricidad. Esto plantea serias dificultades a la tesis de que se puede establecer una sinonimia (o correferencialidad) entre dos lenguajes diferentes.

Putnam propone el siguiente ejemplo [11] (pg. 138): “Supongamos que estoy interpretando un término T_1 de una teoría A sostenida por los marcianos y un término T_2 de una teoría B sostenida por los venusinos. Supongamos que las teorías A y B son diferentes, pero que no es inconcebible que los términos T_1 y T_2 se refieran a la misma entidad. Supongamos, además, que la similitud entre las dos teorías es suficientemente grande, de modo que si los entornos son tales que los términos T_1 y T_2 son correferenciales, entonces ningún intérprete normal juzgaría la diferencia en la creencia entre marcianos y venusinos como una diferencia en el significado”.

En este ejemplo, el problema de la correferencialidad se reduce a las dos preguntas siguientes: ¿A qué se refiere T_1 en el entorno marciano? y ¿a qué se refiere T_2 en el entorno venusino? Si las teorías A y B fueran cosmológicas, para determinar si son aproximadamente verdaderas, -tal como lo entienden y utilizan las respectivas comunidades-, se necesitaría información sobre el resto del universo, de manera que sería falsa la hipótesis de que es posible decir a que se refiere un término usado en un entorno dado a partir de una descripción completa del mismo en términos de algún lenguaje normado, a menos que el entorno incluya todo el universo. Por otro lado, cualquier teoría que pretenda establecer la correferencialidad de otras teorías tiene que examinar todas las teorías posibles porque un algoritmo, merced al cual una máquina de Turing puede resolver estas cuestiones a partir de una descripción de las situaciones en Marte y Venus, debe anticipar los procesos de fijación de creencias de que depende la comprensión de las teorías. Pero, ¿es posible examinar el razonamiento y las creencias de todos los seres y las sociedades humanas posibles? Para verter creencias de un lenguaje natural al cálculo de predicados primero hay que interpretarlas, incurriendo en una petición de principio.

Pero no solo existen problemas para traducir los lenguajes naturales a un lenguaje normado sino que también se dan entre estos. Imaginemos que se quiere traducir la mecánica cuántica al lenguaje formal del cálculo de predicados. Según la interpretación de la escuela de Copenhague, - rechazada por el propio Putnam [13]-, la mecánica cuántica no considera al mundo como compuesto de objetos y propiedades independientes del observador, sino que está compuesto de dos reinos: el reino macroscópico de los aparatos de medición, descrito por la física clásica, y el reino microscópico de los estados atómicos y moleculares descrito por vectores del espacio de Hilbert y operadores de proyección sobre el espacio de Hilbert. El corte entre ambos reinos depende del observador, algo que el simbolismo del cálculo de predicados no puede representar y por tanto la traducción es imposible. El lenguaje de la teoría cuántica de la Escuela de Copenhague no es convencional y se pregunta Putnam: “¿cuántos lenguajes de este tipo que podrían ser utilizados están esperando que alguien los invente?” [11] (pg. 142).

Cualquier sociedad humana es superable y existe otra sociedad posible y más sofisticada cuya manera de describir el mundo no puede ser comprendida por los miembros de la primera sociedad. Por tanto, si existiera una teoría sobre todos los discursos humanos solo podría escribirla un ser mucho más inteligente que los seres humanos de todas las posibles sociedades humanas que en el tiempo finito de existencia de dichas sociedades no va a aparecer. A pesar de este obstáculo insalvable, como la filosofía se interesa más por las condiciones de posibilidad que por la instanciación, sigue siendo pertinente la pregunta de si, a pesar de no existir esperanzas para la obtención de un algoritmo maestro para la interpretación de las teorías, este algoritmo podría existir. El principio de incompletitud de Gödel también prohíbe esta posibilidad porque existen más teorías a interpretar (problemas a resolver) que algoritmos se pueden construir en un lenguaje que contenga un número finito de caracteres. La demostración es simple [14]:

Sea X un conjunto cualquiera, finito o infinito, el *conjunto de las partes* de X , denotado $P(X)$, es el conjunto de todos los conjuntos que se pueden formar tomando cero o más elementos de X . Si X es finito y contiene n elementos, su cardinal es n y el cardinal de $P(X)$ es 2^n . Si X es infinito, entonces el cardinal de $P(X)$ es un infinito superior al de X . Para ello, Cantor demostró que no era posible establecer ninguna correspondencia biyectiva entre X y $P(X)$. Esta observación le condujo a afirmar que existe un número infinito de infinitos distintos, cada uno de ellos estrictamente mayor que el anterior. Para verlo, basta con formar la siguiente secuencia

cardinal (N) < cardinal($P(N)$) < cardinal($P(P(N))$) < ... donde N representa el conjunto de los números naturales.

El número de algoritmos que es posible expresar es a lo sumo un infinito numerable, es decir, el cardinal de los naturales porque cualquier algoritmo consistirá en un texto finito. Dado que el número de caracteres con los que se confeccionan dichos textos

también es finito, podríamos construir la siguiente lista con la seguridad de que aparecerá en ella cualquier algoritmo pasado, presente o futuro

todos los textos posibles de un carácter, por orden alfabético, todos los textos posibles de dos caracteres, por orden alfabético, todos los textos posibles de tres caracteres, por orden alfabético, etc³.

La conclusión es que no pueden existir más algoritmos que números naturales. Si ahora se procede al cálculo del número de teorías posibles, -que son los problemas a resolver-, resulta que, restringiéndose exclusivamente a aquellas teorías que se pueden traducir al cálculo de una función matemática, ésta produce una salida que es función de los datos contenidos en su argumento. Sin pérdida de generalidad, se puede suponer que dicha entrada y dicha salida son dos números naturales porque todos los datos alfanuméricos, gráficos, visuales o sonoros se pueden codificar mediante números y una secuencia de números, incluyendo los separadores entre ellos, se puede representar mediante un número más grande. Por tanto, el número de problemas que son candidatos a ser solubles mediante un algoritmo son las funciones $f: P(N) \rightarrow N$ de naturales en naturales.

Existe, por tanto, tantas funciones computables f como el cardinal de $P(N)$ y como el número de algoritmos es el cardinal de N , resulta que hay más problemas candidatos a ser computables que algoritmos para resolverlos. Luego han de existir problemas no computables y en el ámbito general de las teorías será imposible encontrar un algoritmo que permita decidir si son sinónimas, es decir, correferenciales, o no.

Para terminar este apartado acerca de Putnam y la revisión que hizo de su propio funcionalismo, comentar que uno de los aspectos de los fenómenos mentales que provoca una mayor perturbación, -en la búsqueda de una explicación en términos de estados y sucesos neurales-, es la intencionalidad. En el cap. 7 de su texto [11] Putnam aborda esta cuestión y opina que el paradigma computacional hizo surgir la necesidad de investigar la reducibilidad de la intencionalidad a los términos de las ciencias de la naturaleza pero que, él personalmente, no ve la posibilidad de una teoría científica del reino de lo intencional.

Corrientes reduccionistas

A pesar de las dudas de Putnam, en el panorama de la filosofía de la mente se encuentran bastantes corrientes que defienden posiciones reduccionistas que niegan la existencia de la mente como una realidad distinta del cerebro y tratan de explicar los fenómenos mentales en términos físicos o biológicos. Las posturas reduccionistas consideran, en general, que la distinción entre la mente y el cerebro es debida a la insuficiencia actual de nuestros conocimientos acerca de los procesos cerebrales y en este grupo se pueden situar cuatro planteamientos distintos

³ Muchos de esos textos no representarán algoritmo alguno; otros representarán el mismo algoritmo al que se han cambiado los nombres de las variables.

a) Conductismo filosófico

Ya presentado al comienzo de este trabajo, conviene explicarlo con un poco más de detalle ya que, pese a estar totalmente desacreditado en la actualidad, fue el paradigma predominante en los comienzos de la ciencia cognitiva. Desde sus inicios a fines del siglo XIX la psicología tenía como objeto la mente y sus contenidos (sensaciones, sentimientos e imágenes) pero, a comienzos del siglo XX, John B. Watson fundó una nueva escuela que consideraba la conducta como el tema exclusivo de estudio de la psicología, motivo por el cual fue denominada conductismo. En los años 50, Skinner amplió el conductismo de Watson, radicalizando sus presupuestos. El conductismo reduce todos los fenómenos psicológicos a términos de estímulo-respuesta observables y medibles; su objeto era trasladar a la psicología el método general de las ciencias naturales de la época. Los procesos intermedios entre el estímulo y la respuesta que ocurren en la mente (concebida como una caja negra) no tienen ningún interés para la investigación psicológica para ellos, ya que consideran que es posible explicar la totalidad de la conducta sin referirse a supuestos estados mentales. Como ya se comentó, esta fue la causa de su fracaso.

Para el conductismo lógico todo lo mental se reduce a disposiciones para la conducta, siendo la mente justamente la aptitud y disposición para realizar unas determinadas acciones. Propugnan abandonar el *lenguaje mentalista* (es decir, términos como deseo, creencia, sentimiento o propósito) de las explicaciones psicológicas y sustituirlo por un *lenguaje descriptivo* de los estímulos y respuestas, así como de las leyes que gobiernan la conducta. Una variante más radical del *conductismo metodológico* de Watson y Skinner es el *conductismo ontológico*, que negaba taxativamente la existencia de la mente.

b) Materialismo reduccionista (o Teoría de la identidad mente-cerebro)

El conductismo suponía la renuncia implícita a investigar la mente y sus fenómenos interiores, haciendo imposible la distinción entre los estados mentales y sus manifestaciones conductuales; pero continuamente se experimentan procesos mentales que no van acompañados de ninguna conducta alguna y para resolver estas dificultades y otras, derivadas del análisis restringido del conductismo, algunos filósofos como D. Lewis [16] y J.J.C. Smart [15] propusieron la teoría de que los procesos mentales son idénticos a los procesos cerebrales. Admiten la existencia de los procesos mentales como causa interna de la conducta pero, para esta corriente, la única explicación de la conducta humana y animal que es posible establecer en términos científicos es la debida al funcionamiento físico-químico del sistema nervioso central. Por consiguiente, los estados mentales son idénticos a los estados puramente físicos del sistema nervioso central y la psicología debe reducirse a la neurofisiología.

La teoría de la identidad mente-cerebro se enfrenta también a dificultades insuperables, como se verá algo más adelante.

c) *Materialismo eliminativo*

En su libro *Materia y conciencia* [17] Paul Churchland pasa revista a los temas principales de la filosofía moderna de la mente: el *problema ontológico*, dirigido a averiguar la naturaleza de los procesos mentales; el *problema semántico*, encaminado a esclarecer el origen del significado de los términos utilizados en psicología; el *problema epistemológico*, que se ocupa del conocimiento de las actividades internas de la mente y se divide en dos partes: el estudio de la conciencia de alguien diferente a uno mismo y el del conocimiento directo de los propios procesos mentales o autoconciencia; y finalmente el *problema metodológico*, dirigido a resolver cuál es la mejor metodología científica para elaborar una genuina ciencia de la mente. Su postura es la denominada *eliminativista* que plantea que el proyecto de la teoría de la identidad no es plausible debido a que nuestro marco psicológico basado en el sentido común es una concepción falsa y engañosa de las causas del comportamiento humano y de la naturaleza de la actividad cognitiva.

Como el conductismo, sobre todo en su versión radical, Churchland vuelve a eliminar los estados mentales. Considera que la creencia popular en la existencia de la mente es una teoría primitiva precientífica y que los estados mentales de los que hablamos en el lenguaje ordinario (creencias, deseos, sentimientos, intenciones) no existen realmente. Tal psicología popular debe ser sustituida por una neurociencia estricta, que parta de la idea de que las actividades cognitivas son en última instancia actividades del sistema nervioso y propone una inversión del procedimiento habitual de la investigación de los procesos cognitivos, que parte de las actividades cognitivas de los seres humanos (pensar, hablar, recordar, aprender, etc.) y luego indaga las operaciones cerebrales que pueden producirlas (aproximación de arriba abajo), por una aproximación de abajo arriba: empezar por comprender el comportamiento físico, químico, eléctrico y de desarrollo de las neuronas y sólo después tratar de comprender lo que sabemos intuitivamente sobre nuestras actividades cognitivas.

Reduccionismo de J. Kim

Dentro de la corriente reduccionista el filósofo de la mente quizá más influyente en la actualidad es Jaegwon Kim y por ello se le dedica un apartado exclusivo. En su libro *Mind in a physical world* [18] empieza por señalar como la *teoría de los estados cerebrales*, basada en las investigaciones neurológicas expuestas anteriormente, puso las bases para los debates posteriores sobre la filosofía de la mente y que la primera propuesta reduccionista que parecía surgir de forma natural, la teoría de la identidad, tuvo una vida corta debido, por un lado, al funcionalismo de Putnam, que establecía que las funciones mentales podían tener realizaciones múltiples en distintos sistemas, y, por otro, al monismo anómalo de Donald Davidson [19] que defendía que todo suceso mental es un suceso físico no sometido a leyes “estrictas”. Kim admira el trabajo de Davidson, aunque no esté de acuerdo con él, y en particular le reconoce la introducción

del concepto de *superveniencia* [19] (pg. 117), ampliamente utilizado en el campo de la filosofía de la mente. Pero la consecuencia principal del monismo anómalo es que, al negar que existan leyes para los fenómenos mentales, está afirmando que carecen de regularidad y que no se pueden establecer relaciones claras entre los sucesos físicos y los mentales y con ello se niega la posibilidad del máspreciado objeto de búsqueda del reduccionismo: las leyes psicofísicas.

Un paréntesis terminológico

A propósito del término superveniencia, conviene hacer una pequeña parada para comentar, aunque sea muy someramente, que uno de los aspectos que hacen más complicada la comparación y distinción entre las distintas teorías de filosofía de la mente es la existencia de un conjunto de términos de sentido equívoco a los que es necesario prestar atención y delimitar su significado con la máxima precisión posible. Como ejemplo, se pueden comparar dos de estos términos: la superveniencia y la reducción. El monismo anómalo que defiende la superveniencia, al negar la existencia de leyes psicofísicas, está protegiendo la autonomía de lo mental y negando la reducción de este ámbito al de lo físico; en este contexto, *superveniencia* y *reduccionismo* son términos opuestos y esta es la relación más habitual entre ambos términos, pero en la literatura científica se encuentran documentos donde su relación es distinta⁴. Por otro lado, la superveniencia no es un término de uso exclusivo del ámbito de la filosofía de la mente, sino que se aplica a la relación entre ciencias de distinto nivel, particularmente entre lo que se denominan ciencias especiales, como la química o la biología, y el sustrato inferior que es la física. La superveniencia permite la existencia de leyes propias en cada nivel y es contraria a un reduccionismo de una esfera del conocimiento a otra; esta relación entre niveles de complejidad se encuentra muy bien documentada y explicada en el libro de Paul Davies *Proyecto Cósmico*, en el que se llega a justificar el libre albedrío basándose en las leyes del caos [21] (pg. 190); por otro lado, Davies atribuye el concepto de *superveniencia*, que también es conocido por el nombre alternativo de *emergentismo*, a quien considera su fuente de inspiración en este tema, el premio Nobel de Medicina R. W. Sperry, ([22], pg. 201).

Fin del paréntesis

Regresando a la forma en que Kim presenta el concepto de superveniencia, lo define de la siguiente manera:

“Lo mental sobreviene de lo físico porque es necesario que cualquier sistema que tiene una propiedad mental M en el instante t, tenga una propiedad física de base que le subviene que tiene el valor P en el instante t y cualquier sistema que tenga P en t tiene M en t.”

⁴ Para Chalmers, por ejemplo, existen dos tipos de superveniencia: el lógico que es reduccionista y el natural que no lo es [20]

Una formulación equivalente la denomina el principio de *indiscernibilidad* y reza así

“Lo mental sobreviene de lo físico porque necesariamente cualquier par de sistemas (en el mismo mundo o en diferentes mundos posibles) indiscernibles en todas sus propiedades físicas, lo son en sus aspectos mentales.”

Como se deduce de los dos párrafos anteriores, el tratamiento de Kim de la superveniencia es reduccionista, pero su planteamiento no es muy riguroso porque intenta distinguirla de la dependencia con el argumento de que esta es asimétrica [18] (pg. 11) cuando, por el contrario, la relación de dependencia de una propiedad con otra es, en general, simétrica. Según su planteamiento, la superveniencia es una relación simétrica de covarianza que, para transformarse en una relación de dependencia, necesita que se le añada algo que no es capaz de definir. Toda esta parte de la propuesta de Kim es muy confusa pero, no obstante, uno se puede quedar con su pregunta crucial: ¿puede la superveniencia servir como base de una teoría filosófica sobre la forma en que las mentes se relacionan con los cuerpos? Su respuesta es negativa porque dice que la superveniencia es consistente con muchas teorías mente-cuerpo mutuamente excluyentes. Entre estas teorías cita las siguientes [18] (pp. 13-14)

- i. *Epifenomenalismo*, que considera a la superveniencia como una relación de dependencia causal con el sustrato físico que caracteriza a los fenómenos mentales como contingentes, es decir, que podrían no existir para las mismas realizaciones físicas y, sobre todo, niega su capacidad causal. Esta versión es, en verdad, eliminativista.
- ii. *Emergentismo*, que considera la superveniencia como un hecho bruto que debe ser reconocido pero que no admite explicación. Las propiedades que emergen son intrínsecas, con sus propias capacidades causales e irreducibles a las propiedades de los niveles de base. Esta es la postura de Davidson⁵, Davies y Sperry.
- iii. *Realizacionismo físico* que considera a las propiedades mentales como propiedades funcionales de segundo orden definidas sobre las propiedades físicas de primer orden. Esta es, para Kim, la mejor explicación de la superveniencia y lo que hace es sustituir el dualismo cartesiano por un mundo dividido en niveles de complejidad en cada uno de los cuales existen fenómenos y propiedades que emergen en dicho nivel. Este marco multinivel lo considera un buen esquema ontológico donde situar el problema mente-cuerpo y en él la pregunta clave es: ¿cuál es la relación entre las propiedades de cada nivel y las de los niveles adyacentes,

⁵ Creo que esta es la posición que le corresponde a este filósofo, a pesar de la afirmación de Kim de que su monismo anómalo conduce al epifenomenalismo [18] (pg. 34).

particularmente las del nivel inferior? Lo que, desde su punto de vista, se dibuja es la siguiente propuesta

“Si dos propiedades x e y pertenecientes al nivel N son indiscernibles en todos los niveles inferiores a N , lo son en N .”

Para Kim todas estas consideraciones apuntan a una concepción de la reducción que encaja bien con el paradigma reduccionista general de la ciencia, aunque no en el sentido de Nagel para quien la reducción consiste en deducir las leyes matemáticas de un nivel a partir de las del nivel inferior (o teoría base) mediante leyes puente [23]. Para que se pudiera aplicar directamente este tipo de reducción cada propiedad mental debería tener una propiedad física coextensiva en todas las especies y sistemas; como las propiedades mentales tienen realizaciones múltiples, no satisfacen esta condición y Kim debe hacer una propuesta reduccionista distinta que formula en el capítulo 4º de su libro.

Su propuesta básicamente consiste en proponer leyes puente específicas para cada especie y para darle cobertura teórica se vuelve hacia el *funcionalismo* de Putnam, afirmando que la posibilidad de la funcionalización es una condición necesaria para la reducción. Por ello, la cuestión central en el debate sobre las relaciones entre la mente y el cuerpo es si las propiedades mentales son funcionalizables⁶ o no. Para decidir sobre esta cuestión, Kim agrupa los fenómenos mentales en dos grandes categorías: a) los *intencionales*, que se manifiestan en las actitudes proposicionales, y que están referidos a objetos exteriores y b) los *qualia* que se refieren a impresiones internas (como el color, el sonido o el olor).

Siempre según este filósofo, y frente a otros como Putnam o Searle, que la intencionalidad es funcionalizable se explica por el hecho de que es inconcebible una réplica exacta del mundo físico existente que careciera de intencionalidad. Sin embargo, se puede concebir un mundo físicamente idéntico en que los *qualia* tengan una relación diferente con los sistemas físicos. En otros términos

- La intencionalidad es funcionalizable porque no se sabe cómo diseñar y construir sistemas intencionales; sólo se podría duplicar los seres vivos lo que significa que la intencionalidad es una función específica y exclusiva de ciertos sistemas vivos.
- Por el contrario, se sabe cómo diseñar y construir sistemas (distintos a los vivos) capaces de tener percepciones o de procesar esta información (los robots), luego, según Kim, no son funciones exclusivas de los seres vivos y, por tanto, no serían funcionalizables. Desde mi punto de vista, lo correcto sería decir que existen distintas realizaciones de dichos fenómenos (caso iii de los modelos de superveniencia), pero Kim lo

⁶ Recordando el símil computacional del apartado sobre el funcionalismo, que una propiedad o un estado, mental en este caso, sea funcionalizable significa que consiste en una función u operación que realiza un sistema físico determinado.

interpreta como un ámbito en el que el emergentismo podría ser aplicable (caso ii de los modelos de superveniencia).

Kim reconoce pues que existe un problema con las propiedades supervinientes como los qualia que no son reducibles a sus propiedades de base y ve las siguientes salidas [18] (pp. 118- 120):

- 1^a.- Si se mantiene el funcionalismo, se llega a un punto en que debe elegirse una opción entre dos
 - a) Mantener el fisicalismo, respetando al menos el cierre causal de la física.
 - b) Abandonar el fisicalismo y optar por el dualismo y el interaccionismo.
- 2^a.- Si uno se mantiene en el fisicalismo, debe elegir entre
 - a) Una superveniencia irreducible, es decir, no funcionalizable, pero aceptando su impotencia causal.
 - b) Abrazar el materialismo eliminativo.

Termina concluyendo que esto no debe sorprender porque el fisicalismo es una doctrina ontológica general acerca de toda la realidad y mantenerse en él exige pagar un alto precio. Sin embargo, es partidario de conservar lo mental como parte del mundo físico, pero dentro de lo que denomina un *fisicalismo robusto*. Para él, las posiciones intermedias como el monismo anómalo de Davidson o el fisicalismo no reduccionista⁷ no son aceptables dentro de este fisicalismo robusto⁸.

Vuelta a Putnam

En la segunda parte de su libro *La trenza de tres cabos* [24], compuesta por tres conferencias, Hilary Putnam revisa las posiciones reduccionistas, muy particularmente las de Kim⁹. En la primera de las conferencias retoma el viejo experimento mental de William James acerca de la existencia de una persona con todas las características físicas de una persona normal, pero carente de propiedades mentales. Acerca de este experimento, dice que existen dos posturas que resumirían las dos corrientes de opinión extremas sobre las relaciones mente-cuerpo

- i. AUTOMATA: Aunque ciertas personas no tuvieran propiedades mentales, si sus propiedades físicas y sus entornos físicos fueran iguales, ocurrirían los mismos sucesos físicos.
- ii. NO-AUTOMATA: Si las personas no tienen propiedades mentales, pero sus propiedades físicas y sus entornos físicos fueran iguales, podrían ocurrir sucesos físicos distintos. Esta postura era la de Descartes para quien la

⁷ No deja claro si se refiere a Putnam

⁸ No aclara qué significa *fisicalismo robusto*.

⁹ Hay que tener presente que cuando Putnam escribe este libro ya ha renunciado al funcionalismo

mente era la causa de que nuestros cuerpos actuasen de una determinada manera.

El interaccionismo fuerte de la segunda postura es inadmisibles para Putnam, pero rechaza también la primera y dentro del panorama de la filosofía de la mente encuentra tres posiciones intermedias entre los dos extremos, -el fisicalismo reduccionista (postura que identifica con Kim), el conductismo lógico, y el verificacionismo-, que rechaza para hacer su propuesta personal. Esta parte de un análisis lingüístico del antecedente del silogismo AUTOMATA, -es decir, *ciertas personas no tienen propiedades mentales, pero todas sus propiedades físicas son iguales a cómo serían si las tuviesen y fueran iguales sus entornos físicos-*, y considera que no es inteligible. Para que lo fuera deberíamos saber qué es lo que se está preguntando cuando se pregunta si una determinada propiedad mental es o no reducible a otra propiedad física.

Como Putnam no cree que las propiedades tengan un sentido de identidad particular aislado, sino que lo tienen dentro de una teoría [12], la relación entre propiedades de ámbitos distintos solo se puede dar en el marco de teorías previamente definidas, la reducida y la reductora. Mientras estas teorías no existan, la pregunta sobre la reducibilidad, la dependencia o la independencia de lo mental con respecto a lo físico carece de sentido. Para explicar con más claridad esta opinión, comenta la existencia de dos concepciones distintas de la semántica del lenguaje natural: la clásica, de la que toma como representante destacado a P. Grice, y otra, menos familiar y atribuida a Austin y a Wittgenstein, sensible al habla y que defiende que lo fundamental para la comprensión e incluso para la misma posibilidad del lenguaje es el sentido compartido de lo que es y no es una proyección natural de los usos previos de una palabra dentro de un nuevo contexto y, en este sentido, concluye Putnam que, comprendiendo lo que significa que “los miembros de una tribu sean autómatas sin alma”, no comprende lo que significa en el contexto del argumento de J. Kim.

En la tercera de sus conferencias, Putnam ataca directamente el problema clave de la posibilidad de existencia de leyes psicofísicas. Comienza definiendo con precisión la tesis reduccionista de J. Kim que denomina de *superveniencia estricta* y que considera la conjunción de dos tesis

- a) *Tesis de la superveniencia.*- Todo estado psicológico interno de un organismo sobreviene a su estado físico interno sincrónico
- b) *Tesis de la dependencia.*- Toda propiedad mental depende de su correspondiente propiedad física

La segunda tesis es más fuerte que la primera porque no se da sólo entre estados sino propiedad a propiedad o, lo que es lo mismo, la segunda tesis presupone la primera, lo que significa en definitiva que existe un isomorfismo total entre lo mental y lo cerebral. De forma similar al debate sobre los silogismos “autómata” y “no-autómata”, Putnam va a defender la ininteligibilidad de la tesis de la superveniencia y,

en consecuencia, la de la dependencia. Según el planteamiento de Kim, aquella exige que si se pudiera replicar a una persona y estuviera en el mismo estado neural interno¹⁰ que el original, los estados psicológicos internos de ambos serían exactamente los mismos. Según Kim, todo estado psicológico que no sea interno se puede descomponer en un estado interno y un conjunto de relaciones externas y la conducta es una función, -en el sentido matemático-, cuyos argumentos son los estímulos exteriores y los estados internos de manera que, en consecuencia, dos sistemas físicamente idénticos tienen la misma psicología. Lo que cuestiona Putnam es la inteligibilidad de considerar a las creencias como estados psicológicos internos.

Para ello se apoya en una postura que denomina *externalismo semántico*, que dice compartir con la mayoría de los filósofos del lenguaje y de la mente de la actualidad, y que consiste en esencia en defender que el contenido de las oraciones, de las creencias y de otros contenidos psicológicos dependientes del lenguaje dependen de la referencia (su extensión) de sus términos en el contexto particular del habla y que la referencia depende de factores externos al cerebro del hablante, de manera que las condiciones neurológicas del hablante no son suficientes para determinar el significado de lo que dice. Para Stanley Cavell lo que convierte en correcto el uso de las palabras es la razonabilidad de su proyección dentro del contexto y esta razonabilidad no está sostenida por ningún estado interno sino que va cambiando para adaptarse a aquel. De manera que es tan inútil buscar un objeto científico denominado *contenido estricto* en el caso del significado como un estado psicológico interno en el de las creencias.

Para dejar bien sentada su afirmación previa, Putnam lleva a cabo una demostración contra la idea de las representaciones interiores. El argumento, en esencia, es el siguiente: “Si tengo la misma experiencia despierto y en sueños, deduzco que tienen algo en común que debe ser interno y lo llamo representación”. Apoyándose en Austin, sostiene que no se está autorizado a decir que exista algún objeto significante (un dato sensorial) presente en ambos casos porque no está claro *si se ve algo o parece que se está viendo ese algo* y hay una diferencia radical entre ambas situaciones, ya que en el primer caso se puede afirmar que se está en el mismo estado psicológico, lo cual es una relación transitiva, mientras que la indistinguibilidad de las apariencias no satisface la propiedad transitiva. Para demostrarlo presenta el experimento realizado por un experto en la lógica difusa, R. Parikh, que consiste en presentarle a un sujeto una colección de 100 tarjetas, T_1 a T_{100} bañadas en un bote de pintura blanca a la que se le iba añadiendo una gota de pintura roja al bañar la siguiente tarjeta. Evidentemente, al sujeto dos tarjetas consecutivas le parecen idénticas mientras que un par de tarjetas suficientemente separadas en la serie la parecen distintas. Si a la cualidad del color que presentan dos tarjetas consecutivas

¹⁰ De la lectura del texto [24] (pg. 133) se deduce que un estado psicológico interno se corresponde con una creencia y un estado psicológico externo con un conocimiento.

se le denomina $Q_{i/i+1}$ se tiene que, según lo manifestado por el sujeto, $Q_{i-1/i} = Q_{i/i+1}$ de manera que aplicando esta igualdad desde la primera hasta otra suficientemente alejada en la serie, la última si se quiere, resultaría que $Q_{1/2} = Q_{i/i+1} = Q_{99/100}$ y la última tarjeta sería indistinguible de la primera, cosa totalmente falsa, y de este razonamiento se concluye que es posible encontrarse en dos estados distintos sin que seamos capaces de notarlo, es decir, que al mismo estado mental (la apariencia) le corresponderían dos estados neurales (físicos) diferentes. Concluye Putnam que no se trata de que no existan las leyes psicofísicas, sino de que plantear su existencia no tiene sentido porque las características psicológicas resultan extremadamente sensibles al contexto y dependen de factores externos (objetos) y sociales y de las proyecciones que en cada contexto resulte natural hacer o no hacer.

Conclusión

Para concluir este trabajo me basaré en las opiniones del investigador y profesor de neurociencia Antonio Damasio, plasmadas en su libro *El error de Descartes* [25] y en el cual devuelve la mente al lugar del que nunca debió salir: el cuerpo. En él nos dice que la mente ha sido desplazada del lugar etéreo que ocupaba en el s. XVII hasta algún lugar situado en el cerebro y que surge de la actividad de circuitos neuronales modelados a lo largo de la evolución y que representan al organismo de forma global y continua y son perturbados por los estímulos del medioambiente (físico y sociocultural). Una mente normal solo tiene lugar si dichos circuitos tienen representaciones básicas de todo el organismo, de manera que el cuerpo no solo proporciona sustento vital al cerebro sino que suministra también contenidos a la mente.

La aparición de las imágenes mentales se puede justificar en términos evolutivos como un aumento de las posibilidades de supervivencia del organismo, mediante una mejor apreciación de las condiciones ambientales y una mejor predicción de los escenarios futuros y de las acciones a emprender en ellos gracias a la posibilidad de imaginarlos. Una sugerencia muy atractiva de Damasio es que una forma de aumentar la supervivencia del cuerpo, en su momento, fue la representación del mundo en términos de los cambios ocasionados en el cuerpo, aunque en la actualidad predominen las imágenes no corporales.

Elementos fundamentales en la filosofía de la mente han sido los conceptos de *conciencia* y de *yo*; Damasio ofrece explicaciones de los mismos que considero bastante convincentes. El *yo* refleja la idea de algo estable en el organismo y lo polémico ha sido establecer la identidad de ese “algo estable”. Él lo explica a partir de una base neuronal consistente en la reactivación de dos tipos de representaciones: las relacionadas con acontecimientos clave en la vida del individuo y las relacionadas con elementos primordiales del cuerpo del individuo que comprenden estados corporales de fondo y estados emocionales. Afirma que el dispositivo neuronal mínimo para producir la

subjetividad no exige la existencia del lenguaje y, por tanto, es accesible a los animales de orden superior; el lenguaje sería una capacidad narrativa de segundo orden.

En definitiva, el *yo* no es un homúnculo situado dentro del cerebro que contempla nuestras representaciones mentales, sino que es un estado biológico continuamente reconstruido, de manera que poseer un *yo* es perfectamente compatible con la idea de no poseer un teatro cartesiano dentro de un nuestro cerebro. Existe un *yo* para cada organismo, excepto en aquellas situaciones en que la enfermedad ha creado más de uno (personalidad múltiple) o ha reducido o abolido el *yo* normal (anosognosia). Cita el famoso experimento del cerebro en la tina de Putnam y opina que si a los nervios cerebrales desconectados del cuerpo se les introdujeran señales idénticas a las que proviene del cuerpo, podría existir actividad mental en el cerebro pero que es altamente improbable que estas entradas pudieran imitar el comportamiento del cuerpo activado por el propio cerebro.

Síntesis personal

Para finalizar, deseo presentar un resumen que muestre las distintas posturas dentro del campo de la filosofía de la mente que considero suficientemente nítidas tras la revisión bibliográfica efectuada. En una gradación que se mueva de las posiciones más mentalistas a las más fiscalistas, las propuestas que se ofrecen hoy se pueden agrupar, desde mi punto de vista y aunque sea un tanto groseramente, de la siguiente manera

1ª.- Dualismo.- Contempla la existencia de dos mundo distintos; aquí se sitúan, aparte del propio Descartes, filósofos como K. Popper,- que, de hecho, cree en la existencia de tres mundos [26]-, o J. Eccles.

2ª.- Emergentismo.- Las actividades mentales necesitan el soporte material del cerebro, pero tienen sus leyes propias no reducibles a la neurofisiología. Aquí se situarían Davidson, Davies y Sperry.

3ª.- Reduccionismo.- Las actividades mentales no sólo necesitan el soporte material del cerebro, sino que sus leyes son reducibles a las de la neurofisiología, aunque no se haya encontrado su relación precisa. Esta opción presenta, por lo menos, las tres variantes siguientes

3.1.- Funcionalismo.- Las actividades mentales son las funciones que distintos sistemas físicos pueden realizar, por lo que se puede denominar asimismo *realizacionismo físico*. Es la postura del Putnam de 1960.

3.2.-Fiscalismo robusto.- Es la postura de J. Kim que, honradamente, la considera confusa y me siento incapaz de definir con precisión.

3.3.-Reduccionismo matemático.- Es la postura de Nagel, para quien la relación entre los niveles superiores e inferiores debe manifestarse a través de leyes matemáticas

4ª.- Eliminativismo.- Los fenómenos mentales no existen o son epifenómenos de los procesos cerebrales. Aquí se situaría Churchland.

BIBLIOGRAFIA

1. R. Descartes. *Meditaciones metafísicas*. Alianza, Madrid, 2005; *Meditaciones* 2ª, 5ª y 6ª.
2. R. Descartes. *El tratado del hombre*. Alianza, Madrid, 1990.
3. H. Gardner. *La nueva ciencia de la mente* Paidós, Barcelona, 2000 (orig.: 1985); caps. 2 y 3.
4. A. M. Turing. *Computing machinery and intelligence*. *Mind*, vol. 59, 1950, pp. 433-460.
5. M. A. Arbib. *Cerebros, máquinas y matemáticas*. Alianza, Madrid, 1976 (orig.:1964); cap. 1.
6. W. S. McCulloch and W. Pitts. *A logical calculus of the ideas immanent in neurons activity*. *Bull. Math. Biophys.*, vol. 5, 1943, pp. 115–133.
7. J. E. Eccles. *The Physiology of Imagination*. *Scientific American*, 1959, pp. 135-146.
8. K. Lashley. *In search of the engram*. *Symposia of the Society for Experimental Biology*, vol. 4, 1950, pp. 454-482.
9. U. Neisser. *Psicología cognitiva*. Trillas, Mexico, 1979 (orig.: 1967); caps. 1 y 11.
10. H. Putnam. *Mentes y máquinas* en “*Mentes y máquinas*”, M. Garrido (ed.). Tecnos, Madrid, 1985; pp. 61-101 (orig.: 1960).
11. H. Putnam. *Representación y realidad*. Gedisa, Barcelona, 2000 (orig.: 1988); caps. 5 y 7.
12. W. V. Quine. *Two dogmas of empiricism*. *Philosophical Review*, vol. 60, 1951, pp. 20-42.
13. H. Putnam. *A philosopher looks at quantum mechanics*. *Brit. J. Phil. Sci.*, vol. 56, 2005, pp. 615-634.
14. V. A. Uspensky. *Gödel’s incompleteness theorem*. Mir Publishers, Moscow, 1987.
15. J. J. C. Smart. *Sensations and brain processes*. *Philosophical Review*, vol. 68, 1959, pp. 141-156.
16. D. Lewis. *An argument for the identity theory*. *Journal of philosophy*, vol. 63, 1966, pp. 17-25.
17. P. Churchland. *Materia y conciencia: introducción contemporánea a la filosofía de la mente*. Gedisa, Barcelona, 1999 (orig: 1984).
18. J. Kim. *Mind in a physical world*. MIT Press, Cambridge, 1998; caps. 1 y 4.
19. D. Davidson. *La mente material* en “*Mentes y máquinas*”, M. Garrido(ed.). Tecnos, 1985; pp. 103-126 (orig.: 1971).
20. D. Chalmers. *La mente consciente. En busca de una teoría fundamental*. Gedisa, Barcelona, 1999 (orig.: 1996); cap. 2.

21. P. Davies. *The cosmic blueprint*. Simon and Schuster, N. Y., 1988
22. R. W. Sperry. *Mind-brain interaction: mentalism, yes; dualism, no*. *Neuroscience*, vol. 5, 1980, pp. 195-206.
23. E. Nagel. *La estructura de la ciencia*. Paidós, Barcelona, 1981 (orig.: 1961); cap. 11.
24. H. Putnam. *La trenza de tres cabos*. Siglo XXI, Madrid, 2001 (orig.: 1999); 2ª parte.
25. A. Damasio. *El error de Descartes*. Drakontos bolsillo, Barcelona, 2009 (orig.: 1994); cap. 10.
26. K. Popper. *Conocimiento objetivo*. Tecnos, Madrid, 1988 (orig.: 1974); pg. 148.