

Alumno: Carlos Ruiz Jiménez
Asignatura: Filosofía de la mente
Profesor: Eduardo de Bustos Guadaño

Enero de 2020

CONDUCTISMO, FUNCIONALISMO Y ELIMINATIVISMO

“La soledad absoluta es el destino inevitable del alma. Solamente nuestros cuerpos se pueden encontrar.”

(Ryle 1949, p. 29)

“...todo argumento filosófico que haya sido empleado en relación con el problema de la mente y del cuerpo (...) tiene su exacta contraparte en el caso del “problema” de los estados lógicos y estructurales en las máquinas de Turing.”

(Putnam 1960, p. 82)

“Si, finalmente, el materialismo está en lo cierto, entonces es el marco conceptual de una neurociencia madura que encarnará la sabiduría esencial sobre nuestra naturaleza interna.”

(Churchland 1984, p. 255)

Índice general

| | |
|--|-----------|
| INTRODUCCIÓN | 1 |
| 1. CONDUCTISMO FILOSÓFICO | 3 |
| 1.1. Contra la leyenda intelectualista | 3 |
| 1.2. Sentimientos y sensaciones | 7 |
| 2. FUNCIONALISMO | 9 |
| 2.1. Realizabilidad múltiple | 9 |
| 2.2. Sensaciones: el dolor | 11 |
| 3. ELIMINATIVISMO | 15 |
| 3.1. De la Identidad al Eliminativismo | 15 |
| 3.2. Sensaciones: el color | 18 |
| CONCLUSIONES | 21 |
| BIBLIOGRAFÍA | 26 |

INTRODUCCIÓN

En este trabajo se exponen algunos de los aspectos más importantes de las teorías filosóficas en torno a la mente que se han dado en llamar conductismo filosófico, funcionalismo y eliminativismo. El motivo por el cual nos ha parecido oportuno escoger estas tres corrientes es porque creemos que se trata de una evolución de cierta clase de pensamiento con ideas coincidentes en origen, aunque con desarrollos diferentes. Las tres teorías quieren escapar del influjo de lo que se ha dado en llamar *psicología popular*, un “...conjunto de conocimientos que todos los seres humanos poseemos *de facto* acerca de los procesos mentales en sí mismos y en su relación con la conducta” (Martínez-Freire 1995, p. 64) o “...un cuerpo de explicaciones y predicciones de la conducta que se caracteriza esencialmente por su apelación a estados mentales con contenido –deseos, creencias, miedos, intenciones, intuiciones y, en general, estados psicológicos contruidos en términos proposicionales– como causas de tal conducta” (Toribio Mateas 1995, p. 246), y las tres se alejan, aunque con matices, del dualismo ontológico cartesiano y del *mentalismo* entendido como la aceptación de la existencia de procesos internos de pensamiento que no pueden ser explicados ni compartidos por la comunidad.

El primer capítulo está dedicado, como no podía ser de otra forma, a ciertos aspectos de la obra *El concepto de lo mental*, del filósofo británico Gilbert Ryle (Ryle 1949), como el precursor del conductismo filosófico, también llamado conductismo lógico. Se trata de una tradición heredera del positivismo lógico y la filosofía analítica que no hay que confundir con el conductismo psicológico de Watson y Skinner, basado en una metodología empírica de estudio de la conducta. Los rasgos en común son bastante evidentes, pero el conductismo filosófico pretende analizar la conducta relacionándola con la verbalización real o potencial de ciertas disposiciones a ella y con circunstancias observables, y revisita el problema mente-cuerpo calificándolo como un falso problema fruto de errores lógicos. En la primera parte de este capítulo se tratan estas cuestiones. Aunque los tratados de filosofía de la mente añaden siempre como otra de las bases del conductismo lógico el *Investigaciones filosóficas* de Wittgenstein (Wittgenstein 1953), la dificultad de su análisis y la brevedad que imponen las condiciones de este estudio han obligado a incluirlo solo tangencialmente en algún comentario. Los dos autores coinciden en alejarse del mentalismo, pero el caso de Ryle, quizás más relacionado con la materia de la filosofía de la mente que nos ocupa, es

un alejamiento total del uso de este concepto en las explicaciones del comportamiento, y el segundo Wittgenstein sigue admitiendo en algunos casos los procesos mentales internos como causas de conducta, y creemos que su análisis quizás sea más pertinente en un curso de filosofía del lenguaje. En la segunda parte del capítulo se explican una serie de consideraciones que el conductismo filosófico aporta sobre el problema de las sensaciones y los sentimientos.

En el segundo capítulo del escrito analizamos el funcionalismo tal como lo concibió el filósofo norteamericano Hilary Putnam en los años sesenta (Putnam 1960; Putnam 1963; Putnam 1967). Se trata de un escritor difícil que además a lo largo de su trayectoria se fue alejando de sus posturas iniciales, pero dejamos claro cuáles son las ideas fundadoras de esta corriente en la filosofía de la mente y cuáles han sido algunas de sus reconsideraciones posteriores (Putnam 1988; Putnam 1999), aunque estas últimas sólo han sido apuntadas tímidamente, ya que se alejan del objetivo de este resumen. En la primera parte describimos la esencia de su programa funcional computacional, sintetizada en el concepto de la realizabilidad múltiple de los estados mentales, que el funcionalismo sí admite como causas de conducta, y apuntamos brevemente algunas ideas del llamado funcionalismo analítico, de David Lewis. Introducimos aquí también el problema de los *qualia* o cualidades experienciales particulares (Hierro-Pescador 2005, p. 194). En la segunda sección se resumen las ideas que sobre la sensación de dolor hizo Putnam en sus primeros escritos, y por ende la solución que propone la teoría funcionalista a las experiencias subjetivas y la crítica de la introspección como fuente de conocimiento, lo que le acerca al conductismo de Ryle.

Por último, dedicamos un capítulo al eliminativismo, desarrollado en los años ochenta. Nos hemos enfocado principalmente en el análisis de la obra *Materia y conciencia*, de Paul Churchland (Churchland 1984), y del artículo *¿Puede la neurobiología enseñarnos algo sobre la conciencia?*, de Patricia Churchland (Churchland 1993). En primer lugar relatamos la evolución del eliminativismo a partir de las teorías de identidad de tipos, así como la confianza en que una eliminación o revisión de los términos psicológicos facilite la labor reduccionista, y en segundo lugar nos centramos en el análisis reduccionista concreto que propone Paul Churchland a propósito de las sensaciones de color. Además, por indicación del profesor Armero, completamos esta parte con las teorías eliminativistas del concepto de color como propiedad física que han hecho varios filósofos a partir de los experimentos de percepción del mismo de la segunda mitad del siglo XX. Veremos cómo este tipo de eliminativismo llevado al extremo, en aras de una correcta interpretación de los resultados empíricos, pretende eliminar también conceptos normalmente asociados con propiedades físicas. En este caso, no simplemente se trata de eliminar el vocabulario psicológico en torno a las sensaciones de color, sino la misma propiedad física del color como algo objetivo. Paradójicamente, contra este tipo de eliminativismo reaccionará el propio Paul Churchland.

1. CONDUCTISMO FILOSÓFICO

1.1. Contra la leyenda intelectualista

Descartes en su religiosidad no podía aceptar que la mente humana fuera una complicada variedad de un artefacto mecánico, debería diferir en esencia de mecanismos como el de un reloj. Las diferencias entre las causas mentales y los efectos físicos tenían que ser de diferente naturaleza. Según Ryle, Descartes defendía una hipótesis *paramecánica* del problema mente-cuerpo. Las operaciones mentales ni están en el espacio ni en el tiempo ni se asociaban a partículas materiales ni podían observarse públicamente. El cuerpo humano ya no sería una máquina al uso, sino una máquina gobernada por un espectro invisible dentro de ella. Las leyes que rigen las mentes debían ser no mecánicas. Pero como los cuerpos seguían leyes deterministas, las leyes de las mentes también deberían serlo. Siendo esto así, no se podría hablar de conceptos como, por ejemplo, *elección* o *responsabilidad*, a menos que las leyes mecánicas que gobiernan las mentes sean menos estrictas que las corporales. El libre albedrío sólo se puede concebir si se acepta que la conducta humana no responde a las mismas leyes causales del funcionamiento de las máquinas. Pero si no se puede acceder a las causas inmatriciales de la conducta, no podríamos distinguir entre un comportamiento automático y otro intencional, no se podría establecer la distinción entre hombre y robot. Ni siquiera podríamos caracterizar los comportamientos de las personas como inteligentes o tontos, porque no se podría establecer una hipótesis causal que sirva como fundamento.

Para Ryle el origen de esta paradoja se encuentra en un error en la lógica del problema. Descartes se debió preguntar por los criterios que se utilizan en la práctica para distinguir los comportamientos de las personas en lugar de buscar una causación inmaterial de ellos. Buscaba alguna disciplina análoga a la mecánica que estuviera detrás de las conductas humanas, pero no se pueden poner en conjunción términos de tipos diferentes. Ese era el error del dogma del «fantasma en la máquina», que no se puede admitir que los procesos corporales tengan causas mecánicas y mentales a la vez, porque pertenecen a categorías distintas, no pertenecen al mismo tipo lógico. Esto no quiere decir, eso sí, que se niegue la existencia de procesos mentales.

Cuando hemos de hacer descripciones de los comportamientos de las personas, para el conductismo no estamos describiendo los efectos que generan ciertas causas mentales ocultas que generan expresiones o actos, sino esas expresiones y actos en sí mismas, es decir, las diferencias no se basan en las suposiciones de acciones fantasmales implícitas que construyen esas expresiones, sino en el análisis de ciertos tipos de aserciones verificables que puedan explicar o predecir las conductas. Los supuestos estados mentales son en realidad disposiciones a la conducta que se podrían traducir a la lógica proposicional. En el contexto de los juegos lingüísticos del segundo Wittgenstein, nos estaríamos refiriendo a los comportamientos reconocibles que generan los distintos usos del lenguaje.

Tradicionalmente se ha asociado la principal función de los procesos mentales con actividades de cognición, de teorización. Las actividades más importantes del intelecto humano habrían dado lugar a las matemáticas y las ciencias naturales. Esa sería la superioridad del hombre civilizado sobre los bárbaros o los animales, el descubrimiento de proposiciones verdaderas. En eso consistía el ser racional. Para Ryle, sin embargo, hay muchas actividades no asociadas a ninguna operación intelectual de esta índole. Nuestro pensamiento interno consiste en una serie de imágenes visuales y proposiciones que, en la mayoría de los casos, no articulamos con palabras, pero para desarrollar esta capacidad es necesario que hayamos oído y comprendido las proposiciones de los demás y que hayamos hablado en voz alta como van aprendiendo los niños. La mente no es, para Ryle, el sitio de los pensamientos silenciosos que nos relatamos a nosotros mismos, este es otro error de la concepción tradicional. Según lo que denomina *leyenda intelectualista*, la persona que hace cosas con corrección lo consigue gracias a que sigue una serie de proposiciones internas que le guían acerca de lo que debe hacer. Está siguiendo una teoría previa. Pero esta leyenda es falsa, la teoría es una consecuencia de una práctica eficaz. Antes de Aristóteles, que definió las leyes del buen razonar, ya había gente inteligente.

El argumento principal contra la leyenda intelectualista es que supone una regresión infinita, una petición de principio. Porque si para realizar un acto inteligente uno tiene que seguir una proposición interior, deberá elegir primero entre toda la cantidad de proposiciones de que dispone, y para esa elección ya necesitaría una teoría, que se debería haber creado eligiendo alguna proposición, para la que se necesitaría otra teoría, y así sucesivamente. En definitiva, *saber hacer* no proviene, en palabras de Ryle, de *saber que...* Ryle habla de la versión positiva de *saber hacer*, heredada de la práctica más que de la teoría. Pone el ejemplo de un niño que aprende a jugar al ajedrez por observación de muchas partidas de otros, como se supone que se aprende el lenguaje. Finalmente sabría jugar sin haber aprendido las reglas. Incluso si un niño aprende las reglas para jugar, cuando haya jugado muchas partidas no tendría que pensar en ellas para hacerlo, la *disposición adquirida*

no le haría pensar en las reglas en cada partida, aunque siempre jugara correctamente.

Cuando hablamos de conceptos disposicionales atribuyendo adjetivos de comportamiento a algo lo hacemos con una lógica que consta de un conjunto de proposiciones de tipo hipotético. Por ejemplo, si decimos de una sustancia que es soluble significa que si la sumergiéramos en agua se disolvería. La caracterización de las disposiciones de las personas seguiría ese mismo esquema, aunque el conjunto de proposiciones en muchos casos sería infinito por la complejidad de las operaciones humanas. Las proposiciones disposicionales no son causas ocultas de nuestros comportamientos que parten de una voz interior, sino que son enunciados que, si bien no son leyes, se asemejan a ellas en el sentido de que rigen nuestras conductas. Pero han sido aprendidas a partir de la observación (exterior) de los individuos que nos rodean. En lo que Ryle insiste es en no describir las operaciones mentales de las personas por medio de un conjunto de operaciones ocultas, espectrales, que las definen, sino por medio de estas proposiciones hipotéticas o semihipotéticas: **“No buscamos causas (y a fortiori causas ocultas), sino aptitudes, habilidades, hábitos, propensiones e inclinaciones”**(Ryle 1949, p. 59). Esto no es tirar la toalla en el ejercicio de la comprensión del otro, a un jugador de ajedrez o a un geómetra se les puede comprender perfectamente cuando realizan sus profesiones con el conocimiento de las reglas del ajedrez y de la geometría, sin aludir a ninguna confusa teoría psicológica ni realizando inferencias a partir de la introspección. La aptitud para apreciar un comportamiento y la que se requiere para ejecutarlo van de la mano. El problema de las otras mentes o el solipsismo también se resuelve en esta comprensión, fruto de la observación de las operaciones que hacen los demás: **“Observando el comportamiento público - del que lo que se dice y se escribe es lo más importante - descubro muchas más cosas de las que pretendería saber respecto de las aptitudes, intereses, gustos, rechazos, métodos y convicciones de los demás. Cómo se llevan a cabo esos actos de imaginación, incluyendo los monólogos imaginarios, constituye una cuestión subsidiaria”**(Ryle 1949, p. 75). Descubrimos por observación las reglas a las que alude también el Wittgenstein tardío, las reglas cuyo carácter público enfatizan ambos: **“202. Por tanto ‘seguir la regla’ es una práctica. Y creer seguir la regla no es seguir la regla. Y por tanto no se puede seguir ‘privadamente’ la regla, porque de lo contrario creer seguir la regla sería lo mismo que seguir la regla”**(Wittgenstein 1953, p. 203).

Ryle hace notar que en el comienzo de la psicología se encuentra también el mito del fantasma en la máquina, el de los dos mundos. Se creía que, al igual que Newton y Descartes habían puesto las bases de una teoría mecánica de lo “externo”, debería haber también una ciencia paralela dedicada al mundo de los fenómenos mentales. A menudo los psicólogos defendían la introspección, por ejemplo, como la fuente empírica para estudiar las operaciones mentales. El mito cartesiano así dio lugar también al mito de las dos ciencias. Y, como apunta el inglés, Mendel y Darwin queda-

rían fuera de ese esquema. Hay muchas ciencias, no solo las físicas, el mecanicismo era otro mero espectro. Y al igual que no todos los fenómenos físicos han logrado una explicación con la física, tampoco el comportamiento humano puede explicarse con la psicología. Aunque los psicólogos sigan un programa plausible encaminado a descubrir las causas de los comportamientos, todo a lo que han llegado es a clasificaciones o a meras correlaciones. La psicología no es la única ciencia experimental encaminada a comprender el comportamiento humano, como sería absurdo pretender describir un país utilizando solamente términos meteorológicos.

Otra de las cuestiones en las que fracasa la psicología popular es en la forma de abordar lo que es el “intelecto” humano, en separar lo que es intelectual de lo que no lo es. Para el conductismo de Ryle son de nuevo fronteras ficticias. En el uso del lenguaje todo el mundo acepta que existe un hablar cotidiano, un lenguaje espontáneo, que no se relaciona con el otro lenguaje estudiado, que estaría más relacionado con el intelecto. La idea de enseñanza es un ejemplo que pone Ryle como el que se usa siempre para fijar las diferencias entre lo intelectual y lo que no lo es. Algo que es aprendido mediante lecciones y de lo que se pueden hacer pruebas o exámenes de conocimiento. Pero advierte Ryle que **“nadie puede entender o usar el lenguaje con fines didácticos si no ha aprendido antes a usar el lenguaje para conversar”**(Ryle 1949, p. 306). No se puede reducir el pensamiento superior a la capacidad, por ejemplo, de construir una teoría o un plan, para poder hablar de “pensamiento” esa teoría tiene que ser susceptible de su exposición didáctica, de su defensa ante cualquier objeción, de, al menos, poder dar proposiciones operacionales sobre la misma. Esto es así tanto en la teoría de Newton de la gravedad como en el plan que tenga un ama de casa para instalar determinada alfombra en una habitación. Esta teoría gnoseológica del conductismo diferencia así el aprendizaje de una teoría con su exposición, y afirma que uno de los errores ha sido confundir estos conceptos, una cosa es la utilización de una senda y otra abrir una senda. Es un error pensar que el uso de las oraciones de indicativo refleje actos de «juzgar» o el uso de una condicional refleje un acto de «razonar». Los conceptos, los juicios y las inferencias están relacionados más con el hecho de utilizar o mostrar sendas ya abiertas que con la creación de las mismas. Los teóricos del conocimiento postulan siempre actos cognoscitivos previos y construyen una historia paralela también regida por premisas e inferencias claras que dan lugar a las conclusiones, *dramatizaciones paramecánicas*. Pero más bien se producen otras formas expresivas más pedestres, en forma, sí, no se niega, de soliloquios, pero no inferenciales. Se repiten una cantidad de cosas que pueden ser interesantes para familiarizarse con ellas, o bien se dan órdenes, o se formulan reproches. Para Ryle en lo que los teóricos llaman «pensamiento» en cuanto a crear teorías, se produce de nuevo el mismo error derivado de la hipótesis paramecánica, y una reiteración en soliloquio de lo que luego va a ser mostrado didácticamente en la práctica. Las inferencias y deducciones que se usan para enseñar una teoría no tienen por qué haber estado

antes en un proceso «mental» previo exactamente en la misma forma de inferencias y deducciones.

Ryle narra cómo hubo un conductismo en las teorías psicológicas, el primero en prescindir del estudio de lo “fantasmal”, que desarrolló los principios de la psicología experimental, aprovechando que en aquella época las ciencias físicas ya no eran las únicas que habían compuesto un cuerpo teórico-experimental importante. Fueron los tiempos del desarrollo científico de la biología, una ciencia natural en auge. Por tanto se había diluido la dicotomía entre estudiar el comportamiento humano como consecuencia de su alma fantasmal o integrarlo en el universo mecanicista de la física. Había más mundos.

1.2. Sentimientos y sensaciones

En el mundo de las emociones es en donde más ha calado la idea de la existencia de “fantasmas” dentro de los cuerpos. Pero Ryle vuelve a combatir la idea de una relación causal a partir de procesos ocultos internos de nuestra mente. Según él a menudo se analizan mal los conceptos relacionados con las emociones. Las más de las veces los sentimientos se pueden explicar también como propensiones que van a determinar acciones que se pueden analizar, y los estados de ánimo no son acaecimientos, sino que aglutinan estos sentimientos. Con ellos no se explica el comportamiento, en todo caso habría que diagnosticarlos. Se puede indagar en el estado de ánimo de otra persona de nuevo haciendo inferencias a partir de sus hábitos o padecimientos, y así puedo también descubrir el mío propio.

Es aceptada, según Ryle, la idea del Doble Acceso Privilegiado de la mente a su “interior”, a sus estados inmediatos, «conciencia», como a otros estados mediante percepción no sensorial, «introspección», y que este autoconocimiento es además superior al que obtengo de los demás. Desde el conductismo de nuevo tales ideas no son sino confusiones lógicas. En realidad el autoconocimiento y el conocimiento de los demás deberían situarse en el mismo plano, ya que siguen los mismos esquemas y limitaciones. Ni siquiera el grado de conocimiento garantiza que el conocimiento de uno mismo tenga que ser forzosamente mayor que el de los demás, uno puede equivocarse respecto de sus sensaciones orgánicas e, incluso, percibir sensaciones inexistentes, como a menudo les ocurre a los hipocondríacos. Según el punto de vista tradicional de la conciencia se puede conocer lo que a uno le acontece en el sentido no disposicional de «conocer», es decir, que lo conocemos activamente en el momento. Esto no es así, y de hecho tampoco es coherente uno de los supuestos sobre cómo opera la conciencia, según el cual la mayor parte del tiempo guardamos un registro de lo que nos ocupa. Si fuera así, también deberíamos registrar los actos de registro de los acaecimientos, generando de nuevo una regresión infinita. Para Ryle, “...conocer lo que nos ocurre mentalmente

no implica la vigilancia constante de nuestro hacer y sentir, sino, solamente, la inclinación *inter alia* a relatarlo cuando nos encontramos en un estado de ánimo propicio”(Ryle 1949, p. 184). La introspección además requeriría eventualmente individuos prestando atención a multitud de cosas a la vez, imposible en la práctica. Es más, alude al argumento de Hume según el cual analizar un estado de pánico, por ejemplo, requeriría hacerlo desde un estado de frialdad propio del análisis, bastante alejado del estado de pánico, con lo que no se podría analizar. Y así habría multitud de ejemplos de estados no susceptibles de analizar por introspección.

Para Wittgenstein tampoco la introspección es fuente de conocimiento. Aunque no la niega, duda de que en la mayoría de los casos no sean sino argumentaciones disposicionales sobre comportamientos posibles, como se puede leer en el parágrafo 587 de las *Investigaciones*: “**¿Tiene sentido preguntar «¿Cómo sabes que crees eso? - y es la respuesta; “Me doy cuenta por introspección”?** En algunos casos se podría decir algo por el estilo, en la mayoría no. Tiene sentido preguntar: “**¿La amo realmente, no me lo imagino nada más?”**, y el proceso de introspección es la evocación de recuerdos; de imágenes de situaciones posibles y de sentimientos que uno tendría si...»”(Wittgenstein 1953, p. 367).

En cuanto a las sensaciones conectadas con los órganos de los sentidos, existe también una dificultad lingüística que hace que su transmisión sea más difícil de interpretar, solemos describir estas sensaciones haciendo alusión a objetos fáciles de interpretar por el resto de las personas. Así se puede hablar de un dolor como punzante sin que nadie te haya clavado ningún puñal, o tener ardor en la garganta sin que tenga que ver en absoluto con el fuego, o decir que tienes la vista nublada para que se entienda que ves como si en el exterior hubiera niebla. Se produce otro error lógico cuando uno confunde el concepto de «sensación» con el de «observación» o «percepción». Este error se ha generado por la Teoría de los datos sensoriales, según la cual las sensaciones son el resultado de una aprehensión cognitiva de los datos recogidos por nuestros sentidos, pero “**Tener una sensación no es estar en una relación cognoscitiva con un objeto sensible. Tales objetos no existen, como tampoco existen tales relaciones. No sólo es falso que las sensaciones pueden ser objetos de observación, sino que también es falso que sean, en sí mismas, observación de objetos**”(Ryle 1949, p. 239). De esta forma se confunden los conceptos de sensación con los conceptos de observación, categóricamente diferentes. No hay que torturarse intentando establecer relaciones entre los objetos que se muestran ante nuestros sentidos y algo “privado” que los aprehende. Cuando observamos efectivamente tendremos alguna sensación pero no necesariamente estamos observando cuando tenemos sensaciones. Las sensaciones no son herramientas cognitivas que podamos controlar para nuestro aprendizaje o nuestras percepciones.

2. FUNCIONALISMO

2.1. Realizabilidad múltiple

Con el funcionalismo de Putnam se vuelve a admitir la existencia de estados internos de conciencia en la explicación, si bien es cierto que estos estados serían estados abstractos, que se definirían por funciones análogas a las *tablas de máquina* o programas de una máquina universal de Turing¹ (Putnam 1960). Los procesos mentales internos sí son ahora causas de la conducta, pero se pueden definir de forma totalmente independiente de su concreción material: **“mi “funcionalismo” sostenía que, en principio, una máquina (digamos uno de los robots de Isaac Asimov), un ser humano, una criatura de silicio y un ser puramente espiritual funcionan de la misma manera cuando se los describe en un nivel apropiado de abstracción y que es incorrecto creer que la esencia de nuestra mente es nuestro “hardware””(Putnam 1988, p. 15).** Estos procesos, de naturaleza computacional, constituirían funciones mediadoras entre las entradas sensoriales y las salidas motoras. Por estas razones al funcionalismo de Putnam se le conoce como *funcionalismo de tabla de máquina*. El diseño de esta tabla de máquina corresponde pues al desarrollo de propiedades de segundo orden, es decir, propiedades sobre las propiedades mentales, concepto que sería análogo al desarrollo de “capas” de *software* implementadas sobre distintas estructuras inferiores, como las aplicaciones multiplataforma tan ubicuas en nuestros días.

En (Putnam 1960) se describe una analogía entre los acontecimientos físicos y los acontecimientos mentales del comportamiento humano y las dos posibles descripciones de una máquina de Turing. Cualquier concreción de tal máquina habría de estar diseñada y fabricada por un ingeniero, ese sería su *proyecto estructural*, y la máquina se hallaría en cada momento en un estado estructural concreto (en el que tales válvulas estuvieran funcionando, o tales conexiones falla-

¹Una máquina de Turing es un dispositivo imaginario que ideó en 1936 el matemático inglés para simular procesos computacionales. Consistía *grosso modo* en un cabezal que podía actuar sobre una banda de papel parcelado regido por una unidad de control que podía encontrarse en diferentes estados según el cabezal escribiera, borrara, leyera, se moviera a derecha o izquierda o se parase. Así se pasaba de una situación o configuración inicial a una final en un número finito de pasos, que se podían diseñar con las *tablas de máquina*, describiendo así un procedimiento general o algoritmo.

ran, etc.), definiendo sus *estados físicos*; pero al mismo tiempo la máquina estaría dotada de un *proyecto lógico*, concebido por un matemático, que daría cuenta de las rutinas de control y configuraciones de la máquina, definiendo sus *estados lógicos*. Para Putnam el estudio del proyecto estructural sería la interpretación “conductista”, encaminada a proporcionar una explicación positivista del comportamiento humano, y la comprensión del proyecto lógico sería la meta a la que aspiraba la psicología clásica, que según él falló por razones empíricas: **“los estados mentales y las “impresiones” de los seres humanos no constituyen un sistema causal cerrado, en la medida en que lo constituyen las “configuraciones” de una máquina de Turing”**(Putnam 1960, p. 69). Para el funcionalismo de Putnam la psicología tradicional tenía su propia área de estudio y sus propias leyes irreductibles, aunque dudaba de las bondades de su red teórica causal y admitía otras formas de comprender el proyecto lógico asociado a los estados mentales. Más aún, como afirma en (Putnam 1967, p. 6), su hipótesis no era incompatible con el dualismo desde el punto de vista ontológico.

El llamado *funcionalismo de la psicología popular* de David Lewis (Bechtel 1988, pp. 113-115), de hecho, intenta codificar las actitudes proposicionales de la psicología popular para crear un conjunto de leyes causales que relacionen entradas y salidas. Putnam se alejaba de esta visión, y admitía que el estudio para el diseño del estado funcional podía provenir de varias fuentes, siempre dentro de teorías científicas falibles. Una psicología experimental futura que podría incluir resultados de la neurofisiología ayudaría a especificar estados mentales en forma de programas de ordenador. No obstante, la teoría funcionalista, en sus variantes, no deja de ser una evolución de la conductista, en cuanto a interpretar estados mentales como *disposiciones* a la conducta. La diferencia es que en el conductismo estas *disposiciones* tenían un carácter meramente epistemológico, mientras que en el funcionalismo se asume un compromiso ontológico, se trata de una teoría en este sentido «realista», en cuanto a que las disposiciones son estados internos definidos en las descripciones funcionales, causalmente relacionados: **“Incluso cabe decir que es la teoría psicológica completa (científica o popular) la que define como un todo cada una de esas disposiciones que son los estados mentales, a través del papel causal respectivo que la teoría les asigna”**(García-Carpintero 1995, p. 68). Cualquier forma de funcionalismo se basa en un armazón sintáctico que desarrolle los mecanismos de computación, y su vertiente semántica, en donde la noción de *referencia*, o conexión entre el lenguaje y la realidad subyacente, jugará un destacado papel.

Uno de los argumentos contra la realizabilidad múltiple tiene que ver con *el problema de los qualia ausentes*. Podríamos construir un organismo compuesto por una multitud de gente dentro de un gran recito robótico que respondiera conjuntamente a estímulos externos. Tal ente sería un sujeto de estados mentales y, sin embargo, esos estados, que cumplirían las funciones relacionadas con el dolor, el color, etc, no tendrían estados internos como los nuestros, no se podría hablar de

verdaderos estados mentales. Putnam salva en parte esta objeción en (Putnam 1967) al analizar el dolor no admitiendo que se puedan definir organismos divisibles en partes autónomas, como un enjambre de abejas. El problema de los *qualia* en general es sin embargo el talón de Aquiles de distintas formas del *fisicalismo*², y en este caso viene a afirmar que la organización funcional no puede definir la conciencia fenoménica.

Aunque no es objeto de este trabajo analizar toda la evolución del pensamiento de Putnam, hay que decir que la tesis funcionalista original también se fue abandonando. Y fue precisamente Putnam quien se dio cuenta ya en los setenta de que en realidad no estaba defendiendo sino un programa reduccionista de corte científicista alejado de sus principios filosóficos. El programa funcional, pensaba, era una versión edulcorada de las teorías de la identidad, en este caso identificando propiedades mentales y funcionales. Además, si la tesis funcionalista fuera cierta, se podrían establecer relaciones de equivalencia entre sistemas físicos con las mismas creencias, y en un argumento que recuerda al teorema de Gödel afirma que esa relación, de existir, sería imposible de descubrir para todo sistema inteligente (Putnam 1988, p. 18). La clave está en que el concepto de *referencia* trasciende los sistemas físicos concretos y solo puede ser entendido en términos culturales o relacionales más amplios, lo que invalidaría el diseño de estructuras computacionales individuales. Lo que se pone en duda es que se pueda llegar a realizar una interpretación reduccionista de los conceptos de la mente humana: **“Jamás podremos llegar a poseer el Algoritmo Maestro para la interpretación porque es de todo punto imposible llevar a cabo la “valoración” de todos los posibles modos de conceptualización de los seres humanos en todos sus lenguajes y todas sus culturas, y en todas sus formas de fijación de creencias”**(Rodríguez 2006, p. 63). O, en palabras del propio Putnam: **“En mi opinión, lo que tenemos que abandonar es la exigencia de que todas las nociones que nos parecen importantes sean reducibles al vocabulario y al aparato conceptual de las ciencias exactas. Creo que es el reduccionismo el que está en problemas y no la intencionalidad”**(Putnam 1988, p. 117). Aún así, Putnam admite como éxito del funcionalismo el sostener que los distintos estados físicos no son necesariamente fruto de los mismos estados físico-químicos internos, y se pregunta si no tendría éxito alguna forma más complicada de funcionalismo que tuviera en cuenta estas objeciones.

2.2. Sensaciones: el dolor

Pasando a la posible introspección de nuestros estados mentales internos, el análisis inicial de Putnam se dirigía a investigar el concepto de dolor. Desde el funcionalismo se discutía también

²El término *fisicalismo* está definido como la tesis metafísica de que todo es físico o que todo sobreviene de algo físico (Stanford). Aunque el concepto admite varias interpretaciones (Montero 2011), normalmente se asocia al estudio de los fenómenos en cuanto sucesos espacio-temporales relacionados con partículas materiales susceptibles de ser observados experimentalmente.

la teoría del Doble Acceso Privilegiado a la mente que criticaba el conductismo pero, no obstante, cualquier concreción física de una máquina de Turing podría tener ciertos órganos sensoriales electrónicos que permitieran cierto nivel de autoconocimiento. En cuanto al proyecto estructural, los seres humanos tampoco tenemos la posibilidad de detectar muchas de nuestras anomalías físicas (un tumor, por ejemplo) pero sí que la evolución ha desarrollado ciertos niveles de alarma en algunos casos de falla física. El funcionalismo de tabla de máquina no contempla la evolución (ni filogenética³ ni ontogenética), dado que se admiten construcciones no biológicas equivalentes, pero si la máquina poseyera alguna de estas características sensoriales, el hecho de la detección de una anomalía implicaría también al proyecto lógico de la misma. Aquí entrarían en juego posibles errores de las dos naturalezas. Si la máquina imprime “la válvula X ha fallado”, puede ser cierto, pero también puede ser un error estructural (de los órganos sensoriales fabricados) o un error de computación en la lectura e interpretación de la cinta. Una máquina sin órganos sensoriales que simplemente esté programada para imprimir: “Estoy en el estado A” cuando en efecto el cabezal esté en el estado A no es susceptible de errores computacionales, como mucho, *verbales*. Esta analogía le da pie a Putnam para afirmar que en efecto los estados mentales, análogos a los estados lógicos de una máquina de Turing, tienen mucho que ver con la *verbalización*.

Preguntas como “¿cómo sé que tengo un dolor?” son preguntas *desviantes*, perversas, lógicamente extrañas, que se desvían de una regularidad semántica, un término acuñado por Paul Ziff⁴, y no así lo serían preguntas del estilo de “¿cómo sabe X que Y tiene un dolor?”, porque una máquina virtual X, dotada con aparatos sensoriales, podría conocer información del estado de otras máquinas. El problema radica en la aceptación de la identidad entre los estados estructurales y los lógicos, que en la analogía de Putnam para el problema de la mente serían los estados cerebrales y los mentales. No se pueden afirmar como identidades proposiciones que son de carácter sintético, como “estoy en el estado A sí y sólo sí la válvula 36 está conectada”, en el caso de la máquina, o algo como “tengo un dolor sí y sólo sí, se encuentran estimuladas mis fibras C”, en el caso de un ser biológico. La teoría de la identidad defendería el carácter analítico de estas proposiciones, y en realidad se están relacionando términos pertenecientes a categorías distintas. Cuando se diseña una tabla de máquina, se están definiendo una serie de estados por los que tiene que pasar la máquina para realizar un determinado cálculo en cuanto a sus relaciones entre ellos y de lo que aparece en cada paso de cinta, es una secuencia de estados que se tendrá que recorrer, pero no tiene sentido preguntarse en qué estado está la máquina ya que no hay una secuencia de estados, no hay

³En (Bechtel 1988, pp. 136-140) se dan algunas versiones del funcionalismo, más cercanas a la tradición del término en psicología, que admiten armazones evolucionistas en su construcción.

⁴Un ejemplo de pregunta desviante más claro podría ser “¿de qué color es el 4?”, dado que los números no están dotados de propiedades cromáticas. La traducción del original, *deviant*, a veces es *desviada*, aunque a mi me suena un poco raro. No sé si sería mejor un término más pedestre, como *engañoso*.

algoritmo, que la máquina pueda recorrer, para responder sobre si está en un estado individual. Si fuera así, habría que sobreprogramar indefinidamente en cada estado una secuencia que responda también al estado en cuestión llegando a las regresiones infinitas que condenaba Ryle. La información sobre en qué estado se encuentra un organismo procede directamente del estado en el que se encuentra, no está intermediada por ningún proceso de introspección u observación anterior a la verbalización. Esta es una de las coincidencias más notables con el conductismo, antes de cualquier tipo de proposición del estilo de “me duele” no ha habido una reflexión mental introspectiva, y para Putnam este matiz fue pasado por alto por los epistemólogos clásicos, de Hume a Russell (Putnam 1960, p. 63).

Putnam advierte que las leyes científicas fundamentales solo pueden ser clasificadas como analíticas o sintéticas desde un punto de vista sincrónico. La proposición “la luz es una onda electromagnética” era una proposición desviante en el Renacimiento, pero a partir del siglo XIX ya no, porque se produjo una *identificación teórica* entre las leyes de la óptica y las del electromagnetismo, que dio lugar a numerosas predicciones. De la misma forma “el estado mental ψ es idéntico al estado cerebral Φ ”, es actualmente una pregunta desviante, porque la identificación de estados mentales con estados cerebrales no se da, pero podría no serlo en algún momento: “...**lo único que quiero significar cuando hablo de “explicar causalmente el comportamiento humano” es: explicar causalmente ciertos acontecimientos físicos (nociones de cuerpos, producción de ondas sonoras, etcétera), que, en el sentido que acaba de ser mencionado, son los “mismos” que los acontecimientos que constituyen el comportamiento humano. Y ningún grado de “ryleismo” puede derrotar con sus argumentos a lo que evidentemente es una posibilidad: que la ciencia física podría lograr hacer eso**”(Putnam 1960, p. 78). Insistiendo en que el acto de identificación teórica de esas leyes no se da, y no puede ser algo que pueda ser declarado “a voluntad”, no existe actualmente tal identificación, manteniendo, eso sí, la analogía entre el problema de la mente y el cuerpo con los estados lógicos y estructurales de una máquina de Turing. A lo que se aspira es, en definitiva, a una analogía entre el lenguaje de la máquina y el lenguaje humano, a una lingüística “conductista”. El problema de la identidad carece de sentido. Esta postura crítica sobre la identificación la ha mantenido en trabajos posteriores: “...**los sucesos mentales no son ni «idénticos» ni «no idénticos» a los sucesos físicos; no creo que la noción de «identidad» tenga aquí ningún sentido**”(Putnam 1999, p. 91).

En (Putnam 1967) se desarrolla más este programa afirmando la imposibilidad de dotar o quitar el significado a frases como “los dolores son estados cerebrales”, e intentando encontrar otros enunciados comprobables sobre bases metodológicas o empíricas. Así, Putnam propone la hipótesis de que el dolor “...**es un estado funcional de todo un organismo**”(Putnam 1967, p. 4). Esta hipótesis le parece menos vaga que la de sostener que el dolor es un estado fisicoquímico del cere-

bro, y la desarrolla sobre la base de la definición de Autómata Probabilista, una generalización de la máquina de Turing en la que ahora las transiciones entre estados pueden tener una probabilidad de suceder distinta a 0 y 1, definido por lo que denomina Descripción que desarrolla su organización funcional. La hipótesis de la identidad fisicoquímica del dolor le parece mucho más fuerte que la funcional, puesto que restringe ese estado cerebral causal posible a que sea el mismo en el caso de un mamífero, de un reptil, etc, y, es más, evoluciones paralelas extraterrestres deberían converger al mismo correlato físico del dolor. A Putnam le parece sumamente improbable. Y más el hecho de que todo estado psicológico sea un estado cerebral, puesto que bastaría encontrar un contraejemplo de algún predicado psicológico, como “hambriento”, que tuviera diferente correlato fisicoquímico en un pulpo y un hombre para desmontar la teoría. La teoría funcional, sin embargo, está apoyada por manifestaciones conductistas. Las similitudes en la conducta de dos sistemas que identifican sensaciones como el dolor, el hambre, etc., hacen sospechar más fuertemente similitudes en la organización funcional análoga al diseño de un Autómata Probabilista.

Aunque el funcionalismo sea una evolución del conductismo, las explicaciones de este a partir de disposiciones a Putnam le parecen irremediablemente vagas, porque al decir “la disposición de X a comportarse como si X sintiera dolor” ponemos en juego el concepto mismo de dolor, que se enmarca irremediablemente en el contexto de nuestras creencias, deseos, actitudes ideológicas, etc. El programa funcionalista no iría tan lejos, y el estado funcional se definiría como el que recibe entradas sensoriales que juegan cierto papel en su organización, que detectaran, por ejemplo, un daño concreto, o una subida de temperatura, y el organismo le asignara cierto valor negativo para proceder a la supuesta conducta de respuesta a partir del resto o Estado Total de la máquina. Así se definiría en el funcionalismo el análogo a los estados disposicionales del conductismo, y en este caso estos estados son reales, son físicos. La teoría de la disposición conductual de Ryle parece tener en cuenta tan solo la conducta periférica y la verbalización, y esto es un error para él claro, como describe imaginando una comunidad de “super-espartanos”, o “super-estoicos”: **“They do not wince, scream, flinch, sob, grit their teeth, clench their fists, exhibit beads of sweat, or otherwise act like people in pain or people suppressing the unconditioned responses associated with pain. However, they do feel pain, and they dislike it (just as we do)”** (Putnam 1963, p. 29). Incluso imagina todavía unos individuos, los “super-super-espartanos”, que han sido “super-espartanos” por tanto tiempo que han suprimido también las manifestaciones verbales del dolor, dejando al conductista sin ninguna herramienta para evaluar la existencia de un dolor que, de hecho, existe, y que revela que hay algo más allá de las disposiciones con que el conductista pretendía identificarlo.

3. ELIMINATIVISMO

3.1. De la Identidad al Eliminativismo

La *teoría de la identidad* defendida por algunos filósofos descansa en la confianza de que los estados mentales son en efecto estados fisicoquímicos del cerebro. Como apuntaba Putnam, es una fe en que nuestros conocimientos algún día establezcan esa biyección, más que un hecho de que se den en la actualidad. Se quiere que en algún momento se puedan afirmar frases como “el dolor X es el estado cerebral Y” como se puede afirmar que “la luz son ondas electromagnéticas”, y, para ello, es evidente que se requiere de una labor de reducción interteórica entre la teoría psicológica y las neurociencias, se está tratando de un *materialismo reduccionista* o de una de las caras del fisicalismo. Paul Churchland explica bien este programa: **“Todo lo que se requeriría sería que alguna neurociencia con una buena capacidad explicativa se desarrollara hasta el punto en que se pudiese elaborar una “imagen refleja” adecuada de los supuestos y principios que constituyen nuestro marco conceptual corriente para los estados mentales”**(Churchland 1984, p. 52). Este programa basado en una confianza sobre una ciencia futura, como se puede leer en autores como Paul Noordhof, hace que un fisicalismo así concebido devenga trivialmente verdadero, y en realidad habría que buscar un fisicalismo cuyas conclusiones permitieran reflejar de alguna forma la posibilidad de que pudiera haber desarrollos científicos que lo falsificasen (Noordhof 2011, pp. 241,248).

No vamos a desarrollar en este trabajo las complicaciones que surgen al realizar estas reducciones interteóricas, pero de este hecho se dieron cuenta muchos de los filósofos materialistas que aquí se nombran. Una forma de salvar las dificultades que se encuentran al pretender establecer semejanzas entre los términos psicológicos y los de la neurociencia fue precisamente negar los primeros, los conceptos de las teorías psicológicas al uso debían ser reemplazados por otros provenientes de la ciencia cognitiva (Churchland 1981), de este modo se definió el *eliminativismo*: **“A juicio del materialismo eliminativo, no podrán encontrarse las correspondencias biunívocas, y no se podrá efectuar una reducción interteórica del marco de referencia psicológico corriente, porque el marco de referencia psicológico que utilizamos corrientemente es una concepción falsa y radicalmente engañosa sobre las causas de la conducta humana y la naturaleza de la actividad cognitiva”**(Churchland

1984, p. 75). Y no se puede construir una teoría científica sobre un marco teórico inadecuado, es más bien a la inversa, las proposiciones de una psicología científica nueva deben ser desarrolladas a partir de los resultados experimentales de la neurociencia. Se salva así la dificultad de reducción interteórica, eliminando los conceptos de la psicología tradicional, hecho que se apuntala con una serie de paralelos en la historia de la ciencia, como la eliminación del “flogisto” en las combustiones, o el “calórico” como transmisor de la temperatura, o la esfera de las estrellas fijas en la cosmología antigua. Paul Churchland justifica esta eliminación basándose además en la enumeración de fracasos explicativos de la psicología popular a lo largo de su historia, incapaz de explicar el sueño, o la memoria, o las enfermedades mentales, admitiendo, eso sí, que su éxito se debe a que los fenómenos de los que se ha ocupado son de una complejidad manifiesta, y muy difíciles de manejar. No obstante, Patricia Churchland en (Churchland 1993, p. 263) habla de revisión más que de eliminación, debiéndose eliminar algunas categorías psicológicas, mantener otras, o dividir alguna antigua en partes, como en el ejemplo de las clases de sueño que ha revelado la técnica electroencefalográfica. Una vez redefinido el contexto en psicología, el programa reduccionista seguiría vigente.

Hay que decir que el eliminativismo tampoco niega los estados mentales internos, no podemos tratar a los humanos como *cajas negras* y negar la existencia de procesos internos a los que podemos de alguna manera acceder introspectivamente. Sin ir más lejos, el conductismo lógico en principio partía también de la crítica de las inconsistencias y carencias científicas de la psicología popular, para emprender un programa positivista de mejora de la misma, para intentar convertirla en una verdadera ciencia. El problema para Paul Churchland es que quizás los conductistas fueron demasiado lejos al negar la posibilidad de estudiar conceptos teóricos relacionados con los estados mentales como causas, y restringirse a los relacionados con las manifestaciones empíricas de la conducta y el hábito en términos operacionales: **“La idea positivista de que todo término teórico significativo debe admitir una una definición operacional en términos de elementos observables se reconoció rápidamente como un error”**(Churchland 1984, p. 137). No calificar como científica una materia que emplee términos sin asociación directa con manifestaciones empíricas de la realidad dejaría fuera muchas disciplinas consideradas generalmente dentro de las ciencias, o al menos partes importantes de ellas. Un ejemplo se puede dar en muchos conceptos de la física teórica, en los que no se pueden dar definiciones operacionales, relacionados con una red extensa de otros conceptos también teóricos y sólo indirectamente con otros conceptos claramente mensurables. Más bien hoy en día se acepta que el significado de cada término teórico en las ciencias viene precisamente determinado por su lugar en la red de conceptos que se ha establecido en el sistema de creencias en el que aparece, sea un concepto experimental o puramente teórico. Es la relación con los demás conceptos de la teoría la que le provee de un significado que da coherencia a un todo,

y sería muy difícil construir ciencias basadas únicamente en términos que deriven de definiciones operacionales, se trata de un programa que de hecho ni los conductistas pudieron llevar a cabo.

Patricia Churchland en (Churchland 1993, pp. 264-276) justifica el programa de reducción neurobiológica respondiendo sucintamente a las principales críticas que se le han realizado por parte de algunos filósofos, que según ella son cinco y resumiré a continuación con las respuestas que aporta:

- 1) El objetivo es absurdo, jamás se podrán extraer conceptos pertenecientes a la conciencia humana de la neurobiología por ser estancos independientes. **Respuesta:** El argumento es falaz, porque no conocemos todavía el alcance de los mecanismos neuronales. Antes de que se desarrollara la biología molecular tampoco se imaginó nadie que se pudieran explicar “cosas vivas” como la célula a partir de moléculas “muertas”. El argumento solo refleja un estado actual de ignorancia y una falta de imaginación. Así se ventila Churchland aquí el problema de los *qualia*, confiando en que algún día la neurociencia los desvele. Para ella, a veces se emplea también con suficiencia el concepto de “error categorial”, pero se olvidan de que muy a menudo en ciencia los significados cambian.
- 2) Es inconsistente con la «realizabilidad múltiple», no se pueden identificar macrofenómenos resultado de un solo mecanismo si se admite que puede haber macrofenómenos parecidos con otros mecanismos. Aquí se alude a contraejemplos típicos del funcionalismo como los famosos seres basados en el silicio. **Respuesta:** La objeción carece de interés científico y no invalida el logro reduccionista. A menudo en biología se utiliza una especie mamífera concreta, por ejemplo, para estudiar su corazón, y las teorías son válidas, con cambios mínimos, para estudiar la mayoría de las especies mamíferas. O el descubrimiento del mecanismo de réplica del ADN, solución que llevó a estudiar diferencias fenotípicas en multitud de especies, o gracias al estudio de las conexiones neuronales se establecieron diferencias en los perfiles de reacción de distintos tipos de neuronas.
- 3) Entre los estados de conciencia y los cerebrales como mucho puede haber correlaciones, no identidades. John Searle afirma que cualquier reducción entre cerebro y conciencia es imposible porque la conciencia abarca todo, apariencia y realidad. A mi juicio es una variante de la primera objeción. **Respuesta:** Searle no entiende las razones por las cuales los científicos hacen identificaciones. Por ejemplo, el calor *es* movimiento molecular. Si pensamos que el calor es otra cosa causada por el movimiento molecular es cuando surgen los problemas y tendríamos que dar otra naturaleza al concepto de calor para explicarlo. Si se acierta en la idea de que en efecto la conciencia *es* algún patrón de actividad neuronal, entonces esa actividad cerebral *es* la realidad detrás de la experiencia, aunque yo siga teniendo experiencias a la manera regular.

- 4) La conciencia surge cuando los humanos aprenden el lenguaje y empiezan a hablarse a sí mismos. Este punto de vista, sostenido por Dennett, implica que los patrones neuronales constituyen en realidad una máquina paralela que en realidad simula una máquina virtual secuencial que es la que conforma la conciencia. Dennett realiza la analogía con un simulador de vuelo, la conciencia tiene la relación con nuestro cerebro que tendría el simulador con las interioridades del computador, y jamás podremos saber cómo funciona el simulador estudiando estas, sino estudiando su ejecución. La psicología experimental sería la única que podría estudiar la conciencia. **Respuesta:** Para empezar se duda de su afirmación sobre los orígenes de la conciencia, porque indicaría que ni los bebés preverbales ni los chimpancés serían conscientes, o los sujetos sin hemisferio izquierdo o con afasia global. Dennett descarta cualquier tipo de cognición no lingüística y olvida que para obtener una secuenciación temporal no es necesaria una máquina serial virtual, las redes neuronales con bucles recurrentes pueden proporcionarla, pudiéndose obtener además conducta seguidora de reglas (Elman 1991). Churchland está convencida de que la conciencia no envuelve una sola operación a la vez, podríamos percibir varios colores simultáneamente. No hay razones para pensar que Dennett está en la pista correcta al afirmar que debe existir una máquina virtual.
- 5) El problema está por encima de nuestra inteligencia. **Respuesta:** No podemos afirmar todavía que la complicación de nuestro cerebro supere a nuestra inteligencia, dado que estamos haciendo pequeños avances y, por tanto, no hay razón para abandonar este programa.

3.2. Sensaciones: el color

Una de las críticas que desde la teoría de la identidad y el materialismo reduccionista se le hacen al funcionalismo y al conductismo es que no resolvería ejemplos como el experimento de la *sensación del espectro invertido* (Churchland 1984, p. 68), relacionado con los *qualia* del color y heredado de Locke. Este experimento mental consiste en imaginarse que nuestras sensaciones internas sobre el color están perfectamente invertidas en espectro respecto de otra persona. Como en el funcionalismo las sensaciones internas están ocultas por la estructura funcional, una persona puede tener la sensación del verde mientras que otra pudiera tener la del rojo y no hay ningún modo de averiguar si esto en efecto está pasando o no, siempre que la persona que sintiera el verde siendo rojo lo verbalizara como rojo, porque el isomorfismo funcional se seguiría manteniendo: si las dos personas están de acuerdo en que el color que ven es rojo no hay forma de saber que lo que está experimentando una en realidad es el color verde¹.

¹Representacionistas como Michael Tye defienden que el hecho de que exista una diferencia a nivel fenoménico, aunque haya una identidad a nivel funcional, no implica que exista una diferencia a nivel personal. Según estas teorías, las propiedades de representación *determinan* el contenido fenoménico (Noordhof 2011, p.260).

Parece que este experimento mental solo se puede aceptar en su postulación lógica, y no hay razones empíricas para pensar que se pueda dar esa situación en la práctica (Hierro-Pescador 2005, p. 197). Aún así, Churchland afirma que este problema lo resolvería, supuestamente, la neurobiología cognitiva como única explicación defendida por el fisicalismo. Así, se alude a las tres clases de células sensitivas al color que posee nuestra retina, los conos, que detectan longitudes de onda cortas, medias y largas, respectivamente. Cada tipo de cono responde a una cantidad de fotones de un rango de longitudes de onda considerable y, si representamos por S al nivel de estimulación de cada célula sensorial, podremos definir un vector tridimensional de *codificación sensorial*, $(S_{\text{corta}}, S_{\text{media}}, S_{\text{larga}})$, que generaría un *espacio de sensaciones cromáticas*, el espacio de *qualia* de los colores. Estas teorías se conocen como *teorías vectoriales* de la codificación sensorial o *modelos de fibras combinadas*. Así la sensación del naranja estaría claramente definida, y además se hallaría entre las sensaciones del rojo y el amarillo. Del mismo modo, las personas con algún déficit de sensibilidad hacia algún rango de longitudes de onda, ceguera de color, verían reducido dimensionalmente su espacio de colores.

No se tardaron en ver las insuficiencias de esta teoría. En 1988 Larry Hardin publicó *Color for Philosophers: Unweaving the Rainbow*, defendiendo un eliminativismo de color más extremo. El color, según esta teoría, no se puede interpretar como una cualidad de una sustancia, y por tanto se asemejaría a los términos psicológicos que el eliminativismo filosófico debe desechar. La principal objeción la constituye el llamado fenómeno del *metamerismo*, según el cual bajo una determinada iluminación un variado número de superficies con distintas distribuciones de reflectividad espectral pueden ser visualmente indistinguibles para un observador. Esto es así porque diferentes combinaciones de luces con todas las longitudes de onda pueden producir una respuesta equivalente en el receptor, es decir, el mismo punto en el espacio de sensaciones cromáticas de Churchland puede ser obtenido en condiciones muy diferentes. Esto anima a pensar que el color no es una propiedad objetiva de la materia y debería ser eliminado. La energía que absorbe el ojo humano es fruto de la combinación de la sensibilidad de los conos de su retina a cada longitud de onda y del espectro de la luz recibida, que a su vez depende del espectro del iluminante y de la reflectividad de la superficie del material. Si cualquiera de estos factores cambia también cambiará la composición de la luz que llega al ojo. Sin embargo, como afirma Hardin, se observa una notable constancia en la percepción de los colores bajo una amplia gama de cambios de iluminancia (Hardin 1988, p. 46).

De esta manera, los eliminativistas del color se opondrían a la clase de fisicalismo que dotaría de realidad objetiva a los colores, e invalidaría cualquier teoría de la Identidad que pretendiera establecer isomorfismos entre los espacios de color y los resultados neurobiológicos de su percepción.

Al contrario que en el “optimista” materialismo reduccionista, de corte realista, manifestado por Patricia Churchland, o el calificado por Hardin “funcionalismo” de las ideas de Paul Churchland, que asocia los *qualia* a entradas sensoriales, estaríamos ante un pesimismo epistemológico basado en la imposibilidad de discernir isomorfismos entre alguna combinación de marcadores neurológicos y los conceptos experimentales asociados a la física del color: **“Spectral energy, reflectance and transmittance are quite obviously physically specifiable and significant. They are causally central to color perception and play a fundamental role in all parts of color science . They are doubtless the characteristics of the physical world which are in some fashion picked out by perceived color; if any physical properties deserve to be identified with color, it is they. Unfortunately, however, they can’t fill that role-basically because of the phenomena of metamerism”** (Hardin 1988, p. 64).

Putnam también combatió la visión reduccionista de Churchland del problema del color y, aunque ya alejado del funcionalismo inicial, discutió la pertinencia del experimento mental del espectro invertido precisamente con los conocimientos de fisiología y neurología que él tenía. Según él, Jerome Lettvin había mostrado en los sesenta que la información del color es obtenida después de una «computación» realizada en el ojo y, por tanto, la información que transmite el nervio óptico está codificada, no es el producto directo de ningún «cableado». Resultaba muy improbable admitir que malformaciones en esos fibrados cambiaran o invirtieran el espectro electromagnético correspondiente a los colores, las enfermedades asociadas al color son fruto más bien de determinadas causas relacionadas con errores puntuales en esa información codificada (Putnam 1999, p. 199). Por otra parte, fisicalistas como Dennett han negado directamente que el problema de los *qualia* sea un problema real, son **“un producto de un mal proceso teórico”** (Dennett 2017, p. 321), postularlos duplicaría el trabajo cognitivo a elaborar, serían un falso problema. No hay necesidad de imaginar cosas misteriosas, son efectos de los impulsos nerviosos del cerebro y así han de estudiarse.

No podemos aquí profundizar en este apasionante debate, pero hay que aceptar que sigue abierto. Aunque Paul Churchland redefinió su espacio de color para responder a las críticas de Hardin y seguir manteniendo sus posturas realistas tal como las concibió, (Churchland 2007), su propuesta fue también muy discutida, (Wright 2009), y actualmente parece que la mayoría de los filósofos fisicalistas defienden asociar a las propiedades del color un carácter más amplio, identificándolas como clases de distribuciones espectrales de reflectividad, o incluyen en la explicación la interacción con el receptor en la teoría, como los llamados *relacionalistas* (Cohen 2010).

CONCLUSIONES

En este trabajo se han analizado tres de las grandes teorías que a lo largo del siglo XX se desarrollaron en torno a la filosofía de la mente. Hemos visto la oposición al dualismo cartesiano que supuso la teoría conductista y cómo a partir de su declive, y con el desarrollo de las ciencias computacionales y de la neurofisiología, se dio paso a teorías de la mente que volvían a incluir los estados mentales internos en las explicaciones del comportamiento humano. Estos estados aparecen en forma de programas lógicos abstractos cuya estructura física podía admitir múltiples implementaciones (funcionalismo) o en forma de resultados concretos empíricos relacionados con la estructura material del cerebro humano (identidad y eliminativismo). Aunque el funcionalismo de Putnam era compatible con el dualismo desde el punto de vista ontológico, en nuestra opinión las nuevas teorías postconductistas analizadas en este escrito tenían un matiz materialista también alejado del dualismo cartesiano original. Como reacción a la teoría reduccionista de la identidad se desarrolló el programa funcionalista, y como evolución de la misma se llegó al eliminativismo.

Otro de los puntos en común en los que más se ha insistido es en que las tres teorías pretenden alejarse de las explicaciones y los conceptos que se usaban para interpretar el comportamiento humano desde la llamada psicología popular, un conjunto de términos teóricos en torno a la mente que eran normalmente aceptados por la comunidad, que para el conductista lógico están basados en errores categoriales y en suposiciones erróneas, para el funcionalista no se podían reducir nunca a términos físicos porque muchos tipos de sistemas físicos podrían dar lugar a los mismos comportamientos externos, y para el eliminativista simplemente era una teoría que no cumplía con las condiciones que requerían las teorías nomológico-deductivas para una explicación científica rigurosa, y tendría que ser eliminada o, como mucho, adaptada a los nuevos resultados de las neurociencias.

De las tres teorías estudiadas, en nuestra opinión la que no ha resistido el paso del tiempo y de las diferentes críticas que se le han hecho es el conductismo, y de las otras dos el eliminativismo quizás sea la que goce de más salud, aunque esto sea debido a cierta vaguedad en sus postulados. En primer lugar, el programa conductista no parece viable principalmente por obviar

la dependencia que debe darse entre los estados físicos concretos internos del cerebro humano y el comportamiento. El conductismo filosófico no nos parece que sea un programa no *científico*, o al menos hay que aceptarle cierto carácter experimental, fisicalista, porque uno de sus principios es el análisis de hechos susceptibles de ser observados empíricamente: la misma conducta y circunstancias antecedentes de los comportamientos observados. Lo que sí nos parece es que la tesis de que las proposiciones relacionadas con lo mental se puedan traducir en proposiciones relacionadas con la conducta o con disposiciones hacia ella se ha demostrado impracticable y errónea. Impracticable por el hecho de que las transcripciones implican contemplar circunstancias a menudo inabarcables, y errónea por el hecho de que las conductas externas pueden estar relacionadas con distintos estados psicológicos internos, es decir, no se resuelve el problema de la posible simulación, uno puede estar fingiendo un dolor sin tenerlo, ni tampoco el control de las sensaciones, uno puede estar sintiendo dolor y no manifestarlo conductualmente, como hemos visto con la objeción “super-espartana” de Putnam. No obstante, creemos que parte de la crítica de Ryle a la leyenda intelectualista y a los mecanismos de introspección aceptados sigue vigente, dado que en algunos casos se trata de muchas de las críticas que se hacen a la llamada psicología popular que era comúnmente aceptada.

El funcionalismo, curiosamente, tampoco da importancia a la relación que debe existir entre los mecanismos fisiológicos concretos del cerebro y las acciones físicas externas a las que dan lugar, pero esta consecuencia es fruto de causas radicalmente distintas a las que defendía el conductismo. Es consecuencia más bien del hecho de admitir que los estados mentales que dan lugar a la conducta pueden realizarse físicamente de múltiples maneras, es decir, que distintas implementaciones pueden dar lugar a las mismas respuestas conductuales, la llamada realizabilidad múltiple. Esta idea de Putnam concebía las propiedades psicológicas como de segundo orden, es decir, propiedades de las propiedades de los estados mentales y, por tanto, podría diseñarse una “capa” funcional, a la manera en la que lo hacen los analistas informáticos, que diera respuesta a los fenómenos causal/funcionales de la conducta cualquiera que fuera la naturaleza física de los estados mentales. Esta “capa” de “software” tiene cierto aroma conductista, en el sentido en que hay que traducir un “lenguaje de pensamiento” a un programa funcional, aunque en este caso no se corta, como en el conductismo, la comunicación entre “capas”. Esta teoría sirvió para desbancar las teorías iniciales de la identidad de tipos con argumentos curiosamente de corte fisicalista: parecía muy fácil admitir la existencia de alguna propiedad psicológica común, como el dolor, en dos especies distintas, para derribar la creencia en la identificación entre tipos psicológicos y neurofisiológicos, puesto que en este caso un mismo tipo psicológico podría tener dos correlatos físicos distintos. Los descubrimientos sobre la plasticidad cerebral, en los que ni los mismos neurofisiólogos pueden establecer claramente correlatos de este tipo, han acabado por hacer buenas algunas

de las tesis del funcionalismo. El funcionalismo nos parece por tanto también una forma de fisicalismo, puesto que sí es compatible con la teoría de la identidad entendida como una identidad de *casos* o *instancias* (*token*). En efecto, que niegue una identidad uno-uno no implica que no pueda haber una identidad de uno a muchos, y hemos visto que el mismo Putnam admite que en última instancia las propiedades de primer orden sobre las que se asienta la “capa” funcional pudieran estar relacionadas con una buena teoría neurofisiológica del cerebro, aunque de esta concesión al reduccionismo se arrepintiera en escritos posteriores, en los que se alejaría de su visión cientificista y desarrollaría el llamado *realismo interno*.

Hemos visto también en el escrito la posición más drástica respecto a la crítica de la psicología tradicional defendida por el eliminativismo. La tesis de Paul Churchland es que ni se puede negar la existencia de estados mentales causales, como hace el conductismo, ni se pueden aceptar de nuevo ciertos postulados de la psicología tradicional para el diseño funcional de una estructura mental, como hace el funcionalismo. Hay que echar abajo todo el entramado teórico psicológico y cimentar de abajo a arriba nuevas proposiciones conductuales a partir de los resultados de la neurociencia. Se admite por tanto también la imposibilidad de una reducción interteórica, con la salvedad de que ahora esta imposibilidad es fruto del hecho de que una de las dos teorías es errónea y habrá de ser reconstruída, no es una imposibilidad epistemológica. El eliminativismo de Churchland sí es compatible con el reduccionismo, pero las propiedades psicológicas han de ser redefinidas. En este sentido, nos hemos sentido más identificados con las apreciaciones que sobre el tema del reduccionismo hace Patricia Churchland, hablando de una revisión de términos psicológicos, más que de una eliminación, y dejando claro que para ella la reducción de las teorías no es una identificación sino una relación, y que tampoco estas relaciones tienen siempre que ser de abajo a arriba. Se trata de un eliminativismo más moderado que permite, en nuestra opinión, más caminos a explorar.

También se ha visto cómo las mismas ideas de eliminación de términos psicológicos confusos o erróneos han dado lugar a teorías como el eliminativismo del color, en donde un término que normalmente se considera una propiedad física objetiva de los objetos también ha sido cuestionado por ciertos filósofos que proponen su eliminación, si se quiere hacer una teoría más acorde con los resultados de los experimentos de percepción. Los intentos de refutación de estas ideas han dado lugar a un fructífero debate entre los propios partidarios de las teorías materialistas eliminativas. Con esto podemos percibir la complicación que supone admitir o no la relevancia de cada uno de los términos que se suelen asociar con nuestras sensaciones, y particularmente la dificultad que tiene el análisis de los *qualia* para cualquier forma de fisicalismo.

En suma, en nuestra opinión, el reto del materialismo, consistente en poder dar explicaciones

y predicciones sobre la fenomenología de la mente humana y nuestra conducta en términos puramente físicos, sigue vigente. Sin embargo, creemos que ninguna forma de fisicalismo puede pasar por alto la importancia de la interpretación de los estados mentales por medio del análisis del lenguaje, y ahí la visión funcionalista del problema tendrá mucho que decir. No compartimos, por tanto, la optimista visión que autores como Churchland tienen de la esperanza en una reducción a partir de una nueva ciencia psicológica que recoja los resultados experimentales de la neurofisiología ni, de hecho, la visión reduccionista general que tiene mucha parte de la comunidad científica. En este sentido son relevantes las palabras del Putnam tardío: **“Decir que algún día la ciencia podrá encontrar la manera de reducir la consciencia (o la referencia, o lo que sea) a la física, aquí y ahora, es lo mismo que decir que algún día la ciencia puede que haga no-sabemos-qué de manera que no-sabemos-cómo”** (Putnam 1999, p. 204). Pero tampoco deseamos, no podemos, el hecho de que la respuesta debe venir de alguna forma de fisicalismo. Como apunta Paul Noordhof en su artículo, es razonable continuar con una perspectiva fisicalista mientras que los dualistas no expliquen bien cómo funciona su alternativa (Noordhof 2011). Ni siquiera en el terreno fangoso de los *qualia* tenemos evidencias para afirmar que dos cerebros exactamente iguales en las mismas condiciones físicas y ambientales experimenten cosas distintas. A quien afirma incumbe la prueba. Quizás haya que aceptar un compromiso entre el fisicalismo no-reduccionista funcional y el eliminativismo entendido de la forma comedida en que lo hace Patricia Churchland. En este sentido, hacemos nuestras las palabras del profesor Rabossi, adscribiéndonos a la receta que propone como resumen programático: **“...si al ingrediente de base que proporciona el fisicalismo se le agregan cantidades adecuadas de funcionalismo, antirreduccionismo e identidad de casos, se obtendrá el condimento que sazona la dieta teórica actual de muchos filósofos de la mente”**(Rabossi 1995, p. 37).

Bibliografía

- Bechtel, William. 1988. *Philosophy of mind*. New Jersey: Lawrence Erlbaum Associates.
- Churchland, Patricia S. 2001(1993). “¿Puede la neurobiología enseñarnos algo sobre la conciencia?” *Contrastes* Suplemento VI:259–289.
- Churchland, Paul. 1981. “Eliminative Materialism and the Propositional Attitudes.” *The Journal of Philosophy* 78 (2): 67–90.
- . 1999(1984). *Materia y conciencia*. Barcelona: Gedisa.
- . 2007. “On the Reality (and Diversity) of Objective Colors: How Color-Qualia Space Is a Map of Reflectance-Profile Space.” *Philosophy of Science* 74:119–149.
- Cohen, Jonathan. 2010. “It’s Not Easy Being Green: Hardin and Color Relationalism.” En *Color Ontology and Color Science*, editado por Jonathan Cohen y Mohan Matthen, 229–244. Cambridge: MIT Press.
- Dennett, Daniel C. 2017. *De las bacterias a Bach*. Barcelona: Pasado & Presente.
- Díez, José A., y C. Ulises Moulines. 2008(1997). *Fundamentos de Filosofía de la Ciencia*. Barcelona: Ariel.
- Elman, J. L. 1991. “Distributed representations, simple recurrent networks, and grammatical structure.” *Machine Learning* 7:195–225.
- García-Carpintero, Manuel. 2007(1995). “El funcionalismo.” En *La mente humana*, editado por Fernando Broncano, 43–76. Madrid: Trotta, CSIC.
- Hardin, L. 1988. *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.
- Hierro-Pescador, José. 2005. *Filosofía de la mente y de la Ciencia cognitiva*. Madrid: Akal.
- Martínez-Freire, Pascual F. 2015(1995). *La nueva filosofía de la mente*. Barcelona: Gedisa.
- Montero, Barbara. 2011. “Physicalism.” En *The Continuum Companion to Philosophy of Mind*, editado por James Garvey, 239–279. London: Continuum.
- Noordhof, Paul. 2011. “Current Issues in the Philosophy of Mind.” En *The Continuum Companion to Philosophy of Mind*, editado por James Garvey, 239–279. London: Continuum.

- Putnam, Hilary. 1985(1960). “Mentes y máquinas.” *MATHESES I* (2): 55–83.
- . 1965(1963). “Brains and Behavior.” En *Analytical Philosophy*, vol. 2, editado por R. J. Butler, 24–36. Oxford: Blackwell.
- . 1981(1967). *La naturaleza de los estados mentales*. México: Instituto de Investigaciones Filosóficas.
- . 2014(1988). *Representación y realidad*. Barcelona: Gedisa.
- . 2001(1999). *La trenza de tres cabos*. Madrid: Siglo XXI.
- Rabossi, Eduardo. 2007(1995). “La tesis de la identidad mente-cuerpo.” En *La mente humana*, editado por Fernando Broncano, 17–42. Madrid: Trotta, CSIC.
- Rodriguez, Mariano. 2006. “Desmontando la máquina: las razones de Putnam contra el funcionalismo.” *LOGOS. Anales del Seminario de Metafísica*. 39:53–76.
- Ryle, Gilbert. 2005(1949). *El concepto de lo mental*. Barcelona: Paidós.
- Toribio Mateas, Josefa. 2007(1995). “Eliminativismo y el futuro de la psicología popular.” En *La mente humana*, editado por Fernando Broncano, 245–272. Madrid: Trotta, CSIC.
- Wittgenstein, Ludwig. 2008(1953). *Investigaciones filosóficas*. Barcelona: Crítica.
- Wright, W. 2009. “The physical unnaturalness of Churchland’s ellipses.” *Philosophy of Science* 76:391–403.
- . 2020. “Eliminativism.” Penúltimo borrador, a aparecer en *The Routledge Handbook of Philosophy of Colour*.