UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE FILOSOFIA Y LETRAS

EL CONCEPTO INCOMPLETO DE LA MENTE EN EL CONDUCTISMO, LA TEORIA DE LA IDENTIDAD Y EL FUNCIONALISMO

T E S I S

QUE PARA OBTENER EL TITULO DE:

LICENCIADO EN FILOSOFIA

P R E S E N T A :

SERGIO ARMANDO GALLEGOS ORDORICA

DIRECTORA DE TESIS: DRA. SALMA SAAB HASSEN

MEXICO, D. F. MAYO DEL 2001.

UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

Agradecimientos

Al igual que muchos autores que al terminar de redactar un libro descubren con sorpresa que su obra difiere del proyecto que imaginaron en un inicio, he constatado con cierta sorpresa que media una cierta distancia entre mi proyecto inicial de tesis y el resultado obtenido después de varios meses de trabajo. En mi caso, este distanciamiento ha sido positivo en la medida que el proyecto original contenía una serie de intuiciones correctas, aunque muy vagas, que era menester precisar así como una serie de concepciones erróneas que era menester eliminar. El análisis reiterado y cuidadoso del texto de la tesis a medida que era redactado me permitió precisar algunas de estas intuiciones así como eliminar algunas de las ideas erróneas mas evidentes. Sin embargo, la mayoría de las intuiciones precisadas así como la mayoría de las ideas erróneas eliminadas sólo pudieron serlo a través del concurso de varias personas a las cuales reitero mi más caluroso agradecimiento.

En primera instancia, quisiera expresar mi gratitud a la Dra. Salma Saab, mi directora de tesis, que ha revisado y corregido con suma paciencia varias versiones previas de este trabajo, proporcionándome siempre valiosas observaciones y sugerencias para mejorarlo. También me gustaría expresar mi gratitud a la Dra. Lourdes Valdivia y al Dr. Silvio Pinto por los numerosos comentarios y críticas (muchas veces despiadadas y sin dar ningún cuartel) que me hicieron durante las largas horas que pasamos en *tête-à-tête* revisando el texto, así como a la Dra. Laura Benítez y al Dr. José Antonio Robles en cuyo seminario hace algunos años comencé a forjar las herramientas para realizar esta tarea.

En segunda instancia, quisiera agradecer a mis compañeros, los Estudiantes Asociados del Instituto de Investigaciones Filosóficas, la ayuda que me han brindado para mejorar la tesis a través de sus críticas y de sus comentarios, tanto en la sala de seminarios como fuera de ella. Entre ellos destacan principalmente Alberto Fonseca, Laura Duhau, Sandra Ramírez, Alicia Pazos, Andrea Pozas, Ángeles Eraña, Fernando Morett y Merlín Sosa. Espero poder seguir discutiendo libremente mis inquietudes filosóficas con todos ellos en años venideros como lo hemos venido haciendo estos últimos meses.

Last, but not least, quisiera expresar mi gratitud y mi cariño a mis padres y a mi hermana, sin cuyo

Primum animum dico, mentem quam saepe vocamus,
in quo consilium vitae regimenque locatum est,
esse hominis partem nihilo minus ac manus, et pes,
atque oculei partes animantis totius extant.
Sensum animi certa non esse in parte locatum,
verum habitum quendam vitalem corporis esse,
harmoniam Grai quam dicunt, quod faciat nos
vivere cum sensu, nulla cum in parte siet mens (...)

Lucrecio, *De Natura Rerum*, III, 94-101

El CONCEPTO INCOMPLETO DE LA MENTE EN EL CONDUCTISMO, LA TEORÍA DE LA IDENTIDAD Y EL FUNCIONALISMO.

Índice

Introducción

Uno de los problemas que ha marcado de manera central el desarrollo de la filosofía desde sus orígenes hasta los tiempos modernos ha sido la determinación de la naturaleza de la mente. La discusión sobre la naturaleza de la mente ha sido tradicionalmente un campo de batalla entre aquellos que sostienen que el cuerpo y la mente son dos entidades con naturalezas totalmente distintas y entre aquellos que sostienen que la naturaleza de ambas entidades es la misma. En esta discusión, vieja de varios siglos, han existido eventos de importancia capital que, en tanto constituyen una ruptura con respecto a la tradición, han determinado la evolución ulterior del debate. Uno de estos eventos capitales es la aparición de Descartes en la escena histórica en tanto su pensamiento marcó un hito en la discusión sobre la naturaleza de la mente. Para entender la importancia de la obra de Descartes sobre el debate acerca de la naturaleza de la mente, acaso convenga recordar brevemente la naturaleza y los objetivos de su proyecto filosófico.

Como la mayoría de las figuras más importantes en filosofía, Descartes se propuso realizar un proyecto de proporciones gigantescas: su objetivo era construir una ciencia universal de la naturaleza que abarcase todas las áreas del conocimiento humano (física, biología, matemáticas, etc.) que se habían desarrollado separadamente hasta entonces. El proyecto de construcción de esta ciencia universal partió del siguiente hecho. Tras haber constatado que el conocimiento se encuentra disperso en distintas ciencias particulares y que contiene muchas contradicciones y errores, Descartes dedujo de ello que era imposible construir la ciencia universal de la naturaleza tomando como base parte de estas ciencias particulares. Podía darse el caso de que la parte que se tomara para realizar el proyecto estuviese viciada por contradicciones o errores que, aunque invisibles a primera vista, provocasen ulteriormente el colapso de la ciencia universal de la naturaleza al ser introducidos como cimientos en la obra. La solución propuesta por Descartes para remediar esto consistió en la estrategia de la duda metódica que consistía en eliminar sistemáticamente todo conocimiento de naturaleza dudosa hasta llegar a una serie de verdades indubitables *a priori*. Según él, estas verdades indubitables serían la base apropiada sobre la cual se podría erigir eventualmente una ciencia universal de la naturaleza.

Una de las verdades a las cuales llegó Descartes por eliminación sistemática de todo el conocimiento dudoso –verdad que constituye uno de los puntos culminantes de las *Meditaciones Metafísicas*– consiste en el establecimiento de un dualismo de sustancias en el cual la mente y el cuerpo se consideran como pertenecientes a categorías ontológicas distintas. A continuación, realizaremos una reconstrucción del argumento que le permite a Descartes derivar, a partir de algunas hipótesis que tienen el estatuto de premisas, esta verdad central. Una de estas hipótesis consiste en la creencia de que Dios puede hacer cualquier cosa que entiendo de manera clara y distinta.[1] Otra de las hipótesis esenciales de las cuales se deriva el dualismo de sustancias consiste en la creencia de que, como Dios es perfecto, es imposible que desee engañarnos en la medida que el engaño es un tipo de imperfección que no puede existir en un ente que, por definición, es perfecto.[2] Con base en estas dos hipótesis, podemos derivar uno de los principios rectores del pensamiento cartesiano que es habitualmente conocido en la literatura como Principio Epistemológico Fundamental (o PEF):

(1) Dios puede hacer cualquier cosa que entiendo clara y distintamente tal como la entiendo.

(2) Como Dios es perfecto, es imposible que desee engañarnos en la medida que el engaño es un tipo de imperfección.

(3) De las premisas (1) y (2) se deriva que *todo aquello que entiendo clara y distintamente existe tal como es entendido*.

El PEF tiene una importancia capital en el pensamiento cartesiano en la medida que es la piedra angular sobre la cual se construye el dualismo de sustancias sostenido por Descartes. Para ello, es necesario añadir algunas otras premisas al argumento –premisas que Descartes asume como hipótesis. El cuarto paso del argumento consiste en aseverar que sé que existo (i.e., sé que mi esencia es el ser una cosa pensante) y

[1] AT VII, 78: "Quoniam scio omnia quae clare & distincte intelligo, talia a Deo fieri posse qualia illa intelligo (...)"
[2] AT VII, 53: "Agnosco fieri non posse ut ille me unquam fallat; in omni enim fallacia vel deceptione aliquid imperfectionis reperitur, & quamvis posse posse fallere, nonnullum esse videatur acuminis aut potentiae argumentum, proculdubio velle fallere, vel malitiam vel imbecillitatem testatur, nec proinde in Deum cadit"

que dicho conocimiento es algo que se da de manera clara y distinta así como en aseverar que también sé que tengo un cuerpo (cuya esencia es el ser una cosa extensa) y que dicho conocimiento se da de manera clara y distinta.[3]

(4) Entiendo de manera clara y distinta que soy una cosa cuya esencia consiste en pensar (y no en ser extensa) y también entiendo de manera clara y distinta que tengo un cuerpo cuya esencia consiste en ser una cosa extensa (y no una cosa que piensa).

(5) En virtud de (3) y (4), se deriva que soy una cosa cuya esencia consiste en pensar y que tengo un cuerpo cuya esencia consiste en ser una cosa extensa.

En este punto del argumento, disponemos de los dos términos que nos permiten establecer el dualismo de sustancias. Lo que resta por hacer para derivar el dualismo de sustancias es mostrar que ser una cosa pensante y ser una cosa extensa son propiedades mutuamente excluyentes. Esto es implícitamente asumido por Descartes como un hecho.[4]

(6) Si algo es una cosa extensa, entonces no puede ser una cosa pensante y viceversa.

(7) En virtud de (5) y (6), se deriva que soy realmente distinto de mi cuerpo y que puedo existir sin él (i.e., tenemos un dualismo de sustancias).

Ahora bien, es importante notar que el PEF tiene tanto importantes ventajas como serias desventajas. La primera gran ventaja del PEF es que nos permite presentar un baluarte contra el escepticismo global que

[3] AT, VII, 78: "Ac proinde, ex hoc ipso quod sciam me existere, quodque interim nihil plane aliud ad naturam sive essentiam meam pertinere animadvertam, praeter hoc solum quod sim res cogitans, recte concludo meam essentiam in hoc uno consistere, quod sim res cogitans. Et quamvis fortasse (vel potius, ut postmodum dicam pro certo) habeam corpus, quod mihi valde arcte conjunctum est, quia tamen ex una parte claram & distinctam habeo ideam mei ipsius, quatenus sum res cogitans, non extensa, & ex alia parte distinctam ideam corporis, quatenus est tantum res extensa, non cogitans (...)"

[4] AT, VII, 27: "Sum autem res vera, & vere existens: sed qualis res? Dixi, cogitans. Quid paeterea? (...) non sum compages illa membrorum, quae corpus humanum apellantur."

amenaza destruir el proyecto de la ciencia universal de la naturaleza. A través del PEF, podemos ver que existe algo que es indubitable (el hecho que yo sea una cosa pensante) en la medida que siempre entiendo de manera clara y distinta que no puedo dudar del hecho que soy una cosa pensante, puesto que al dudar no estoy haciendo otra cosa que pensar. La segunda gran ventaja del PEF consiste en que nos permite dar cuenta de los estados mentales con contenido fenoménico (sensaciones de dolor, de hambre, emociones, etc.) como entidades realmente existentes en la medida que, cada vez que tenemos dichos estados mentales (i.e., cada vez que experimentamos esas sensaciones o emociones), tenemos un entendimiento claro y distinto de ellos. Por lo tanto, cada vez que sentimos hambre o dolor, realmente tenemos hambre o dolor: no puede haber ningún engaño o ilusión así como en el caso de la indubitabilidad de la existencia del sujeto

A pesar de estas dos ventajas, existe una objeción que ha sido determinante en el rechazo del dualismo de sustancias cartesiano. Para examinar esta objeción, asumamos por un momento que el PEF es válido. Por un lado, con base en esta asunción, en la medida que Descartes entiende clara y distintamente el cuerpo como una entidad cuya propiedad esencial es la extensión y la mente como una entidad cuya propiedad esencial es el pensamiento, es evidente para él que el cuerpo debe de ser algo esencialmente extenso mientras que la mente debe de ser algo esencialmente pensante. Además, en la medida en que el pensamiento y la extensión son propiedades mutuamente excluyentes, es evidente que la mente no puede ser extensa y el cuerpo no puede ser pensante. Así pues, se hace patente que el PEF cartesiano apunta sin reservas hacia un dualismo de sustancias en el cual la mente y el cuerpo pertenecen a distintas categorías ontológicas mutuamente excluyentes. Por otro lado, es menester recordar que Descartes se adhirió, como muchos de sus contemporáneos, a una visión mecanicista del mundo material en la cual se explicaba la interacción de los distintos cuerpos por medio de un vínculo causal. Dados estos dos condicionamientos, el problema de Descartes consistió en intentar explicar cómo podía haber un vínculo causal entre mente y cuerpo si ambos elementos pertenecían a distintas categorías ontológicas. Acaso el mejor intento de Descartes por resolver el problema de la interacción causal mente-cuerpo sea el que encontramos plasmado en la Sexta Meditación *Grosso modo*, la propuesta de Descartes es la siguiente: cuando ocurre

algo que daña nuestro cuerpo (e.g., cuando recibimos una pedrada arrojada por alguien en una pierna), una información especial es transmitida por ciertos "espíritus animales" que viajan a través de los nervios hasta el cerebro de acuerdo con las leyes de la física. Habiendo llegado ahí, esta información particular es transmitida por los "espíritus animales" a la glándula pineal donde se establece el puente entre cuerpo y mente.

Los críticos del dualismo cartesiano han señalado que, si bien Descartes se extiende largamente en sus obras con respecto a los mecanismos de transmisión de la información de los miembros del cuerpo al cerebro y del cerebro a la glándula pineal, permanece sospechosamente silencioso en cuanto a lo que ocurre exactamente en la glándula pineal. Este silencio se debe, según los críticos, al hecho que Descartes no podía dar cuenta de la naturaleza de los mecanismos que conectaban la mente con el cuerpo en tanto que dichos mecanismos implicaban "saltos ontológicos". Los herederos putativos de Descartes idearon varias estrategias para poder resolver este problema, pero ninguna de ellas tuvo realmente éxito. Esto explica que el dualismo cartesiano declinara lentamente con el correr de los siglos hasta finales del siglo XIX donde acaso su última gran exposición y defensa consistió en la tesis de doctorado de Bergson, *Essai sur les données immédiates de la conscience* (1889). Sin embargo, a pesar de su empeño y de su perseverancia, Bergson nunca pudo presentar una versión del dualismo cartesiano que diese cuenta de la naturaleza de la interacción entre mente y cuerpo.

Ahora bien, el dualismo cartesiano no era la única corriente de pensamiento filosófico que pretendía en el siglo XVII resolver el problema de la naturaleza de la mente. Frente a Descartes y a sus seguidores, se alzaban varios pensadores materialistas que rechazaban la interpretación cartesiana del PEF en el caso de la mente. Para rechazar la idea de una mente inextensa e inmaterial, estos adversarios de Descartes se apoyaron en el paradigma del atomismo mecanicista que constituía la mejor manera de dar cuenta de los procesos naturales para muchos filósofos y científicos del siglo XVII. Según este paradigma, todos los procesos naturales como la caída de los cuerpos, la reproducción de los organismos, la absorción de los alimentos o la transmisión del calor podían ser explicados postulando una serie de átomos que constituían los componentes últimos de la realidad y que se relacionaban entre sí por medio de varias leyes, una de

las cuales era la ley de causalidad. En el ámbito del atomismo mecanicista del siglo XVII, el movimiento de una piedra lanzada por una persona al aire se explicaba por el hecho que había otro objeto distinto de la piedra (e.g., el brazo de la persona que la lanza) que causaba el movimiento de la piedra. Es importante notar que hay una condición implícita para que una ley causal como la que se da cuenta del movimiento de la piedra en el ejemplo anterior sea valida: los cuerpos que liga la relación causal (y, por ende, los átomos que los componen) deben de pertenecer a una misma categoría ontológica ya que, de otra manera, se llegaría al mismo callejón sin salida de "saltos ontológicos" al que llegó Descartes. Así pues, en tanto es evidente para la mayoría de las personas que la mente se relaciona de manera causal con los objetos del mundo (lo cual se hace patente a través del uso cotidiano que hacemos de oraciones como "María pensaba que iba a llover, por lo que tomó un paraguas al salir de casa" para explicar el comportamiento de los demás) y que dichos objetos del mundo son eminentemente materiales, los materialistas concluyen que la mente debe de ser también algo de naturaleza material.

Es importante destacar que, si bien el materialismo parece no poder dar cuenta de ciertos rasgos de la mente y de los estados mentales (en particular, de la no-extensión, de la indubitabilidad y de su carácter intrínseco que hace que sean precisamente lo que son), posee a cambio una ventaja que muchos juzgan decisiva: permite explicar claramente aquello que Descartes dejó siempre sin explicar, i.e., cómo puede haber una interacción entre mente y cuerpo y, más precisamente, una interacción causal. Como podemos apreciar, esto plantea un serio dilema para aquellos interesados en resolver el problema de la naturaleza de la mente: o bien se aceptan las exigencias del PEF, lo cual parece eliminar toda posibilidad de explicar las propiedades causales de la mente respecto al cuerpo (a menos de postular un milagro), o bien se aceptan las exigencias de una serie de leyes causales entre mente y cuerpo, lo cual parece eliminar toda posibilidad de explicar las propiedades fenoménicas de la mente y de los estados mentales. Este es el dilema central alrededor del cual se ha articulado la discusión de la relación mente-cuerpo de la época de Descartes hasta nuestros días.

Del siglo XVII a las postrimerías del siglo XIX, la discusión en torno al dilema estuvo marcada por un debilitamiento progresivo de las propuestas de tipo cartesiano y por un desarrollo también progresivo de

las propuestas materialistas. Tanto el debilitamiento del dualismo como el desarrollo del materialismo se explican en buena medida por el progreso y la sistematización del conocimiento científico durante este período. Ambos procesos hicieron que disciplinas como la psicología, que tenían un carácter meramente cualitativo y teórico, adquieran poco a poco un carácter cuantitativo y práctico. Este carácter cuantitativo y práctico de la psicología se apreciaba en el siglo XIX en el interés creciente de ciertos psicólogos por encontrar en el cuerpo (y, en particular, en el cerebro) el origen de ciertos desordenes mentales. Este interés se benefició del perfeccionamiento paulatino de un instrumento como el microscopio que permitió descubrir las neuronas y estudiar su estructura. Dicho estudio proveyó a los pensadores materialistas con datos empíricos precisos sobre el funcionamiento del cerebro que reforzaron la plausibilidad de sus tesis.

Después de que la última gran exposición y defensa de una versión del dualismo cartesiano encarnada en la persona de Bergson fracasase en las dos últimas décadas del siglo XIX, el materialismo se convirtió a los ojos de muchos en la única opción viable para resolver el problema de la naturaleza de la mente, por lo cual el siglo XX presenció un increíble desarrollo del materialismo en el ámbito de la filosofía de la mente, a tal grado que poco a poco se ha convertido en la corriente dominante. Ahora bien, aunque es cierto que se articula en torno a una tesis básica, es menester notar que el materialismo dista mucho de ser una corriente de pensamiento homogénea y uniforme. A lo largo del siglo XX, se han sucedido muchas corrientes de pensamiento materialista tanto en filosofía como en psicología (conductismo psicológico, conductismo filosófico, teoría de la identidad mente-cerebro, etc.) que, enfrentando las limitaciones y los errores de las teorías materialistas precedentes, han buscado aportar una solución definitiva al problema de la naturaleza de la mente. Hoy en día, la corriente de pensamiento materialista más popular entre muchos filósofos, psicólogos y neurocientíficos es conocida como *funcionalismo*. Es importante destacar que tampoco el funcionalismo es una corriente de pensamiento uniforme y homogénea, sino que existen distintos tipos de funcionalismo con distintos rasgos. Sin embargo, hay un elemento común, una tesis básica que todos comparten y que consiste en lo siguiente: para un funcionalista, un estado mental M es una función que mapea un tipo de objetos P (estímulos y "estados mentales" distintos de M) a otro tipo de objetos Q (conductas y otros "estados mentales" distintos de M) de acuerdo a una estructura causal, sin

que importe la composición química o física de aquello que realiza la función. Las ligeras diferencias de posición que encontramos presentes de un funcionalista a otro (diferencias que son responsables de la existencia de funcionalismos teleológicos, homunculares, etc.) serán en general pasadas por alto en este trabajo en aras de una percepción unitaria de la teoría.

Partiendo de la hipótesis que una buena teoría de la mente debe explicar tanto los vínculos causales que parecen existir entre mente y cuerpo (en la medida que la inmensa mayoría de la gente tiende a dar una interpretación causal a oraciones del tipo "Como María pensaba que iba a llover y no quería mojarse, tomó un paraguas al salir de casa") así como las propiedades fenoménicas (en la medida en que son indubitables de acuerdo con el PEF) y la intencionalidad de los estados mentales (en la medida que ciertos estados mentales tienen una cierta direccionalidad con respecto a los objetos del mundo), el propósito central de esta tesis es demostrar que ninguna de las principales teorías materialistas desarrolladas en el siglo XX puede ser una teoría satisfactoria de la mente. Para ver esto, procederé en varias etapas. En el primer capítulo, realizaré un breve estudio de las dos primeras grandes corrientes de pensamiento materialista que surgieron en los inicios del siglo XX: el conductismo psicológico y el conductismo filosófico. En este primer capítulo, estudiaré las principales motivaciones detrás del desarrollo de ambas teorías, examinaré los supuestos sobre los cuales reposan y determinaré cuáles son sus éxitos y sus alcances así como sus errores y sus limitaciones. En la primera sección, analizaré exclusivamente el conductismo psicológico. En el caso del conductismo filosófico, estudiaré las propuestas de Carnap así como de los demás positivistas lógicos en la segunda sección y las propuestas de Ryle en la tercera sección del capítulo. Esta división se justifica por el hecho que, aun cuando Ryle es considerado tradicionalmente como un conductista filosófico, su posición difiere significativamente de la de los positivistas lógicos en ciertos puntos (como habré de mostrarlo en la tesis), por lo cual requiere de un tratamiento separado. Finalmente, en una cuarta y última sección, realizaré un balance general de las distintas versiones de conductismo.

En el segundo capítulo, realizaré un estudio breve de la corriente de pensamiento materialista que sucedió a las dos versiones principales de conductismo y que precedió al funcionalismo: la teoría de la

identidad mente-cerebro. En la primera sección de este capítulo, estudiaré los orígenes de la teoría de la identidad mente-cerebro. En la segunda sección, presentaré los distintos argumentos que se desarrollaron para sostener la teoría de la identidad mente-cerebro. En la tercera sección, examinaré las distintas objeciones que se han presentado para rechazar la teoría de la identidad. Finalmente, en la cuarta y última sección de este segundo capítulo, haré un balance general de la teoría de la identidad mente-cerebro.

El objetivo de estos dos primeros capítulos es doble. Por un lado, deseo replantear en una perspectiva histórica el desarrollo del funcionalismo para entender cómo llegó a ser la teoría más popular de la mente en la actualidad. Por otro lado, deseo hacer constatar bien a mis lectores los errores y las limitaciones de estas teorías materialistas de la mente que precedieron al funcionalismo porque, en algunos casos, son también errores y limitaciones propios del funcionalismo como habré de mostrar a lo largo de la tesis.

La primera parte del tercer capítulo estará consagrada a un estudio detenido del funcionalismo en el cual analizaré sus orígenes en la década 1960-1970 principalmente a través de las propuestas de tres autores: Lewis, Putnam y Dennett. Este análisis tendrá por lo menos tres objetivos: mostrar cuáles eran las motivaciones que propiciaron el desarrollo de la teoría, exponer las distintas versiones elaboradas por Lewis, Putnam y Dennett para determinar sus puntos comunes y sus divergencias, y mostrar cuáles son sus ventajas y sus desventajas. En una segunda parte, mostraré que las tres versiones de funcionalismo estudiadas anteriormente (y, en particular, la versión sostenida por Dennett) son incapaces de explicar de manera satisfactoria la dimensión intencional de los estados mentales, por lo cual estas tres versiones (así como cualquier otra propuesta funcionalista basada en los supuestos que las tres propuestas estudiadas comparten sobre el holismo semántico) están condenadas a ser teorías incompletas de la mente. Finalmente, en una tercera y última parte, asumiendo que existen ciertas versiones de pensamiento funcionalista que pueden dar cuenta satisfactoriamente de la intencionalidad de los estados mentales (en especial, la propuesta de Fodor que se encuentra basada en un atomismo semántico), mostraré, por un lado, que cualquier definición funcional de un estado mental (sin importar que se encuentre en el contexto de un holismo o de un atomismo semántico) es incapaz de recuperar las propiedades fenoménicas de los estados mentales por razones de principio y, por otro lado, que las propiedades fenoménicas de los

estados mentales no pueden ser eliminadas puesto que constituyen una parte esencial del fenómeno de la conciencia, sin el cual la mente no es mente. Este será el argumento decisivo para rechazar al funcionalismo como teoría de la mente.

En las conclusiones del presente trabajo, realizaré una evaluación global de las diversas corrientes de materialismo con base en los argumentos presentados en los tres capítulos, señalando lo que debe ser conservado así como lo que debe ser eliminado en mi opinión para el desarrollo eventual de una buena teoría de la mente que llene las tres condiciones básicas: dar cuenta de los vínculos causales entre la mente y el cuerpo, dar cuenta de la dimensión intencional y dar cuenta de las propiedades fenoménicas de los estados mentales. Esta evaluación servirá para determinar algunas posibles líneas de investigación ulterior

Capítulo I

Conductismo psicológico y conductismo filosófico. Motivaciones, aciertos, errores y limitaciones en la explicación de la naturaleza de la mente.

El conductismo, que es considerado como la primera gran corriente de pensamiento materialista del siglo XX, tuvo dos expresiones principales, cada una en campos distintos: una en psicología, que es denominada conductismo metodológico o psicológico, y otra en filosofía, que es denominada conductismo analítico o filosófico. A primera vista, ambas corrientes son virtualmente indistinguibles en la medida en que comparten la misma tesis central así como muchos supuestos teóricos. Sin embargo, cuando las dos corrientes son examinadas con la atención debida, se perciben varias divergencias notables en ciertos puntos claves que separan a los conductistas psicológicos de los conductistas filosóficos. Por ejemplo, los conductistas psicológicos sostienen que la tesis central sobre la cual reposa su propuesta (tesis que será enunciada algunas líneas más abajo) es una hipótesis de carácter empírico que debe ser demostrada mediante la experimentación mientras que los conductistas filosóficos sostienen que se trata de una verdad analítica *a priori* que no requiere demostración. Si bien el propósito del presente trabajo es de carácter filosófico (por lo cual habremos de analizar con sumo detenimiento el conductismo filosófico), es pertinente dedicar al menos una sección de este capítulo al estudio del conductismo psicológico por la siguiente razón. Como los conductistas filosóficos sostienen que la tesis central del conductismo es una verdad analítica *a priori*, de ello se deduce que debe de ser verdadera no sólo en el mundo actual, sino en todos los mundos posibles donde existen seres humanos que presentan conductas. En cambio, el alcance de la aseveración de los conductistas psicológicos es mucho más limitada en la medida que su hipótesis se encuentra restringida al mundo actual. Así pues, si conseguimos mostrar que la tesis del conductismo no es verdadera en el mundo posible actual (lo cual haría al conductismo psicológico erróneo), tendremos buenas razones para suponer que tampoco es válida como una verdad analítica *a priori*.

I.1 El conductismo psicológico

El conductismo psicológico surgió por primera vez en las principales obras del psicólogo americano J. B. Watson, *Psychology as the Behaviorist views it* (1913) y *Behaviorism* (1924), y fue posteriormente

retomado por B. F. Skinner en *Science and Human Behavior* (1953) y *Verbal Behavior* (1957) como una teoría que pretendía explicar la naturaleza de la mente rechazando las tesis del dualismo de sustancias cartesiano y apoyándose en las observaciones empíricas realizadas por la psicología experimental, que se encontraba entonces en pleno desarrollo. Expuesta a grandes rasgos, la tesis central del conductismo metodológico tiene dos versiones básicas: (1) aquello que la gente conoce como "estados mentales", y que constituye el objeto de estudio de la psicología tradicional, en realidad no existe, y (2) la expresión "estados mentales" no es sino una manera de designar la conducta de las personas, por lo que un estado mental es idéntico a una conducta particular Como los conductistas psicológicos utilizan ambas versiones de manera alternativa, nosotros las emplearemos también de manera alternativa en nuestro trabajo. Para los conductistas psicológicos, la manera cómo la gente habla de los "estados mentales" es particularmente engañosa puesto que induce a la gran mayoría de la gente a pensar que existe algo más allá de las conductas de las personas, algo que determina estas conductas. Las oraciones del tipo "Pensó que iba a llover, por lo que tomó un paraguas al salir de casa", que habitualmente usamos para explicar y/o predecir la conducta de los demás, y que son esgrimidas como argumentos para justificar la existencia de un vínculo causal entre la mente y el cuerpo, eran consideradas por los conductistas psicológicos como ejemplos de un mal uso del lenguaje. Dichas oraciones debían ser eliminadas según ellos y remplazadas por oraciones donde sólo hubiese descripciones de conductas como "Tomó un paraguas al salir de casa".

Originalmente, el conductismo psicológico pretendió limitarse a la descripción de las meras conductas para dar cuenta de la "mente". Sin embargo, los conductistas psicológicos constataron pronto que, si bien las conductas no podían ser causadas por "estados mentales", debía de existir algo de lo cual dependieran porque, de otra manera, cualquier individuo podía tener virtualmente en cualquier momento cualquier conducta. Así pues, introdujeron la noción de *estímulo* que servía para determinar el tipo de conducta que presentaba un individuo sometido a dicho estímulo (e.g., el estímulo que consiste en el olor de un pedazo de queso determina en un ratón la conducta dirigida a obtener el pedazo de queso). El conductismo psicológico se considera habitualmente una teoría materialista de la mente en tanto que las únicas entidades que acepta (*estímulos* y *conductas*) parecen ser describibles únicamente en términos físicos. Sin embargo, es menester notar que las nociones de *estímulo* y *conducta*, tal como son entendidas por los

conductistas psicológicos, son en extremo confusas y que, al ser sometidas a un análisis cuidadoso, revelan ser totalmente vacuas y carentes de sentido, puesto que existen cosas que son habitualmente consideradas como estímulos o conductas, pero que no son describibles en meros términos físicos. Veremos este problema más adelante. Nos limitaremos a aceptar estas dos nociones por el momento y continuaremos con la exposición.

La razón primaria por la cual los conductistas metodológicos no admitían sino estímulos y conductas como los únicos elementos aceptables en el ámbito de la psicología consistía en su creencia de que dichos elementos eran los únicos que permitían la elaboración de una teoría psicológica que fuese empíricamente comprobable. El motivo central de suspicacia de los conductistas psicológicos respecto a las teorías de la mente inspiradas en el dualismo sustancial consistía en que dichas teorías apelaban a "estados mentales" que se encontraban más allá de las posibilidades explicativas de la conducta. Ahora bien, como los datos de la conducta son los únicos empíricamente comprobables para los conductistas psicológicos, es evidente que cualquier teoría de la mente que apele a otras cosas que no sean los meros datos de la conducta para ofrecer una explicación psicológica carece de todo valor científico. Como los "estados mentales" son entidades que se encuentran más allá de lo que es empíricamente comprobable, toda explicación psicológica en la cual aparecen es por demás dudosa e incierta.

Es interesante también notar que los conductistas psicológicos incluso se muestran escépticos respecto al hecho de introducir en las teorías psicológicas estados físicos del cerebro por dos razones principales. Por un lado, como dichos estados caen mas allá del ámbito de lo que es públicamente observable (que se limita a los datos de la conducta), para los conductistas psicológicos siempre existirá una duda *prima facie* en cuanto a la existencia misma de una relación entre lo neuronal (que, por ser invisible, no es empíricamente comprobable) y la conducta (que, por ser visible, si lo es) mientras que, por otro lado, aun cuando algún día se desarrolle una ciencia del cerebro completa, no existe ninguna garantía de que tenga un carácter experimental y empírico (podría sólo ser una ciencia meramente teórica, por lo cual no tendría absolutamente ningún valor para la psicología experimental con capacidad predictiva que los conductistas psicológicos buscaban desarrollar):

> Eventualmente una ciencia del sistema nervioso basada en la observación directa más que en la inferencia describirá los estados y sucesos neuronales que preceden inmediatamente a las instancias de conducta. (...) Sin embargo, podemos notar que no tenemos y podemos nunca llegar a tener este tipo de información neurológica en el momento que la necesitamos para predecir una instancia específica de conducta. Es incluso aún más improbable que podamos alterar el sistema nervioso directamente para establecer las condiciones antecedentes de una instancia particular.[5]

Así pues, de acuerdo con el criterio de validez del conductismo psicológico, la única manera correcta de describir a los organismos es como "auténticas cajas negras cuya estructura interna se encuentra para siempre cerrada a la investigación psicológica."[6] Incluso si existe algo que sirva de mediador entre estímulos y conductas (e.g., ciertos estados físicos del cerebro), esto no es relevante para un conductista psicológico puesto que la discusión sobre la naturaleza física o química de este mediador e incluso la discusión sobre su existencia hacen referencia a entidades distintas de los datos de la conducta, lo cual las torna dudosas e inciertas automáticamente. Aun cuando la neurociencia eventualmente descubra el mecanismo neuronal que media entre el hecho de tocar con la mano una parrilla al rojo vivo y el hecho de retirar la mano y gemir, este mecanismo neuronal no tiene valor explicativo alguno para un conductista psicológico puesto que no pertenece a la categoría de eventos empíricamente comprobables que constituyen la evidencia.

Si deseamos entender mejor el concepto de "estado mental" en el contexto del conductismo metodológico, conviene probablemente establecer una analogía con la matemática. Para los conductistas

---

[5] Skinner, 1953 en Block, 1980, p. 38: "Eventually a science of the nervous system based upon direct observation rather than inference will describe the neural states and events which immediately precede instances of behavior. (...) However, we may note here that we do not have and may never have this sort of neurological information at the moment it is needed in order to predict a specific instance of behavior It is even more unlikely that we shall ever be able to alter the nervous system directly in order to set up the antecedent conditions of a particular instance."

[6] Kim, 1998, p. 43: "[On this principle, organisms have to be construed as] veritable black boxes whose internal structure is forever closed to psychological investigation."

metodológicos, los estados mentales son como las variables de las ecuaciones con las cuales trabaja un matemático: para un matemático, no importa no conocer el valor de las variables en la medida en que nos ayudan a descubrir las regularidades en la matemática (i.e., a demostrar los teoremas matemáticos).[7] Para un conductista psicológico, no importa no saber cuál es la naturaleza de los "estados mentales" (incluso no importa saber si existen realmente o no) si podemos ligarlos con las constantes que conocemos (i.e., estímulos y conductas) para sentar las regularidades que rigen la psicología como bien señala Spence al destacar lo siguiente:

> Los únicos significados que estos constructos teóricos involucrados tienen actualmente son provistos por las ecuaciones que los relacionan con las variables experimentales conocidas - las medidas experimentales por un lado y las medidas del comportamiento por otro.[8]

El conductismo psicológico tiene una ventaja central que es importante destacar para comprender el éxito que tuvo en su tiempo: permite dar cuenta de la mente sin apelar (como el dualismo cartesiano lo hace) a entidades oscuras o de existencia dudosa. Las únicas entidades que el conductista reconoce son objetos del dominio público que son describibles en términos físicos como los estímulos (que pueden ser de carácter visual, auditivo, etc.) o las conductas resultantes de los estímulos, lo cual hace que su teoría sea empíricamente comprobable a través de la predicción de la conducta El conductismo psicológico puede vanagloriarse entonces de haber dado por primera vez en la historia de la psicología un giro decisivo que permitía dejar de considerarla como una mera ciencia cualitativa para admitirla en el grupo de las ciencias empíricas cuantitativas, donde el progreso a través de la experimentación es esencial.

---

[7] Es importante que nuestros lectores sepan que esta concepción de las variables (y de los números en general) no es la concepción general de la matemática, sino sólo de un pequeño grupo de matemáticos conocidos como Formalistas. En general, la mayoría de los matemáticos son Realistas, por lo que sí les importa saber si las entidades con las que trabajan existen o no.

[8] Spence, 1948 en Feigl y Brodbeck, 1953, p. 580: "The only meanings that these theoretical intervening constructs have at the present time is provided by the equations which relate them to the known experimental variables -the experimental measurements on the one hand and the behavior measures on the other."

Una segunda razón que explica el éxito del conductismo metodológico en los inicios del siglo XX es el hecho que ofrece una explicación plausible de la diferencia entre las conductas de individuos distintos en presencia de un mismo estímulo, apelando a distintas historias de *reforzamiento* en los individuos. La noción de reforzamiento es, al igual que las nociones de estímulo y conducta, esencial para entender el conductismo psicológico aunque, al igual que ellas, no resiste un análisis detallado como veremos más adelante. Aceptémosla por el momento y prosigamos con nuestro análisis de las ventajas del conductismo psicológico. Si las conductas de dos individuos A y B son distintas en presencia de un mismo estímulo (e.g., si en presencia de un dulce, la conducta de un individuo A consiste en tomarlo mientras que la conducta de un individuo B consiste en ignorarlo), esto se explica según el conductismo psicológico en la medida en que los individuos A y B han sido sometidos previamente al mismo estímulo varias veces, pero con distintos condicionamientos: cada vez que A veía el dulce e intentaba tomarlo, se le permitía hacerlo sin problema mientras que cada vez que B veía el dulce e intentaba tomarlo, se le aplicaba un castigo.

Una tercera razón por la cual el conductismo metodológico tuvo una gran aceptación en las primeras etapas de su desarrollo es que parecía proporcionar una respuesta afirmativa al problema de la existencia de otras mentes, lo que constituía un avance con respecto al dualismo de sustancia cartesiano que sólo podía responder negativamente al problema (lo que contradecía las intuiciones que tenemos con respecto a las demás personas). En el marco del pensamiento dualista sustancial, la pregunta "¿Tienen los demás seres humanos mentes?" no podía ser contestada afirmativamente a causa del PEF como veremos a través del siguiente razonamiento. Es claro que nuestros sentidos nos proporcionan percepciones de que existen cuerpos externos muy similares al nuestro, y a partir de ciertos signos generalmente tendemos a deducir que dichos cuerpos albergan mentes similares a la nuestra. Lamentablemente, estas percepciones no son claras y distintas, por lo cual son vulnerables ante los ataques del escepticismo radical. El dualismo de sustancias cartesiano tiene la ventaja de proporcionarnos un acceso aparentemente directo e incorregible a nuestras mentes a través de la introspección, pero al mismo tiempo tiene un gran inconveniente que consiste en aislarnos de las mentes de los demás que no puede ser percibidas directamente como percibimos la nuestra, sino sólo a través de un paso inferencial. Por lo tanto, siempre subsistirá la duda en

el marco del pensamiento dualista de sustancias de que aquello que tomamos por signos de la existencia de un ente consciente no sea otra cosa más que la actividad visible de un simple mecanismo artificial diseñado para hacernos creer que tiene mente cuando en realidad no la tiene. En la medida que no podemos deducir con *certeza* de los signos externos la existencia de una mente distinta de la nuestra por un lado y no podemos acceder directamente a la mente de los demás por medio de la introspección, para un dualista sustancial siempre cabe la posibilidad de que las personas con las cuales se intercambian pareceres no sean más que máquinas inventadas por un demonio maligno que se comportan como seres humanos aunque en realidad no lo sean. En el contexto del conductismo psicológico, por el contrario, dicha duda no tiene lugar en la medida en que, como los únicos datos aceptables para elaborar una teoría psicológica son aquellos de la conducta, es claro que todo aquello que se comporte como un ser humano debe necesariamente ser un ser humano [9]

Hemos visto hasta ahora en qué medida el conductismo psicológico puede ser considerado como una teoría materialista de la mente así como las principales ventajas que presenta y que hicieron de él en su tiempo una teoría ampliamente aceptada. Lamentablemente, el conductismo metodológico también tiene un número significativo de limitaciones y errores señalados por muchos autores a lo largo de los años. Acaso la primera objeción que se dirige al conductismo psicológico consiste en la necesidad que la teoría impone de excluir de cualquier explicación psicológica todo aquello que no sea describible con base en la conducta, particularmente los "estados mentales" y los estados físicos internos del cerebro, puesto que no son entidades públicamente observables. Esta estrategia restrictiva acaso sea justificada en el caso de los "estados mentales" que, al ser principalmente subjetivos y privados, evidentemente no pueden constituir una base sólida para elaborar una teoría de la mente que pueda ser empíricamente comprobable a través

[9] En el caso de esta aparente ventaja del conductismo psicológico, es importante destacar algo. El problema de la existencia de otras mentes es eminentemente un problema de tipo filosófico, por lo cual un conductista psicológico no le presta atención en principio en tanto ni siquiera se le plantea. Recuérdese que hablar de "mentes" es engañoso para los conductistas psicológicos. Sólo podemos hablar con seguridad de los datos de la conducta y constituye un auténtico truismo decir que hay conductas en los demás. Sin embargo, lo que quiero decir en estas líneas es que, si bien el problema de las otras mentes no se plantea para un conductista psicológico, su metodología de trabajo le sirve a aquellos que sí se plantean el problema (i.e, los filósofos) para intentar aportar respuestas al problema.

de un proceso de experimentación. Sin embargo, en el caso de los estados físicos internos del cerebro (también conocidos como estados neuronales), esta estrategia no se justifica puesto que los estados físicos internos del cerebro no son entidades privadas y subjetivas, sino que tienen un aspecto cuantitativo y son públicamente observables. Esta aseveración se encuentra justificada por el hecho patente y comprobado que la neurobiología progresa por medio de la experimentación y de la comprobación pública de sus hipótesis. Así pues, la primera objeción de peso contra el conductismo psicológico consiste en el hecho de que las restricciones de la teoría son demasiado estrictas desde un punto de vista metodológico, lo cual resulta contraproducente puesto que excluye datos que son empíricamente comprobables y que podrían tener una importancia capital para la elaboración de explicaciones psicológicas adecuadas.

Dejando de lado el hecho de que las restricciones del conductismo psicológico excluyen datos que podrían ser relevantes para elaborar explicaciones psicológicas correctas, la segunda objeción de peso contra el conductismo psicológico consiste en lo que permite su tesis según la cual todo lo que es mental se reduce a estímulos y conductas. Al aceptar que los únicos datos pertinentes para elaborar una teoría psicológica son los datos de la conducta, el conductismo psicológico se auto-condena a un criterio de mentalidad tan laxo que casi cualquier cosa puede tener mente, a condición de dar una descripción apropiada de ella en términos de meros estímulos y conductas. Por ejemplo, se puede sostener en el ámbito del conductismo psicológico sin demasiados problemas que una lavadora *desea* lavar ropa en la medida en que, dadas ciertas condiciones equivalentes a un estímulo (e.g., que no esté descompuesta, que tenga una carga de ropa con jabón y agua, que esté conectada a la corriente eléctrica, que el interruptor esté en posición "Encendido", etc.), presenta una cierta conducta (i.e., lava la ropa), lo cual es completamente absurdo desde un punto de vista intuitivo.[10] Este hecho sencillo basta para echar por tierra las aspiraciones del conductismo psicológico de ser una buena teoría de la mente, aunque es importante

[10] Un conductista psicológico podría replicar que este ejemplo no constituye un buen argumento para rechazar su teoría puesto que, si bien la acción de la lavadora puede ser descrita en términos de estímulo y respuesta, carece de cualquier tipo de aprendizaje o de reforzamiento que se observa en organismos vivos como las palomas o los ratones. Sin embargo, cuando es estudiada detenidamente, se hace patente que la noción de reforzamiento es problemática para los conductistas psicológicos puesto que apela a entidades mentales como veremos más adelante.

notar que la propuesta teórica también adolece de otros serios errores y limitaciones que estudiaremos a continuación

Además del hecho de que el conductismo metodológico tiene un criterio de mentalidad demasiado laxo por lo cual casi todos los objetos –en particular los aparatos electrodomésticos– pueden ser descritos como entes con mente (a condición de presentar una descripción adecuada de ellos en términos de estímulos y conductas), hay otros problemas que parecen no tener ninguna solución dentro del marco de la teoría. En particular, existen dos situaciones que revelan que una de las dos versiones de la tesis básica del conductismo psicológico –el hecho de que los "estados mentales" no sean otra cosa más que conductas– es muy probablemente errónea. A continuación, describiremos brevemente estas situaciones.

Según los conductistas psicológicos, lo que habitualmente es denominado "dolor" no es otra cosa más que una serie de conductas (gemir, hacer muecas, llorar, gritar, etc.) producidas por ciertos estímulos que pueden ser daños en la dermis producidos por calor excesivo o por objetos punzo-cortantes, dientes cariados, infecciones de origen viral o bacterial en los tejidos del cuerpo, etc. Como una de las dos versiones de la tesis básica del conductismo psicológico es eminentemente restrictiva y reduccionista (el "dolor" no es otra cosa más que una serie de conductas), de ello se deduce que si una persona no muestra ninguna de las conductas típicas del dolor cuando tiene alguno de los estímulos antes mencionados, entonces no debe tener dolor alguno. La validez de esta tesis ha sido puesta en duda repetidamente tanto por experimentos mentales imaginados por los filósofos como por observaciones empíricas realizadas por los científicos (en especial por los médicos). En el caso de los filósofos, Putnam (1963) desarrolló un argumento destinado a rechazar el conductismo psicológico (aunque también una cierta versión de conductismo filosófico) que consiste en lo siguiente: imagina una sociedad similar a la nuestra en la cual es juzgado moralmente degradante expresar cualquier signo de dolor en cualquier momento que sea. En esta sociedad, aunque los niños pequeños expresan sus sensaciones de dolor con conductas como gemidos y llantos (al igual que los niños de nuestra sociedad), son educados desde su más tierna edad a reprimir cualquier síntoma visible de dolor de tal manera que, al llegar a la edad adulta, pueden padecer dolores

agónicos sin dejar de hablar pausadamente, sin respirar agitadamente y sin dejar de sonreír. Si un conductista psicológico se encontrase por casualidad en esta sociedad de "super-espartanos" y solamente pudiese observar a los individuos adultos de esta sociedad, sin duda concluiría que ninguno de los individuos de esta sociedad puede sentir dolor. Sin embargo, si un "super-espartano" es transportado de su sociedad a la nuestra y se le convence progresivamente de expresar sus dolores como nosotros lo hacemos (lo cual es una hipótesis bastante plausible), entonces la tesis básica del conductismo psicológico queda descartada. Como bien señala Putnam sobre este respecto, "hay algo claramente absurdo en la tesis de que uno no puede atribuir a estas personas la capacidad de sentir dolor".[11]

En el caso de los médicos, los argumentos usados para rechazar el conductismo psicológico surgieron de una serie de observaciones provenientes de ciertos errores médicos cometidos en la primera mitad del siglo XX. Estos errores consistieron en que ciertos médicos, profundamente influenciados por el conductismo psicológico, confundieron por desgracia las propiedades de una sustancia paralizante con las propiedades de una sustancia anestésica. Dennett hace en (1978b) un buen recuento de esta serie de errores médicos y deja que su lector saque sus conclusiones respecto a la validez del conductismo psicológico como teoría de la mente:

> El curare (...) es un paralizante que actúa directamente en todas las coyunturas neuromusculares (...) para producir una parálisis total y relajamiento de todos los músculos voluntarios. (...) En la década de los cuarenta, sin embargo, algunos médicos cayeron en el error de pensar que el curare era un anestésico general, y lo administraban como tal en cirugías mayores. Los pacientes se quedaban, por supuesto, tranquilos bajo el cuchillo, sin fruncir el ceño, crisparse o gemir, pero cuando los efectos del curare se disipaban, se

[11] Putnam, 1963 en Putnam, 1975, p. 333: "Yet there is a clear absurdity to the position that one cannot ascribe to these people a capacity for feeling pain."

quejaban amargamente de haber estado conscientes y padecido un dolor atroz, sintiendo cada corte del escalpelo, pero simplemente paralizados y sin poder expresar su sufrimiento.[12]

Ahora bien, es importante destacar que un conductista psicológico muy convencido de la validez de su teoría puede replicar tanto al argumento de Putnam como a las situaciones expuestas por Dennett que no constituyen refutaciones absolutas y definitivas del conductismo psicológico. Un conductista psicológico puede argüir que tanto las declaraciones del "super-espartano" transportado a nuestra sociedad como las del paciente operado con curare no son confiables porque ambas hacen referencia a entidades que no son datos de la conducta, por lo que no son empíricamente comprobables. Por otra parte, en el caso del paciente operado con curare, hay un período de tiempo entre la operación y el reporte del dolor que hace que el reporte sea incierto (recordemos que la memoria no es una facultad infalible y que muchas veces creemos recordar cosas que en realidad nunca sucedieron). Para refutar el conductismo psicológico de manera radical e indiscutible, es necesario refutarlo *desde adentro*, i.e., únicamente a partir de sus propios supuestos, sin apelar a nada externo.

La segunda situación problemática para el conductismo psicológico consiste en la inversa de la primera situación que hemos examinado a través de los escritos de Putnam y Dennett. En el caso de los pacientes operados con curare y de los "super-espartanos" (que constituyen la tercera objeción de peso contra el conductismo psicológico), tenemos estados mentales sin tener ninguna conducta visible. La inversa de esta situación consiste en tener ciertas conductas visibles sin tener estados mentales correspondientes (cuarta objeción). Esta segunda situación es moneda corriente en la vida cotidiana puesto que casi todos nosotros hemos simulado al menos una vez los síntomas de un dolor de muelas para no ir a la escuela o al trabajo. Casi todos hemos presentado en algún momento dado las conductas típicas que corresponden a un cierto "estado mental" sin estar realmente en ese estado mental. Esta situación, conocida como el

[12] Dennett, 1978a en Dennett, 1978b, p. 209: "Curare (...) is a paralytic that acts directly on the neuromuscular junctions (...) to produce total paralysis and limpness of all the voluntary muscles (...) In the 1940's, however some doctors fell under the misapprehension that curare was a general anesthetic, and they administered it as such for major surgery. The patients were, of course, quiet under the knife, and made not the slightest frown, twitch or moan, but when the effects of curare wore off, complained bitterly of having been completely conscious and in excruciating pain, feeling every scalpel stroke but simply paralyzed and unable to convey their distress "

argumento del mentiroso, constituye una refutación del conductismo psicológico desde adentro en la medida que parte solamente de aquello que los conductistas aceptan (los datos de la conducta) y muestra que, en ciertos casos, las conductas pueden no corresponder a nada, con lo cual la identidad establecida por los conductistas ("los estados mentales no son otra cosa que conductas") queda falseada.

La quinta objeción que se puede presentar contra el conductismo psicológico consiste en el hecho de que, en cualquiera de las dos interpretaciones de la tesis central que consideremos, la teoría no puede explicar los vínculos causales que parecen existir entre la "mente" y el cuerpo. Si consideramos la interpretación de la tesis central del conductismo que sostiene una identidad entre "estados mentales" y conductas, es evidente que no puede haber una relación causal porque toda relación causal requiere que sus causas sean distintas de sus efectos. Por lo tanto, el conductismo psicológico en esta interpretación no puede ser una teoría completa de la mente en tanto no puede dar cuenta de los vínculos causales entre la "mente" y el cuerpo por una razón de principio. Si consideramos la interpretación de la tesis central que sostiene que los "estados mentales" no existen, entonces tampoco podemos dar cuenta de los vínculos causales entre la "mente" y el cuerpo en la medida en que no se puede incluir en una relación causal algo que no existe.

Una sexta objeción central dirigida contra el conductismo psicológico parte del hecho que, al restringir la mentalidad a un conjunto de conductas generadas por ciertos estímulos, se deja de lado completamente toda reflexión sobre lo que ocurre entre los estímulos y las conductas. Para un conductista psicológico, no es relevante saber cuál es la composición de lo que hay en la "caja negra", e incluso no importa saber si hay realmente algo en la "caja negra" o si se encuentra vacía. Dichas preguntas y especulaciones apelan a datos distintos de la conducta, por lo cual no pueden ser empíricamente comprobables. Sin embargo, es un hecho comprobado que muchas veces un estímulo simple puede generar una conducta increíblemente rica y compleja. Intuitivamente, la única manera de explicar esto es suponiendo que en la "caja negra" ocurre un cierto procesamiento de la información contenida en el estímulo. Este procesamiento de la información supone a su vez algo que realice el procesamiento (que no puede ocurrir de manera gratuita). Este algo que realiza el procesamiento debe de tener una cierta estructura que le permite realizar dicho

procesamiento, por lo cual es necesario tener un conocimiento mínimo de su estructura para saber cómo se articulan ciertos estímulos con determinadas conductas como señala Chomsky:

> Uno esperaría naturalmente que la predicción de la conducta de un organismo (o máquina) complejo requiriera, además de información sobre las estimulaciones externas, conocimiento de la estructura interna del organismo, de las maneras a través de las cuales procesa la información de entrada y organiza su propia conducta.[13]

Como podemos apreciar claramente, la indiferencia de los conductistas psicológicos respecto a lo que ocurre en la "caja negra" deja de lado algo muy importante, una dimensión necesaria para explicar la articulación de estímulos simples con conductas complejas Por lo tanto, es claro que el conductismo psicológico no puede ser una buena teoría de la mente si deja de lado algo tan importante.

Finalmente, la séptima objeción (o, para hablar con mayor precisión, el séptimo grupo de objeciones) de peso en contra del conductismo metodológico consiste básicamente en el uso indebido de términos clave como *conducta, estímulo* y *reforzamiento* que sirven para designar en las obras de los conductistas (especialmente de Skinner) casi cualquier cosa, incluso las turbias "entidades mentales" como bien señala Chomsky. Por ejemplo, en la medida en que el término *estímulo* es habitualmente entendido como una parte del entorno bajo cuyo dominio se encuentran determinadas conductas, los conductistas tienen que resolver un serio problema que surge de la naturaleza de nuestra percepción. Aún cuando se ha demostrado que ciertos fotorreceptores en nuestros ojos son sensibles al color mientras que otros son sensibles a la forma, la percepción del mundo es en general de carácter holista: percibimos los objetos externos como unidades bien integradas más que como entidades con propiedades desconectadas e independientes unas de otras. Siendo esto cierto, ¿cómo puede saber un conductista a qué tipo de estímulo (o estímulos) está respondiendo el sujeto de un experimento cuando mira un objeto dado (e.g., una

[13] Chomsky, 1959 en Block, 1980, p. 49: "One would naturally expect that prediction of the behavior of a complex organism (or machine) would require, in addition to information about external stimulation, knowledge of the internal structure of the organism, the ways in which it processes input information and organizes its own behavior."

naranja)? ¿Cómo puede predecir con certeza el tipo de conductas que el sujeto habrá de tener respecto a ese objeto dado? Si el conductista responde que obtiene este conocimiento cuando el sujeto pronuncia oraciones del tipo "Veo algo redondo", "Veo una fruta" o "Veo algo anaranjado", entonces la noción de *estímulo*, que los conductistas presentaban como un término con distintas clases de referentes que podían ser estudiados a través de instrumentos y metodologías típicas de la ciencia, pierde toda objetividad como lo señala Chomsky:

> Los estímulos ya no son parte del mundo físico externo; son traídos de vuelta al interior del organismo. Identificamos los estímulos cuando escuchamos la respuesta. Es claro a partir de estos ejemplos –muy numerosos, por cierto– que hablar de estímulos controladores oculta simplemente una retirada hacia la psicología mentalista. No podemos predecir la conducta verbal en términos del entorno en la medida en que no sabemos cuáles son los estímulos hasta que la persona responde.[14]

Al igual que el concepto de *estímulo*, la noción de *conducta* tal como es entendida por los conductistas psicológicos presenta serios problemas. Intuitivamente, todos tenemos una idea más o menos vaga de las acciones que son consideradas como conductas (e.g., gritar, llorar, reír, sudar etc.). Todas estas acciones son completamente describibles en términos físicos y son visibles (hecho muy importante si tenemos en cuenta que los conductistas psicológicos consideraban vital que su teoría de la mente fuese públicamente comprobable). Sin embargo, las descripciones de las acciones en términos de conductas presentan un serio problema: no son capaces de dar cuenta de la dimensión intencional de los estados mentales. Esta dimensión intencional de los estados mentales es esencial puesto que permite diferenciar conductas que de otra manera serían indistinguibles, aunque tengan propósitos totalmente distintos. Para ver esto con mayor claridad, consideremos la situación siguiente. Tenemos dos laberintos A y B tales que en cada uno

[14] *Ibid.*, p. 52. "Stimuli are no longer part of the outside physical organism. We identify the stimulus when we hear the response. It is clear from such examples, which abound, that the talk of stimulus control simply disguises a complete retreat to mentalistic psychology. We cannot predict verbal behavior in terms of the stimuli, since we do not know what the current stimuli are until he responds."

de ellos hay un ratón. El diseño interno de ambos laberintos es idéntico y la posición inicial del ratón en cada laberinto es también la misma. El único punto de divergencia que existe entre A y B consiste en el hecho que hay un pedazo de queso en uno de los extremos del laberinto A y un gato en uno de los extremos del laberinto B (el queso y el gato se encuentran en extremos opuestos). Supongamos ahora que, cuando cada uno de los ratones es soltado en su laberinto respectivo, corre hacia el mismo rincón, siguiendo exactamente la misma ruta. Si un conductista psicológico observa lo que ocurre en los dos laberintos, no puede hacer otra cosa que señalar que cada uno de los ratones exhibe un tipo particular de conducta: la conducta del ratón en A apunta a encontrar el queso mientras que la conducta del ratón en B apunta a huir del gato. Es menester notar que el conductista psicológico puede distinguir las conductas de los dos ratones en tanto sabe qué estímulo determina cada conducta. Asumamos ahora que el psicólogo conductista puede observar la conducta de los ratones pero que, por medio de un mecanismo de ocultamiento instalado en cada laberinto, le está vedado contemplar tanto el queso en A como el gato en B. ¿Puede en este caso el conductista psicológico distinguir las dos conductas? Si los dos ratones siguen exactamente la misma ruta (como suponemos *ex hypothesi* que lo hacen), el conductista no puede sino decir que presentan el mismo tipo de conducta, por lo cual ambos deben de estar en el mismo "estado mental". Sin embargo, nosotros sabemos que las conductas de los ratones son distintas en tanto tienen objetivos distintos. Es razonable suponer además que los objetivos son distintos en tanto los "estados mentales" de los ratones son distintos: uno quiere comerse el queso mientras que el otro quiere alejarse del gato. Así pues, aun cuando pueden exhibir la misma conducta, es claro que dos organismos pueden tener "estados mentales" con dimensiones intencionales distintas. Esta situación muestra claramente que el hecho de que las conductas sean observables –hecho sobre el cual se basa el supuesto carácter científico y empíricamente comprobable del conductismo psicológico– no basta para distinguirlas entre sí: es necesario añadir para ello la dimensión intencional, una dimensión que el conductismo psicológico es incapaz de recuperar. Podemos constatar entonces que la fundación de la psicología sobre la noción de *conducta* no garantiza de ningún modo (como los conductistas psicológicos parecen creerlo) que la disciplina tenga un carácter cuantitativo y experimental.

Finalmente, es posible apreciar que la noción de *reforzamiento* también presenta serios problemas en tanto está íntimamente ligada a la noción de *conducta*. Consideremos por un momento las acciones que consisten en enseñar a un niño a no comerse las uñas. En este caso, se puede apreciar claramente que un estímulo que tienda al reforzamiento (e.g., poner jugo de chile en los dedos del niño) no necesita producir ningún tipo de conducta; en efecto, si el niño no se mete el dedo a la boca, no hay conducta alguna que pueda ser asociada con el reforzamiento. De hecho, en ciertos casos no se necesita siquiera que haya estímulos actuales para que se produzca un reforzamiento (e.g., según Skinner, un escritor puede reforzar su tendencia a escribir mejor si piensa en sus lectores potenciales que seguramente nunca llegará a conocer).[15] De igual manera, al recordar el niño la sensación picante que experimentó al meter los dedos por primera vez a su boca, evitará hacerlo de nuevo. Sin embargo, estos dos ejemplos de reforzamiento presuponen por parte tanto del escritor como del niño la existencia de actividades mentales subjetivas no exhibidas en la conducta: en el caso del escritor el reforzamiento presupone la capacidad de imaginar sus lectores futuros mientras que en el caso del niño el reforzamiento presupone la capacidad de recordar. Podemos apreciar entonces que el término *reforzamiento*, que también aspiraba junto con el término *estímulo* a ser considerado como un término teórico cuyos referentes podrían ser estudiados con los instrumentos de la ciencia, pierde toda objetividad en tanto puede ser aplicado a casi cualquier cosa, incluso a estados y procesos mentales. De hecho, Chomsky sugiere que, así como en el caso de los estímulos, hablar de reforzamientos oculta una retirada de los conductistas hacia la psicología mentalista de la cual buscan constantemente deslindarse:

> La frase "X es reforzado por Y" (estímulo, estado de cosas, suceso, etc.) es usada como un disfraz para "X quiere Y", "A X le gusta Y, "X desea que fuera el caso que Y", etc. Invocar el término reforzamiento no tiene fuerza explicativa alguna, y cualquier idea de que esta

[15] Skinner, 1959, p. 206: "The verbal behavior [of a writer] may reach over centuries to thousands of listeners or readers at the same time. The writer may not be reinforced often or immediately, but his net reinforcement may be great."

paráfrasis introduce alguna luz u objetividad en la descripción del desear, del gustar, etc., es un serio engaño.[16]

Resumiendo lo que hemos visto hasta ahora, el conductismo metodológico tiene un gran mérito que consiste en haber sacudido los cimientos de la tradición dualista cartesiana en plena decadencia (al no poder explicar la naturaleza de los mecanismos de interacción entre la mente y el cuerpo) para hacer de la psicología una ciencia cuantitativa, que pudiese progresar y comprobar sus hipótesis y sus supuestos por medio de la experimentación. Sin embargo, en tanto que teoría de la mente, el conductismo psicológico estaba condenado irremediablemente al fracaso por varias razones: eliminaba todo vínculo causal entre los estados mentales y las conductas en cualquiera de sus interpretaciones (si A es idéntico a B, entonces A no puede ser la causa de B o si A no existe, entonces no puede ser causa de B), no podía dar cuenta de la intencionalidad de los estados mentales (como hemos visto en el ejemplo de los ratones) y sus términos teóricos principales como *estímulo*, *conducta* y *reforzamiento* que pretendían ser objetivos y empíricamente comprobables eran tan vagos en realidad que se podían aplicar casi a cualquiera cosa (incluso hacían referencia a las dudosas entidades mentales de las que se querían desembarazar). Ahora bien, es importante recordar que, paralelamente al desarrollo del conductismo psicológico, también se desarrolló una teoría de la mente conocida como conductismo filosófico. El estudio de dicha teoría será el objeto de estudio de la próxima sección.

I.2 El conductismo filosófico: Carnap y los positivistas lógicos.

Antes de comenzar el análisis del conductismo filosófico por sí solo, acaso convenga primero recordar las diferencias que lo distinguen del conductismo psicológico. La diferencia principal que existe entre ambas teorías es la siguiente: el conductismo psicológico sostiene que la tesis según la cual los estados

[16] Chomsky, 1959 en Block, 1980, p. 56: "The phrase "X is reinforced by Y" (stimulus, state of affairs, event, etc.) is being used as a cover term for "X wants Y", "X likes Y", "X wishes that Y were the case", etc. Invoking the term reinforcement has no explanatory force, and any idea that this paraphrase introduces any new clarity or objectivity into the description of wishing, liking, etc., is a serious delusion "

mentales no son otra cosa que conductas es una hipótesis científica tentativa postulada para dar cuenta de la naturaleza y de los procesos de la mente de manera *empíricamente comprobable*. Lo que aparece en bastardilla en la oración anterior es capital: el criterio decisivo enarbolado por los conductistas psicológicos para sostener su tesis central es que otorga a la psicología un carácter cualitativo y práctico en el cual el progreso sólo puede darse por medio de la experimentación. En cambio, el conductismo filosófico sostiene que la tesis según la cual los estados mentales no son otra cosa que conductas no es una hipótesis científica tentativa que sea necesario introducir para dar cuenta de ciertas observaciones, sino una verdad analítica *a priori* que no es demostrable y que debe ser postulada como el axioma básico a partir del cual es posible construir la psicología. [17] Esta concepción distinta de la misma tesis conductista debe explicarse en el contexto del desarrollo de otra corriente de pensamiento filosófico en la misma época, una corriente que ejerció una influencia decisiva sobre el conductismo filosófico: el positivismo lógico.

El positivismo lógico comenzó a desarrollarse en la década 1920-1930 a través de los escritos de algunos pensadores alemanes y austriacos que trabajaron sobre los mismos temas, primero de manera individual y luego sentando las bases de un grupo de discusión de ideas conocido posteriormente como "el círculo de Viena". Partiendo de una tesis expuesta por Wittgenstein según la cual "la totalidad de las proposiciones verdaderas es la ciencia natural entera"[18], el propósito central de los positivistas lógicos consistía en crear una sola ciencia natural universal a través de la unificación del conjunto de ciencias particulares conocidas hasta entonces. Esto sólo se podía llevar a cabo, según ellos, encontrando ciertos fundamentos (que son verdades analíticas *a priori*) a los cuales se pudiesen reducir las proposiciones de las ciencias particulares y que pudiesen resistir los ataques del escepticismo radical. Así pues, como

[17] La principal razón que presentan los conductistas filosóficos para argumentar que la tesis central de su teoría debe ser una verdad analítica *a priori* (i.e., una proposición que es verdadera únicamente en virtud de su estructura sintáctica), y no una mera verdad empírica (i.e., una proposición que sea verdadera en virtud del contenido que tiene, y de la adecuación de éste al mundo), reside en el siguiente pasaje (Carnap, 1988, p 199): "El contenido depende de los resultados de las ciencias de la realidad (...) Dado que son todavía discutibles los resultados de dichas ciencias, no se puede garantizar que la traducción al lenguaje de constitución sea siempre correcta." Si la tesis central fuese una verdad empírica (i.e., si dependiese para ser verdadera de la adecuación de su contenido con los hechos del mundo), entonces sencillamente no habría seguridad de que pudiésemos elaborar una auténtica ciencia universal de la naturaleza que tuviese como fundamentos verdades necesarias.

[18] Wittgenstein, 1973, p. 65 (Prop. 4.11)

podemos constatar, los positivistas compartían en el fondo la misma ambición que Descartes abrigó al redactar su obra. Ahora bien, existía un problema para realizar dicha unificación que consistía en el hecho de que las diversas ciencias particulares habían desarrollado cada una por su lado un lenguaje particular, muchas veces con términos tomados del lenguaje natural. El lenguaje natural era reconocido por los positivistas lógicos como un instrumento importante en la medida en que nos permite comunicar con nuestros semejantes.[19] Sin embargo, en el campo de la ciencia, dicho instrumento se torna inútil puesto que su ambigüedad inherente choca con los requerimientos de exactitud y precisión del estudio de la naturaleza de las cosas como bien señala Wittgenstein.[20] Dicha ambigüedad se reflejaba en muchas ocasiones, de acuerdo con los positivistas, en los lenguajes particulares de cada ciencia, por lo que la estrategia que consistía en amalgamarlos para construir a partir de ellos el lenguaje de la ciencia natural universal era por demás dudosa.

La solución que idearon los positivistas para resolver este problema consistió en proponer el uso de un "lenguaje sígnico que no use el mismo signo en símbolos distintos, ni use de igual manera signos que designen de modo diferente."[21] Este lenguaje sígnico, que Carnap posteriormente denominó *lenguaje de constitución*, tiene por objetivo sentar las bases de un sistema de constitución que ordene "en un sistema unitario los objetos de todas las ciencias de acuerdo con la reducibilidad de un objeto a otro."[22] Los positivistas lógicos pensaban que si todos los lenguajes de todas las ciencias particulares pudieran ser reducidos a un único lenguaje de constitución, la unificación de las diversas ciencias particulares en una ciencia natural universal sería sólo cuestión de tiempo.

Ahora bien, esta tesis positivista plantea la siguiente pregunta: ¿qué se requiere para que los diversos lenguajes de las diversas ciencias particulares puedan ser reducidos a un solo lenguaje de constitución? La respuesta que Carnap aporta a esta pregunta consiste en señalar que, aun cuando los diversos objetos que pertenecen a todas las ciencias particulares pueden presentarse de un número de maneras muy grande (casi inconmensurable), "resulta que para cada objeto hay un 'hecho básico', a saber: en todos los otros

[19] Esto es una perogrullada, pero no por ello deja de ser verdadero
[20] *Ibid.*, p. 41 (Prop. 3.323 y 3.324)
[21] *Ibid*, p. 43 (Prop. 3.325)
[22] Carnap, 1988, p. 88 (§47)

hechos en que se presenta el objeto, éste se presenta solamente dentro del marco de dicho hecho básico."[23] La tarea del filósofo consiste entonces, según los positivistas lógicos, en determinar el "hecho básico" de cada uno de los objetos de las ciencias particulares, de tal manera que se pueda ulteriormente construir el lenguaje de constitución. Ahora bien, esta tarea plantea la siguiente pregunta: ¿de qué manera se puede determinar el "hecho básico" de un objeto particular? La respuesta aportada por Carnap a esta pregunta consiste en lo siguiente: el modo de presentación de un objeto particular es un "hecho básico" cuando se puede dar una condición necesaria y suficiente para que este "hecho básico" sea precisamente lo que es. Esto constituye el corazón mismo del programa positivista: la ambición de proveer para cada objeto particular de una ciencia las condiciones necesarias y suficientes para que el "hecho básico" de ese objeto sea lo que es, y no otra cosa. Puesto de otra manera, la proposición que exprese estas condiciones necesarias y suficientes debe de constituir imperativamente una verdad analítica *a priori*. De no ser este el caso, se corre el riesgo de que el "hecho básico" no sea realmente un "hecho básico", lo cual echa por tierra toda esperanza de llegar a construir un lenguaje de constitución para los positivistas lógicos. Sólo cuando se tiene una proposición que exprese el "hecho básico" de un objeto perteneciente a una ciencia particular y otra que exprese las condiciones necesarias y suficientes de este "hecho básico" se puede integrar el objeto en el lenguaje de constitución o, como Carnap también dice, se puede constituir el objeto.

El ejemplo que pone Carnap para ilustrar el proceso de constitución de los objetos de las diversas ciencias particulares consiste en lo siguiente: primero plantea el equilibrio de temperatura como uno de los objetos pertenecientes a una ciencia particular (la termodinámica). En segunda instancia, señala que el "hecho básico" de este objeto es: "*x* tiene respecto a *y* un equilibrio de temperatura". En tercera instancia, explica cómo logra determinar que este modo de presentación del objeto es realmente un "hecho básico": esto se realiza a través de la formulación de una proposición que expresa las condiciones necesarias y suficientes para que dicho "hecho básico" sea lo que es, y no otra cosa. Esta proposición es la siguiente: "si los cuerpos *x* y *y* son puestos en contacto espacial (directo o indirecto), no muestran un aumento ni una

[23] *Ibid.*, p. 90 (§48)

disminución en la temperatura" Para Carnap, la prueba decisiva de que esta proposición expresa las condiciones necesarias y suficientes del "hecho básico" expresado en la primera proposición consiste en que ambas proposiciones son coextensionales. Cuando tenemos estas dos proposiciones, se puede realizar el proceso de integración del objeto "equilibrio de temperatura" al lenguaje de constitución definiéndolo por medio de las dos proposiciones de la manera siguiente: "llamamos 'equilibrio de temperatura' a la relación que hay *x* y *y*, la cual se caracteriza en que los cuerpos *x* y *y*, si son puestos en contacto espacial (directo o indirecto), no muestran un aumento ni una disminución en la temperatura."

Como podemos visiblemente constituir los objetos de ciencias como la biología, la química, la geología o incluso la economía, la pregunta que se plantea entonces es la siguiente: ¿podemos hacer lo propio con los objetos de la psicología? Si se considera que la psicología es una disciplina de rasgos eminentemente cualitativos y abstractos (como lo fue hasta la segunda mitad del siglo XIX), es evidente que el proceso de constitución de los objetos no puede ser aplicado con éxito. Para apreciar esta imposibilidad claramente, basta con imaginar lo que puede ser el "hecho básico" de un dolor para la psicología tradicional, marcada por el sello del PEF. Según la psicología tradicional, el dolor se define principalmente por la manera en como es percibido por un sujeto S (i.e., por su "dolorosidad"), lo que sugiere que el "hecho básico" del dolor debe de ser de carácter subjetivo. Sin embargo, esto automáticamente excluye que otras personas distintas del sujeto S puedan sentir dolor puesto que los dolores que dicen sentir no se presentan dentro del "hecho básico" que es privativo de S. En cambio, si se considera (como lo hacen Carnap y los demás positivistas lógicos) que la psicología es una ciencia empírica igual que las demás, una ciencia "sin más parentesco con la filosofía que cualquier otra ciencia natural"[24], entonces el proceso de constitución de sus objetos se vuelve teóricamente realizable.

Aun cuando los objetos de la psicología (i.e., los objetos mentales) puedan en teoría ser constituidos a primera vista con base en la noción de psicología sostenida por los positivistas, existe una pregunta que es menester responder antes de poder llevar realmente a cabo el proceso de constitución: ¿cuáles son los "hechos básicos" de los objetos de la psicología científica? Acerca de esta interrogante, Carnap destaca

[24] Wittgenstein, *op. cit* , p. 65 (Prop. 4.1121)

que "en la psicología todavía no hay unanimidad respecto a los principios fundamentales que deben servir como guía de la investigación."[25] Esta carencia de unanimidad se debe a que existe, en el contexto del positivismo lógico, una tensión entre dos tendencias opuestas, dos tendencias que son elementales para el positivismo lógico como bien ha señalado uno de sus más célebres críticos, Hilary Putnam.[26] Por un lado, en la medida en que los positivistas se dicen herederos del pensamiento empirista británico de los siglos XVII y XVIII, tienden a señalar que los únicos "hechos básicos" a partir de los cuales se pueden construir teorías sobre la naturaleza de las cosas son los "sense-data" o, para ser más exacto, las experiencias sensoriales presentes en una persona, también conocidas como "vivencias personales". Sin embargo, los positivistas reconocen, por otro lado, que la gran mayoría de los experimentos llevados a cabo en el ámbito de las diversas ciencias naturales (química, física, biología, etc.) sugieren abiertamente que los "hechos básicos" a partir de los cuales se deben de construir las teorías son hechos puramente materiales. En la primera etapa del positivismo lógico, cuya expresión más depurada es sin duda *Die logische Aufbau der Welt* (1928), se impuso parcialmente la tendencia empirista (que preconizaba una reducción de lo mental a lo fenoménico) a la tendencia materialista (que preconizaba una reducción de lo mental a lo material). ¿Por qué ocurrió esto? Para contestar esta pregunta, tenemos que recordar que, aunque Carnap admitía que los objetos físicos eran en teoría reducibles a los mentales y viceversa,[27] por lo cual era posible en principio sentar las bases del sistema de constitución tanto en lo mental como en lo material, también se mostraba ligeramente escéptico respecto a la posibilidad de sentar en la práctica las bases del sistema de constitución en lo material (i.e., de legitimar teorías materialistas como el conductismo psicológico) puesto que "es problemático que esté justificada la pretensión del conductismo de que, mediante esa ordenación de los objetos, se pueda reproducir adecuada y precisamente la relación epistemológica"[28] que va de las cosas como las percibimos (clara y distintamente) a las cosas como son. Como podemos apreciar, el Carnap de la etapa fenomenalista debe mucho a Descartes.

---

[25] Carnap, *op. cit.*, p. 235 (§132)
[26] Putnam, 1969 en Putnam, 1975, p 441
[27] Carnap, *op cit*, pp. 105-106 (§57)
[28] *Ibid.*, pp. 110 (§59)

Si bien las aprensiones de Carnap respecto a lo que denomina *conductismo* (que no es otra cosa que el conductismo psicológico de Watson y Skinner) estaban justificadas como hemos mostrado en la sección anterior, esto no implicaba que la postura fenomenalista adoptada en *Die logische Aufbau der Welt* fuese correcta. De hecho, se hizo patente muy rápido que presentaba problemas que no podía resolver. Por ejemplo, al asumir que los "hechos básicos" que debían servir para constituir los objetos de la psicología eran experiencias actuales o potenciales, el positivismo lógico era blanco de críticas severas puesto que "los absurdos del solipsismo metafísico eran equiparados con los absurdos de un lenguaje fenoménico que se encontraba condenado a ser 'privado', 'propio de soliloquios' e 'incomunicable'."[29] Además, el positivismo tampoco podía aportar una solución clara al problema de las otras mentes como el propio Carnap lo señala al declarar que "en principio no es posible observar en otra persona los 'hechos básicos' del problema psicofísico"[30], lo cual arroja una sombra de duda sobre la posibilidad de que las otras personas tengan realmente "hechos básicos".

Gracias a las observaciones críticas de Neurath y Popper, Carnap superó pronto la etapa fenomenalista de su pensamiento, lo cual lo condujo a poner las exigencias de la tendencia materialista del positivismo por encima de las exigencias de la tendencia empirista. Así pues, en esta nueva etapa de su pensamiento conocida como fisicalismo, la dirección de la reducción cambió en la medida en que era se consideraba que los "hechos básicos" de los estados mentales eran hechos físicos. Habiendo llegado a este punto, podemos plantearnos la siguiente pregunta sobre el tipo de reducción propuesta por Carnap. Aún si había señalado en su etapa fenomenalista dos tipos de reducción posibles (una reducción lingüística y una reducción ontológica), es por demás evidente que, en las dos etapas de su pensamiento, aboga por una reducción meramente lingüística (también llamada tesis del fisicalismo) como podemos apreciar en el siguiente pasaje:

[29] Feigl, 1950 en Feigl y Brodbeck, 1953, p. 615: "[And] the absurdities of a metaphysical solipsism were parallelled by the absurdities of a phenomenal language that was doomed to be 'private', 'soliloquistic', 'incommunicable'."
[30] Carnap, *op. cit*, p 309 (§167)

> A efecto de hacer la tesis del fisicalismo tan claramente comprensible como sea posible, me inclino a formularla como sigue: para cada estado mental existe un estado físico correspondiente; este último se relaciona con el primero mediante leyes universales. Por consiguiente, por cada oración psicológica, digamos O_1, existe una oración física correspondiente, digamos O_2, de tal manera que O_1 y O_2 son equipolentes de acuerdo con determinadas leyes válidas; ahora bien, sólo la segunda parte de esta formulación, es decir, la parte relativa a las oraciones O_1 y O_2 es correcta.[31]

Carnap se muestra muy reticente a realizar una reducción ontológica en la medida en que esto puede llevar según él a revivir viejos pseudo-problemas. De hecho, basta con que planteemos el problema de la reducción ontológica (aún si no podemos resolverlo) en los términos del modo material de hablar (i.e., asumiendo que cada cosa que nombramos existe realmente) para otorgar realidad a los estados mentales, lo cual puede llegar a constituir un pseudo-problema. Habiendo precisado esto, es también importante señalar que, para Carnap, proponer una solución al problema mente-cuerpo postulando los estados neuronales como soporte ontológico de los estados mentales es muy problemático (en esto, su reserva es muy similar a la de los conductistas psicológicos). ¿Cómo podemos explicar esta reserva? En mi opinión, existen dos razones principales por las cuales Carnap se muestra muy reticente a postular estados neuronales como soporte ontológico de los estados mentales. La primera razón parece ser la misma que animaba a los conductistas psicológicos: la obsesión de que todo aquello que se usa sea empíricamente comprobable. Carnap reconoce que, en el momento en el que escribe, aún no se dispone de conocimientos suficientes de la fisiología del cerebro, por lo que cualquier intento de constituir a partir de los estados neuronales los objetos de la psicología puede desembocar en fracaso. Así pues, la única opción que queda para constituir los objetos de las psiques ajenas es la que señalaron en un principio los conductistas psicológicos, i.e., sostener que los "hechos básicos" de los estados mentales no son otra cosa sino un conjunto de conductas observables que son producidas por ciertos estímulos:

[31] Carnap, 1998, p. 52.

> (...) no hay en absoluto psiques ajenas sin un cuerpo. Pues (dicho en el lenguaje de constitución:) la psique ajena solamente puede ser constituida con la mediación de un cuerpo, más precisamente, con la mediación de un cuerpo en que se presentan ciertos procesos ("los procesos expresivos"), los cuales son semejantes a los de mi propio cuerpo; (dicho en el lenguaje del realismo:) una psique ajena que no estuviese unida a un cuerpo a través del cual se exteriorizara, sería fundamentalmente incognoscible, y por eso no podría ser objeto de una proposición científica.[32]

Como podemos apreciar, esta primera razón tiene esencialmente motivaciones de índole práctica, no de principio. De hecho, Carnap deja abierta la posibilidad de que, cuando el conocimiento de la fisiología del cerebro haya progresado lo suficiente, puedan ser constituidos los estados mentales a partir de estados neuronales. En esto se aprecia su divergencia con los conductistas psicológicos que sostienen que, aun cuando pueda eventualmente establecerse una ciencia detallada del cerebro, ésta siempre será irrelevante para resolver el problema mente-cuerpo. La segunda razón de la reticencia de Carnap, por el contrario, tiene motivaciones de principio. Como Carnap busca siempre mantenerse en sus obras constantemente en el nivel del puro análisis del lenguaje (puesto que de otra manera se arriesgaría a caer en un pseudo-problema), debe de comprenderse que el lenguaje de constitución al cual busca reducir todas las oraciones de las ciencias particulares es un lenguaje neutro: no se está ontológicamente comprometido ni con el mentalismo ni con el materialismo. Sin embargo, si se constituyen los objetos de la psicología con base en estados neuronales como "hechos básicos", el lenguaje de constitución termina comprometido con una ontología de tipo materialista.[33] De hecho, Carnap va más lejos todavía, sugiriendo que ni el lenguaje de la psicología mentalista ni el lenguaje de la psicología materialista están comprometidos con una ontología de estados mentales o con una ontología de estados materiales. Ambos lenguajes son neutros (y,

[32] Carnap, 1988, pp. 249-250 (§140)
[33] Esto no es deseable para Carnap en (1928), pero posteriormente adopta esta posición en su etapa fisicalista. Cf. infra, nota 35.

por lo tanto, equivalentes) como señala Carnap en la medida en que denotan una sola y misma realidad, *cuando son utilizados en el modo formal de hablar*. El modo formal de hablar consiste en emplear oraciones como "En la psicología el término 'estados mentales' ocurre" en vez de oraciones del modo material como "La psicología versa acerca de estados mentales" (que inducen a la gente a pensar que realmente existen cosas como los estados mentales). Cuando empleamos exclusivamente el modo formal de hablar en la psicología, nos damos cuenta que el término "estados mentales" puede designar vivencias personales o conductas observables, puesto que son los únicos elementos que pueden constituir los "hechos básicos" de los estados mentales. Como el establecimiento del lenguaje de constitución con base en las vivencias personales es susceptible de recibir críticas en tanto las vivencias personales son eminentemente subjetivas, Carnap argumenta lo siguiente:

> Si en el lenguaje psicológico hay un predicado usado originalmente sólo para describir nuestros propios estados mentales mediante introspección, entonces el mero uso de este predicado al hablar o escribir constituye, de hecho, un síntoma de dicho estado. Así, el lenguaje psicológico no puede contener un predicado que designe un estado de cosas para el que no exista síntoma observable alguno.[34]

Esta estrategia de Carnap le permite mostrar que, aun cuando para el lenguaje de constitución se pueden tomar "hechos básicos" distintos, estos hechos básicos denotan en el fondo una misma realidad.[35] Si esto es cierto, ¿por qué la historia de la filosofía de la mente abunda en disputas entre mentalistas y materialistas? El problema según Carnap consiste en que la mayoría de los pensadores han utilizado ya sea el lenguaje mentalista o el lenguaje materialista en el modo material de hablar (que tiende a otorgar una dimensión ontológica a las cosas que expresa), olvidando que los lenguajes sólo proporcionan una

[34] Carnap, 1988, pp. 97-98 (§52)

[35] En (1928), Carnap sostiene que esta realidad es esencialmente neutral en la medida que se encuentra influido por el monismo neutral de Russell, y esto explica que exhiba dudas sobre si el lenguaje de constitución debe ser mentalista o fisicalista (aun cuando se incline más en esa época por un lenguaje de constitución mentalista o fenoménico). Más tarde, ya en su período fisicalista, sostiene que la realidad que denotan todos los lenguajes es una realidad física, por lo cual el lenguaje de constitución debe de ser un lenguaje cuyos "hechos básicos" sean físicos.

visión particular del mundo. Los lenguajes son sólo maneras distintas de recortar la realidad. Así pues, si logramos separar la ontología de la epistemología, habremos dado, de acuerdo con Carnap, el primer gran paso en el camino a la solución del problema mente-cuerpo. Esta separación sólo puede ser realizada a través de la eliminación, no del modo material del hablar, sino de las implicaciones ontológicas de este modo de hablar puesto que son ellas las que nos sumergen en los pseudo-problemas:

> El conductismo lógico no sostiene que las mentes, los sentimientos, los complejos de inferioridad, las acciones voluntarias, etc., no existan ni que su existencia sea al menos dudosa. Hace hincapié en que la misma pregunta sobre la existencia de estos constructos psicológicos es ya un pseudo-problema, puesto que dichas nociones aparecen en su "uso legítimo" sólo como abreviaciones en oraciones fisicalistas.[36]

Como hemos podido constatar en esta breve exposición, la tesis esencial del conductismo filosófico parece apuntar hacia una reducción de los objetos (i.e. de las proposiciones) de la psicología mentalista a ciertos "hechos básicos" (que son principalmente proposiciones sobre conductas visibles) a través de una reducción de los términos psicológicos a los términos del lenguaje de constitución. Ahora bien, una de las condiciones principales para que pueda haber una reducción del lenguaje de la psicología mentalista al lenguaje de constitución (o lenguaje físico) a partir del cual se puede construir la ciencia universal de la naturaleza consiste en lo siguiente:

> Las proposiciones acerca del objeto puedan ser transformadas en proposiciones acerca de los objetos básicos del sistema, o transformadas en proposiciones acerca de los objetos previamente constituidos. Es decir, que hay que establecer una regla que haga posible

[36] Hempel, 1935 en Block, 1980, p. 20: "Logical behaviorism claims neither that minds, feelings, inferiority complexes, voluntary actions, etc., do not exist, nor that their existence is in the least doubtful. It insists that the very question as to whether these psychological constructs really exist is already a pseudo-problem, since these notions in their legitimate use appear only as abbreviations in physicalistic statements."

eliminar el nombre del nuevo objeto en todas las proposiciones en que se pueda presentar. En otras palabras, hay que establecer una definición del nombre del objeto.[37]

La necesidad de los positivistas de desarrollar definiciones para poder traducir las proposiciones de un lenguaje a otro constituye una muestra clara de su postura verificacionista, según la cual el significado de una oración se agota en las condiciones que deben ser verificadas si consideramos que verdadera.[38] Esta condición constituye la primera premisa de un argumento desarrollado por Hempel en (1935) y citado por Kim en (1998) para mostrar la validez del conductismo psicológico.

(1) El significado de una proposición está únicamente dado por las condiciones que verifican dicha proposición.

(2) Si una proposición tiene un contenido inter-subjetivo (i.e., un contenido que puede ser compartido por varias personas), entonces las condiciones que las verifican deben ser públicamente observables.

(3) Sólo los datos de la conducta y los fenómenos físicos son públicamente observables.

(4) En consecuencia, el contenido de cualquier enunciado psicológico debe poder ser determinado a través de proposiciones que contengan únicamente datos de la conducta y descripciones de fenómenos físicos. Estas proposiciones deben ser verdaderas si el contenido del enunciado psicológico en cuestión es considerado verdadero también.

A medida que el positivismo lógico evolucionó a lo largo de los años, se hizo patente para todos sus promotores y partidarios que las definiciones explícitas requeridas para la traducibilidad de un lenguaje al otro (definiciones que constituyen las condiciones de verificación de las proposiciones)[39] no podían ser

[37] Carnap, *op. cit.*, pp. 69-70 (§38)

[38] *Ibid.*, p. 294 (§161): "La información que se dé acerca de la esencia de un objeto [i.e., la definición de un objeto], o lo que es lo mismo, la información acerca de la referencia del signo de un objeto consiste en la información acerca de los criterios de verdad que valen para aquellos enunciados en que puede aparecer el signo para ese objeto."

[39] Para poder apreciar esto claramente, consideremos la siguiente situación: imaginemos que alguien tiene un dolor y pronuncia la oración "Tengo un dolor". Un positivista lógico como Carnap transformaría esta oración O_1 en la siguiente oración O_2 "En un momento dado T y en una ubicación espacial dada E, la pronunciación de la oración

construidas, ni siquiera en un sólo caso, por razones de principio como veremos a continuación. En el caso concreto de la psicología, como bien señala Feigl, se intentó sucesivamente dar una definición explícita de las entidades mentales en términos de los estímulos, de los elementos directos de la percepción (o sense-data) y de las conductas. En el caso de los estímulos, el programa positivista fracasó en tanto es bastante obvio que "no podemos decir que una sensación de color es idéntica a la radiación (de una cierta intensidad y frecuencia) que bajo ciertas condiciones simplemente produce esta sensación."[40] En el caso de los elementos directos de la percepción, hemos visto previamente que no se pueden reducir las proposiciones de los diversos lenguajes de las ciencias naturales a un mero lenguaje de vivencias personales, so pena de cometer los mismos errores del solipsismo metafísico (e.g., carecer de toda posibilidad de comprobar públicamente la verdad de los enunciados que contienen vivencias personales). Finalmente, en el caso de las conductas, el programa positivista también fracasó en tanto que "los síntomas visibles y la conducta que indican una emoción como la ansiedad son confirmables y medibles en términos de temperatura de la piel, de secreciones endocrinas, de reflejos psicogalvánicos, de respuestas verbales, etc., pero no deben ser confundidos con la emoción misma."[41] Así pues, el conductismo filosófico comete una seria confusión entre los síntomas visibles de un estado mental y el carácter intrínseco del estado mental mismo, lo cual arroja serias dudas respecto a su pretensión de ser una buena teoría de la mente.

Si la traducción de las proposiciones del lenguaje de la psicología mentalista a las proposiciones del lenguaje físico de las conductas pudiese ser llevada a cabo sin alteración del valor de verdad ni del sentido de las proposiciones originales, no habría entonces problema alguno en sostener que todas las sensaciones

---

'Tengo un dolor' ocurre en un individuo." Ahora bien, cuando se da la definición de un dolor como aquello que tiende a provocar gemidos, llantos y otras conductas particulares, al mismo tiempo estamos dando las condiciones de verificación de la oración "Tengo un dolor" en la medida que "Tengo un dolor" sólo sería verificada según un positivista cuando el individuo se encontrase en un estado que tendiese a provocar gemidos, llantos y otras conductas particulares.

[40] Feigl, 1950 en Feigl y Brodbeck, *op. cit*, p. 620: "[Obviously] we cannot say that a color sensation is identical with the radiation (of a certain intensity and frequency pattern) which, under certain conditions merely elicits that sensation."

[41] *Ibid.*, p. 618: "The overt symptoms and behavior that indicate an emotion, like e.g., anxiety, are confirmable and measurable in terms of skin-temperature, endocrine secretions, psychogalvanic reflexes, verbal responses, etc., but must not be confused with the emotion itself."

de dolor pueden ser consideradas como conjuntos de conductas sin ningún problema. Pero esto no es el caso, puesto que todo proyecto de traducción choca con la imposibilidad de encontrar *algo* a lo que las proposiciones psicológicas puedan ser uniformemente traducidas. Además de esto, existe otro grave problema que muestra de manera clara las limitaciones del conductismo filosófico. Hemos señalado que algunos conductistas psicológicos puristas sólo admiten como conductas determinadas respuestas fisiológicas y meros movimientos corporales. Influenciados por este purismo, existen también algunos conductistas filosóficos que admiten solamente como conductas los mismos elementos. Ahora bien, si la tesis del conductismo filosófico es verdadera, entonces cualquier estado mental (en particular, la creencia de un individuo S de que la democracia representativa es el mejor régimen político inventado hasta ahora y su creencia de que los números reales que hay entre 0 y 10 son tantos como los números reales que hay entre 0 y 1) debe poder ser traducido sin alteración del sentido o del valor de verdad a un conjunto de proposiciones que contengan únicamente datos de la conducta y descripciones de fenómenos físicos. ¿Cuál es entonces la respuesta fisiológica o la combinación de meros movimientos corporales que puede traducir cualquiera de estas dos creencias? Una de las propuestas sostenidas por los conductistas psicológicos consiste en el hecho de asumir que, cuando se le pregunta a S si piensa que la democracia representativa es el mejor régimen político inventado hasta ahora y responde a esta pregunta pronunciando la oración "Si, creo que la democracia es el mejor régimen político inventado hasta ahora", su respuesta verbal constituye una conducta. Sin embargo, lo que implica su respuesta, como bien señala Kim, es que "S debe comprender lo que estas palabras significan y tratar de que sean comprendidas por su oyente con el significado propuesto."[42] Esto implica una serie de presupuestos psicológicos y una serie de acciones internas que no pueden ser recuperadas por medio de la definición de conductas como respuestas fisiológicas o meros movimientos corporales. Esta definición deja fuera algo capital para poder entender el comportamiento de S, por lo cual el conductismo filosófico parece no ser una buena teoría de la mente.

[42] Kim, 1998, p 33: "S must understand what these words mean and intend them to be understood by his or her hearer with the intended meaning."

Frente a la imposibilidad patente de establecer las definiciones relevantes para poder reducir las proposiciones del lenguaje de la psicología mentalista a las proposiciones del lenguaje de la ciencia universal de la naturaleza, algunos positivistas lógicos como Feigl sugirieron ciertas soluciones que, al ser analizadas con detenimiento, revelan los gérmenes de la siguiente etapa del pensamiento materialista en el siglo XX: la teoría de la identidad mente-cerebro. Sin embargo, antes de abordar el estudio de este punto de transición, acaso sea conveniente analizar brevemente la obra clásica de Ryle, *The Concept of Mind* (1949), que aun cuando es considerada hoy en día como la expresión más depurada del conductismo analítico, posee rasgos únicos que la distinguen de las obras de los positivistas como Carnap. Así pues, hemos dejado deliberadamente el pensamiento de Ryle de lado hasta ahora en la medida en que no podía ser tratado de manera paralela al positivismo lógico. Las razones por las cuales Ryle constituye un caso separado son varias y merecen una breve mención. En primera instancia, *The Concept of Mind* apareció en una época relativamente tardía en la cual tanto el conductismo filosófico como el positivismo lógico empezaban a declinar a causa de una serie de críticas virulentas, lo cual motivó a Ryle a dar un tono más mesurado a su exposición (e.g., remplazando el concepto de conducta por el concepto de disposición de conducta). En segunda instancia, los positivistas lógicos tenían la ambición gigantesca de elaborar un lenguaje de constitución al cual pudiesen ser reducidos todos los términos de las ciencias empíricas como la física, la química, la biología, etc. Si bien los términos de la psicología mentalista debían ser también reducidos en su opinión al lenguaje de constitución, esto no constituía sino una muy pequeña parte de su programa general. En cambio, Ryle tenía ambiciones mucho más modestas. No planteaba la creación de una ciencia universal de la naturaleza, sino simplemente la elaboración de una teoría completa de la mente por medio de una reformulación del problema central de la filosofía de la mente (¿qué es la mente?) en términos distintos. La realización de un minucioso análisis lingüístico en este problema central revelaría, de acuerdo con Ryle, que el problema surge básicamente de una confusión (llamada por Ryle "el mito de Descartes") y que, cuando es planteado de manera correcta, sencillamente se desvanece.

I.3 El conductismo lógico: el caso de Ryle

Al igual que los conductistas psicológicos y que los conductistas filósoficos, Ryle tiene la convicción de que hablar de estados mentales es totalmente erróneo. Sin embargo, más que propugnar por una reducción del lenguaje de la psicología mentalista a un lenguaje de constitución a la manera de Carnap, Ryle parece sugerir que hay que erradicar los términos mentales (i.e., ni siquiera usarlos como abreviaciones cómodas de las descripciones del lenguaje físico) puesto que son sólo ficciones dañinas que no denotan nada real y engañan a las personas. Como el mito del "fantasma en la máquina" no es otra cosa que una ficción (y ni siquiera una ficción conveniente), hablar de mentes para Ryle es prácticamente equivalente a hablar de unicornios y sirenas. Ahora bien, si hemos de entender la razón por la cual mucha gente vivió durante siglos bajo el dominio de este mito, debemos de entender cómo se originó para después poder enmendar nuestras creencias. De acuerdo con la opinión de Ryle, el origen del mito reside en un error categorial realizado por la mayoría de la gente que, al usar el modo material de hablar, otorga una dimensión ontológica a cosas que en realidad no existen. Para que sus lectores puedan entender bien este concepto de error categorial, Ryle presenta el siguiente ejemplo:

> A un extranjero que visita Oxford o Cambridge por primera vez se le muestra un cierto número de colegios, librerías, campos de juego, museos, departamentos científicos y oficinas administrativas. Pero entonces pregunta: "¿Dónde está la universidad? He visto dónde los miembros de los colegios viven, dónde está el Registro, donde los científicos experimentan y todo el resto. Pero no he visto la universidad en la que residen y trabajan los miembros de vuestra universidad." Se le tiene que explicar entonces que la universidad no es otra institución colateral, otra contraparte adicional a los colegios, laboratorios y oficinas que ha visto. *La universidad es sólo la manera en la cual está organizado todo lo que ha visto previamente* [43]

[43] Ryle, 1949, p 16: "A foreigner visiting Oxford or Cambridge for the first time is shown a number of colleges, libraries, playing fields, museums, scientific departments and administrative offices He then asks 'But where is the

A través de este ejemplo, Ryle sugiere que la gente que está convencida que la mente existe y que se pregunta dónde se halla comete el mismo error que el extranjero. La mente no es una entidad más allá de los síntomas visibles (i.e., estímulos y conductas) a partir de los cuales se deduce tradicionalmente su existencia: lo que denominamos tradicionalmente "mente" es sólo la manera en la cual están organizados estos síntomas visibles.

A continuación, expondremos el rasgo central que hace Ryle merezca un lugar especial en el análisis del conductismo. En los inicios del conductismo psicológico, sus expositores más importantes (Watson, en especial) sostenían que la mente de un sujeto no era otra cosa más que el conjunto de conductas actuales, i.e., aquello que se podía ver y constatar de manera patente. En cambio, en la época en que Ryle comenzó a redactar su obra, el argumento anticonductista según el cual había personas que sentían dolor aun cuando no podían expresarlo a través de ninguna conducta visible, comenzaba a cobrar vigor. Para evitar este tipo de críticas, Ryle propuso que la mente fuese considerada, no sólo como el conjunto de conductas actuales, sino como el conjunto de conductas potenciales, o de *disposiciones de conducta*. Esta ingeniosa alteración de la tesis conductista original tenía una ventaja esencial. Permitía dar cuenta de los estados mentales aun cuando no hubiera conductas visibles actuales. De la misma manera en que podemos decir que un vidrio es frágil sin tener que comprobar su fragilidad al romperlo (basta con saber que tiene tendencia a romperse dadas ciertas condiciones), o que la sal es soluble en agua sin necesidad de echar una cucharada en un vaso de agua, podemos decir que un hombre "quiere" fumar un cigarrillo sin necesidad de verlo fumar un cigarrillo en un momento dado como bien señala Ryle al declarar que:

University? I have seen where the members of the Colleges live, where the Registrar works, where the scientists experiment and the rest. But I have not yet seen the University in which reside and work the members of your University.' It has then to be explained to him that the University is not another collateral institution, some ulterior counterpart to the colleges, laboratories and offices he has seen. *The University is just the way in which all that he has already seen is organized.*" La bastardilla es mía.

> El que yo sea un fumador habitual no implica que esté fumando en este u otro momento; es mi tendencia permanente a fumar cuando no estoy comiendo, durmiendo, dando una conferencia o asistiendo a un funeral, y cuando no he estado fumando recientemente.[44]

Desde un punto de vista intuitivo, resulta simple constatar que la posición adoptada por Ryle es mucho más inclusiva y poderosa que la posición del conductismo psicológico tradicional que es demasiado simplista y restrictiva. Según los conductistas psicológicos, el hecho que Pedro quiera comer un helado sólo puede ser comprobado si se le ve comprando en una heladería un helado y después llevárselo a la boca. En cambio, el criterio de verdad de esta aseveración para Ryle consiste en la verificación de una serie de oraciones condicionales contrafácticas del tipo (1) "Si le preguntara a Pedro si comerá una manzana o un helado, respondería que comerá el helado", (2) "Si Pedro estuviese en un mercado con una heladería y una frutería, iría a la heladería", (3) "Si Pedro tuviese un billete de veinte pesos, compraría un helado", etc. En otros términos, para un conductista psicológico, basta con ver en el mundo actual que Pedro compra un helado y se lo lleva a la boca para decir que quiere comer un helado. Para Ryle, es necesario que, en cada mundo posible donde Pedro puede comer un helado (i.e., en todo mundo posible donde existen tanto Pedro como los helados y en que los helados están al alcance de Pedro), Pedro se compre un helado y se lo lleve a la boca para poder decir que comerá un helado.

En la medida en que Ryle apela a las disposiciones de conducta para dar cuenta de los estados mentales prescindiendo del lenguaje mentalista y, por ende, en la medida en que necesita de la verificación de las oraciones contrafácticas relacionadas con las conductas relevantes, la versión del conductismo que propone es mucho más sólida y coherente que la de los conductistas metodológicos y la de los positivistas lógicos. Sin embargo, existen por lo menos tres problemas serios en la teoría de Ryle que no han podido ser resueltos hasta ahora. El primero consiste en el hecho de que no se puede dar una lista exhaustiva de las oraciones contrafácticas que deben ser verificadas para poder dar cuenta de un estado mental dado.

[44] Ryle, *op cit.*, p. 43: "My being an habitual smoker does not entail that I am at this or that moment smoking; it is my permanent proneness to smoke when I am not eating, sleeping, lecturing or attending funerals, and have not quite recently been smoking."

Aún si nos concentráramos en listar el conjunto de oraciones contrafácticas que definen una sola disposición de conducta de una sola persona, dejando de lado todas las demás, la vida del universo no bastaría. Por lo tanto, cualquier término del lenguaje mentalista (e.g., cualquier término que denote un estado mental) no puede ser bien definido en los términos del lenguaje de las disposiciones de conducta en tanto resulta imposible dar las condiciones necesarias y suficientes para reducir la oración "Pedro quiere comer un helado" a "Pedro comerá un helado". Ahora bien, un conductista puede responder que este problema es solamente de índole práctica. Puede argumentar que si alguien estuviese en condiciones ideales (e.g., si fuese eterno como Dios), entonces sin duda sería posible para dicha persona enumerar todas y cada una de las oraciones contrafácticas que definen un estado mental dado. Es en este punto donde se presenta el segundo problema que consiste en lo siguiente: asumiendo que se puedan enumerar todas las oraciones contrafácticas que definen según Ryle un estado mental, dicha enumeración es incapaz de recuperar la intencionalidad del estado mental original. Para observar esto, imaginemos por un momento que Ryle alcanza a enumerar todos y cada uno de los condicionales contrafácticos que definen el estado mental de María expresado en la proposición "María quiere ir de vacaciones a Hawai". Estos condicionales sólo expresan lo que María haría en un cierto momento, dadas tales o cuales condiciones. Sin embargo, hay una posibilidad de que María quiera engañarnos para hacernos creer que quiere ir a Hawai, cuando en realidad no desea esto. Puede darse el caso de que, si ponemos a María en condiciones de verificar todos y cada uno de los condicionales contrafácticos que definen según Ryle su deseo de ir a Hawai, María los verifique todos. Por ejemplo, en el caso del condicional contrafáctico "Si a María se le preguntase dónde iría de vacaciones, respondería que le gustaría ir a Hawai", se pone a María en condiciones de verificarlo preguntándole dónde iría de vacaciones. Sin embargo, aún si María verifica todos los condicionales, esto no implica que realmente *quiera* ir a Hawai. Así pues, vemos que el argumento del mentiroso, que tanto daño causó al conductismo psicológico, también se aplica al conductismo de Ryle, lo que arroja serias dudas serias dudas sobre la validez de la teoría de la mente que propone. El listado de oraciones contrafácticas sobre María y la verificación subsecuente de las mismas no sirve para determinar si hay realmente una dimensión intencional en ella o cuál es esta dimensión intencional (si María desea ir a Hawai o si desea engañarnos haciéndonos creer que desea ir a Hawai) en

la medida en que, aun cuando asumamos que nos encontramos en condiciones ideales, siempre existe la posibilidad de que cuando todos los condicionales contrafácticos hayan sido verificados, María nos diga que estuvo mintiendo todo el tiempo y que no quería ir a Hawai. Ahora bien, como hemos visto anteriormente a través del ejemplo de los dos ratones en los dos laberintos, la dimensión intencional es fundamental para poder identificar un estado mental. Por lo tanto, como las disposiciones propuestas por Ryle no nos permiten (y este es un problema de principio, no meramente de índole práctica) recuperar la intencionalidad de los estados mentales, es razonable suponer que su propuesta no puede ser una buena teoría de la mente.

El tercer problema parte del razonamiento siguiente: asumiendo que pudiese darse un listado completo de las oraciones contrafácticas que definen una disposición de conducta dada, cada oración contrafáctica exige a su vez que se especifiquen las condiciones bajo las cuales es verdadera. Retomando el ejemplo presentado anteriormente, (1) será verdadera en el caso en que Pedro *piense* que el helado no está envenenado, (2) será verdadera en el caso en que Pedro *sepa* dónde se encuentra ubicada la heladería en el mercado, (3) será verdadera en el caso en que Pedro *crea* que no tiene necesidad de comprar algunas piezas de pan dulce para cenar puesto que su hermana ya lo hizo, etc. Esto es en extremo problemático en tanto lesiona las bases mismas del conductismo como bien señala Churchland al declarar que:

> Reparar cada condicional añadiendo la condición relevante sería reintroducir una serie de elementos mentales en la estructura de la definición, y ya no podríamos definir lo mental sólo en términos de circunstancias públicamente observables y de conducta.[45]

Como podemos notar, la reparación de cada condicional contrafáctico exige la introducción de una actitud proposicional (creer, saber, pensar, etc.) Como estas actitudes proposicionales no son otra cosa más que estados mentales, podemos constatar que la teoría conductista de Ryle se encuentra condenada de

[45]Churchland, 1981 en Churchland, 1989, pp. 24-25: "To repair each conditional by adding in the relevant qualification would be to reintroduce a series of mental elements into the business end of the definition, and we would no longer be defining the mental solely in terms of publicly observable circumstances and behavior."

antemano al fracaso puesto que padece del mismo tipo de error señalado por Chomsky[46] en el caso del conductismo psicológico: las disposiciones de conducta, al igual que los estímulos y los refuerzos, ocultan en última instancia actividades y procesos mentales.

Ahora bien, aunque por las razones antes mencionadas el conductismo de Ryle estuviese condenado al fracaso como teoría de la mente, es muy importante señalar varios hechos que constituyen contribuciones muy interesantes del pensamiento de Ryle, y que posteriormente recuperaron sus sucesores. Uno de los rasgos más interesantes de su pensamiento es su crítica al PEF que, según él, no es sino una aberración. Según él, pensar que las cosas son cómo las entendemos cuando las entendemos clara y distintamente lleva a la gente a pensar que puede conocer las cosas de manera indubitable e inmediata, lo cual es un error puesto que:

> No hay contradicción alguna en afirmar que alguien puede no reconocer los contenidos de su mente en lo que son. En efecto, es notorio que la gente lo hace constantemente Suponen equivocadamente que conocen cosas que son erróneas; se engañan sobre sus motivaciones (...) no saben qué están soñando, cuándo están soñando y algunas veces no están seguros de no estar soñando cuando están despiertos.[47]

Ahora bien, esta crítica de Ryle al PEF cartesiano no es del todo atinada en la medida que Descartes previó algunas de las objeciones que son presentadas por Ryle. Descartes tomó en cuenta, por ejemplo, la posibilidad de que una persona crea equivocadamente que está despierta cuando en realidad está dormida y soñando. En este tipo de situaciones, Descartes señaló que, aun cuando una persona puede equivocarse al pensar que ve un perro (cuando en realidad lo está soñando), no puede equivocarse al pensar que piensa que lo está viendo. Aun cuando pueda dudarse que los contenidos mentales correspondan a cosas reales

[46] Vid. supra pp. 24-27

[47] Ryle, *op. cit.*, p. 162: "There is no contradiction in asserting that someone might fail to recognize his frame of mind for what it is; indeed, it is notorious that people constantly do so. They mistakenly suppose themselves to know things which are actually false; they deceive themselves about their own motives (...) they do not know that they are dreaming, when they are dreaming, and sometimes they are not sure that they are not dreaming, when they are awake."

en el mundo, los contenidos mentales por sí solos así como el hecho que el sujeto los reconozca como propios (i.e., la auto-adscripción) son siempre indubitables para Descartes. Sin embargo, es importante notar que la tesis de la indubitabilidad de los contenidos mentales es algo que ciertas ciencias modernas han puesto en cuestión: las diversas corrientes de pensamiento psicoanalítico han establecido que existen ciertos estados mentales (especialmente estados mentales intencionales como creencias, deseos, etc.) de los cuales no somos conscientes. La vieja tesis cartesiana de la indubitabilidad de los contenidos mentales sólo parece sostenerse en el caso en que los contenidos mentales colapsan con el hecho que el sujeto los reconozca como propios (e.g., en el caso de los qualia en tanto que no se puede sentir dolor sin sentir el dolor como propio).[48] Ahora bien, donde Ryle sí tiene razón en criticar la propuesta cartesiana derivada del PEF según la cual todo pensamiento es consciente,[49] es en el caso de los estados mentales con contenido intencional como deseos, intenciones, creencias, etc, pero no en el caso de los estados mentales con contenido fenoménico. Así pues, es menester notar que, aun cuando la crítica de Ryle al PEF no es del todo atinada como hemos señalado en la medida que no se aplica a ciertos estados mentales, se encuentra parcialmente justificada en tanto sí se aplica a otros estados mentales, por lo que se torna imperativo para Ryle desarrollar un concepto de conciencia distinto al cartesiano.[50] Ryle halla este concepto alternativo de conciencia en el concepto de atención ("heed"), que es bastante más limitado, como veremos en las líneas siguientes. Al contrario de la percepción interna cartesiana que se da de manera continua, la atención sólo puede darse en ciertos momentos en la medida en que requiere de ciertas condiciones. La primera condición que requiere es que exista un cierto acto de voluntad por parte de la persona que presta atención puesto que la atención no es algo que ocurra naturalmente:

> Hay muchas cosas que no podemos hacer, o hacer bien, a menos que prestemos atención a las instrucciones apropiadas y oportunas, aun cuando nosotros somos los autores de estas

[48] Esto será examinado con mayor detenimiento en la tercera sección del tercer capítulo.
[49] AT, VII, 160 y AT, VIII, 7-8 citado en Wilson, 1990., p. 228
[50] Este hecho es esencial puesto que ha permitido que se desarrollen conceptos de conciencia distintos del concepto cartesiano, incluso entre adversarios del conductismo como Rosenthal (1986) y Block (1995) que en la actualidad proponen una división del concepto conciencia en dos conceptos distintos: una conciencia fenoménica de auto-adscripción (que corresponde al viejo cartesiano de conciencia) y una conciencia de acceso.

instrucciones. En tales casos, intentar hacer algo implica intentar darse las instrucciones correctas en el momento preciso e intentar seguirlas.[51]

La segunda condición que requiere la atención para poder darse en el caso de un individuo es una cierta preparación o disponibilidad ("readiness"). Si una persona hace algo con atención, ello implica que "hace lo que hace con preparación para hacer justamente eso en esta situación y está listo para hacer algunas otras cosas que tenga también que hacer."[52] Teniendo en cuenta estas dos condiciones, podemos realizar ahora un estudio detenido de las ventajas así como de los problemas que presenta el concepto de conciencia sostenido por Ryle. En el caso de las ventajas, quizás la más importante sea que la propuesta de Ryle deja un lugar importante para el inconsciente (o las cosas a las cuales no prestamos atención), con lo cual rechaza la tesis cartesiana de la continuidad ininterrumpida de la percepción interna Este rechazo es importante en tanto nos permite explicar ciertas situaciones que observamos cotidianamente como el hecho que muchos automovilistas experimentados cambien de velocidad al manejar automáticamente, sin ser conscientes de ello.

La segunda ventaja central del concepto de conciencia de Ryle consiste en el hecho que permite un acercamiento al estudio de la conciencia desde una perspectiva objetiva y públicamente comprobable, mientras que el concepto cartesiano de conciencia (e incluso el concepto de conciencia como conjunto de vivencias personales del propio Carnap en su etapa fenomenalista) sólo permitía un acercamiento desde una perspectiva meramente subjetiva. La atención, en cambio, puede ser estudiada de manera objetiva a través de las conductas que presenta un individuo cuando se prepara para realizar algo. Por ejemplo, podemos señalar que un individuo se encuentra atento al hecho de estar comiendo un helado en tanto su conducta nos revela que está preparado para no dejar que el helado derretido que escurre por el cucurucho ensucie sus dedos y/o que está preparado para no dejar que alguna mosca atraída por el olor del azúcar

[51] *Ibid.*, p. 144: "There are many things which we cannot do, or do well, unless we paid heed to appropriate and timely instructions, even when we ourselves have to be the authors of those instructions. In such cases, trying to do the thing involves both trying to give oneself the right instructions at the right time and trying to follow them."
[52] *Ibid.*, p. 147: "[He both] does what he does with readiness to do just that in just this situation and is ready to do some of whatever else he may be called on to do."

venga a posarse sobre su helado. Sin embargo, el concepto de conciencia de Ryle cuenta también con varios problemas que surgen tanto de aquello que el concepto deja sin explicar como de los rasgos que Ryle atribuye al concepto.

En el caso del primer tipo de problemas, es patente que el concepto de conciencia según Ryle no puede capturar las propiedades fenoménicas de los estados mentales. La atención no constituye un criterio lo bastante seguro para establecer una taxonomía de los estados mentales. Si bien es cierto que existen varios grados de atención posibles y que cada grado de atención revela cosas distintas (e.g., se requiere un grado de atención más elevado para percibir las sátiras anticlericales en *Las tentaciones de San Antonio* de Bosch que para percibir la visión de un mundo perverso al borde de la destrucción), resulta en extremo dudoso pensar que estos grados de atención permitan separar una sensación particular de rojo de otra sensación ligeramente distinta. Para esta tarea requerimos de una taxonomía que sólo las propiedades fenoménicas (o qualia) de los estados mentales parecen poder proveer.

En el caso del segundo tipo de problemas, es claro que, al hablar de los actos de voluntad y de la preparación como condiciones de la atención, Ryle no hace otra cosa más que replegarse de nueva cuenta hacia el ámbito de la psicología mentalista que tanto critica. Por ejemplo, si presto atención a cerrar las puertas de casa con llave cuando salgo, es porque no *quiero* que nadie entre en mi ausencia y *temo* que algún ladrón pueda introducirse y robar algo. Si presto atención a comprar un billete de lotería cada semana, es porque *espero* algún día ganar el premio mayor y *creo* que el dinero del premio me servirá para realizar con menor esfuerzo mis proyectos. Podemos ver entonces que la conciencia entendida como atención, en el contexto del conductismo analítico de Ryle, no hace otra cosa que ocultar en última instancia estados mentales como Chomsky lo señala a propósito del conductismo metodológico de Skinner. Así pues, podemos constatar que el análisis de la conciencia propuesto por Ryle adolece de serios errores puesto que "cada término mental será contextualmente explicado en términos de los otros (...) pero la mente *tout court* permanecerá *sui generis* y autónoma."[53] Toda explicación de la conciencia

[53] Lycan, 1987, pp. 6-7: "Each mental term will be contextually explained in terms of the others (..) but the mind *tout court* will remain *sui generis* and autonomous."

entendida como atención se encuentra entonces condenada a replegarse al ámbito de la psicología mentalista, lo cual condena de manera inapelable las pretensiones del conductismo analítico de convertirse en una buena teoría de mente, en tanto no puede dar cuenta de la conciencia.

Otro rasgo importante del pensamiento de Ryle consiste en el hecho siguiente: a la inversa de muchos de sus predecesores (principalmente los conductistas psicológicos, aunque también algunos conductistas filosóficos) que sólo admitían como conductas respuestas fisiológicas de los organismos (transpiración, salivación, aumento del ritmo cardíaco, de la presión arterial, etc.) y meros movimientos corporales, Ryle admitía también como conductas ciertas *acciones* que implicaban movimientos corporales (ir a la tienda a comprar queso, redactar una carta, jugar fútbol, llenar un cheque, etc.). Esta categoría de elementos eran rechazados como conductas por los conductistas puristas en tanto que, si bien involucraban movimientos corporales, involucraban también componentes psicológicos internos como bien señala Kim:

> Consideremos el acto de llenar un cheque. Sólo alguien con ciertas capacidades cognitivas, creencias, deseos y una comprensión de las instituciones sociales relevantes puede llenar un cheque. Se debe tener el deseo de hacer un pago y la creencia de que llenar un cheque es un medio dirigido hacia ese objetivo. Se debe tener también alguna comprensión del intercambio de dinero por bienes y servicios y del banco como institución.[54]

Algunos comentadores de Ryle, que son en extremo puristas respecto a lo que debe ser considerado como conducta, consideran que esta decisión de Ryle constituye una traición respecto a los postulados básicos del conductismo. Otros, como Hornsby (1997), sostienen al contrario que esta decisión de Ryle es una de las más certeras y más productivas que jamás haya tomado. La interpretación que hace Honrsby del pensamiento de Ryle para justificar su aseveración es la que presentamos a continuación. Aun cuando Ryle consideraba como los demás conductistas que la única manera de dar cuenta de la naturaleza de la

[54] Kim, *op. cit.*, p. 28-29: "Consider the act of writing a check. Only someone with certain cognitive capacities, beliefs, desires, and an understanding of relevant social institutions can write a check. You must have a desire to make a payment and the belief that writing a check is a means toward that end. You must also have some understanding of exchange of money for goods and services and the institution of banking."

mente era a través de las conductas (o de las disposiciones de conducta), tenía una intuición filosófica más aguda y penetrante que la gran mayoría de sus antecesores. Esta intuición le reveló que limitar la noción de conducta a las respuestas fisiológicas del cuerpo así como a los meros movimientos corporales era una estrategia demasiado restrictiva. El resultado último de esta estrategia consistía en la eliminación de todas las acciones que presuponían componentes psicológicos internos por medio de su traducción a una serie de descripciones que solo incluían términos conductuales. Hornsby sostiene que Ryle percibió que estas descripciones de movimientos corporales serían siempre incapaces de recuperar la dimensión intencional de las acciones. Ahora bien, en tanto la dimensión intencional es claramente necesaria para diferenciar ciertas acciones que de otra manera serían indistinguibles (como vimos en el ejemplo de los ratones), se hace patente que la tesis según la cual las conductas son sólo respuestas fisiológicas o meros movimientos corporales nos lleva a un callejón sin salida. La gran originalidad de Ryle respecto a los conductistas psicológicos como respecto a sus colegas filósofos (y Hornsby sostiene que también esta originalidad ocurre con respecto al funcionalismo) consiste entonces en haber apreciado la importancia de la intencionalidad para individuar acciones y en haber intentado introducirla en su teoría de la mente sin violar o destruir el espíritu del conductismo. El juicio de Hornsby a este respecto se encuentra expresado en el siguiente pasaje:

> Lo que tendríamos que utilizar para entender la complejidad que experimentamos en la psicología de las actitudes proposicionales no es la complejidad del cerebro, sino nuestro conocimiento de que la psicología del sentido común nos permite explicar mucho más que por qué hay los movimientos corporales que hay en la gente. El paso del conductismo del tipo ryleano al funcionalismo parecerá entonces haber sido, de cierta manera, un paso hacia atrás. Si los estados mentales deben ser considerados como disposiciones de algún tipo (...), entonces en el entendido de que son disposiciones de conducta (...), la noción relevante de

conducta es la amplia, la que los conductistas filosóficos usaban y la que los funcionalistas abandonaron.[55]

Si aceptamos la interpretación de Hornsby respecto a la presencia de las *acciones* en el conjunto de las conductas, Ryle no es entonces tan vulnerable a la objeción presentada por Churchland según la cual las disposiciones de conducta ocultan en última instancia actividades y procesos mentales intencionales, en la medida que las acciones que considera ser conductas presentan una dimensión intencional. Sin embargo, existen sin duda ciertas personas que consideran que esta interpretación de Ryle no es correcta en tanto rompe con las premisas tradicionales del conductismo según las cuales lo único que cuenta para considerar que algo es una conducta es que sea pública y directamente observable. Y la intencionalidad nunca es pública y directamente observable en el caso de los demás, sino que siempre es derivada. Para un conductista purista, no observamos nunca que una persona firme un cheque: sólo observamos una serie de movimientos de los dedos de esta misma persona con respecto a un pluma y a un papel. La dimensión intencional no es algo inherente a esos movimientos corporales que pueda ser observado al mismo tiempo que ellos, sino es algo que le atribuimos a la persona ulteriormente. Ahora bien, ¿puede una propuesta legítimamente seguirse llamando conductista cuando acepta elementos explicativos que no son pública y directamente observables, lo cual contradice una de las motivaciones básicas del conductismo? Dejo esto a la consideración de mi lector. Sin embargo, independientemente de la respuesta que podamos dar a esta pregunta, no podemos dejar de reconocer que la introducción de las *acciones* en el conjunto de las conductas realizada por Ryle es, al igual que su crítica al PEF, un rasgo de su pensamiento de importancia capital en tanto influenció a muchos pensadores materialistas que le sucedieron. Por ejemplo, una de las tesis básicas que cimientan buena parte de la obra de Davidson (1980) consiste que la marca distintiva de

[55] Hornsby, 1997, p. 121: "What we should then have to exploit in understanding the felt complexity of propositional-attitude psychology is not the brain's complexity, but our knowledge that commonsense psychology enables us to explain so much more than why there are the movements of peoples' bodies that there are. The step from a Rylean sort of behaviorism to functionalism will then seem to have been, in a way, a retrograde step. If mental states are to be thought as dispositions of any sort (...) then, to the extent that they are dispositions to behave (...) the relevant notion of behavior is the broad one that the philosopher-behaviorists used and the functionalists left behind."

una acción es la intención con la cual es realizada.[56] Con esto damos por concluida nuestra exposición del conductismo filosófico de Ryle para pasar a la última sección de este capítulo en el cual haremos un balance general de las tres versiones de conductismo que hemos visto.

I.4 Un balance general del conductismo

Habiendo llegado al término de nuestra exposición sobre las ventajas y las limitaciones que presenta el conductismo en tanto que teoría de la mente, acaso sea conveniente hacer un balance de las observaciones que hemos realizado hasta ahora. Este balance es mitigado en el sentido de que, si bien se ha mostrado que ninguna de las tres versiones principales de conductismo que hemos visto parece ser una buena teoría de la mente, esto no implica que la tesis básica del conductismo carezca completamente de valor y deba ser desechada. Sobre este punto, compartimos la opinión de Campbell que, al intentar señalar el logro más importante del pensamiento conductista, escribe lo siguiente:

> [El conductismo] expresa, en forma distorsionada, una verdad de capital importancia. Esta verdad consiste en que hay una conexión conceptual entre las descripciones de los seres en términos mentales y las descripciones en términos conductuales Resulta imposible comprender o explicar los términos mentales sin hacer ningún tipo de referencia a las disposiciones conductuales.[57]

Aun cuando el conductismo tuvo razón en señalar la conexión conceptual entre las descripciones en términos mentales y las descripciones en términos conductuales, acaso éste no sea su logro mayor o más

[56] Es interesante señalar que, si bien para Davidson las intenciones son marcas distintivas de las acciones, no son propiamente "causas" de las acciones. El propósito de las intenciones es sólo el de "racionalizar" las acciones dando cuenta de ellas a partir de deseos, creencias, temores, expectativas, etc. Sin embargo, esta "racionalización" de las acciones no puede ser un vínculo causal sino "cuasi-causal" porque, de ser causal, cada vez una cierta persona desea dormirse se dormiría (hecho que no es cierto porque hay personas que, por más que deseen dormir y tomen pastillas soporíferas, padecen de insomnio crónico)

[57] Campbell, 1987, p. 70

importante. El logro más importante del conductismo probablemente no radica en sus aciertos sino en sus más rotundos fracasos, pues fueron éstos los que abrieron el camino para que se desarrollaran nuevas y mejoradas versiones de pensamiento materialista. Para ver esto con claridad, es menester recordar que al principio de la década de los cincuenta, cuando comenzaba a observarse con claridad que el conductismo en cualquiera de sus versiones principales no podía ser una buena teoría de la mente, los positivistas lógicos llevaron a cabo un arduo trabajo de reflexión sobre sus escritos anteriores para determinar con exactitud los supuestos erróneos del conductismo así como la mejor manera de corregirlos. Entre todos los positivistas lógicos, acaso Feigl probablemente intuyó con mayor claridad que el principal error del conductismo (en la versión que sostiene que los estados mentales son idénticos a las conductas) había consistido en pensar que era posible establecer una identidad lógica entre los estados mentales y las conductas (o al menos entre las proposiciones sobre los estados mentales y las proposiciones sobre las conductas) El establecimiento de esta identidad implicaba la imposibilidad de una relación causal entre los estados mentales y las conductas, lo cual iba contra las intuiciones plasmadas en el lenguaje ordinario usado por la gente para dar cuenta de sus acciones o de las acciones de los demás.[58] Después de haber criticado severamente durante años el modo material de hablar en tanto pensaban que inducía a la gente al engaño, los positivistas comenzaron a reconocer que la psicología del sentido común (que explicaba nuestras conductas con base en nuestros deseos, creencias y temores, i.e., con base en nuestros estados mentales) revelaba un hecho verdadero e ineludible: no se podía prescindir enteramente de las relaciones causales en la explicación de la conducta sin caer en errores como los que hemos señalado previamente:

> Si hemos de evitar los errores de una reducción fenomenalista y, con más generalidad, del negativismo del positivismo ortodoxo, entonces todas las relaciones mencionadas no son

[58] Un conductista ryleano que afirme la identidad entre estados mentales y disposiciones de conducta puede sostener que, si bien su estrategia le hace sacrificar la causalidad entre estados mentales *qua* estados mentales y conductas (que denominamos causalidad vertical), le permite recuperar una causalidad entre disposiciones de conducta (que se asume son idénticas a los estados mentales) y conductas (que denominamos causalidad diagonal). Sin embargo, el problema es que la causalidad diagonal que va de las disposiciones de conducta a las conductas no puede recuperar la intencionalidad de la acción (i.e., no se puede naturalizar la intencionalidad por medio de las meras disposiciones de conducta). Para ver esto con mayor claridad, cf. infra nota 150

> identidades sino –en el mejor de los casos– conexiones legaliformes (causales) entre estados o sucesos distinguibles. La equivalencia de enunciados sobre cada par de estados sólo puede ser entonces de tipo empírico.[59]

Al abandonar los intentos de establecer una identidad lógica entre las proposiciones sobre los estados mentales y las proposiciones sobre las conductas, para reintroducir la noción de causalidad en la relación entre objetos mentales y objetos físicos, Feigl argüía en realidad por el establecimiento de una identidad *a posteriori*, empírica, basada en hechos científicos contingentes. Según él, si podían presentarse evidencias empíricas lo bastante sólidas y consistentes para decir que dos oraciones pertenecientes a dos lenguajes distintos (e.g., una oración de la psicología mentalista y otra de la psicología conductista) denotan una sola y misma cosa, podía establecerse una identidad empírica que era lo óptimo a lo cual podía aspirarse (puesto que establecer identidades analíticas en este terreno era imposible). Ahora bien, de la época en que el conductismo daba sus primeros pasos a la década 1950-1960, el estudio del cerebro había progresado de manera impresionante, a tal grado que numerosos estudios y experimentos sugerían que los estados neuronales del cerebro jugaban un papel causal importante respecto a las conductas. De hecho, algunas personas iban un poco más lejos en tanto sostenían que las oraciones de la psicología mentalista, las oraciones de la psicología conductista y las oraciones que contenían términos de neurofisiología denotaban en realidad los mismos estados y procesos.

> Afirmamos que los denotados del lenguaje mentalista son idénticos con los objetos descritos por el lenguaje conductista, y que ambos son idénticos con los denotados del lenguaje neurofisiológico. (...) Podemos decir que la referencia factual de algunos de estos términos en

[59] Feigl, 1950 en Feigl y Brodbeck, *op. cit.*, p. 620: "If we are to avoid the errors of phenomenalist reduction and quite generally of the negativism of orthodox positivism then all the relationships mentioned are not identities, but - at best- lawful (causal) connections between distinguishable states or events. The equivalence of statements about each pair of states or events can therefore be only of the empirical type "

> cada uno de estos lenguajes (o vocabularios) distintos puede ser la misma mientras que sólo difieren sus bases evidenciales.[60]

En este pasaje podemos apreciar que la posición de Feigl en 1950 era todavía la de un filósofo sin una convicción clara, como él mismo lo reconoce. Sin embargo, Feigl tuvo la precaución de señalar que, si se acumulase eventualmente más evidencia empírica clara que apoyase la coextensionalidad de los denotados del lenguaje mentalista y las descripciones neurofisiológicas del cerebro, podría llegar a establecerse una identidad empírica mucho más sólida entre estados mentales y estados neuronales. Esta es la posición que adoptó algunos años más tarde, pero veremos esto con más detenimiento en el próximo capítulo que está dedicado a la corriente de pensamiento materialista que sucedió al conductismo: se trata de la identidad mente-cerebro (también conocida como materialismo del estado central).

[60] *Ibid.*, p. 623: "We contend that the designata of the mentalistic language are identical with the descripta of the behavioristic language, and that both are identical with the designata of the neurophysiological language (...) We may say that the factual reference of some of the terms in each of these different languages (or vocabularies) may be the same while only their evidential bases differ."

Capítulo II

La teoría de la identidad mente-cerebro. Motivaciones, aciertos, errores y limitaciones en la explicación de la naturaleza de la mente.

II.1 Los orígenes de la teoría de la identidad mente-cerebro

La teoría de la identidad mente-cerebro puede ser considerada como la teoría según la cual la mente y el sistema nervioso central son una y la misma cosa, i.e., todo estado mental M (sea sensación, intención, deseo, juicio, dolor, etc.) es idéntico un estado físico del cerebro, un estado neuronal N.[61] Esta definición es sólo una primera aproximación a lo que realmente constituye la esencia de la teoría de la identidad mente-cerebro, pero podemos servirnos de ella como punto de partida de nuestro estudio. La caracterización de los detalles y rasgos particulares de la teoría vendrán paulatinamente a medida que progresemos.

Nuestros lectores recordarán sin duda que, en la última sección del pasado capítulo, señalamos que la teoría de la identidad mente-cerebro había surgido en la década de los cincuenta como producto de los errores y de los fracasos del conductismo. Esta aseveración es correcta, pero no debemos olvidar que la tesis básica de la teoría de la identidad mente-cerebro no apareció en los inicios de la década 1950-1960, sino que existía desde antes como opción teórica. Con la decadencia progresiva del conductismo en sus distintas versiones, la opción teórica que constituía la teoría de la identidad se tornó mucho más atractiva para los materialistas de toda índole, especialmente para los positivistas lógicos. De hecho, una versión de la tesis básica de la teoría de la identidad era aceptada por algunos positivistas como Carnap a manera de una hipótesis empírica:

> Por problema psicofísico no entendemos aquí la pregunta de si a todos los procesos psíquicos corresponde un proceso simultáneo del sistema nervioso central (más precisamente: si corresponde de tal manera, que a los procesos psíquicos semejantes pertenezcan procesos fisiológicos semejantes). Esto lo presuponemos aquí como una hipótesis empírica.[62]

[61] Smart, 1959, p. 145: "All it [the central-state materialism thesis] claims is that in so far as a sensation statement is a report of something, that something is in fact a brain process."
[62] Carnap, *op cit*, p 307 (§166)

Es interesante observar que, en la formulación propuesta por Carnap, no aparece tanto una identidad entre procesos físicos del cerebro y procesos mentales, sino una correspondencia entre estados neuronales y estados mentales. La mayoría de los positivistas lógicos compartían con Carnap la idea según la cual había algún tipo de correspondencia estrecha entre los estados mentales y los estados neuronales. Esta correspondencia se hacía patente por medio de ciertas correlaciones entre mente y cerebro. Aun cuando no existía en las décadas 1920-1930 y 1930-1940 el conocimiento detallado que ahora tenemos sobre cómo ciertos químicos actúan sobre el cerebro, alterando nuestros estados mentales (e.g., sabemos que el litio produce relajación mientras que las dopaminas atenúan el dolor), Carnap y sus contemporáneos tenían ciertas evidencias de que la composición de los estados neuronales determinaba ciertos estados mentales. Por ejemplo, sabían por medio de la observación empírica que una lobotomía provocaba la perdida irreparable de las funciones elevadas de la mente como la conciencia y que los cambios químicos en el cerebro inducidos por la ingestión de alcohol alteraban los estados mentales de las personas (e.g., generando desinhibiciones o provocando depresiones). Para ellos, la única manera de dar cuenta de dichas observaciones consistía en proponer como hipótesis explicativa una correlación entre mente y cerebro. Es por ello que Carnap señala que se trata de una hipótesis empírica.

En tanto el objetivo último del programa positivista consistía en establecer una ciencia universal de la naturaleza cuya estructura estuviese basada en un lenguaje de constitución compuesto por "hechos básicos" definidos por medio de sus condiciones necesarias y suficientes (lo cual generaba una serie de proposiciones analíticas *a priori*),[63] los positivistas se mostraban en principio recelosos respecto a la posibilidad de sentar las bases de la psicología en una proposición empírica como aquella que establece la identidad entre estados mentales y estados neuronales. De acuerdo con el programa, la psicología debía poder ser integrada como las demás ciencias en el cuerpo de la ciencia universal de la naturaleza, lo que implicaba que sus cimientos teóricos debían estar constituidos sólo por verdades analíticas *a priori*

[63] *Ibid*, p. 192 (§103): "La forma del sistema y las formas de los objetos del sistema se determinan empíricamente; es decir, estas formas se rigen por la realidad y por los objetos individuales que se suponen conocidos en la experiencia. Sin embargo, debe depender de algo, más precisamente, de ciertas propiedades formales, el hecho de que en una situación empírica determinada de un nivel, se deba proceder de tal y cual manera (..)"

(aunque ulteriormente los enunciados que se deriven de estos cimientos sean proposiciones empíricas) Cuando se hizo patente que el programa era irrealizable puesto que las premisas sobre las cuales reposaba eran erróneas, comenzó a ser admitido por los filósofos que la psicología podía ser una ciencia sólida aún si sólo podía apoyarse en cimientos empíricos (y, por ende, contingentes según ellos), y que podía elaborarse una buena teoría de la mente que estuviese basada en una identidad entre la mente y el cerebro, aunque dicha identidad (que es el basamento de la teoría) no fuese en realidad una verdad analítica *a priori*. A continuación, presentaré algunos de los argumentos que apoyan esta propuesta.

II.2 Los argumentos en pro de la teoría de la identidad mente-cerebro

El primer argumento que se presenta para sostener la validez de la teoría de la identidad mente-cerebro consiste en que la teoría se ajusta al criterio de observabilidad pública y empírica que exigen muchos científicos (entre los cuales se cuentan los conductistas psicológicos) para poder determinar si algo es correcto o no. Aun cuando los conductistas psicológicos rechazaban la inclusión de los datos neuronales por considerar que, en tanto eran internos debían ser inobservables (por lo que cualquier teoría elaborada a partir de ellos no podía ser empíricamente comprobable), es por demás claro que los datos neuronales son públicamente observables y empíricamente comprobables porque, de otra manera, la neurobiología no sería una ciencia empírica. Si bien la comprobabilidad empírica de los datos no es en general un criterio propio de los filósofos para determinar la validez de las teorías sino más bien de los científicos (recordemos que buena parte de los argumentos que se presentan en filosofía son *a priori*, sin que por ello dejen de ser válidos),[64] el hecho de que las aseveraciones de la teoría de la identidad mente-cerebro sean empíricamente comprobables constituye un hecho positivo que habla en favor de la teoría.

[64] Soy consciente de que esta idea es bastante controversial en la actualidad donde muchos miembros de la comunidad filosófica (especialmente los filósofos de la ciencia) se encuentran convencidos de que la verificabilidad empírica de sus hipótesis es esencial para que puedan ser consideradas verdaderas. Sin embargo, desde mi punto de vista, estos filósofos hacen más ciencia que auténtica filosofía. El gran problema con el criterio de verificabilidad empírica de las hipótesis es que lo único que puede proveer es la certidumbre de que las hipótesis presentadas son verdaderas *en este mundo posible*. Esto sin duda basta a los científicos, pero un filósofo debe de tener estándares más altos, y debe exigir que las hipótesis presentadas sean metafísicamente necesarias, i.e., que puedan ser verificadas en todos los mundos posibles relevantes. Podemos apreciar esto claramente en el tratamiento distinto que tienen Chomsky y Descartes de la hipótesis del innatismo de las estructuras profundas del lenguaje del pensamiento·

El segundo argumento que ciertos autores presentan para afirmar que la relación entre estados mentales y estados neuronales es una relación de identidad consiste en mostrar que la teoría es acorde con el pensamiento de Frege (por esta razón los partidarios de la teoría de la identidad sostienen que su tesis básica se inscribe dentro de la línea del neo-fregeanismo). Para entender el argumento, es menester recordar que, en *Uber Sinn und Bedeutung* (1892), Frege se propuso resolver el problema de la naturaleza de la identidad y, para hacerlo, presentó una hipótesis (también conocida como "hipótesis de la referencia mediada") según la cual existen tres niveles distintos en un término o en una proposición: el signo, el sentido y el significado (también llamado referente). Según Frege, estos tres elementos se relacionan entre sí de la siguiente manera: "un nombre propio (palabra, signo, combinación de signos, expresión) expresa su sentido, y significa o designa su significado "[65] El papel del sentido o modo de presentación consiste en realizar una mediación entre signos y significados, y esta mediación tiene las siguientes características. De acuerdo con Frege, un signo particular puede expresar solamente un sentido, y un sentido particular sólo puede hacer referencia solamente a un significado Sin embargo, esta regla no excluye que haya varios sentidos (expresado cada uno por un signo distinto) que puedan hacer referencia al mismo significado. Como para Frege el sentido de un nombre propio está dado por una descripción definida (de la cual el nombre en cuestión no es sino una abreviación), es evidente que los nombres propios "Molière" y "Jean-Baptiste Poquelin" deben tener sentidos distintos si la descripción de "Molière" es "el hombre que escribió *El Tartufo*" y la descripción de Jean-Baptiste Poquelin es "el hombre que fundó *El Ilustre Teatro*". Pero aunque los sentidos de estos dos nombres sean distintos, hacen referencia a un mismo individuo. Este mismo razonamiento es aplicado por los partidarios de la teoría de la identidad mente-cerebro (entre los cuales destaca principalmente Feigl) para justificar la validez de su postura:

---

mientras Chomsky (1965) sostiene que se trata de una hipótesis empírica que debe ser introducida para explicar ciertos fenómenos y que puede ser verificada a través de la observación, Descartes sostiene que se trata de una tesis *a priori* que debe ser postulada como condición de posibilidad de la representación.

[65] Frege, 1892 en Frege, 1970, p. 61: "A proper name (word, sign, sign combination, expression) expresses its sense, means or designates its meaning "

> Usando la distinción de Frege entre *Sinn* ("significado", "sentido", "intensión") y *Bedeutung* ("referente", "denotado", "extensión"), podemos decir que los términos neurofisiológicos y los términos fenomenales correspondientes, aunque difieren profundamente en su sentido y, por ende, en los modos de confirmación de las proposiciones que los contienen, tienen referentes idénticos.[66]

El tercer argumento que es presentado para sostener que la relación entre estados mentales y estados neuronales es una relación de identidad apela a la simplicidad de la teoría de la identidad mente-cerebro respecto a otras alternativas teóricas. Este argumento es expuesto por Smart en los términos siguientes:

> Me parece que la ciencia nos presenta cada vez en mayor medida un punto de vista a través del cual los organismos pueden ser vistos como mecanismos fisicoquímicos: parece que algún día la conducta del hombre será explicable en términos mecánicos. No parece haber, en lo que concierne a la ciencia, nada en el mundo mas que organizaciones cada vez más complejas de constituyentes físicos, con excepción de un lugar: la conciencia. Esto es, para tener una descripción completa de lo que ocurre en un hombre se tendría que mencionar no sólo los procesos físicos en sus tejidos, glándulas, sistema nervioso (...) sino también sus estados de conciencia: sus sensaciones visuales, auditivas, táctiles (...) El hecho de que estos últimos deban estar relacionados con estados cerebrales no ayuda, en tanto decir que están relacionados es decir que son algo "por encima". (...) El hecho de que todo deba ser explicable en términos de la física (...) con excepción de la ocurrencia de las sensaciones, me parece francamente increíble.[67]

---

[66] Feigl, 1960 en Feigl, 1981, p. 347: "Utilizing Frege's distinction between *Sinn* ('meaning', 'sense', 'intention') and *Bedeutung* ('referent', 'denotatum', 'intension'), we may say that neurophysiological terms and the corresponding phenomenal terms, though widely differing in *sense*, and hence in the modes of confirmation of statements containing them, do have identical referents."

[67] Smart, *op. cit.*, p. 142: "It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms: it seems that even the behavior of man will one day be explicable in mechanistic terms. There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. That is, for a full description of what is going on in a man you would have to mention not only the physical processes in his tissues,

Este argumento tiene de hecho dos interpretaciones como señala Kim: puede verse desde una perspectiva ontológica o desde una perspectiva lingüística.[68] Desde una perspectiva ontológica, la tesis del argumento consiste en que es más plausible una teoría como el materialismo del estado central que una teoría como el dualismo cartesiano en tanto permite dar cuenta de los estados mentales apelando a un sólo tipo de entidades (que son entidades físicas), lo cual implica una ontología más simple. En cambio, como el dualismo cartesiano postula que los estados mentales son entidades "por encima" de los estados materiales, debe asumir el compromiso de explicar los estados físicos así como los estados mentales, lo cual (aún si es realizable) implica una ontología más compleja, en la cual las relaciones entre elementos son más difíciles de explicar.

Desde una perspectiva lingüística, la tesis del argumento consiste en lo siguiente: si hay una identidad entre estados mentales y estados físicos del cerebro, entonces el vocabulario de la psicología mentalista debe poder ser eliminado en principio y remplazado por el vocabulario de la neurobiología. Esta línea de pensamiento ha sido seguida sobre todo por eliminativistas como Quine (1952), Feyerabend (1963a, 1963b) y Rorty (1965) en la década 1960-1970 y actualmente todavía encuentra partidarios como los Churchland (1989) o Stich (1983). La discusión del materialismo eliminativista rebasa muy ampliamente los objetivos de esta tesis, por lo cual será dejada de lado. Sin embargo, consideramos necesario señalar que cualquier teoría materialista que proponga una eliminación sistemática del vocabulario de la psicología mentalista y su remplazo por el vocabulario de la neurociencia no puede constituir, *en el estado actual de las cosas*, una buena teoría de la mente por una sencilla razón: el vocabulario de la neurociencia todavía se encuentra muy poco desarrollado, por lo cual no hay seguridad de que pueda sustituir al vocabulario de la psicología mentalista sin que haya perdidas explicativas. En cambio, el vocabulario de la psicología mentalista nos permite dar cuenta de nuestros estados mentales como causas

---

his glands, nervous system (...) but also his states of consciousness: his visual, auditory, tactual sensations (...) That these should be *correlated* with brain processes does not help, for to say that they are *correlated* is to say that they are something 'above and over'. (...) That everything should be explicable in terms of physics (...) except the occurrence of sensations seems to me frankly unbelievable."

[68] Kim, *op. cit.*, pp. 53-54

de conductas particulares sin ningún problema. Por lo tanto, resulta muy arriesgado, *en el estado actual de las cosas*, pretender sustituir algo que sirve para dar cuenta de las conductas por algo que puede revelarse erróneo o insuficiente en último término. Esto constituye sin duda sólo una razón de hecho, no de principio, para rechazar el materialismo eliminativista, pero ello bastará en la medida que no es nuestro objetivo discutir a fondo el materialismo eliminativista.

El cuarto argumento presentado para sostener por algunos autores para sostener que la relación entre mente y cerebro es una relación de identidad consiste en el hecho que la teoría de la identidad mente-cerebro puede dar cuenta en términos claros de la interacción causal entre mente y cuerpo. Para poder apreciar el peso de este argumento, acaso convenga recordar unos hechos sobre el dualismo cartesiano. Al declarar que la mente y el cuerpo pertenecen a dos categorías ontológicas distintas, el dualismo cartesiano no podía dar cuenta consistentemente de la interacción entre ambos, en particular de la manera en que la mente actúa sobre el cuerpo. Si la mente es en realidad una cosa distinta de la materia, algo que se encuentra "por encima" de ella, cabe plantearse la siguiente pregunta. ¿cómo puede algo que no es material actuar sobre algo que sí lo es y moverlo? La coherencia de las respuestas aportadas por los dualistas cartesianos a este problema ha sido comparada habitualmente por ciertos materialistas con la coherencia de una tira cómica de Gasparín, el fantasma amigable, en la cual vemos a Gasparín de un cuadro a otro atravesar paredes y asir objetos.[69] De esta incoherencia en la concepción cartesiana de la mente, los materialistas deducen que sólo aquello que es material puede actuar sobre lo material. Usando esta conclusión como premisa, se puede construir entonces un argumento que demuestre la necesidad de que los estados mentales sean materiales:

(1) Los estados mentales se definen a través de su papel causal respecto a la conducta (o respecto a otros estados mentales).

(2) Nada que no sea material puede ejercer un papel causal sobre lo material.

[69] Dennett, 1991, p 35

(3) Por lo tanto, los estados mentales deben ser estados materiales.[70]

Si bien este argumento no demuestra que los estados materiales de los cuales se habla sean en realidad estados neuronales, esta identidad puede ser comprobada, arguyen los teóricos de la identidad mente-cerebro, a través de una serie de investigaciones empíricas. Para demostrar esto, Armstrong presenta el argumento que citamos a continuación. En primera instancia, Armstrong introduce el concepto de estado mental como algo que se encuentra definido por las relaciones causales que mantiene con otros elementos (en especial, con estímulos y conductas):

> El concepto de un estado mental es primariamente el concepto de *un estado de una persona apto para producir un cierto tipo de conducta*. Sacrificando toda exactitud en aras de la brevedad podemos decir que, aunque la mente no es conducta, es la *causa* de la conducta. En el caso de algunos estados mentales sólo, también son *estados aptos para ser producidos por un cierto tipo de estímulos*.[71]

Partiendo de esta hipótesis, Armstrong sostiene que es tarea de la ciencia (mas no de la filosofía) determinar la naturaleza de aquello que juega el papel de mediador causal entre estímulos y conductas. Para mostrar la validez de su estrategia, Armstrong no se priva de señalar que ha sido aplicada en otros casos con éxito como, por ejemplo, el establecimiento de la identidad entre el gen y la molécula de ADN

[70] La validez de la primera premisa de este argumento de Dennett (y, por lo tanto, la validez del argumento mismo) es bastante cuestionable en la medida que hay ciertos estados mentales con propiedades fenoménicas o qualia (como el dolor) que también tienen propiedades causales, pero en los que las propiedades causales no son las características definitorias y esenciales que sea precisamente un estado mental dado, sino las propiedades fenoménicas. Examinaremos esto con más detalle en la tercera sección del tercer capítulo. Sin embargo, aceptémoslo por el momento por mor del argumento. Además, es importante notar que la primera premisa expresa una tesis *a priori* mientras que la segunda expresa algo que sólo es verificable por medio de la experiencia (i.e., una tesis *a posteriori*). Sin embargo, esto no significa que necesariamente la primera tenga que ser verdadera y la segunda pueda ser falsa. De hecho, aunque es *a priori*, la primera premisa es falsa mientras que la segunda, aunque es *a posteriori*, es necesariamente verdadera.

[71] Armstrong, 1969, p. 82: "The concept of a mental state is primarily the concept of *a state of the person apt for bringing about a certain sort of behavior*. Sacrificing all accuracy for brevity we can say that, although the mind is not behavior, it is the *cause* of behavior. In the case of certain mental states only they are also *states of the person apt for being brought about by a certain sort of stimulus*."

dar cuenta de los estados mentales con contenido fenoménico de manera neutral. Esto constituye a primera vista una ventaja de gran importancia puesto que, tradicionalmente, el problema central que las principales corrientes de pensamiento materialista han tenido que resolver consiste en dar cuenta del carácter intrínseco y subjetivo de los estados mentales (e.g., la "dolorosidad" del dolor). Este carácter intrínseco sobre el cual versa la fenomenología se resiste a ser explicado por medio de un vocabulario puramente físico. Para dar cuenta de las experiencias inmediatas, Smart propone una estrategia de análisis tópico neutral de los objetos ("topic-neutral analysis") que consiste en lo siguiente:

> Cuando una persona dice, "Tengo una sensación amarillenta-anaranjada", está diciendo algo como esto: "*Hay algo que está ocurriendo que es como lo que ocurre* cuando tengo mis ojos abiertos, estoy despierto, y hay una naranja convenientemente iluminada frente a mí, esto es, cuando realmente veo una naranja."[73]

La ventaja principal del análisis neutral de los objetos según Smart consiste en lo siguiente. Tanto los dualistas cartesianos como los materialistas clásicos se encuentran en principio comprometidos a dar explicaciones de las experiencias conscientes, de las experiencias como son para la persona que las está experimentando, ya sea en términos de procesos que ocurren en una sustancia inmaterial (como en el caso de los dualistas), ya sea en términos de procesos materiales (como en el caso de los materialistas), lo cual los compromete, desde un inicio, con una ontología dualista o con una ontología materialista. En cambio, la estrategia propuesta por Smart nos permite hablar de las experiencias inmediatas como objetos neutros, independientes en un primer momento tanto de los compromisos ontológicos del materialismo como del dualismo:

[73] Smart, *op. cit.*, p. 149: "When a person says, 'I see a yellowish-orange after-image', he is saying something like this: '*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange' "

hecha por los biólogos. Como el concepto de gen para un biólogo es el concepto de un elemento interno de los organismos que es responsable de la transmisión de ciertas características de una generación a otra, y las investigaciones realizadas en biología y bioquímica han revelado que aquello que es responsable de la transmisión de ciertas características de una generación a otra no es otra cosa que la molécula de ADN, entonces la identidad del gen con la molécula de ADN se sigue de manera inmediata. De igual manera, los teóricos de la identidad (que por esta razón también son llamados teóricos causales de la mente) sostienen que:

> La ciencia moderna declara que este mediador entre estímulos y respuestas es de hecho el sistema nervioso central o, puesto más cruda y burdamente, pero también más simplemente, el cerebro.[72]

Este argumento es uno de los más conocidos para sostener la interpretación según la cual la relación que existe entre estados mentales y estados neuronales es una identidad. El peso de este argumento se deriva del hecho que, por medio de él, la teoría de la identidad mente-cerebro llena el primer requisito básico para poder ser una buena teoría de la mente: dar cuenta de la causalidad que parece existir entre la mente y el cuerpo. Como los materialistas del estado central asumen que los estados mentales son idénticos a los estados neuronales, es posible presentar en la perspectiva de su teoría una explicación causal de los mecanismos biológicos que permiten la interacción entre el cerebro y el resto del cuerpo. Además, como estos mecanismos no implican ningún tipo de "salto ontológico" a la usanza del dualismo cartesiano y como son públicamente observables, la teoría de la identidad mente-cerebro se presenta como una opción plausible.

El quinto argumento presentado por los partidarios del materialismo del estado central (en particular por Smart) para sostener la validez de su teoría consiste en que, según ellos, por medio de ella se puede

[72] *Ibid.*, p. 79. "Modern science declares that this mediator between stimulus and response is in fact the central nervous system, or more crudely and inaccurately, but more simply, the brain."

> En la medida en que nuestro hablar de la experiencia inmediata se encuentra en términos de una situación de estímulos típicos (y en el caso de los dolores y de otras sensaciones similares puede estar en términos de una situación de respuestas típicas), podemos ver que nuestro hablar de la experiencia inmediata es neutral entre el materialismo y el dualismo. Reporta nuestras ocurrencias internas como lo que internamente ocurre (o no ocurre) en situaciones típicas, pero el dualista construiría estas ocurrencias como ocurrencias en una sustancia inmaterial mientras el materialista construiría estas ocurrencias como teniendo lugar en nuestros cráneos.[74]

Además de los cinco argumentos que he presentado anteriormente, existen dos razones de hecho (mas no dos argumentos) que abogan por la validez de la teoría de la identidad en tanto que teoría de la mente. La primera de estas razones consiste en el hecho que la teoría de la identidad parece proveer una solución sencilla al problema de la existencia de otras mentes. Cuando trepanamos el cráneo de una persona común y corriente (o cuando examinamos lo que su cráneo contiene con métodos menos bárbaros y primitivos como la tomografía por resonancia magnética), por lo general descubrimos un cerebro. Si la mente es idéntica al cerebro, entonces es evidente que las demás personas tienen mentes. Más aún, como sus cerebros son numéricamente distintos al nuestro (tienen propiedades espacio-temporales distintas), los materialistas del estado central pueden sostener que las mentes de los demás son numéricamente distintas a la nuestra.

La segunda razón de hecho empleada para mostrar que la teoría de la identidad es probablemente una buena teoría de la mente consiste en el hecho que es consistente con la explicación biológica (y, por ende, material) del origen del hombre. Según la biología, los hombres nacen como resultado de la unión de dos células reproductoras que, al combinar sus materiales genéticos respectivos, crean una nueva célula que se

[74] Smart, 1963, p. 654: "Since our talk of immediate experience is in terms of a typical stimulus situation (and in the case of some words for aches and pains and the like it may be in terms of some typical response situation), we can see that our talk of immediate experience is itself neutral between materialism and dualism. It reports our internal goings on as like or unlike what internally goes on in typical situations, but the dualist would construe these goings on as goings on in an immaterial substance, whereas the materialist would construe these goings on as taking place inside our skulls."

desarrolla y se divide por acreción de moléculas con base en su programa genético. Como en este proceso no entra ningún elemento inmaterial, y como las etapas sucesivas del desarrollo del individuo del estado embrionario a la edad adulta son explicables en base a meros procesos materiales, resulta absurdo pensar que la mente puede ser algo distinto de la materia, distinto del cuerpo. De ser este el caso, ¿de dónde surgiría? ¿Cómo podrían una serie de procesos materiales generar algo que no es material? Como ninguna de estas dos preguntas admite una respuesta sencilla, una alternativa más segura consiste en pensar que la mente es en realidad un objeto material y que es idéntica al cerebro.

Así pues, resumiendo lo que hemos visto en esta sección, la teoría de la identidad mente-cerebro cuenta con varios argumentos que hicieron que la teoría fuese ampliamente aceptada en la década 1950-1960, e incluso en la década posterior. Sin embargo, aún cuando existen algunas buenas razones para aceptar que la teoría de la identidad mente-cerebro puede ser una buena teoría de la mente, es importante destacar que existen numerosas objeciones a la teoría, algunas de las cuales pudieron ser contestadas a través de argumentos convincentes por los materialistas del estado central mientras que otras, al no poder ser resueltas, provocaron que la teoría cayera poco a poco en el descrédito. En la próxima sección, analizaremos algunas de estas objeciones.

II.3 Algunas objeciones a la teoría de la identidad mente-cerebro

De las múltiples objeciones que se han hecho contra la teoría de la identidad, la mayoría se encuentran basadas en una crítica de la noción de identidad. ¿Qué significa exactamente la tesis según la cual los estados mentales son idénticos a los estados neuronales? De acuerdo con Smart, la relación de identidad es este caso debe ser entendida como una identidad estricta. Ahora bien, ¿qué quiere decir Smart al sostener que la identidad entre estados mentales y estados neuronales es una identidad estricta? Hay varias nociones de identidad que habitualmente son entendidas en un sentido estricto como, por ejemplo, la noción matemática de identidad. Según los matemáticos, la identidad es una relación de equivalencia R

entre objetos que es reflexiva, transitiva y simétrica.[75] Esta noción con estas tres características es útil en la medida en que nos dice cómo se encuentra estructurada la identidad (i.e., nos dice que un objeto aRa, que si aRb y bRc, entonces aRc, y que si un objeto aRb, entonces bRc), pero para un filósofo esto es insuficiente puesto que la noción matemática de igualdad no señala cuáles son las condiciones para que un objeto sea idéntico a si mismo (lo cual es una interrogante metafísica que los matemáticos por lo general no se plantean). Así pues, además del concepto matemático de identidad, hay otros conceptos de identidad desarrollados con motivaciones filosóficas que buscan establecer las condiciones necesarias y suficientes para que un objeto sea idéntico a sí mismo. Acaso el concepto de identidad con motivaciones filosóficas más conocido e intuitivamente más plausible sea el que propuso Leibniz. Este concepto, conocido como "principio de identidad de los indiscernibles" (PII), consiste en lo siguiente: dos entidades x e y son idénticas si y sólo si comparten todas y cada una de sus propiedades (i.e., si para cada propiedad P o bien ambas la tienen, o bien ambas carecen de ella). La traducción de esta noción de identidad en un lenguaje formal nos proporciona la formula que reproducimos a continuación:

$$(\forall x)(\forall y)\ [x=y \leftrightarrow (\forall \varphi)\ (\varphi x \leftrightarrow \varphi y)]$$

Si aceptamos el PII como noción de identidad estricta, automáticamente surgen una serie de objeciones a la teoría de la identidad que plantean que un estado mental no puede ser idéntico a un estado neuronal en tanto ambos tienen propiedades distintas. Una primera objeción a la teoría de la identidad mente-cerebro sostiene que la teoría no puede ser verdadera en tanto los estados mentales tienen varias propiedades fenoménicas que los estados neuronales parecen no compartir. Por ejemplo, una sensación de dolor puede ser "aguda", "lancinante", "punzante", "agónica", etc., pero resulta por demás absurdo decir

[75] De hecho, un matemático tiene que especificar mucho más para definir la identidad puesto que la relación de paralelismo entre rectas también es una relación de equivalencia (reflexiva, transitiva y simétrica), pero no es por ello una relación de identidad. Para mostrar que una relación R es una relación de identidad, el matemático debe mostrar que, además de ser una relación de equivalencia, R es también relación de orden (reflexiva, transitiva y anti-simétrica). El problema es que la definición de anti-simetría contiene en su formulación una instancia de la relación de identidad que buscamos definir por medio de ella. De ello, se deriva la tesis de que la identidad no es un concepto que pueda ser definido satisfactoriamente en matemáticas, sino que debe ser postulado como axioma.

de una fibra C activada que es "lancinante" o "aguda". Así pues, con base en el PII, la tesis básica teoría de la identidad *parece* ser refutada por esta objeción.

Es importante señalar en este punto que el uso del verbo *parecer* en la oración anterior está justificado en la medida que, antes de poder decir que hemos refutado la tesis básica de la teoría de la identidad, es necesario determinar exactamente *qué* es lo que la tesis básica dice. Al principio de este capítulo, hemos señalado que la tesis básica de la teoría de la identidad mente-cerebro afirma que todo estado mental M es en realidad idéntico a un cierto estado neuronal N. Esta tesis básica puede ser interpretada de dos maneras distintas: (1) puede decirse que afirma que la *propiedad* de ser el estado mental M es idéntica a la *propiedad* de ser el estado neuronal N o (2) puede decirse que afirma que el *objeto* que es el estado mental M es idéntico al *objeto* que es el estado neuronal N. Estas dos interpretaciones de la tesis básica de la teoría de la identidad hacen aparecer los conceptos de identidad de tipo ("type identity") y de identidad de instancia ("token identity") sobre los cuales reposa la distinción (aceptada comúnmente en la literatura filosófica actual) entre el fisicalismo de tipo ("type physicalism") y el fisicalismo de instancia ("token physicalism").

En el ámbito de las identidades estrictas, podemos tener identidades estrictas entre clases así como identidades estrictas entre objetos. Para entender la noción de identidad de clase, debemos determinar en primera instancia *qué* son las clases Según la mayoría de los filósofos en la actualidad, una clase $\mathcal{P}$ se encuentra definida por una propiedad P; esta propiedad es nombrada por un cierto predicado y *puede* ser realizada por uno o varios objetos X en el mundo de tal manera que se hace verdadera la oración "X es P" (e.g., la clase de los perros se encuentra definida por la propiedad de *ser perro* –propiedad que se caracteriza por el hecho de que es nombrada por el predicado "es perro" y de que *puede* ser realizada por varios objetos X que hacen verdadera la oración "X es perro"). Teniendo en mente esta caracterización de clase, la pregunta que podemos plantearnos automáticamente es la siguiente: ¿cómo podemos saber si dos clases son idénticas? Puesto que las clases se establecen en base a propiedades, que las propiedades son nombradas por predicados y que pueden ser realizadas por objetos del mundo, acaso uno de los criterios que debemos de exigir para que dos clases A y B sean idénticas es que ambas sean coextensionales, i.e., que la clase definida con base en la propiedad A se encuentra realizada por los mismos objetos que

realizan la clase definida con base en la propiedad B, y viceversa.[76] ¿Podemos aplicar este criterio de identidad de clases a la primera interpretación de la tesis básica de la teoría de la identidad? ¿Puede acaso ser la clase de estados mentales M ser idéntica a la clase de estados neuronales N? Existe una razón muy poderosa para buscar proporcionar una respuesta afirmativa a esta pregunta para los teóricos de la identidad: si asumimos que hay una identidad estricta entre estados mentales y estados neuronales, dicha identidad debe de tener el estatus de una ley. Pero una ley sólo puede ser tal si los elementos que liga son *clases* o *propiedades*. En el caso en que una ley ligara *objetos*, perdería automáticamente el carácter normativo y generalizador que hace precisamente que sea una ley. Es por ello que "la formulación clásica de la teoría de la identidad propuesta por Smart y Feigl es un fisicalismo de tipo."[77] Sin embargo, si asumimos que la tesis básica de la teoría de la identidad expresa una identidad entre clases (i.e., una identidad de tipo), entonces no tenemos otra opción que reconocer que la teoría de la identidad es errónea por la razón que expondremos a continuación. Si asumimos que la propiedad de ser un cierto estado mental M (e.g., una sensación de dolor) es idéntica a la propiedad de ser un cierto estado neuronal N (e.g., una fibra C activada), es claro que nos comprometemos a que la propiedad mental esencial nombrada por el predicado "es M" sea realizada por los mismos objetos que realizan la propiedad neuronal esencial nombrada por el predicado "es N", y viceversa. Examinemos cada una de las partes del compromiso anterior por separado. En el caso en que proporcionemos a un teórico de la identidad el tiempo suficiente así como las herramientas adecuadas, es posible que logre determinar que la propiedad de ser una fibra C activada nombrada por el predicado "es una fibra C activada" es realizada por exactamente los mismos objetos que realizan la propiedad de dolorosidad nombrada por el predicado "es un dolor". Sin embargo, resulta imposible determinar que la propiedad de dolorosidad nombrada por el predicado "es un dolor"

[76] De hecho, este criterio es insuficiente para asegurar una identidad de clases por sí solo en la medida que, como bien dice Quine (Quine, 1961, p. 31): "That 'bachelor' and 'unmarried man' are interchangeable *salva veritate* assures us of no more than that 'All bachelors are unmarried is true'. There is no assurance here that the extensional agreement of 'bachelor' and 'unmarried man' rests on meaning rather than merely on accidental matters of fact, as does the extensional agreement 'creature with a heart' and 'creature with kidneys'." Sin embargo, si bien es insuficiente, es necesario que este criterio sea verificado para que podamos hablar de identidad entre propiedades. En el presente trabajo nos concentraremos solamente en este requisito, puesto que si podemos demostrar que si la identidad propuesta por los teóricos de la identidad entre estados mentales y estados neuronales no llena ni siquiera el requisito de la coextensionalidad, entonces sencillamente no tiene sentido hablar de ella.

[77] Kim, *op. cit.*, p. 60: "The classic formulation of the identity theory due to Feigl and Smart is type physicalism."

sea realizada por exactamente los mismos objetos que realizan la propiedad de ser una fibra C activada nombrada por el predicado "es fibra C activada". Esto implicaría que algo sólo puede ser un dolor cuando es también una fibra C activada, lo cual restringe claramente la existencia del dolor a criaturas como nosotros. La imposibilidad de sostener esta tesis será presentada con más detalles al exponer la séptima objeción a la teoría de la identidad más adelante.

En el caso en que interpretamos la tesis básica de la teoría de la identidad como sosteniendo que el *objeto* que es el estado mental M es idéntico al *objeto* que es el estado neuronal N, tenemos una identidad estricta entre objetos. De igual manera que hicimos anteriormente, para entender la noción de identidad de objeto (o identidad de instancia), es necesario determinar primero *qué* son los objetos. Un objeto es cualquier entidad que reconocemos como algo particular y ontológicamente independiente de todas las otras entidades que reconocemos como tales. Un cigarrillo es un objeto en la medida que lo reconocemos como algo particular y ontológicamente independiente de todos los demás cigarrillos que hay en el mundo o de la cajetilla en la cual está guardado. Ahora bien, cuando ciertos teóricos de la identidad postulan que la tesis básica de la teoría de la identidad debe de entenderse como afirmando una identidad entre objetos, no están sosteniendo que todas las propiedades de los estados mentales deban ser exactamente las mismas que las propiedades de los estados neuronales, lo cual sería muy problemático. Según ellos, el origen del problema radica en el hecho de que cualquier discusión sobre propiedades implica de manera automática una discusión sobre tipos en tanto que las propiedades son los criterios que empleamos para establecer las clases o los tipos. Así pues, cuando algunos teóricos de la identidad postulan una identidad de objetos o identidad de instancia entre estados mentales M y estados neuronales N, aquello que postulan realmente es una identidad donde M y N no comparten exactamente los mismos predicados, por lo cual no pueden ser idénticos en cuanto al significado o a las propiedades, aunque a pesar de esto pueden ser *objetivamente* idénticos. De esta noción de identidad de instancia se deriva el fisicalismo de instancia (que es una versión más débil de la teoría de la identidad clásica basada en tipos) cuya tesis es expresada por la proposición "Todo estado que tiene una propiedad mental también tiene una cierta propiedad material." Esta versión particular de la teoría de la identidad es inmune a la objeción de las propiedades fenoménicas. Según la tesis básica del fisicalismo de instancia, un estado del cerebro puede tener

propiedades fenoménicas (e.g., ser sentido como un dolor agónico) con la condición de que tenga también ciertas propiedades materiales (e.g., tener ciertos grupos de neuronas particulares activados con ciertos potenciales eléctricos). Como podemos apreciar claramente en este ejemplo, una teoría de la identidad mente-cerebro basada en un fisicalismo de instancia admite, no un dualismo de sustancias como aquel que propone Descartes, pero sí un dualismo de propiedades. Esta posición teórica (un fisicalismo de instancia no reduccionista aunado a un dualismo de propiedades) ha sido adoptada por muchos materialistas que no están dispuestos a aceptar una teoría de la identidad mente-cerebro basada sobre un fisicalismo de tipo. El principal problema de un fisicalismo de tipo consiste en el hecho que, al sostener una identidad estricta de tipo entre estados mentales y estados neuronales, aboga por una reducción de los primeros a los segundos, *pero también implícitamente por una reducción de los segundos a los primeros* en tanto la identidad es una relación simétrica. En cambio, el fisicalismo de instancia, del cual constituye un buen ejemplo el monismo anómalo de Davidson, tiene la gran ventaja de respetar una cierta independencia de lo mental al no proponer ningún tipo de reducción.[78] Por ello, varios materialistas han considerado (sobre todo en las década 1960-1970 y 1970-1980) que un fisicalismo de instancia puede ser una buena teoría de la mente. Sin embargo, en años recientes, algunos autores han presentado varias objeciones contra el fisicalismo de instancia, pero sólo serán examinadas más adelante. Prosigamos por ahora nuestra exposición de las objeciones sobre la validez de la teoría de la identidad tipo mente-cerebro como teoría de la mente.

La segunda objeción que se presenta habitualmente contra la teoría de la identidad parte del hecho que los estados mentales parecen tener ciertas propiedades epistemológicas que los estados neuronales no comparten. Como para los teóricos de la identidad una sensación de dolor no es otra cosa más que una fibra C activada, resulta difícil explicar como, cuando una persona tiene un dolor, es inmediatamente consciente de que lo tiene, aunque es en extremo probable que no sea consciente de que en su cerebro hay una fibra C activada. Por lo tanto, de acuerdo con el PII, el dolor no puede ser idéntico a la fibra C, con lo cual la tesis básica de la teoría de la identidad mente-cerebro (en la versión que propone una identidad de tipo) parece ser refutada. Ahora bien, los materialistas del estado central tienen una manera de responder a

[78] Esta independencia de lo mental respecto a lo físico no es total ni absoluta para Davidson en la medida en que sostiene que, aun cuando no hay leyes causales, hay una relación de superveniencia de lo mental respecto a lo físico

esta objeción arguyendo lo siguiente: el hecho que una sensación de dolor no sea otra mas que una fibra C activada no implica que todas las personas *deban* saber que una sensación de dolor *es* en realidad una fibra C activada. Lo que ocurre es que, como Russell señaló en (1905), esta objeción se sitúa dentro de un contexto opaco. De la misma manera en que resulta erróneo inferir de las proposiciones "Jorge IV quería saber si Scott era realmente el autor de *Waverley*" y "Scott es el autor de *Waverley*" la proposición "Jorge IV quería saber si Scott era realmente Scott", resulta erróneo inferir de las proposiciones "X sabe que tiene un dolor" y "Un dolor es idéntico a una fibra C activada" la proposición "X sabe que tiene una fibra C activada."

La tercera objeción que habitualmente se presenta a la teoría de la identidad tipo mente-cerebro consiste en el hecho que los estados neuronales tienen una localización precisa mientras que los estados mentales parecen no tenerla.[79] Si asumimos que mi creencia de que los Yankees van a ganar la Serie Mundial este año es un estado neuronal, tenemos un problema: podemos determinar por medio de una investigación cuál es la ubicación precisa de un cierto estado neuronal, pero no tiene demasiado sentido decir que mi creencia de que los Yankees van a ganar la Serie Mundial este año tiene una ubicación espacial precisa en mi cerebro, a menos que se pueda corroborar de manera independiente. Asumamos por un momento que esta corroboración independiente es posible y que podemos llegar a determinar con precisión la ubicación espacial de todo estado mental. Sin embargo, esta tesis contradice el hecho que ciertos estados mentales (en particular, los dolores) parecen tener un espacio "fenoménico" que difiere del espacio físico de los estados neuronales (este espacio "fenoménico" corresponde al lugar donde los dolores son sentidos como tales). La posición de los teóricos de la identidad respecto a este espacio "fenoménico" consiste en sostener que debe ser eliminado o por lo menos reducido al espacio físico de los estados neuronales. Sin embargo, ninguna de estas dos estrategias es correcta en la medida en que violan una de las condiciones que hemos establecido para que algo pueda ser una buena teoría de la mente: poder dar cuenta del carácter intrínseco de los estados mentales. Si sostenemos que el espacio "fenoménico" en el que ocurren los dolores es parte constitutiva de su carácter intrínseco, es claro que ni

[79] Cf. Baier, 1964 y Taylor, 1965.

una reducción ni una eliminación del espacio "fenoménico" pueden dar cuenta del carácter intrínseco de los estados mentales de manera tal que la independencia parcial de lo mental sea respetada.

Para responder a los tres argumentos contra la teoría de la identidad que hemos señalado anteriormente (propiedades fenoménicas, propiedades epistemológicas y propiedades locativas), algunos teóricos de la identidad han razonado de la siguiente manera. Como la teoría de la identidad basada en un fisicalismo de tipo es vulnerable ante las tres objeciones mencionadas anteriormente (en especial ante la objeción sobre las propiedades fenoménicas), algunos teóricos de la identidad han sugerido que la teoría de la identidad no debe ser establecida con base en un fisicalismo de tipo (en el que hay identidad tanto de contenido como de referente), sino en un fisicalismo de instancia (en el que sólo hay identidad de referente). A continuación, expondremos brevemente la propuesta de Davidson, mostrando cuál es su motivación (i.e., mostrando cuáles son los errores o las limitaciones que pretende subsanar), cuáles son las ventajas que presenta y cuáles son sus debilidades y sus puntos vulnerables que la hacen (en mi opinión) insostenible. Después de haber expuesto las razones que nos llevan a rechazar el fisicalismo de instancia como teoría de la mente, continuaremos con nuestra exposición de las objeciones que habitualmente son presentadas contra el fisicalismo de tipo.

La motivación central detrás del surgimiento del monismo anómalo surge del rechazo de Davidson de la conclusión del argumento que reproducimos a continuación –un argumento que es implícitamente asumido como verdadero por los partidarios de la teoría de la identidad tipo (así como por buena parte de los materialistas contemporáneos):

> (1) Las relaciones causales requieren de leyes. No puede haber una relación causal que no sea la instancia de una ley bajo alguna descripción.[80]
>
> (2) Hay relaciones causales entre ciertos estados mentales y ciertos estados físicos.

[80] Armstrong, 1969, p. 84: "But in speaking of the sequence as a causal sequence, we imply that there is some description (not necessarily known to us) that falls under a law."

(3) Por lo tanto, se deriva de (1) y (2) que deben de existir leyes que conecten los estados mentales con los estados físicos (también llamadas leyes psicofísicas).

Ahora bien, es importante observar que Davidson rechaza la conclusión de este argumento en tanto se encuentra en contradicción abierta con una de las premisas básicas sobre la naturaleza de la mente que Davidson asume, a saber, que la atribución de estados intencionales como deseos y creencias para dar cuenta de las acciones está regulada por ciertos principios de racionalidad y coherencia.

(4) La atribución de estados intencionales como deseos y creencias para dar cuenta de las acciones está regulada por ciertos principios de racionalidad y coherencia, no por vínculos causales.[81]

(5) Estos principios de racionalidad y coherencia constituyen la esencia de lo mental.[82]

(6) El mundo físico no se encuentra sometido a estos principios de racionalidad y coherencia.[83]

(7) Supongamos por un momento que (3) es verdadera, i.e., que tenemos una serie de leyes que conectan nuestras creencias con nuestros estados neuronales.

(8) Por lo tanto, para determinar si una persona tiene una creencia C, basta con determinar si tiene un cierto substrato neuronal N; no se necesita determinar si C es racional y coherente con respecto a otras creencias y otros estados mentales.

(9) Sin embargo, (8) no es compatible con (4); por lo tanto, tenemos dos opciones: o bien (4) es errónea o bien (3) lo es.

---

[81] Davidson, 1995, p. 293: "Para que un deseo y una creencia expliquen correctamente una acción, tienen que causarla de una manera adecuada, tal vez mediante una cadena o proceso de razonamiento que se ajuste a estándares de racionalidad."

[82] *Ibid.*, p. 62: "Nosotros percibimos una criatura como racional en la medida que somos capaces de ver sus movimientos como parte de un modelo racional que también comprende pensamientos, deseos, emociones, y voliciones"

[83] *Ibid.*, p 292: "Estas condiciones [de coherencia, racionalidad y consistencia] no tienen eco en la teoría física, y es por eso que no podemos buscar más que correlaciones burdas entre los fenómenos psicológicos y los físicos."

En la medida que Davidson asume que (4) es verdadera, de ello se sigue que (3) debe de ser errónea Ahora bien, ¿cómo podemos deducir del hecho que (8) sea errónea que la teoría de la identidad clásica en de Smart, Feigl y Armstrong es errónea? Para comprender esto, es importante recordar algo que hemos señalado anteriormente: una ley sólo puede ser una ley en la medida en que vincula, no sucesos o estados particulares, sino sucesos o estados tipo (o, en otros términos, clases de sucesos o estados). Por lo tanto, si sostenemos que (3) es errónea, lo que queremos decir, según Davidson, es que "puede haber enunciados generales verdaderos que relacionen lo mental y lo físico (...) pero no son legaliformes."[84] A partir de esta propuesta causalidad no-nomológica de Davidson, se puede construir el siguiente argumento:

(10) Puede haber enunciados generales verdaderos que relacionen lo mental y lo físico, pero estos enunciados no son leyes.

(11) La relación de reducción tiene un carácter nomológico en la medida que es habitualmente considerada como una relación entre teorías. No se reduce una entidad mental M a una entidad neuronal N, sino la propiedad de ser la entidad M a la propiedad ser de la entidad N.

(12) Se deriva de (10) y (11) que lo mental es nomológicamente irreducible a lo neuronal y, por ende, a lo físico.

(13) El objetivo de los teóricos de la identidad tipo al establecer una identidad entre estados mentales y estados neuronales es poder reducir los primeros a los segundos.

(14) Se deriva de (12) y (13) que la teoría de la identidad tipo es errónea en tanto es reduccionista.

Ahora bien, a partir de la tesis de la irreductibilidad de lo mental con respecto a lo físico, Davidson deduce otra característica importante de su propuesta que consiste en el hecho de que, si bien los estados mentales no tienen vínculos causales con los estados físicos, al menos resulta plausible sostener que los estados mentales se encuentran en una relación de superveniencia con respecto a los estados físicos (i.e.,

[84] *Ibid.*, p. 274

que no puede haber dos sucesos iguales en todos sus aspectos físicos y distintos en algún aspecto mental sin que se altere algún estado físico):

(15) La tesis de superveniencia es consistente con la inexistencia de leyes psicofísicas.

(16) La inexistencia de leyes psicofísicas implica la irreductibilidad de lo mental con respecto a lo físico en virtud de (10) y (11).

(17) Por lo tanto, de (15) y (16) se deriva que la tesis de superveniencia es consistente con la tesis de la irreductibilidad, i.e., ⊢—TS ∧TI

(18) En virtud de la ley de simplificación, podemos derivar de (17) lo siguiente: ⊢—TS

Así pues, el corazón de la propuesta de Davidson consiste en sostener tres tesis básicas: (1) la tesis de que todos los estados (o sucesos) son físicos, (2) la tesis de que los sucesos mentales se encuentran relacionados causalmente con los sucesos físicos y (3) la tesis de que las leyes psicofísicas no pueden existir (i.e., la tesis de que todas las instancias de causalidad entre mente y cuerpo deben ser no-nomológicas). Es importante observar que, de la tercera tesis básica de Davidson, se deducen dos consecuencias esenciales que son la irreductibilidad de los conceptos mentales respecto a los físicos y la postulación de una relación de superveniencia de los estados mentales en los estados físicos del cerebro (en tanto que no puede haber un vínculo causal nomológico entre estados mentales y estados neuronales). Como podemos constatar, las dos principales ventajas del monismo anómalo como teoría de la mente consiste en que, por un lado, afirma que todos los sucesos (incluyendo los sucesos mentales) son sucesos físicos, por lo cual llena las expectativas de muchos materialistas con respecto a la necesidad de explicar la interacción de entre cuerpo y mente en el ámbito de un solo y mismo ámbito ontológico (sin necesidad de hacer "saltos ontológicos" como Descartes), y, por otro lado, afirma la irreductibilidad de los conceptos mentales respecto a los conceptos físicos, por lo cual llena las expectativas de algunos fenomenólogos. Esta segunda ventaja resulta decisiva para explicar el número significativo de adhesiones al monismo anómalo por parte de la comunidad filosófica. En efecto, como los proyectos filosóficos materialistas previos habían o bien ignorado sistemáticamente los rasgos fenoménicos de la mente (como

hicieron los conductistas) o bien fracasado en sus intentos de dar cuenta de ellos por medio una naturalización que pretendía describirlos a través de predicados tópico neutrales (como hicieron ciertos partidarios del fisicalismo de tipo como Smart),[85] se consideró que el monismo anómalo, que postulaba que los estados mentales eran en última instancia estados físicos con la salvedad de que algunas de sus propiedades no eran reducibles a propiedades físicas, podía satisfacer plenamente tanto las demandas de los materialistas (preocupados por la necesidad de eliminar todo rastro de "saltos ontológicos" de sus explicaciones de las interacciones entre cuerpo y mente) como de los fenomenólogos (preocupados por la necesidad de mantener una cierta independencia de la mente con respecto al cuerpo), por lo cual era muy plausible que la propuesta de Davidson constituyese una buena teoría de la mente. De hecho, muchos filósofos siguen sosteniendo en la actualidad que el monismo anómalo es una alternativa válida respecto al conductismo y la teoría de la identidad tipo. Sin embargo, en las dos últimas décadas, varios autores entre los que destacan principalmente Schiffer (1987), Kim (1993, 1998) y Haugeland (1998) han examinado los argumentos de Davidson a favor del monismo anómalo, sacando a la luz varias contradicciones y errores que se derivan de su razonamiento. Estas críticas han arrojado muy serias dudas acerca de la posibilidad de considerar el monismo anómalo como una buena teoría de la mente

Acaso la objeción más dura contra la propuesta de Davidson sea la que presenta Schiffer, que habremos de exponer a continuación. Esta objeción tiene la estructura básica de una demostración por reducción al absurdo: se asume en un principio que el monismo anómalo es verdadero para derivar ulteriormente de él consecuencias inaceptables. Schiffer imagina la siguiente situación: una mujer llamada Ava se dispone a cruzar la calle, observa que un auto viene sobre ella y regresa de nuevo a la acera. El argumento que, según Schiffer, presentan la mayoría de los teóricos de la identidad (sean de tipo o de instancia) para sostener que el hecho de que Ava crea que un auto va a embestirla es en realidad un estado físico (específicamente, un estado neuronal) es el siguiente (Schiffer, *op. cit.*, p. 146-149):

[85] Para observar la razón por la cual la tentativa de explicación de los qualia por medio de una naturalización que pretende describirlos a través de predicados tópico neutrales, mi lector puede referirse a la cuarta objeción a la teoría de la identidad que encontrará algunos párrafos más abajo. Cf. infra, p. 88

(1) Hubo en Ava una cadena ininterrumpida de sucesos neuronales que empezó con la estimulación de las células fotorreceptoras en sus ojos por la luz relejada por el auto y terminó con su paso atrás, y cada uno de esos sucesos neuronales fue una causa de aquel que le seguía y, por lo tanto, de su paso atrás.

(2) El hecho de que Ava creyese que un auto venía sobre ella fue una causa de su paso atrás.

(3) Por lo tanto, o bien el hecho de Ava creyese que un auto venía sobre ella es idéntico al suceso neuronal simultáneo que fue causa de su paso atrás, o bien el paso atrás de Ava fue causalmente sobredeterminado (i.e., hubo dos cadenas causales que provocaron su paso atrás, una de las cuales contiene meros sucesos causales mientras que la otra contiene al menos un suceso mental irreducible –el hecho de que Ava creyese que auto venía encima de ella.)

(4) Pero no hubo una sobredeterminación causal en (3). (Esta premisa se argumenta con base en el principio de economía ontológica)

(5) Por lo tanto, el hecho de que Ava creyese que un auto venía sobre ella es idéntico a un estado neuronal, i.e., hubo un solo evento que fue al mismo tiempo la adquisición de la creencia y una instancia de algún estado neuronal tipo.

Ahora bien, tanto los fisicalistas de tipo (como Place y Smart) como los fisicalistas de instancia (como Davidson) aceptarían la conclusión (5) del argumento anterior. Lo que distingue a Davidson es que, mientras para Smart y Place todas las propiedades del estado mental deben poder ser reducidas en última instancia a ciertas propiedades físicas, Davidson sostiene que algunas de estas propiedades mentales son irreducibles. Es precisamente a partir de esta tesis de irreductibilidad de las propiedades mentales aunada a la conclusión (5) que Schiffer puede construir un argumento para mostrar que la posición de Davidson (que implica un dualismo de propiedades) no puede dar cuenta del papel causal de las propiedades mentales (*Ibid.*, p. 150):

(6) Ava está en una instancia de estado neuronal N. N tiene b, la propiedad mental de ser una creencia de que un auto viene sobre uno; y b no es idéntica con ninguna propiedad que sea

intrínsecamente especificable en un idioma no intencional y no mentalista (i.e., b no es reducible).

(7) N tiene una propiedad neurofisiológica p tal que p es la propiedad más comprensiva incluida en la explicación del hecho que n es una causa de que Ava haya dado un paso atrás, por lo que p es necesaria y suficiente para N sea la causa del paso atrás.

(8) Pero si hay una propiedad mental b, entonces b también es una propiedad causal esencial de N con respecto al hecho de que N sea una causa de que Ava haya dado un paso atrás (i.e., si N no tuviera b, N no habría podido causar el paso atrás de Ava).

Como podemos ver, la propuesta de Davidson que consiste en postular que las propiedades mentales son irreducibles con respecto a las propiedades físicas conlleva el grave problema de que, bajo esta óptica, las propiedades mentales pierden toda eficacia causal y se convierten en entidades vagas y superfluas. Esto evidentemente atenta contra todas nuestras intuiciones que sostienen que es precisamente porque un estado mental M tiene una cierta propiedad mental Q (e.g., la propiedad de ser una creencia de que va a llover) que está inmerso en una relación causal con un cierto estado físico F con una propiedad física P (e.g., tomar un paraguas antes de salir a la calle).

El argumento que Kim presenta contra el monismo anómalo también apunta hacia la misma dirección. Según Kim, si asumimos por un momento que el monismo anómalo (que es para Davidson la única alternativa válida para constituir una buena teoría de la mente ante el rechazo de la teoría de la identidad tipo) es correcto, esto implica que no podemos decir que un estado mental M sea causa de un estado neuronal N: lo más que podemos decir es que M tiene una cierta propiedad Q tal que existe una ley física adecuada que conecta a Q con una propiedad física F de N. En la medida que no existen leyes psicofísicas en el ámbito del monismo anómalo, la propiedad Q sólo puede ser una propiedad física. Ahora bien, como una de las consecuencias del monismo anómalo es la irreductibilidad nomológica de lo mental, es muy plausible suponer que M tiene ciertas propiedades mentales distintas de Q e irreductibles a propiedades físicas. El problema central del monismo anómalo surge de la caracterización del estatus de estas propiedades mentales como Kim lo señala al escribir lo siguiente:

> Parece entonces que, en el monismo anómalo, las propiedades mentales son ociosos causales sin ningún trabajo que hacer. Sin duda alguna, el monismo anómalo no es epifenomenalismo en el sentido clásico en tanto que permite que los eventos mentales sean causas de otros eventos. El punto es que es un epifenomenalismo sobre propiedades mentales (...) puesto que hace las propiedades y los tipos mentales sean causalmente irrelevantes. Al menos, no tiene nada que decir sobre cómo las propiedades mentales pueden ser causalmente relevantes mientras que afirma que todo evento que se encuentra en una relación causal lo está en virtud de sus propiedades físicas.[86]

Ahora bien, si el monismo anómalo niega que las propiedades mentales sean causalmente relevantes, esto implica que los estados de dolor podrían tener una propiedad mental distinta de la "dolorosidad" sin que por ello dejasen de producir conductas como gemidos o muecas. Esta conclusión resulta absurda y completamente contra-intuitiva respecto a la necesidad que hemos establecido de que una buena teoría de la mente de cuenta del vínculo causal entre estados mentales y estados físicos. Por lo tanto, es claro que el monismo anómalo no puede ser una buena teoría de la mente.

Antes de abandonar el monismo anómalo, es conveniente que precisemos un último punto. Hemos visto que, si bien Davidson rechaza que exista un vínculo causal *nomológico* entre los estados mentales y los estados físicos, admite que existe un cierto tipo de relación causal mucho más débil que la causalidad nomológica entre las instancias de estados mentales y las instancias de estados físicos: la relación de superveniencia de los estados mentales en los estados físicos. Considerando que dispone de este instrumento teórico, un partidario del monismo anómalo podría argüir lo siguiente: "No importa que las objeciones de Schiffer y Kim demuestren que las propiedades mentales no tienen eficacia causal, en la

[86] Kim, 1998, p. 138: " It seems, then, that under anomalous monism mental properties are causal idlers with no work to do. To be sure anomalous monism is not epiphenomenalism in the classic sense, since mental events are allowed to be causes of other events. The point is that it is an epiphenomenalism about mental properties (...) in that it renders mental properties and kinds causally irrelevant. At least it has nothing to say about how mental properties might be causally relevant, while affirming that every event that enters into a causal relation does so on account of its physical properties."

medida en que es posible sustituir la noción de eficacia causal de las propiedades mentales (basada en el concepto nomológico de causalidad) por la noción de superveniencia. Aun cuando la noción de superveniencia es más débil que la causalidad nomológica, sin duda es lo bastante fuerte para rechazar las críticas que hacen del monismo anómalo un tipo de epifenomenalismo." Esta estrategia de defensa del monismo anómalo con base en la noción de superveniencia ha sido criticada por Schiffer que señala que la noción de superveniencia es terriblemente confusa y que, al ser introducida en un argumento filosófico, lo único que hace es oscurecer la interpretación del argumento de tal manera que se torna imposible derivar con certeza algo interesante de él. Para mostrar que la noción de superveniencia es terriblemente confusa, no sólo en el terreno de la filosofía de la mente, sino en cualquier campo de la filosofía, Schiffer cita un caso muy célebre en el cual la introducción de la noción de superveniencia para la explicación de la naturaleza de ciertas propiedades morales en el ámbito de la ética provoca según él confusión en vez de contribuir a la solución del problema:

> G.E. Moore era un no naturalista en ética: sostenía que las propiedades morales no podían ser identificadas con propiedades naturales; y sostenía, en una faceta positiva, que estas propiedades eran simples, irreducibles, no analizables y no naturales y que, al ser no naturales, eran discernidas por medio de una facultad de especial de la intuición moral. Los fisicalistas más testarudos (incluyendo muchos positivistas lógicos) aceptaban que las propiedades morales no podían ser reducidas a propiedades naturales, pero no tenían ninguna simpatía respecto a la tesis positiva de Moore, que postulaba un reino de propiedades y hechos no naturales. Se pensaba que no podía darse cuenta de estas propiedades dentro de la imagen científica del mundo, que eran oscurantistas y que producían más problemas de los que resolvían. Al mismo tiempo, los filósofos que aborrecían las propiedades irreducibles y no-naturales de Moore sabían que también tenía esta tesis acerca de ellas: no era posible para dos cosas o sucesos idénticos en todos sus aspectos físicos diferir con respecto a alguna propiedad moral. Nadie pensaba que la teoría positiva de las propiedades morales de Moore era de alguna manera mitigada por esta tesis ulterior de superveniencia ¿Cómo podría el

> hecho de decir que las propiedades morales no naturales se encuentran en una relación de superveniencia con respecto a las propiedades físicas hacerlas más digeribles? Al contrario, invocar una primitiva y especial relación metafísica de superveniencia para explicar cómo propiedades morales no naturales están relacionadas con propiedades físicas es sólo agregar un misterio a un misterio, cubrir una maniobra obscurantista con otra.[87]

Por lo tanto, si la superveniencia es en verdad una noción oscura y confusa que, al ser introducida en un argumento, sólo contribuye a confundir las cosas en tanto no tiene ningún valor explicativo real, es evidente que un partidario del monismo anómalo no puede sostener que la noción de superveniencia constituye una herramienta teórica eficaz para poder rechazar las acusaciones de epifenomenalismo presentadas por Schiffer y Kim. De hecho, Schiffer sostienen que, en última instancia, "la superveniencia es sólo epifenomenalismo sin causación."[88] Por lo tanto, es evidente que, aun cuando se intente sostener el monismo anómalo como teoría de la mente con base en la noción de superveniencia, ello no basta para escapar a las acusaciones de epifenomenalismo que hacen mella en la teoría. Tras haber expuesto las razones las motivaciones que hicieron surgir el fisicalismo de instancia como alternativa, las ventajas que presenta como teoría de la mente así como los errores y las contradicciones que se derivan de él y que nos permiten rechazarlo como teoría de la mente, regresamos a la exposición de los argumentos en contra del fisicalismo de instancia.

---

[87] Schiffer, *op cit.*, p. 153-154: "G. E. Moore was a non-naturalist: he held that moral properties could not be identified with natural properties; and he held, on the positive side, that they were simple, irreducible, unanalyzable, non-natural properties, and that, being non-natural, they were discerned through a special faculty of moral intuition. Tough-minded physicalist types (including many Logical Positivists) agreed that moral properties could not be reduced to natural properties, but had no sympathy at all with Moore's positive thesis, which postulated a realm of non-natural properties and facts. These properties, it was felt, could not be made sense of within a scientific world view; they were obscurantist and produced more problems that they solved At the same time, philosophers who abhorred Moore's irreducibly non-natural properties knew that he also held this thesis about them: that it was not possible for two things to be alike in all physical respects while differing in some moral property. No one thought that Moore's positive theory of moral properties was in any way mitigated by this further supervenience thesis. How could being told that non-natural moral properties stood in the supervenience relation to physical properties make them any more palatable? On the contrary, invoking a special primitive metaphysical relation of supervenience to explain how non-natural moral properties were related to physical properties was just to add mystery to mystery, to cover an obscurantist move with another."

[88] *Ibid*, p. 154: "Supervenience is just epiphenomenalism without causation."

El cuarto argumento contra la teoría de la identidad tipo surge de una línea de pensamiento muy interesante según la cual, en tanto los análisis tópico neutrales parecen constituir un cimiento importante sobre la cual reposa la validez de la tesis básica de la teoría de la identidad, acaso el señalamiento de que los análisis neutrales de la mente son imposibles en los términos que los plantean los materialistas del estado central baste para mostrar que la teoría de la identidad mente-cerebro es incorrecta. A continuación, reproducimos el razonamiento de Coder destinado a demostrar la imposibilidad de realizar análisis neutrales de la mente.

A pesar de que existen poderosas evidencias empíricas proporcionadas por la ciencia según las cuales el materialismo del estado central es probablemente verdadero, los teóricos de la identidad reconocen que aún no tienen un instrumento que nos permita traducir integralmente nuestro lenguaje de estados mentales a un lenguaje científico que verse sobre los estados materiales del cerebro (como el concepto estricto de identidad que tienen lo exige). Así pues, al no disponer de este instrumento, los teóricos de la identidad no pueden responder a la pregunta que interroga sobre la naturaleza de los estados mentales en los términos que desean, que serían en términos de meros estados neuronales, sino que sólo pueden hacerlo de manera indirecta a través del análisis tópico neutral de conceptos como los de pensamiento, intención, creencia, sensación, experiencia, etc.

Ahora bien, cuando traducimos una experiencia a un predicado tópico neutral, ¿qué estamos haciendo realmente? Para poder apreciar esto, consideremos el siguiente caso de traducción de una experiencia a un lenguaje tópico neutral: asumamos que la proposición "X siente un dolor agudo en su pulgar derecho" es equivalente a la proposición "Algo le ocurre a X que es como aquello que le ocurre cuando se clava una aguja en su pulgar derecho y gime." Si la segunda proposición expresa en verdad un predicado tópico neutral, ¿por qué este predicado contiene términos físicos? ¿Acaso la tópico neutralidad no implica que es menester eliminar tanto los predicados mentales como los predicados físicos al realizar la traducción de las proposiciones que expresan estados mentales a las proposiciones tópico neutrales? Esto es verdadero sin duda alguna, pero la imposibilidad de llevar a cabo la traducción en los términos que queremos arroja muy serias dudas sobre la validez de la noción de tópico neutralidad como señala Coder en el pasaje que citamos a continuación:

> Sea '*Fx*' un predicado tópico neutral que define algún tipo de estado mental· '*x* es un estado de depresión', por ejemplo, significa '*Fx*', donde '*Fx*' es tópico neutral del tema. No puede ser entonces el caso que aquello que hace que un estado de depresión sea un estado mental sea algo además de lo que lo hace ser un estado de depresión. Por lo tanto, la idea de lo mental debe estar ya contenida en '*Fx*', que tiene la forma '*x* es el estado de una persona que se encuentra en tal y tal relación respecto al comportamiento y los estímulos'.[89]

Como podemos constatar, el rechazo de Coder a los análisis neutrales de la mente parte del hecho que los materialistas, al llevarlos a cabo, necesitan ir más allá de lo tópico neutral para intentar explicar la naturaleza de los estados mentales, lo que es una contradicción puesto que la definición de lo mental debe estar encerrada en la mera naturaleza de lo tópico neutral. De hecho, lo que Coder muestra es que hay en los predicados tópico neutrales una petición de principios. Cuando se define una experiencia en términos de situaciones típicas de estímulos o de conductas (i.e., cuando una experiencia es definida en términos extra-conceptuales), se introduce subrepticiamente en el predicado tópico neutral un término físico que corresponde a las situaciones típicas. Así pues, al ir más allá de la tópico neutralidad en su intento de establecer la naturaleza de los estados mentales, los materialistas del estado central no hacen otra cosa más que separar la noción de estado mental en dos elementos distintos de los cuales retoman uno como bien señala Rorty que sigue el razonamiento de Coder:

> Podemos decir que la falta de un minucioso análisis neurológico promovió la noción de que hay algo distintivo acerca de la mente -que debe ser algo espiritual- pero esta táctica simplemente divide la noción tradicional de lo mental en dos partes: el papel causal y la

[89] Coder, 1973, pp. 292-293: "Let '*Fx*' be topic neutral predicate that defines some sort of mental state '*x* is a state of depression', let us say, means '*Fx*', where '*Fx*' is topic neutral. Now it cannot be that what makes a state of depression a mental state is something in addition to what makes it a state of depression So the idea of mentality must already be contained in '*Fx*', which has the form '*x* is a state of a person that stands in such and such relation to behavior and stimuli.'

esencia transparente que se piensa juega este papel causal. Los análisis neutrales del tema obviamente no pueden, y no quieren aprehender esta última parte. Pero parece ser mera prestidigitación el dividir nuestro concepto de 'estado mental' en la porción que es compatible con el materialismo y la porción que no lo es, y luego decir que sólo la primera es esencial para el concepto.[90]

El quinto argumento habitualmente presentado contra la teoría de la identidad consiste en mostrar que un mismo estado neuronal puede tener contenidos intencionales distintos -dichos contenidos apuntan a la existencia de estados mentales distintos. Se ha comprobado que cuando tenemos un sujeto dado al cual se presenta un dibujo que contiene una ilusión óptica de la psicología de la *Gestalt* (e.g., el florero cuya silueta también es la silueta de dos personas a punto de besarse), puede declarar en un momento que el dibujo representa un florero y en otro momento que el dibujo representa la silueta de dos personas a punto de besarse. Como en los momentos precisos en que la persona en cuestión ha hecho estas declaraciones han sido llevados a cabo sendos exámenes neurológicos minuciosos (que incluyen electroencefalografía y TRM) del cerebro del individuo, y estos exámenes han revelado que sus ondas cerebrales son idénticas y que las mismas áreas neuronales se encuentran activadas en las dos situaciones, de ello se deduce que un mismo estado neuronal puede tener dimensiones intencionales distintas.[91] Como la intencionalidad es uno de los criterios básicos para reconocer un estado mental particular como tal (así como pudimos constatar en el capítulo anterior a través del ejemplo de los ratones), es claro que, si hay dimensiones intencionales

[90] Rorty, 1979, p. 116: "We may say that the lack of a fine-grained neurological account promoted the notion that there is something distinctive about the mind -that it must be something ghostly- but this tactic simply splits the traditional notion of the mental into two parts: the causal role and the Glassy Essence believed to play this causal role. Topic-neutral analyses obviously cannot capture, and do not want to capture the latter. But it seems mere gerrymandering to split our concept of a 'mental state' into the portion which is compatible with materialism and the portion which is not, and then say that only the former is essential to the concept."

[91] Flanagan, 1992, p. 117: "Strictly speaking, there is only one configuration before one's eyes [and hence, only one brain state], but it can be seen in either two ways, as a vase or as a pair of faces (...) Gestalt illusions show that something metaphysically unproblematic may be seen, known or described in two different ways (...) But remember that Gestalt illusions are illusions. There is just one thing there." Para demostrar que un estado neuronal no basta para instanciar un estado intencional, también se puede recurrir al argumento de la Tierra Gemela presentado por Putnam, 1975, pp. 223-227.

distintas, tiene que haber estados mentales distintos, lo cual constituye una refutación clara de la teoría de la identidad (de tipo).

El sexto argumento presentado habitualmente para rechazar la teoría de la mente consiste en mostrar que la noción de la identidad sostenida por los materialistas del estado central es errónea. Para los materialistas del estado central, en tanto la identidad entre mente y cerebro se establece por medio de una investigación empírica (al igual que la identidad entre un gen y una molécula de ADN), se puede deducir de ello que se trata de una *identidad contingente*:

> Si hay algo cierto en filosofía, es cierto que "La mente es el cerebro" no es una verdad lógicamente necesaria. (...) Si es verdad que la mente es el cerebro, entonces un modelo debe ser encontrado entre los enunciados de identidad contingente. Se debe comparar el enunciado a "La estrella matutina es la estrella vespertina" o a "El gen es la molécula de ADN", o a alguna otra aserción de identidad contingente.[92]

A primera vista, el concepto de *identidad contingente* parece tener una cierta justificación si apelamos al concepto de mundos posibles propio de la lógica modal. Retomando la idea de Leibniz según la cual un enunciado es necesario si es verdadero en todo mundo posible, podemos apreciar que un enunciado como "1+2=3"es necesario puesto que, según la definición axiomática de los números naturales presentada por Peano,[93] "1+2=3" es verdadero en todo mundo posible donde el número 2 sea el sucesor del número 1 y el número 3 sea el sucesor del número 2. Como el número 2 no puede ser lo que es si no es sucesor del número 1 (y el número 3 no puede ser lo que es si no es sucesor del número 2), entonces en todo mundo posible el enunciado "1+2=3" debe ser verdadero. Por otro lado, un enunciado de identidad como "Benjamín Franklin es el inventor de los bifocales" *parece* expresar una identidad contingente en tanto

[92] Armstrong, *op. cit.*, pp. 76-77: "If there is anything certain in philosophy, it is certain that 'The Mind is the brain' is not a logically necessary truth. (...) So if it is true that the mind is the brain, a model must be found among contingent statements of identity. We must compare the statement to 'The morning star is the evening star' or 'The gene is the DNA molecule', or some other contingent assertion of identity."

[93] De hecho, el establecimiento de lo que se tradicionalmente se conoce como los "axiomas de Peano" es en realidad producto de un matemático alemán, Richard Dedekind.

que es posible imaginar un mundo posible en el cual exista (o haya existido) Benjamín Franklin sin que este individuo sea (o haya sido) el inventor de los bifocales. Los materialistas del estado central (así como los demás partidarios de las identidades contingentes) argumentan que "Benjamín Franklin es el inventor de los bifocales" es una identidad contingente en la medida que el descriptor "el inventor de los bifocales" no nombra una propiedad esencial de Benjamín Franklin (i.e., no es una propiedad que Benjamín Franklin deba o hubiera debido tener para ser Benjamín Franklin) mientras que el número 2 debe ser sucesor del número 1 para poder ser el número 2. Sin embargo, Kripke critica esta distinción señalando lo siguiente.

> El que un particular tenga necesaria o contingentemente una propiedad depende de la manera como se lo describa Esto se halla quizás estrechamente relacionado con la tesis de que la manera como nos referimos a las cosas particulares es mediante una descripción. (..) Si consideramos el número 9. ¿tiene la propiedad de ser necesariamente impar? ¿Tiene que ser impar en todos los mundos posibles? Ciertamente es verdadero que en todos los mundos posibles que el nueve es impar; digamos que no podría haber sido de otra manera. Pero, por supuesto, el 9 podría haber sido seleccionado como el número de planetas. No es necesario, no es verdadero en todos los mundos posibles que el número de planetas sea impar [94]

Como podemos ver claramente en este pasaje, la distinción entre propiedades esenciales y propiedades contingentes sobre la cual se encuentra basada la distinción entre identidades necesarias e identidades contingentes parece presentar un problema. Aun cuando el número 9 sea idéntico al número de planetas en nuestro sistema solar, es por demás claro que no podemos sustituir "X" por las expresiones "el número de planetas de nuestro sistema solar" y "el número 9" en la expresión "X es necesariamente impar" de manera uniforme. De acuerdo con Kripke, esto se debe a que la expresión "el número 9" es un designador rígido (i.e., es algo que en todo mundo posible designa al mismo objeto) mientras que la expresión "el número de planetas de nuestro sistema solar" es un designador no-rígido puesto que, si bien designa en el

[94] Kripke, 1995, pp. 43-44

estado de cosas actual el número 9, puede designar en un cierto mundo posible (e g., un mundo posible en el cual Júpiter se convierte en un sol por intervención de una raza extraterrestre, como en *2010 Odisea Dos*), el número 8.

A partir de esta distinción entre designadores rígidos y designadores no-rígidos, Kripke señala que, aun cuando enunciado de identidad puede expresar un hecho contingente en el caso en que al menos una de las expresiones que liga la relación de identidad sea un designador no-rígido, el enunciado siempre es necesario.[95] Ahora bien, en el caso de los enunciados de las diversas ciencias particulares que expresan identidades teóricas (e.g., "El agua es H_2O", "El calor es movimiento molecular", "La luz es un grupo de partículas llamadas fotones que vibran a una determinada longitud de onda", "El relámpago es una descarga eléctrica", etc.), tradicionalmente se asume que son identidades contingentes en la medida que han sido descubiertas por medios empíricos.[96] Sin embargo, Kripke sostiene que, aun cuando estas identidades son establecidas por medios empíricos, esto no implica que sean contingentes. Para demostrar esto, Kripke presenta el siguiente argumento que reproducimos a continuación:

> ¿Podría algo ser oro sin tener el número atómico 79? Supongamos que los científicos han investigado la naturaleza de oro y han encontrado que es parte de la naturaleza misma de esta sustancia, por así decirlo, que tiene el número atómico 79. Supongamos ahora que encontramos otro metal amarillo, u otra cosa amarilla, con todas las propiedades mediante las

[95] *Ibid.*, pp. 97-98: "Si es verdad que el hombre que inventó los lentes bifocales era el primer director general de correos de Estados Unidos -que éstos eran uno y el mismo- esto es contingentemente verdadero. Es decir, podría haber sido el caso que un hombre inventara los lentes bifocales y otro fuese el director general de correos de Estados Unidos. Así, ciertamente, cuando hacemos enunciados de identidad usando descripciones, cuando decimos que 'el x tal que φx y el x tal que ψx son uno y el mismo'; esto puede ser un hecho contingente."

[96] En este punto, tocamos la raíz de uno de los errores más importantes tradicionalmente cometidos en filosofía - error que probablemente debemos a Kant. Kant sostenía que los juicios de identidad necesarios coincidían con lo que es *a priori* (i.e., con lo que puede ser comprendido por medio del mero análisis conceptual) mientras que los juicios de identidad contingentes coincidía con lo que es *a posteriori* (i.e., con lo que puede sólo ser comprendido por medio de la experiencia). Ambas categorías de juicios eran mutuamente exclusivas para Kant Sin embargo, ni todas las verdades necesarias son *a priori* (como bien muestra Kripke) ni todas las verdades contingentes son *a posteriori*. La raíz de este error kantiano radica, según algunos autores, en el siguiente hecho (Coffa, 1991, p 20): "[Kant] confused conceptual knowledge with definitional knowledge; that is, he confused what can be grounded on concepts with the much smaller subclass of what can be grounded on definitions."

cuales identificamos originalmente el oro y muchas propiedades adicionales que hemos descubierto posteriormente. Un ejemplo de una cosa con muchas propiedades iniciales es la pirita de hierro, el "oro de los tontos." Como he dicho antes, no diríamos que esta sustancia es oro. Hasta aquí estamos hablando del mundo real. Ahora consideremos un mundo posible. Consideremos una situación contrafáctica en la que, digamos, se encontrara realmente pirita de hierro u "oro de los tontos" en varias montañas de Estados Unidos o en partes de la Unión Soviética y de Sudáfrica. Supongamos ahora que todas las áreas que realmente contienen oro contuviesen pirita en lugar de oro, o alguna otra sustancia que simulara las propiedades superficiales del oro, pero careciese de su estructura atómica. ¿Diríamos de esta situación contrafáctica que en esa situación el oro no habría sido ni siquiera un elemento (...)? Me parece que no lo diríamos. Describiríamos esa situación más bien como una en la que se hubiera encontrado, en las mismas montañas que de hecho contienen oro, una sustancia, digamos la pirita de hierro, que no es oro, y que tendría las mismas propiedades mediante las que comúnmente identificamos el oro. Pero no sería oro, sería algo diferente.[97]

Este argumento muestra, como bien señala Kripke, que un enunciado de una teoría científica como "el oro es el metal que tiene por número atómico 79" –un enunciado del cual tradicionalmente se piensa que expresa una identidad contingente– expresa en realidad una identidad necesaria. Al aplicar el mismo tipo de razonamiento a las *identidades contingentes* establecidas por los materialistas del estado central entre estados mentales y estados neuronales, Kripke señala que dichas identidades no pueden ser contingentes, sino necesarias.

El ejemplo que toma Kripke para mostrar esto de manera patente es la identidad "El dolor es una fibra C estimulada", en la cual tanto el término "dolor" como el término "fibra C estimulada" son designadores rígidos según él. Por un lado, la expresión "fibra C estimulada" debe ser un designador rígido puesto que "un mundo en el cual ninguna fibra C excitada ocurre es un mundo en el cual este evento, que es una fibra

[97] Kripke, *op. cit.*, pp. 121-122

C excitada, no ocurre."[98] Por otro lado, la expresión "dolor" también debe ser un designador rígido puesto que "si algo es un dolor, lo es esencialmente y parece absurdo suponer que el dolor podría haber sido algún fenómeno distinto del que es."[99] De todo lo anterior se deduce que la identidad expresada en el enunciado "El dolor es una fibra C excitada" debe ser necesaria, con lo cual el materialismo del estado central que se encuentra basado en el establecimiento de una serie de identidades contingentes entre estados mentales y estados neuronales queda descartado como opción válida. Ahora bien, acaso sea posible que un teórico de la identidad acepte los argumentos de Kripke y sostenga que la identidad mente-cuerpo es necesaria Sin embargo, esta tesis plantea un serio problema a partir del cual se constituye otra importante objeción a la teoría de la identidad que examinaremos a continuación.

Antes de analizar el séptimo y último argumento presentado habitualmente contra el materialismo del estado central, acaso sea relevante recordar primero un rasgo del conductismo analítico de Ryle. Cuando Ryle propuso la sustitución de los estados mentales por disposiciones de conducta, percibió la necesidad de mantener el antiguo papel causal de los estados mentales en el nuevo esquema explicativo de la mente que estaba desarrollando. Ahora bien, una de las tesis tradicionalmente más importantes de la psicología mentalista consistía en el hecho que los estados mentales (las actitudes proposicionales, en especial) no tienen una realizabilidad unívoca, sino múltiple. Dicho en otras palabras, cualquier estado o suceso mental complejo se puede expresar por medio de una gran variedad de conductas distintas. Por ejemplo, el deseo que tengo de un helado se puede expresar a través de la conducta verbal que consiste en que pronuncie la oración "Quiero un helado", se puede expresar a través del hecho de ir caminando a la heladería más cercana o de varias otras maneras. Ryle intentó recuperar en su teoría la noción de la realizabilidad múltiple de los estados mentales a través de las llamadas *disposiciones multi-track* que consisten en lo siguiente:

[98] Kim, *op. cit.*, p. 69: "A world in which no C-fiber excitation ever occurs is a world in which this event, which is a C-fiber excitation, does not occur."
[99] Kripke, *op cit.*, p. 144

> Las disposiciones elevadas de la gente de las cuales esta investigación se ocupa no son, por lo general, disposiciones unívocas sino disposiciones cuyas realizaciones son indefinidamente heterogéneas. Cuando Jane Austen quiso mostrar el tipo específico de orgullo que caracteriza a la heroína de *Orgullo y prejuicio*, tuvo que representar sus acciones, palabras, pensamientos y sentimientos en miles de situaciones distintas. No hay ningún tipo de acción o reacción estándar tal que Jane Austen pudiese decir "El tipo de orgullo de mi heroína tenía justo la tendencia de hacer esto, cuando una situación de este otro tipo ocurría".[100]

Como podemos apreciar claramente, las disposiciones multi-track constituían una propuesta interesante en la medida que no ligaban de manera restrictiva un estado mental con una conducta particular, sino con un conjunto de conductas. Lamentablemente, el materialismo del estado central no pudo recuperar esta noción de realizabilidad múltiple de los estados mentales en la medida en que postulaba una identidad estricta entre los estados mentales y los estados neuronales. Como esta identidad no puede ser contingente sino que debe de ser necesaria (según los argumentos de Kripke que aceptamos plenamente), la conclusión de esto consiste en el hecho que, para tener estados mentales, es menester tener estados neuronales, i.e., es menester tener un cerebro biológico. Al aceptar esta identidad estricta, la teoría de la identidad mente-cerebro rechazaba la posibilidad de realizar una de las grandes aspiraciones del pensamiento materialista clásico que consistía en la creación de autómatas que duplicasen los procesos mentales del hombre. Esta limitación se convirtió en el talón de Aquiles del materialismo del estado central que se presentó a los ojos de sus detractores como una teoría chauvinista que negaba la posibilidad de que sistemas carentes de cerebros (e.g., computadoras o extraterrestres con constituciones biológicas completamente distintas a la nuestra) tuviesen mentes. La crítica más severa contra la teoría de la identidad desde la perspectiva de la carencia de una noción de realizabilidad múltiple surgió de la obra de

[100] Ryle, *op. cit.*, p. 44: "The higher-grade dispositions of people with which this inquiry is largely concerned are, in general, not single-track dispositions, but dispositions the exercises of which are indefinitely heterogenous. When Jane Austen wished to show the specific kind of pride which characterized the heroine of *Pride and Prejudice*, she had to represent her actions, words, thoughts and feelings in a thousand different situations. There is no one standard type of action or reaction such that Jane Austen could say 'My heroine's kind of pride was just the tendency to do this, whenever a situation of that sort arose'."

Putnam que señala lo siguiente acerca de las implicaciones de una identidad estricta entre mente y cerebro:

> Consideremos lo que el teórico de los estados cerebrales tiene que hacer para dar validez a sus afirmaciones. Tiene que especificar un estado fisicoquímico tal que cualquier organismo (no solamente un mamífero) siente dolor si y sólo si (a) posee un cerebro con una estructura fisicoquímica adecuada y (b) su cerebro está en ese estado fisicoquímico. Esto significa que el estado fisicoquímico en cuestión tiene que ser un estado posible de un cerebro de mamífero, de un cerebro de reptil, de un cerebro de molusco (...) Al mismo tiempo, tiene que ser un estado que no sea posible (físicamente posible) para el cerebro de ninguna criatura físicamente posible que no pueda sentir dolor.[101]

Como podemos constatar, ninguno de los dos requerimientos citados por Putnam puede ser llenado por un teórico de la identidad tipo, lo cual constituye un serio problema. El requerimiento de identidad propuesto por el materialismo del estado central en sus exposiciones clásicas (Feigl, Smart, Place, etc.) es tan estricto que basta con que encontremos dos organismos (e.g., un perro y un pulpo) a los cuales se pueda aplicar un mismo predicado psicológico (e.g., "tiene dolor") en una situación dada, pero que tengan estados físico-químicos distintos para socavar la tesis básica de la teoría de la identidad mente-cerebro. Retomando la primera objeción que hemos presentado, podemos entonces constatar que la teoría de la identidad no puede ser una teoría satisfactoria de la mente en la medida que si lo fuera, quizás se podría determinar que la propiedad de ser una fibra C activada nombrada por el concepto de *fibra C activada* es realizada por exactamente los mismos objetos que realizan la propiedad de dolorosidad nombrada por el concepto de dolor. Sin embargo, resultaría imposible determinar que la propiedad de dolorosidad

[101] Putnam, 1967b in Putnam, 1975, p. 436: "Consider what the brain state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (no just a mammal) is in pain if and only if (a) it possesses a brain of suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusk's brain (.) At the same time, it must not be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain."

nombrada por el concepto del dolor sea realizada por exactamente los mismos objetos que realizan la propiedad de ser una fibra C activada nombrada por el concepto de *fibra C activada* a menos de sostener una perspectiva chauvinista en la cual sólo los seres humanos puedan sentir dolor. Como la identidad es una relación simétrica, y no se puede demostrar que todas las instancias de dolor son instancias de fibra C activada a causa de la realizabilidad múltiple, entonces se deduce que la teoría de la identidad no puede ser una teoría satisfactoria de la mente

Ahora bien, un materialista del estado central puede intentar responder a la objeción de Putnam señalando que un estado mental es idéntico con la disyunción de los estados materiales (sea en un cerebro de hombre, en un cerebro de pulpo, en una computadora o en un marciano) que pueden realizar el estado mental en cuestión.[102] Putnam rechaza terminantemente esta estrategia señalando que, en caso de ser adoptada, la disyunción permanece siempre abierta: en el caso en que aparezcan sistemas nuevos con realizaciones distintas de la nuestra de un estado mental dado (e.g., dolor), el materialista del estado central puede ampliar la disyunción con nuevos supuestos *ad hoc*, haciendo entonces la disyunción infinita. Ahora bien, el problema de la infinitud de la disyunción es un problema práctico, no de principio. En el caso en que un materialista del estado central se sitúe en condiciones ideales (e.g., si imagina por un momento ser omnipotente, omnisciente y eterno como Dios), acaso la disyunción pueda ser cerrada. Sin embargo, además del problema práctico, existe un problema de principio en la estrategia señalada por Putnam que consiste en lo siguiente: la ley de correspondencia entre estados mentales y sus realizaciones a través de estados materiales se torna entonces tan amplia, i.e., con tantos casos particulares, que deja de tener interés desde un punto de vista científico. Si imaginamos que un estado mental (e.g., dolor) puede tener un número infinito de realizaciones materiales distintas, entonces se torna imposible elaborar una teoría científica que pueda dar cuenta de manera sistemática y unitaria del fenómeno del dolor. Esta objeción a la teoría de la identidad es de importancia capital puesto que, no sólo a partir de ella se desarrolló la corriente de pensamiento materialista que sucedió a la teoría de la identidad mente-cerebro

[102] *Ibid.*, p. 437: "The brain-state theorist can save himself by *ad hoc* assumptions (e g., defining the disjunction of two states to be a single 'physical-chemical state)."

(corriente conocida como Funcionalismo), sino que también constituye la base de una de las críticas más duras que se han elevado contra el funcionalismo también (principalmente contra la propuesta de Lewis), como habremos de ver en el próximo capítulo. Con la exposición del argumento de Putnam, damos por terminada esta tercera sección para pasar a la cuarta y última sección del presente capítulo donde realizaremos un balance general de la teoría de la identidad.

II.4 Un balance general de la teoría de la identidad mente-cerebro

Al hacer un balance general de la teoría de la identidad mente-cerebro, podemos apreciar que la teoría presenta varias ventajas y aciertos que es necesario recuperar si queremos desarrollar eventualmente una buena teoría de la mente. Acaso uno de los aciertos más importantes del materialismo del estado central consiste en haber desplazado el interés de los materialistas de las meras conductas observables a los estados neuronales. Este desplazamiento del interés del pensamiento materialista de las conductas a los estados neuronales es de gran importancia por dos razones. Por un lado, reivindica el carácter empírico y de observabilidad pública de las ciencias que estudian el cerebro (de la neurobiología, en particular), rechazando así la tesis chauvinista del conductismo que sostenía que sólo los datos de la conducta son confiables puesto que son los únicos datos empíricos y públicamente observables. Por otro lado, este desplazamiento de interés permite sentar bases sólidas sobre las cuales se puede estudiar las correlaciones causales entre mente y cerebro. Sobre esta observación, acaso sea necesario hacer una precisión. Como la teoría de la identidad sostiene que los estados mentales son idénticos con los estados neuronales, resulta virtualmente imposible sostener que hay un vínculo causal entre estados mentales y estados neuronales en un sentido interesante. En efecto, si aseveramos la verdad de A=B, ¿cómo podemos sostener la verdad de A es la causa de B, a no ser que estemos dispuestos a violar el principio que afirma que, en una relación causal, la causa debe ser distinta del efecto? Sin embargo, los teóricos de la identidad rescatan un cierto tipo de vínculo causal (causalidad diagonal) sosteniendo que éste ocurre, no entre estados mentales y estados neuronales, sino entre estados neuronales y conductas. Si aceptamos la sustitución del vínculo causal entre mente y cuerpo por un vínculo causal cerebro-cuerpo (con base en la identidad que existe de

acuerdo con los materialistas del estado central entre estados mentales y estados neuronales), podemos dar cuenta con mucho mayor claridad de las interacciones entre la "mente" y el cuerpo en la medida en que podemos explicarlas en un único ámbito ontológico. En particular, podemos dar cuenta de las oraciones de la psicología popular (e.g., "María creía que iba a llover, por lo que tomó un paraguas al salir de casa") remplazando los términos de la psicología mentalista tradicional por términos que designen estados neuronales (lo que nos daría algo como "María estaba en el estado neuronal S-453, por lo que tomó un paraguas al salir de casa"). Así pues, el materialismo del estado central presenta la gran ventaja de llenar el requisito de un vínculo causal entre mente y cuerpo que es necesario para constituir una buena teoría de la mente. Además, la teoría presenta otras pequeñas ventajas adicionales como el hecho que puede dar cuenta del problema de la existencia de otras mentes de manera clara.

Sin embargo, a pesar de estas ventajas, la teoría de la identidad mente-cerebro cuenta con múltiples errores y limitaciones que arrojan serias dudas respecto a su aceptación en tanto que una buena teoría de la mente. Hemos visto que los estados neuronales parecen no tener características fenoménicas intrínsecas como tienen los estados mentales. Esto sugiere que la teoría de la identidad mente-cerebro no puede dar cuenta del carácter intrínseco de los estados mentales, lo cual la inhabilita para llegar a constituir una buena teoría de la mente. Ahora bien, hemos visto que es posible adoptar una posición teórica en la cual podemos tener estados que tengan al mismo tiempo propiedades físicas y propiedades fenoménicas que sean irreducibles a las propiedades físicas del estado (nos referimos al fisicalismo de instancia). Por lo tanto, quizás sea posible sostener una versión de la teoría de la identidad basada en un fisicalismo de instancia como teoría de la mente. Sin embargo, hemos presentado anteriormente dos argumentos (el de Schiffer y el de Kim) que rechazan esta alternativa.

Aun cuando asumamos que podemos sostener una versión de la teoría de la identidad que pueda dar cuenta tanto de la causalidad entre la mente y el cuerpo como del carácter intrínseco de los estados mentales, existen otros problemas. Al igual que el conductismo en sus dos versiones principales, la teoría de la identidad parece no poder dar cuenta del carácter intencional de los estados mentales. Es posible que un individuo se encuentre en dos momentos distintos en un mismo estado neuronal, pero que los contenidos intencionales de este estado neuronal particular (y, por ende, los estados mentales

correspondientes) sean completamente distintos. Esto basta para mostrar que el materialismo del estado central no puede ser una buena teoría de la mente. Sin embargo, es posible refutar de manera más radical la teoría de la identidad mente-cerebro sin usar nuestros requerimientos (que, en última instancia, pueden ser considerados por un teórico de la identidad testarudo como *ad hoc*). Esta manera radical de refutar la teoría de la identidad consiste en mostrar que su tesis básica se encuentra en contradicción con las aspiraciones comunes de las corrientes clásicas de materialismo. Para llevar esto a cabo, se demuestra en primer lugar por medio de los argumentos de Kripke que si hay realmente una identidad entre estados mentales y estados neuronales, esta identidad es necesaria y no contingente (como pensaban los materialistas del estado central). Después de haber mostrado esto, se arguye que la consecuencia inmediata de esta identidad consiste en la aseveración de que es necesario tener un cerebro biológico para tener mente. Finalmente, se demuestra que esta conclusión es incompatible con la idea materialista según la cual los procesos y estados mentales de una persona dada pueden ser en principio reproducidos por sistemas no-humanos (como computadoras o extraterrestres con constituciones biológicas distintas a la nuestra). La incompatibilidad entre la tesis básica de la teoría de la identidad tipo, la tesis de la necesidad de la identidad y la tesis de la realizabilidad múltiple constituye el último clavo en el ataúd del materialismo del estado central.[103] Sin embargo, esta limitación apunta directamente a la nueva corriente de pensamiento materialista que fue desarrollada para resolver el problema de la realizabilidad múltiple de los estados mentales: el funcionalismo. El propósito del próximo capítulo será examinar esta corriente de pensamiento en detalle.

[103] De hecho, veremos en el capítulo siguiente que la tesis básica de la identidad tipo y la tesis de la realizabilidad múltiple no son completamente incompatibles entre sí como hemos señalado en este punto. Lewis argumenta que la teoría de la identidad tiene cabida dentro del funcionalismo en la medida que piensa el papel causal que constituye la característica definitoria de un estado mental puede ser realizado por otras entidades materiales distintas de los estados neuronales de los seres humanos La tesis básica de la identidad tipo y la tesis de la realizabilidad múltiple sólo resultan incompatibles cuando se piensa (como los materialistas del estado central lo hacen) que la única entidad que puede llenar el papel causal atribuido al estado mental es el estado neuronal de un ser humano

Capítulo III

El funcionalismo. Motivaciones, aciertos, errores y limitaciones en la explicación de la naturaleza de la mente.

En los dos capítulos anteriores, hemos señalado que el conductismo (en sus dos versiones principales) así como el materialismo del estado central, a pesar de sus virtudes como teorías explicativas de las relaciones entre la naturaleza de la mente, cuentan con problemas y limitaciones de diverso tipo que, al ser estudiados con sumo cuidado, revelan que ninguna de las dos teorías es propia para sentar las bases de una teoría de la mente satisfactoria. Por un lado, el conductismo se revela una teoría demasiado amplia en la medida en que sus criterios otorgan una mente a cosas que patentemente carecen de ella. Por otro lado, al intentar establecer identidades contingentes entre estados físicos del cerebro y estados mentales, el materialismo del estado central acepta, de manera implícita, que una de las condiciones necesarias para tener estados mentales es tener un cerebro. Esto excluye que entidades carentes de cerebros biológicos (e.g., computadoras o extraterrestres con constituciones biológicas completamente distintas a la nuestra) puedan tener estados mentales, lo cual constituye una posición chauvinista que contradice una de las grandes aspiraciones del materialismo clásico: la creación de autómatas que reproduzcan los procesos así como los estados mentales de los seres humanos. Si la realización de esta aspiración parecía bastante lejana en la época del materialismo clásico (donde los autómatas creados no eran sino máquinas muy simples como la sumadora de Pascal), el siglo XX presenció la creación de autómatas mucho más complejos basados en las ideas de Turing –autómatas que se acercaban mucho más al ideal del materialismo clásico. Aun cuando no se ha logrado todavía crear un autómata que reproduzca de manera exacta los procesos y los estados mentales de un ser humano, los progresos realizados en esta dirección desde la creación de las primeras computadoras en la década 1940-1950 sugieren a muchos materialistas que es menester desarrollar una propuesta filosófica que pueda dar cuenta de la posibilidad de que las computadoras reproduzcan algún día nuestros estados y procesos mentales. Como ni el conductismo en sus dos versiones principales ni la teoría de la identidad mente-cerebro parecen poder explicar esta posibilidad y respetar al mismo tiempo los tres requerimientos que hemos establecido para que algo pueda ser considerado una buena teoría de la mente, el funcionalismo, cuya tesis básica consiste en sostener que un estado mental M es una función cuya naturaleza consiste en mapear un conjunto particular de estados F a otro conjunto de estados P de acuerdo a una estructura causal, parece constituir una alternativa viable para ser una teoría satisfactoria de la mente, al menos en primera instancia, en la medida que permite dar

cuenta de la posibilidad de la realizabilidad múltiple. Así pues, partiendo de la hipótesis según la cual el funcionalismo puede ser (en principio, al menos) una buena teoría de la mente, veremos en la primera sección de este capítulo las tres versiones originales del funcionalismo que se desarrollaron en la década 1960-1970: las propuestas de Lewis, Putnam y Dennett. En esta primera sección, estudiaremos con cuidado los rasgos de las propuestas de los tres autores, intentando determinar en particular cuáles son los puntos de contacto entre ambos y cuáles son las divergencias que los separan. También intentaremos mostrar cuáles son las principales de ventajas así como las desventajas o limitaciones del funcionalismo en sus tres exposiciones principales, en la medida en que estas ventajas y estas limitaciones se encuentran también presentes en versiones ulteriores del funcionalismo. Habiendo precisado claramente los objetivos de esta primera sección, podemos comenzar con nuestra exposición de las tres propuestas originales.

III.1 Los orígenes del funcionalismo: las propuestas de Lewis, Putnam y Dennett

Para entender la aparición del funcionalismo en el ámbito de la discusión filosófica sobre la naturaleza de la mente en la década 1960-1970, es menester comprender que el funcionalismo no surgió de manera espontánea, sino que tiene varios antecedentes en las teorías que posteriormente criticaría y buscaría sustituir. De la misma manera que la teoría de la identidad se encuentra en germen en las tesis de Carnap, el funcionalismo se encuentra también en germen en los escritos de muchos teóricos de la identidad (sobre todo en la década 1960-1970, cuando la tesis básica de la teoría de la identidad comenzaba ser sometida a severas críticas por algunos autores).

III.1.1 La propuesta de Lewis

Sobre este respecto, el caso de Lewis es paradigmático como veremos a continuación. Al igual que la gran mayoría de los teóricos de la identidad (en particular, al igual que Armstrong), Lewis sostiene en un artículo titulado *An Argument for the Identity Theory* (1966) una teoría de la mente según la cual "la característica definitiva de cualquier (tipo de) experiencia [o estado mental] como tal es su papel causal,

su síndrome de causas y efectos más típicos."[104] Además de esta tesis causal, Lewis también afirma una segunda tesis básica que consiste en "la hipótesis plausible de que hay un cierto cuerpo unificado de teorías científicas (...) que proveen juntas una explicación verdadera y exhaustiva de todos los fenómenos físicos."[105] Con base en estas dos tesis, Lewis construye el siguiente argumento:

> (1) Si un estado mental está definido por su papel causal, y este papel causal es realizado por una serie de estados físicos, entonces existe una teoría científica que provee una explicación verdadera y exhaustiva de estos estados (i e., estos estados caen bajo el dominio de una ley causal particular)
>
> (2) O bien los estados mentales caen bajo el dominio de una ley causal particular, o bien no lo hacen.
>
> (3) Los estados mentales no pueden tener un estatus de funciones causales no-nomológicas porque, de ser así, la psicología no podría ser considerada una ciencia auténtica.
>
> (4) Por lo tanto, como los estados mentales se encuentran definidos por sus papeles causales, entonces caen bajo el dominio de una ley causal universal.

La conclusión de este argumento sirve a Lewis para sostener, en última instancia, que la teoría de la identidad es la mejor alternativa que tenemos para resolver el problema mente-cuerpo en tanto no disponemos actualmente de una alternativa mejor:

> Estamos lejos de establecer de manera positiva que los estados neuronales ocupan los papeles causales definitivos de las experiencias, pero no tenemos noción de otros fenómenos físicos que puedan ocuparlos, consistentemente con lo que sabemos. Entonces si los fenómenos no

[104] Lewis, 1966 en Lewis, 1983, p. 100: "The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects."
[105] *Ibid*, p. 105. "[My second premise is] the plausible hypothesis that there is some unified body of scientific theories (. .) which together provide a true and exhaustive account of all physical phenomena."

> físicos son descartados por nuestra confianza en la explicación física, sólo quedan los estados neuronales.[106]

Teniendo en cuenta lo que hemos señalado anteriormente, nuestros lectores pueden preguntarse porque la propuesta de Lewis no fue examinada en el capítulo anterior donde se estudiaron con detenimiento los rasgos centrales de la teoría de la identidad mente-cerebro. La respuesta a esta pregunta consiste en el hecho que Lewis afirma que su tesis supera las limitaciones del conductismo y de la teoría de la identidad en la medida que "nos permite incluir otras experiencias entre las causas y los efectos típicos a través de los cuales una experiencia es definida."[107] Este rasgo de la propuesta de Lewis es muy importante puesto que, por medio de él, se puede dar cuenta de la accesibilidad introspectiva de la experiencia que consiste en "su propensión a causar de manera constante otras (futuras o simultáneas) experiencias dirigidas sobre ellas intencionalmente, en las cuales somos conscientes."[108] La posibilidad de definir las experiencias (i.e., los estados neuronales) entre sí apunta a la existencia de familias de experiencias definidas entre sí (i.e., a tipos de estados mentales). Es precisamente la existencia de familias de experiencias definidas entre sí lo que constituye en la obra de Lewis un principio de propuesta funcionalista (y, por ende, una divergencia ligera respecto a la versión clásica de la teoría de la identidad) como podemos apreciar en el siguiente pasaje:

> Cualquier cosa que ocupe el papel causal definitivo de una experiencia en una familia como ésta lo hace en virtud de su pertenencia a un isomorfismo causal de la familia de experiencias, esto es, a un sistema de estados que tienen el mismo patrón de conexiones causales entre sí y

[106] *Ibid.*, p. 106: "We are far from establishing positively that neural states occupy the definitive causal roles of experiences, but we have no notion of any other physical phenomena that could possibly occupy them, consistent with what we do know. So if nonphysical phenomena are ruled out by our confidence in physical explanation, only neural states are left."
[107] *Ibid.*, p. 103: "[Second,] it allows us to include other experiences among the typical causes and effects by which an experience is defined."
[108] *Ibid.*, loc. cit.: "[The introspective accessibility of an experience is] its propensity reliably to cause other (future or simultaneous) experiences directed intentionally upon it, wherein we are aware of it."

las mismas conexiones causales con estados fuera de la familia, es decir, estímulos y conducta.[109]

En la cita anterior, aquello que sin duda alguna tiene una gran importancia es el uso de la expresión "cualquier cosa". Esta expresión revela que, aun cuando Lewis sostenga claramente una identidad entre estados mentales (que para él son experiencias) y estados neuronales, se encuentra abierto a la posibilidad de que haya otras cosas distintas de los estados neuronales de los seres humanos que ocupen el papel causal definitivo de las experiencias (i.e., se encuentra abierto a la posibilidad del funcionalismo). Ahora bien, ¿cómo podemos dar cuenta del hecho según el cual Lewis parece sostener al mismo tiempo una teoría de la identidad (que tiende a sustentar la tesis de la realizabilidad única) y la posibilidad de la realizabilidad múltiple de los estados mentales (que es un rasgo típico de una propuesta de carácter funcionalista)?

Para responder la pregunta anterior, retomemos un momento la definición que dimos de funcionalismo como la teoría que sostiene que un estado mental M es una función cuya naturaleza consiste en mapear un conjunto particular de estados F a otro conjunto de estados G, de acuerdo a una estructura causal. Esta última precisión es muy importante, puesto que existen funciones que no pueden ser interpretadas de manera causal (e.g., en la ley de hidrostática "La fuerza horizontal hacia arriba que ejerce el agua sobre un cuerpo inmerso en ella es función del volumen de agua desplazado por el cuerpo" no se puede remplazar el predicado "es función" por el predicado "es causada" sin cometer un serio error). Como existen ciertas funciones que no son automáticamente causales, es menester entender primero cuál es la naturaleza de una función para después determinar en cómo una función puede tener una estructura causal. Una función es una entidad abstracta definida por el hecho que existe un conjunto de objetos A $[A_1, A_2...A_n]$ llamado dominio y un conjunto de objetos B $[B_1, B_2...B_n]$ llamado contra-dominio tal que los objetos del dominio son mapeados uno a uno con los objetos del contra-dominio (que también es

[109] *Ibid.*, loc. cit.: "Whatever occupies the definitive causal role of an experience in such a family is does so by virtue of its own membership is a causal isomorph of the family of experiences, that is, in a system of states having the same pattern of causal connections with one another and the same causal connections with states outside the family, viz., stimuli and behavior."

conocido como imagen del dominio). Una función se encuentra entonces definida por un conjunto de pares ordenados [(A_1, B_1); (A_2, B_2)...(A_n, B_n)] que determinan el rango en el cual la función se aplica o es verdadera (e.g., los números reales para la función $y = x^2$, el conjunto de los seres humanos para la función "es hijo de", etc.). A este concepto abstracto de función, Lewis añade el concepto de experiencia (o estado mental) definida por su papel causal (su síndrome de causas y efectos más típicos) que hemos mencionado anteriormente. Una función causal se encuentra entonces definida por un conjunto de pares ordenados [(A_1, B_1); (A_2, B_2)... (A_n, B_n)] tales que $A_1, A_2...A_n$ no sólo son las ante-imágenes de $B_1, B_2...B_n$ respectivamente, sino que son causas de que ocurran $B_1, B_2...B_n$. Así pues, cuando Lewis señala que los estados mentales son funciones causales, lo que quiere decir es que se encuentran definidos por un dominio de estados físicos F (que son en general estímulos o estados neuronales) y un contra-dominio de estados físicos G (que son en general otros estados neuronales o conductas) tales que los elementos de F constituyen causas de que ocurran los elementos de G. Es importante destacar que algunos elementos de G pueden a su vez constituir causas de que ocurran otros estados físicos H, por lo cual pueden ser considerados como dominio de otra función causal distinta de la que mapeaba los elementos de F a G (que, por ende, es otro estado mental distinto del primero). La posibilidad de cadenas de funciones causales mutuamente definidas entre sí revela un rasgo importante de la teoría de la mente de Lewis: posee un carácter holista según el cual la naturaleza de todo estado mental particular se encuentra definida por las relaciones causales que mantiene, no sólo con estímulos y conductas, sino con otros estados mentales. Ahora bien, este carácter holista de la propuesta de Lewis sólo puede ser entendido en el caso que asumamos (como Lewis lo hace) que los elementos que liga la función causal (que es idéntica con un estado mental particular) son, al igual que la función misma, entidades abstractas (i.e., tipos o clases), y no objetos particulares. Lewis muestra esto de manera clara al poner el siguiente ejemplo de su teoría de la mente:

> Consideremos los candados de combinación cilíndrica para cadenas de bicicleta. La característica definitiva de su estado de estar abiertos es el papel causal de ese estado, el síndrome de sus causas y efectos más típicos: es decir, que marcar la combinación

típicamente causa que el candado sea abierto, y que estar abierto típicamente causa que el candado se abra cuando se jala ligeramente. Esto es todo lo que necesitamos para adscribir al candado el estado de estar o no estar abierto. Pero podemos aprender que, como una cuestión de hecho, el candado tiene una línea de discos con muescas; marcar la combinación típicamente causa que las muescas estén alineadas, y el alineamiento de las muescas típicamente causa que el candado se abra cuando se jala ligeramente. Entonces el alineamiento de las muescas ocupa precisamente el papel causal que adscribimos a estar abierto por necesidad analítica, como la característica definitiva de estar abierto (para estos candados). El alineamiento de muescas es idéntico a estar abierto (para estos candados).[110]

Como podemos apreciar en el ejemplo, el hecho de que el candado se encuentre abierto es un tipo de estado que se encuentra definido por las relaciones causales que mantiene con otros tipos de estados que son el marcar la combinación y el abrirse al jalar ligeramente. Con base en la precisión que hemos hecho, podemos entonces explicar como la realizabilidad múltiple del funcionalismo y la afirmación de la teoría de la identidad son compatibles en la propuesta de Lewis a través del siguiente argumento:

(1) Un estado mental M es una función (i.e., una entidad abstracta) que se encuentra definida por su papel causal (i.e., por su síndrome más típico de causas y efectos), mapeando un grupo determinado de estados tipo F a un grupo de estados tipo G.

(2) Un estado neuronal N es una función (i.e., una entidad abstracta) que se encuentra definida por su papel causal (i.e., por su síndrome más típico de causas y efectos), mapeando un grupo determinado de estados tipo F a un grupo de estados tipo G.

[110] Ibid., p. 100: "Consider cylindrical combination locks for bicycle chains. The definitive characteristic of their state of being unlocked is the causal role of that state, the syndrome of its most typical causes and effects: namely that setting the combination typically causes the lock to unlocked when gently pulled. That is all we need now in order to ascribe to the lock the state of being or of not being unlocked. But me may learn that, as a matter of fact, the lock contains a row of slotted discs; setting the combination typically causes the slots to be aligned; and alignment of the slots typically causes the lock to open when gently pulled. So alignment of slots occupies precisely the causal role that we ascribe to being unlocked by analytical necessity, as the definitive characteristic of being unlocked (for these locks) Therefore, alignment of slots is identical with being unlocked (for these slots) "

(3) De (1) y (2) se deriva que M=N donde M y N son tipos o clases.

(4) Como M y N son tipos o clases, pueden ser realizados por distintas instancias materiales con composiciones químicas y estructuras biológicas distintas, sin que se lesione la identidad M=N. (Esta premisa ha sido criticada por Fodor, como veremos más adelante, pero aceptémosla por el momento por mor del argumento)

(5) De (3) y (4) se deriva que la tesis de la realizabilidad múltiple del funcionalismo es compatible con la teoría de la identidad.

La ventaja que presenta esta definición de tipo de estado mental con base en las relaciones causales típicas que mantiene con otros tipos de estados (estímulos, conductas y otros estados mentales definidos a su vez por sus relaciones causales con otros estados mentales) es que, en la medida en que es una entidad abstracta, puede ser realizada de varias maneras así como el estado de que el candado esté abierto puede ser realizado de varias maneras (entre las cuales se cuenta el alineamiento de muescas para el caso de los candados de bicicleta) y, al mismo tiempo, puede ser idéntica con un estado neuronal que también es definido como una entidad abstracta. Así pues, podemos constatar que la propuesta de Lewis acepta la posibilidad de la realizabilidad múltiple de los estados mentales y busca dar cuenta de ella, manteniendo al mismo tiempo una teoría de la identidad. Para esto, sólo se requiere que cualquier tipo de estado neuronal P que realiza un cierto estado mental tipo M tenga exactamente el mismo patrón de relaciones causales respecto a otros tipos de estados Q, R, S... (estímulos, conductas y estados mentales) con los cuales M mantiene esas relaciones causales.

Ahora bien, ¿podemos imaginar un sistema distinto de nosotros que cumpla con el requisito anterior? En un artículo posterior titulado *Mad Pain and Martian Pain* (1980), Lewis responde a esta pregunta de manera afirmativa, adoptando una posición claramente funcionalista que difiere ligeramente con la posición propia de un teórico de la identidad mente-cerebro abierto a la posibilidad de un tipo de funcionalismo que había adoptado en (1966). En este artículo, Lewis propone la situación que citamos a continuación; una situación que sólo puede plantear un partidario del funcionalismo:

> Podría haber un marciano que algunas veces sienta dolor, tal como lo sentimos, pero cuyo dolor difiere mucho del nuestro es su realización. Su mente hidráulica no contiene nada como nuestras neuronas. Más bien, hay distintas cantidades de fluido en varias cavidades inflables, y la inflación de cualquiera de estas cavidades abre ciertas válvulas y cierra otras. Su plomería mental se encuentra en la mayor parte de su cuerpo (...) Cuando se pellizca su piel no se provoca una excitación de las fibras C -puesto que no tiene- sino, más bien, se provoca la inflación de varias cavidades pequeñas en sus pies. Cuando estas cavidades están infladas, tiene dolor. Y los efectos de su dolor son apropiados: su pensamiento y su actividad son interrumpidas, gime y se contorsiona, se encuentra motivado para que se cese de pellizcarlo y para que no se vuelva a hacerlo. En resumen, siente dolor pero carece de los estados corporales que o son dolor o lo acompañan en nosotros.[111]

Después de plantear esta situación, Lewis afirma que una buena teoría de la mente debe de poder dar cuenta de ella. Como esta situación señala que la realización material de un estado mental (e.g., dolor) no es esencial para individuarlo y reconocerlo como tal, es evidente que la situación en cuestión aboga por un rechazo del materialismo del estado central (al menos de las versiones estrictas de la teoría) y por una aceptación de una versión de funcionalismo en el cual un estado mental M sea individuado y reconocido como tal por la función causal particular que juega entre ciertos estímulos, ciertas conductas y ciertos estados mentales antecedentes o subsecuentes respecto a M. Sin embargo, Lewis no se priva de destacar que una teoría funcionalista de la mente que se encuentre articulada alrededor de la mera noción de estado mental como función causal entre estímulos, conductas y otros estados mentales presenta un problema.

[111] Lewis, 1980 en Lewis, 1983, p. 123: "There might be a Martian who sometimes feels pain, just as we do, but whose pain differs greatly from ours in its physical realization. His hydraulic mind contains nothing like our neurons. Rather, there are varying amounts of fluid in many inflatable cavities, and the inflation of any one of these cavities opens some valves and closes others. His mental plumbing pervades most of his body (...) When you pinch his skin you cause no firing of C-fibers -he has none- but, rather, you cause the inflation of many smallish cavities in his feet. When these cavities are inflated, he is in pain. And the effects of his pain are fitting: his thought and activity are disrupted, he groans and writhes, he is strongly motivated to stop you from pinching him and to see it that you never do again. In short, he feels pain but lacks the bodily states that either are pain or else accompany it in us."

Este problema consiste en que esta versión de funcionalismo se halla muy próxima del conductismo, por lo cual recoge sus virtudes (en particular, la posibilidad de dar cuenta de los vínculos causales que existen entre mente y cuerpo) pero también sus limitaciones. Una de las principales limitaciones del conductismo que pudimos constatar en el primer capítulo del presente trabajo consistía en el hecho de que la teoría era incapaz de recuperar el aspecto intrínseco de los estados mentales (e.g., la "dolorosidad" del dolor). En especial, el argumento de los "super-espartanos" presentado por Putnam y los errores médicos descritos por Dennett nos muestran que el conductismo no puede dar cuenta del carácter fenoménico de los estados mentales. Esta limitación se aprecia también en el funcionalismo que sólo toma en cuenta para determinar la naturaleza de los estados mentales el hecho de que mapean ciertos estados F a ciertos estados G de acuerdo a una estructura causal. Es por ello que el funcionalismo no puede dar cuenta de la siguiente situación (que Lewis denomina "dolor del loco") :

> Podría haber un hombre extraño que algunas veces siente dolor, tal como nosotros lo hacemos, pero cuyo dolor difiere en buena medida del nuestro en sus efectos y en sus causas Nuestro dolor es típicamente causado por cortes, quemaduras, presión, y cosas parecidas; el suyo es causado por un ejercicio leve con el estómago vacío. Nuestro dolor generalmente nos distrae; el suyo hace que su mente se torne hacia las matemáticas, facilitando su concentración en ello aunque distrayéndolo de cualquier otra cosa. El dolor intenso no hace que gima o se contorsione, pero provoca que cruce las piernas o que chasquee los dedos. No se encuentra motivado en lo más mínimo para prevenir el dolor o para deshacerse de él. En resumen, siente dolor pero su dolor no ocupa el papel causal típico del dolor.[112]

[112] *Ibid.*, p. 122: "There might be a strange man who sometimes feels pain, just as we do, but whose pain differs greatly from ours in its causes and effects. Our pain is typically caused by cuts, burns, pressure and the like; his is caused by moderate exercise on an empty stomach Our pain is generally distracting; his turns his mind to mathematics, facilitating concentration on that but distracting him from anything else. Intense pain has no tendency whatever to cause him to groan or writhe, but does cause him to cross his legs and snap his fingers. He is not the least motivated to prevent pain or to get rid of it. In short, he feels pain but his pain does not at all occupy the typical causal role of pain."

Como una buena teoría de dolor debe imperativamente también dar cuenta de esta situación según él (al igual que del dolor del marciano), Lewis parece haber encerrado a sus adversarios en una *impasse* teórica puesto que, como no se priva de señalarlo, "una teoría de la identidad resuelve sencillamente el problema del dolor del loco, pero se equivoca acerca del dolor del marciano. Un conductismo simple o un funcionalismo apunta en la otra dirección: es correcto acerca del marciano, y erróneo acerca del loco."[113] Siendo las cosas así, ¿cómo podemos dar cuenta del dolor del marciano y del dolor del loco? ¿Cuál es la teoría de la mente que nos permitiría hacer esto? La propuesta de Lewis para resolver este dilema consiste en utilizar lo que es conocido en la literatura como el método de análisis de Carnap-Ramsey-Lewis. Este método de análisis tuvo como primer objetivo el análisis de los términos teóricos, de los cuales los términos del vocabulario de la psicología mentalista parecen ser una pequeña parte. En un artículo titulado *How to define theoretical terms* (1970), Lewis describe cómo el método puede ser usado en general para definir términos teóricos de manera neutral. En primera instancia, Lewis asume que tenemos una teoría T en la cual se dan varios términos teóricos (t_1, t_2..t_n). La expresión formal de esta teoría consiste en la siguiente expresión:

(1) T (t_1, t_2..t_n)

Una vez que se tiene esta expresión, se remplaza uniformemente en la teoría T todas las ocurrencias de los términos teóricos t_i ($1 \leq i \leq n$) por las variables x_1, x_2...x_n de tal manera que obtenemos la expresión (2) T(x_1, x_2...x_n) que constituye la formula de realización de la teoría T. Si un n-tuplo dado de entidades satisface (2), entonces se dice que este n-tuplo es un modelo de la teoría T. Si queremos sostener que T tiene por lo menos un modelo, entonces aplicamos una serie de cuantificaciones existenciales sobre cada variable de (2), con lo cual obtenemos la oración de Ramsey de T que consiste en lo siguiente:

[113] *Ibid.*, p. 123: "A simple identity theory straightforwardly solves the problem of mad pain. It goes just as straightforwardly wrong about Martian pain. A simple behaviorism of functionalism goes the other way: right about the Martian, wrong about the madman "

(3) $(\exists x_1)(\exists x_2)...(\exists x_n)\ T(x_1, x_2...x_n)$

Al disponer de la oración de Ramsey de T, se puede obtener la oración de Carnap de T que señala que en el caso en que exista por lo menos un modelo de T, entonces T es verdadera, lo cual se traduce en un lenguaje formal por la siguiente expresión:

(4) $(\exists x_1)(\exists x_2)...(\exists x_n)\ T(x_1, x_2...x_n) \rightarrow T(t_1, t_2...t_n)$

Como podemos apreciar, la oración de Carnap de una teoría T constituye el punto esencial para poder definir los términos teóricos empleados en T de manera no-circular. Esto tiene una gran importancia para cualquier persona interesada en desarrollar una buena teoría de la mente puesto que una de las objeciones más duras que habitualmente se hacen contra la psicología mentalista consiste en el hecho de que las definiciones de los términos que establece son circulares. En cambio, si aplicamos el método de análisis de Carnap-Ramsey-Lewis, podemos obtener una teoría de la mente donde, si bien los estados mentales se encuentran definidos entre sí, no hay circularidad puesto que todas las expresiones definidas (i.e., todos los estados mentales) pueden ser eliminadas por medios de sus *definientia*. Para poder apreciar esto con mayor claridad, consideremos el siguiente ejemplo de una teoría T sobre el dolor:

> (T) Para cualquier *x*, si *x sufre daño en su piel* y **está alerta normalmente**, x tiene dolor; si *x está despierto*, x tiende a estar **alerta normalmente**; si x tiene dolor, *x* respinga y gime y **cae en un estado de zozobra**; y si *x* **no está alerta normalmente o está en un estado de zozobra**, *x tiende a cometer más errores al escribir*.[114]

Como podemos apreciar claramente, la definición del dolor se encuentra dada en la teoría T por medio de las relaciones que tiene el estado mental con entidades no-mentales designadas por los predicados en

[114] Kim, *op. cit.* citado en Villanueva, 2000, p 31

cursiva (i.e., predicados que designan propiedades conductuales, físicas, o biológicas) así como con otras entidades mentales designadas por los predicados en negrita. Cuando se aplica el método de análisis de Ramsey a una teoría como T, obtenemos en primera instancia una teoría que denominamos T_{RL} en la cual todos los términos y predicados mentales (M_1, M_2...M_n) son eliminados y remplazados por las variables (y_1, y_2...y_n) sobre las que posteriormente se aplica una generalización existencial:

> (T_{RL}) Existen ciertos estados y_1, y_2 y y_3 tales que para cualquier *x*, si *x sufre daño en la piel* y está en y_1, *x* está en y_2, si *x está despierto*, *x* tiende a estar en y_1; si *x* está en y_2, *x respinga y gime* y pasa al estado y_3; y si *x* no está en y_1 o está en y_3, *x tiende a cometer más errores al escribir*.[115]

Resulta sencillo constatar que el primer paso de la aplicación del método de análisis de Ramsey-Lewis a la teoría T (paso del cual surge la teoría T_{RL}) debe mucho a la tesis de Smart respecto a la necesidad de los análisis tópico neutrales para presentar una buena teoría de la mente. La teoría T_{RL} es completamente neutral en el sentido de que no tiene ningún compromiso ontológico con el materialismo o con el dualismo cartesiano: las variables y_1, y_2 y y_3 pueden ser instanciadas en principio por cualquier tipo de entidades (ángeles, demonios, poltergeists, estados neuronales de seres humanos, etc.) con la condición de que respeten las relaciones causales expresadas en T_{RL}. Asumiendo que el conjunto de relaciones causales que vinculan en T_{RL} las variables con las entidades no-mentales pueden ser simbolizadas por el operador T, entonces la teoría T_{RL} puede ser simbolizada de la siguiente manera:

(5) $(\exists y_1)(\exists y_2)...(\exists y_n)\ T(y_1, y_2...y_n)$

[115] *Ibid.*, loc. cit.

Podemos constatar de manera sencilla que la proposición (5) no es otra cosa que la oración de Ramsey de la teoría T. En tanto disponemos ahora de la oración de Ramsey de T, podemos construir la oración de Carnap de T que consiste en lo siguiente:

$$(6)\ (\exists y_1)(\exists y_2)...(\exists y_n)\ T(y_1, y_2...y_n) \rightarrow T(M_1, M_2...M_n)$$

Es importante ver que la realización de la oración de Carnap de T parece planteas un serio problema: ¿qué tipo de entidades deben ser instanciadas por las variables y_1, $y_2...y_n$? Esta pregunta es importante puesto que la estructura de la oración de Carnap exige que las entidades señaladas por las variables cuantificadas sean tales que puedan dar cuenta de las relaciones causales expresadas en T_{RL}. Como la oración de Carnap de T es un condicional, la única manera de hacer verdadera la oración entera consiste en hacer verdadero el antecedente de la misma, asumiendo que el consecuente es siempre verdadero.[116] Según la interpretación más común de cuantificador existencial, el antecedente de la oración de Carnap sólo puede ser verdadero a condición de que exista *por lo menos* un n-tuplo ordenado de entidades que sea un modelo de la oración de Ramsey, i.e., que exista por lo menos un n-tuplo ordenado de entidades que constituya una realización de la teoría T. Este punto parece engendrar un dilema particularmente espinoso para Lewis como veremos a continuación. El uso del cuantificador existencial exige que exista *por lo menos* un n-tuplo ordenado de entidades. Si la teoría T no tuviese realización alguna, entonces es claro que los estados mentales de dolor no existirían.[117] Como esta conclusión es absurda, es evidente que debe de existir por lo menos una realización de T. En el caso en que existe una única realización de T, parece no haber problema alguno según Lewis a primera vista puesto que "la oración de Carnap presenta exactamente la especificación correcta."[118] Sin embargo, ¿acaso no viola este caso el espíritu mismo del

[116] Esto es sólo una verdad a medias, puesto que existe otro medio de hacer verdadera la oración de Carnap de T que consiste en hacer falso el antecedente. Sin embargo, esta estrategia no es muy interesante puesto que, si se torna falso el antecedente, la oración de Carnap se torna verdadera, pero tiene un valor explicativo nulo.

[117] Lewis, *op. cit.*, p. 82: "The T-terms were introduced on the assumption that T was realized, in order to name components of a realization of T. There is no realization of T. Therefore they should not name anything."

[118] *Ibid.*, loc. cit.: "[In case T it is uniquely realized,] the Carnap sentence clearly gives exactly the right specification."

funcionalismo cuya tesis básica sustenta la realizabilidad múltiple de los estados mentales? Y en el caso en que hay varias realizaciones de T, ¿acaso esto no viola el espíritu de la teoría de la identidad que Lewis dice mantener? Este aparente problema se desvanece cuando recordamos que la identidad que Lewis defiende ocurre no entre objetos, sino entre tipos o clases, y que estos tipos o clases pueden ser realizados por cualquier cosa (sin importar su composición química o su estructura biológica), con tal de que sean respetadas las relaciones causales que los definen. De esta manera, Lewis pretende salvar la posibilidad de la realizabilidad múltiple de los estados mentales por un lado y, por otro lado, pretende evitar que la tesis de la realizabilidad múltiple degenere en un conjunto disyuntivo de estados tan vasto y heterogéneo que no tenga ya sentido sostener que los diversos estados materiales que realizan la misma función constituyen un solo y mismo estado mental (este último objetivo no es realmente cumplido como veremos más adelante):

> Supongamos que el dolor es, en efecto, un cierto estado funcional S_{17} en una apropiada descripción funcional; supongamos que la descripción es realizada *inter alia* por los estados cerebrales humanos $B_1,...B_n$ respectivamente. Estos son los estados que están legalmente relacionados unos con otros, y con estímulos y respuestas humanos apropiados, por medio de apropiadas probabilidades de transición. ¿Por qué no concluir que el dolor = S_{17} = B_{17} en el caso de los humanos [o C_{17} en el caso de las computadoras, o D_{17} en el caso de los Marcianos, etc.]?.[119]

Así pues, como podemos constatar, la propuesta de Lewis constituye un tipo de funcionalismo muy particular puesto que no propone un rechazo radical de la teoría de la identidad-mente cerebro, sino una recuperación de esta teoría por medio del funcionalismo en tanto que un caso particular de realización de

[119] Lewis, 1969 en Block, 1980, pp. 232-233: 'Suppose pain is indeed a certain functional state S_{17} in an appropriate functional description; suppose the description is realized *inter alia* by the human brain states $B_1,...B_n$ respectively Those are the states that are lawfully related to one another, and to suitable human inputs and outputs, by the proper transition probabilities. Why not conclude that pain = S_{17} = B_{17} in the case of humans [or C_{17} in the case of computers, or D_{17} in the case of Martians, etc.]?"

los estados funcionales para los seres humanos. De esto, Lewis deduce que un estado mental M puede ser definido con base en el papel causal que ocupa *en una población dada* sin que por ello se lesione la identidad establecida entre las diversas realizaciones de este estado mental M en poblaciones distintas, en la medida en que la identidad se encuentra basada en el mero papel causal (que se supone es uno y el mismo) que juegan las diversas realizaciones de M. Si bien este intento de sostener tanto una teoría de la identidad mente-cerebro como una versión de funcionalismo a primera vista parece prometedor, se puede apreciar que también presenta varios problemas que arrojan serias dudas sobre la propuesta de Lewis. Algunas de estas objeciones son propias de todos los tipos de funcionalismo, por lo cual sólo serán tratadas en la próxima sección, pero hay una objeción decisiva contra la propuesta de Lewis en particular que examinaremos a continuación en la medida en que dicha objeción servirá para introducir la próxima versión de funcionalismo que estudiaremos.

En la tercera sección del capítulo anterior, hemos destacado que Putnam había mostrado en (1967) que, ante el argumento de la realizabilidad múltiple de los estados mentales, el teórico de la identidad podía intentar defender su tesis básica por medio de asunciones *ad hoc*, definiendo un estado mental como la disyunción de todas las realizaciones materiales que tiene en todas las poblaciones en las cuales se halla presente. Sin embargo, Putnam rechaza esta estrategia declarando que "no debe ser tomada en serio."[120] Lamentablemente, no expone nunca en el mismo artículo las razones por las cuales considera que esta estrategia es errónea. Esta labor ha sido llevada a cabo por otros autores como Fodor que presenta en *The language of thought* (1975) un argumento importante para descartar la posibilidad de considerar un estado mental como la disyunción de sus diversas realizaciones materiales en individuos pertenecientes a poblaciones o a especies distintas. En las siguientes líneas, presentaremos brevemente el argumento.

Tras haber mostrado que es muy poco probable que la psicología sea reducible a la neurobiología en la medida en que esto implicaría (1) que cada tipo de predicado psicológico es coextenso con un tipo de predicado físico y (2) que la generalización que expresa esta coextensión es una ley (i.e., que todo tipo de predicado mental es en principio sustituible por un cierto tipo de predicado físico al cual puede ser

[120] Putnam, 1967b en Putnam, *op. cit.*, p. 437 "(...) this does not have to be taken seriously."

reducido), Fodor señala que "hay una posibilidad abierta de que aquello que corresponde a los predicados de tipo de una ciencia reducida pueda ser una disyunción heterogénea y no-sistemática de predicados en la ciencia reductora."[121] Cuando esta observación se aplica al caso de la psicología, se observa que es posible que los predicados psicológicos correspondan a una disyunción heterogénea y no sistemática de predicados físicos. Si esto ocurre, se asume que la correspondencia debe tener un estatus nomológico, i.e., en el caso en que M es un cierto predicado psicológico y $F_1 \vee F_2 ... \vee F_n$ la disyunción de predicados físicos que corresponden a M, debe ser el caso también que la oración "Es una ley que el estado M es realizado por $F_1 \vee F_2 . . \vee F_n$" sea verdadera. Esta estrategia presenta una importante ventaja, pero también un serio problema que Fodor expone en el siguiente pasaje:

> Podríamos, si quisiéramos, requerir que las taxonomías de las ciencias especiales correspondieran a la taxonomía de la física insistiendo en las distinciones entre los tipos postulados por las primeras cuando corresponden a distintos tipos en la segunda. Esto haría que las leyes de las ciencias particulares fuesen sin excepción, a condición de que las leyes de la ciencia básica también lo sean. Pero también perderíamos las generalizaciones que queremos que las ciencias particulares expresen.[122]

Así pues, aquello que se pierde cuando aceptamos la propuesta de Lewis (o cualquiera otra similar) según Fodor es precisamente el carácter de sistematicidad y homogeneidad que hacen que una propiedad (o tipo de estado) mental realizada en varios objetos u organismos sea precisamente *esa* propiedad (o *ese* tipo de estado). Si una propiedad dada (e.g., dolor) es realizada por un conjunto heterogéneo y no-

[121] Fodor, 1975, p. 20 "(...) there is an open empirical possibility that what corresponds to the kind predicates of a reduced science may be an heterogeneous and unsystematical disjunction of predicates in the reducing science."
[122] *Ibid.*, p. 23: "We could, if we liked, require the taxonomies of the special sciences to correspond to the taxonomy of physics by insisting that upon distinctions between the kinds postulated by the former whenever they turn out to correspond to distinct kinds of the latter. This would make the laws of the special sciences exceptionless if the laws of basic science are But it would also likely loose us precisely the generalizations which we want the special sciences to express."

sistemático de estados materiales (todos los cuales tienen características y propiedades distintas en tanto son realizados en objetos u organismos distintos), entonces es plausible sostener que:

> Estos realizadores físicos de [un estado mental] M cuentan como clases físicas distintas puesto que tienen distintos, y quizás muy diversos, poderes causales. Por esta razón, es imposible asociar un único conjunto de poderes causales con M, cada instancia de M, por supuesto, es una instancia de [cualquiera de los estados físicos] P_1 o P_2 o P_3, y como tal representa un único conjunto de poderes causales, pero M tomado como un tipo o una propiedad, no lo es.[123]

Si los estados mentales carecen de unidad y de sistematicidad nomológica en el pensamiento de Lewis a causa su la teoría de la identidad basada en el papel causal de los estados mentales, se sigue de ello que su concepto de estado mental no puede ser empleado por la ciencia de manera útil. Para entender esto mejor, es menester recordar que, en el ámbito de la ciencia, el criterio empleado para establecer tipos es la similitud causal/nomológica.[124] De acuerdo con este criterio, una proposición A pertenece al mismo tipo o categoría que una proposición B si y sólo si ambas tienen el mismo conjunto de poderes causales y juegan el mismo papel causal bajo las mismas leyes. Con base en lo que hemos dicho anteriormente, podemos construir el siguiente argumento:

(1) La propuesta de Lewis consiste en una teoría de la identidad M=N entre estados mentales y estados neuronales en la cual M y N son clases, y en la cual estas clases pueden ser realizadas en principio por una gran variedad de sistemas con composiciones físicas y químicas distintas.

---

[123] Kim, *op. cit.*, p. 119: "These physical realizers of [a mental state] M count as different physical kinds because they have different, perhaps extremely diverse, causal powers. For this reason, it is not possible to associate a unique set of causal powers with M; each instance of M, of course, is an instance of either P_1 or P_2 or P_3 and as such represents a unique set of causal powers, but M taken as a kind or property does not."

[124] Cf. Fodor, 1987.

(2) Esta propuesta de Lewis nos lleva a definir un estado mental M como la disyunción de todas los realizadores actuales o posibles de la clase que es el estado mental M, i.e., M es idéntico a la realización de las clases de estados F_1 o F_2 o ... o F_n

(3) Los tipos o las clases en la ciencia son individuados con base en sus poderes causales, i.e., dos objetos o sucesos pertenecen a una misma clase cuando tienen los mismos poderes causales.

(4) Si una propiedad mental M es realizada en un momento t por un sistema en virtud de la realización física base P, los poderes causales de esta instancia de M son idénticos con los poderes causales de P.[125]

(5) En virtud de (3) y (4), las instancias de M realizadas por una misma base física P pertenecen a una misma clase causal $\mathcal{M}$. Así mismo, las instancias de M realizadas por bases físicas distintas de P pertenecen a clases causales distintas de $\mathcal{M}$.

(6) Por lo tanto, se concluye de (2) y (5) que las clases mentales realizadas por clases físicas no son aptas para ser clases científicas en la medida que, en la propuesta de Lewis, cada clase mental $\mathcal{M}$ es dividida en tantas clases como existen realizaciones físicas actuales o posibles de $\mathcal{M}$. Esto implica que la psicología en tanto que ciencia con una unidad disciplinaria se encuentra condenada a desvanecerse.

Como esto no es posible en el marco de la propuesta de Lewis, de ello se deduce que su criterio de identidad no puede ser empleado para establecer tipos o categorías de estados mentales que constituyan una auténtica psicología científica. Así pues, el problema central de la posición de Lewis consiste en que "la conjunción del funcionalismo y de la realizabilidad múltiple aparentemente lleva a la conclusión que la psicología está en peligro de perder su unidad y su integridad como ciencia."[126]

[125] Kim, 1993, p. 326: "If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical with the causal powers of P." Cf. también Fodor, 1987.
[126] Kim, 1998, p. 119: "(...) the conjunction of functionalism and the multiple realizabilty of the mental apparently leads to the conclusion that psychology is in danger of losing its unity and integrity as a science."

III.1.2 La propuesta de Putnam

Es precisamente para evitar este peligro que Putnam declara, a contracorriente de las tesis de Lewis, que el funcionalismo y la teoría de la identidad son mutuamente excluyentes, y que la teoría de la identidad es irrecuperable, aún como un mero caso particular de realización de estados mentales en una población determinada. Putnam afirma que todo uso de estados neuronales para establecer la definición de los estados mentales debe ser eliminado sistemáticamente y remplazado por estados funcionales. Pero, ¿qué es un estado funcional para Putnam? Según Putanm, un estado funcional es básicamente un estado de máquina de Turing. Para explicar precisamente en qué consiste un estado de máquina de Turing, es necesario primero explicar la naturaleza de una máquina de Turing, lo que habremos de realizar a continuación.

Una máquina de Turing es un concepto matemático que puede ser descrito intuitivamente de la manera siguiente: imaginemos una cinta de papel que se extiende infinitamente en ambas direcciones dividida en células en las cuales hay un número finito de símbolos impresos $\{A_0, A_1...A_n\}$, uno en cada célula. Estos símbolos pertenecen a un cierto alfabeto A. La máquina de Turing posee un número finito de estados internos $\{E_0, E_1, E_2...E_n\}$ y asumiremos que siempre iniciará su operación en el estado inicial E_0. Pero, ¿en qué consiste exactamente la operación de la máquina? Asumimos que la máquina posee una cabeza de lectura y de impresión que lee en un momento dado (i.e, no continuamente, sino en momentos precisos)[127] el símbolo impreso en el cuadrado de la cinta que se encuentra ante ella. Dependiendo del estado interno en que estaba así como del símbolo leído, la máquina pasa del estado interno en el que estaba a otro (o no), imprime un símbolo en lugar del símbolo anteriormente leído (o no) y finalmente mueve la cinta hacia la izquierda o hacia la derecha para proseguir con la lectura de los demás símbolos impresos en la cinta (o no).

Habiendo presentado una noción bastante intuitiva de lo que es una máquina de Turing, podemos ahora establecer una definición exacta: una máquina de Turing T con un alfabeto A de símbolos de cinta $\{A_0,$

[127] Por esta razón, el funcionamiento de cualquier máquina de Turing supone una concepción discreta del tiempo

$A_1...A_n\}$ y con un conjunto de estados internos $\{E_0, E_1, E_2.. E_n\}$ es un conjunto de cuádruplos que pueden adoptar tres formas: 1) $E_iA_kA_lE_m$, 2) $E_iA_kDE_m$ y 3) $E_iA_kIE_m$ tales que no hay dos cuádruplos en T que tengan los mismos dos primeros símbolos.[128] Un paso en una computación dada (que no es otra cosa más que la descripción de una cierta máquina de Turing en un momento dado) corresponde a un cuádruplo determinado. Como podemos constatar sin problema, el concepto de máquina de Turing recoge un rasgo muy importante de las relaciones entre la mente y el cuerpo tales como las entienden los funcionalistas. Si consideramos que las mentes de las personas pueden ser modeladas por máquinas de Turing, entonces resulta sencillo explicar las relaciones que hay entre estímulos sensoriales y respuestas motoras (que son asimilados a los símbolos de la cinta que la máquina lee o imprime), los estados mentales (que son asimilados a los estados internos de la máquina) y los órganos de percepción y de acción (que son asimilados a la cabeza de lectura y de impresión de la máquina).

Partiendo de esta analogía, Putnam propone como hipótesis de trabajo la existencia de una comunidad de agentes que son máquinas de Turing ligeramente distintas de las que propuso en un principio Turing en la medida que no son meras abstracciones, sino que disponen de medios para recibir datos del entorno y para actuar sobre él:

> Las máquinas de Turing que quiero considerar diferirán de las máquinas de Turing abstractas consideradas por la teoría lógica en que las consideramos equipadas de órganos sensoriales, por medio de los cuales podrán escudriñar su medio ambiente, y de órganos motrices apropiados, que serán capaces de controlar (...) Esta es la generalización natural de una máquina de Turing que permite una interacción con el medio ambiente.[129]

---

[128] Mendelson, 1987, p. 212. El primer tipo de cuádruplo especifica que si la máquina T está en E_i y lee A_k entonces imprime A_l y pasa a E_m. Los dos otros tipos de cuádruplos especifican que si la máquina T está en E_i y lee A_k entonces mueve la cinta hacia la derecha o a la izquierda (esto lo precisan los símbolos D e I respectivamente) sin imprimir ningún símbolo y pasa a E_m.

[129] Putnam, 1967a en Putnam, 1975, p. 409: "The Turing machines I wish to consider will differ from the abstract Turing machines considered in logical theory in that we will consider them to be equipped with sense organs by means of which they can scan their environment, and with suitable motor organs which they are capable of controlling. This is the natural generalization of a Turing machine to allow for interaction with an environment."

Las máquinas de Turing propuestas por Putnam admiten el establecimiento de tablas determinadas (también conocidas como tablas de máquina) como las que proponen los teóricos causales de la mente para describir los estados mentales con respecto a estímulos, conductas y otros estados mentales. En estas tablas de máquina se establece, para cada combinación posible de estados internos específicos de la máquina con determinadas "impresiones sensoriales" (que corresponden a la lectura de un símbolo), las eventuales respuestas motoras (que corresponden a la impresión de un símbolo en la cinta) y el estado interno consecutivo al cual la máquina pasa. La principal ventaja de este modelo consiste en el hecho de que llena el primer requisito que hemos establecido para que algo sea una buena teoría de la mente: permite dar cuenta de los vínculos causales entre mente y cuerpo en la medida que los símbolos de la máquina tienen poderes causales. Aunque un estado interno de una cierta máquina de Turing es en principio algo abstracto, puede ser realizado físicamente de varias maneras (e.g., por medio de una neurona biológica o de un micro-procesador), de tal manera que se pueden explicar los vínculos causales sin "saltos ontológicos". Ahora bien, como Putnam se da cuenta que establecer una tabla de máquina que relacione de manera absoluta estados internos e impresiones sensoriales con respuestas motoras y estados internos consecutivos lo llevaría a cometer uno de los errores del conductismo (i.e., sostener que es posible dar una lista exhaustiva de impresiones sensoriales y respuestas motoras que caracterizan un estado mental dado), debilita voluntariamente su tesis y sostiene que las máquinas de Turing que él propone son autómatas probabilistas, i.e., sistemas cuyas tablas de máquina no relacionan de manera determinista estados internos e impresiones sensoriales con respuestas motoras y estados internos consecutivos, sino sólo a través de una cierta probabilidad. Putnam rechaza considerar a los seres humanos como máquinas de Turing con transiciones bien determinadas de un estado a otro y se inclina más bien por una interpretación probabilista por la siguiente razón:

> Si una de estas máquinas prefiere A a B, no se sigue necesariamente que, en cualquier situación concreta, va a elegir A en vez de B Para decidir si va a elegir A en vez de B, la máquina tendrá que considerar cuáles serán las probables consecuencias de su elección en la

situación concreta, y esto puede muy bien poner en juego otros "valores" de la máquina además de la preferencia que la máquina asigna a A sobre B.[130]

Este hecho es muy importante en tanto que, a través de él, Putnam pretende superar una de las limitaciones del conductismo. En efecto, el conductismo sostenía que un "estados mental" M sólo podía ser definido con base en las relaciones causales que mantenía con ciertas conductas y ciertos estímulos, pero la teoría no planteaba que pudiera haber relación causal entre M y otros "estados mentales. En cambio, la propuesta de Putnam pretende recuperar una noción de interdependencia holista de los estados mentales (que hemos apreciado en Lewis anteriormente) señalando que ciertos estados mentales de una persona (e.g., la preferencia que tiene esta persona de A respecto a B) sólo pueden ser definidos por medio de otros estados mentales (i.e., los otros "valores" de la máquina de Turing a los cuales Putnam hace alusión). Esta situación puede conducir a un problema que citamos a continuación: si los estados mentales sólo pueden ser definidos con respecto a otros estados mentales, ¿acaso no se corre el peligro de caer en un círculo vicioso? Kim señala que esto es una de las dificultades que surgen del intento de recuperación de la interdependencia holista de los estados mentales planteada por el funcionalismo como podemos apreciar en el caso del dolor:

> Entre los efectos típicos del dolor se encuentran otros sucesos mentales como un sentimiento de zozobra o un deseo de ser aliviado de él. (...) Esto parece involucrarnos en un regreso o en una circularidad: para explicar qué es un estado mental dado, necesitamos referirnos a otros estados mentales, y sólo podemos esperar que la explicación de éstos requiera una referencia a otros estados mentales, continuando siempre así -un proceso que puede proseguir en un regreso al infinito o que puede girar sobre sí mismo en un círculo.[131]

[130] *Ibid.*, p. 410: "If one of these machines prefers A to B, it does not necessarily follow that in any concrete situation it will choose A rather than B. In deciding whether to choose A rather than B, the machine will have to consider what the consequences of its choice are likely to be in the concrete situation, and this may well bring 'values' of the machine other than the preference that the machine assigns to A over B."

[131] Kim, *op. cit.*, p. 104-105: "Among the typical effects of pain are further mental events, such as a feeling of distress, and a desire to be relieved of it. (...) This seems to involve us in a regress or circularity: To explain what a given mental state is, we need to refer to other mental states, and explaining these can only be expected to require

Es precisamente para resolver el problema de la circularidad o del regreso *ad infinitum* a la que parece conducir la visión holista de los estados mentales plasmada en el funcionalismo que Putnam propone el establecimiento de tablas de máquina que permitan determinar las relaciones entre estados mentales. La tabla de máquina de una cierta máquina de Turing nos permite dar cuenta de las relaciones entre los estados internos de la máquina en cuestión sin circularidad en la medida en que, como bien señala Kim, los conceptos psicológicos son definidos *en masa* (así como en el caso del método de análisis de Ramsey-Lewis). Para ver como la interdependencia no conduce necesariamente a la circularidad, conviene presentar un ejemplo sencillo. Consideremos la máquina de Turing que modela el comportamiento de una máquina distribuidora de latas de Coca-Cola. Esta máquina de Turing debe de poder dar cuenta de las siguientes características de la máquina distribuidora de latas:

-Entrega una lata de Coca-Cola a cambio de un dólar.

-Acepta billetes de un dólar y monedas de cincuenta centavos.

-En el caso en que se introduzca el equivalente a un dólar con cincuenta centavos, la máquina entrega una lata de Coca-Cola y una moneda de cincuenta centavos.

Para que una máquina de Turing pueda modelar estas características, los funcionalistas sostienen que es necesario que posea por lo menos dos estados internos (E_1 y E_2) respecto a la situación crediticia de la persona que está haciendo uso de ella. El estado E_1 representa el hecho de que la situación crediticia de la persona que hace uso de la máquina es nula (i.e., la máquina no ha recibido nada) mientras que el estado E_2 representa el hecho de que la persona que hace uso de la máquina tiene un crédito de cincuenta centavos. Teniendo en cuenta las tres características mencionadas anteriormente así como los dos estados

reference to further mental states, and so on -a process that can go on in an unending regress, or loop back in a circle."

internos postulados para la realización de éstos, podemos plasmar el funcionamiento de la máquina distribuidora en la siguiente tabla de máquina:

	E1	E2
Insertar 50 centavos	No entregar nada. Ir a E2.	Entregar una lata e ir E1.
Insertar 1 dólar	Entregar una lata y permanecer en E1.	Entregar una lata y cincuenta centavos. Ir a E1.

Los partidarios del funcionalismo de máquina de Turing sostienen que es posible establecer este mismo tipo de tablas de máquina para dar cuenta de los estados mentales de una persona, aunque estas tablas son sin duda mucho más complejas y grandes. Como podemos apreciar claramente en esta tabla, cada uno de los estados internos de la máquina puede ser definido con base en el otro, pero esto no implica que haya un círculo vicioso según los funcionalistas, hecho que justifican poniendo como ejemplo el funcionamiento de un banco que, según ellos, es similar al funcionamiento de una máquina distribuidora de latas así como al funcionamiento de la mente:

> Además de mencionar las entradas y las salidas de dinero, tendremos que mencionar las interconexiones entre cajeros, contadores, responsables de préstamos, gerentes y otros. Sin embargo, a pesar de las interconexiones, no hay un círculo vicioso, y lo que muestra esto es el hecho que se puede usar la historia para identificar quiénes son los cajeros, los contadores, los responsables de prestamos y el gerente. En efecto, se utiliza implícitamente el conocimiento de la historia cada vez que se entra a un banco y se tiene éxito en identificar a los cajeros, los gerentes y otros más.[132]

[132] Braddon-Mitchell y Jackson, 1996, p. 50: "In addition to mentioning the financial inputs and outputs, we will have to mention the interconnections between tellers, accountants, loan officers, managers, and so on. But despite the interconnections, there is no vicious circularity, and what shows this is that you could use the story to identify who the tellers, accountants, loan officers and manager are. Indeed, you implicitly use your knowledge of the story every time you go into a bank and succeed in identifying the tellers, the managers and so on."

De la misma manera en que los estados internos de la máquina distribuidora de latas y los papeles de las personas que laboran en un banco son interdefinibles, los estados mentales de una persona dada pueden ser definidos unos con respecto a otros para los funcionalistas, sin que se establezca un círculo vicioso en la medida que es posible identificar los estados mentales utilizando un cierto conocimiento implícito de la mente que todos tenemos y que se expresa a través de nuestra capacidad de predecir y explicar las conductas que tenemos en determinadas circunstancias en términos de estados mentales. Es muy interesante notar que el establecimiento de tablas de máquina que propone Putnam para definir los estados mentales con respecto a otros estados mentales (sin que caigamos por ello en un círculo vicioso) tiene un paralelo con el método de análisis de Ramsey-Lewis en la medida en que este método de análisis también provee una definición de los estados mentales con respecto a otros estados mentales (sin caer en un círculo vicioso). En tanto la propuesta de Lewis plantea una concepción holista de los estados mentales al igual que el funcionalismo de Putnam, es posible dar cuenta de los estados mentales apelando a "la red completa de relaciones causales que involucran todos los estados psicológicos (...) para anclar las definiciones físico-conductuales de las propiedades mentales."[133] Este rasgo que comparten tanto la propuesta funcionalista de Lewis como la de Putnam es muy importante en tanto nos permite superar una de las limitaciones de la psicología holista (que, al intentar definir los estados mentales con respecto a otros estados mentales, cae en un círculo vicioso), pero también presenta un serio problema que veremos más adelante. Por ahora, examinaremos con detenimiento los problemas y las objeciones que se plantean específicamente a la propuesta de Putnam. Como la tesis básica de Putnam consiste en aseverar que una mente no es otra cosa mas que una realización o instanciación de una máquina de Turing adecuada, una de las mejores maneras de rechazar la tesis consiste en mostrar que las mentes exhiben ciertos rasgos que no pueden ser modelados o reproducidos por las máquinas de Turing. Esta es la línea crítica principal que siguen Block y Fodor en su artículo *What Psychological States are not* (1972) donde esbozan una serie de

[133] Kim, *op. cit.*, p. 105· "[Causal-theoretical functionalism attempts to use] the entire network of causal relations involving all the psychological states (.) to anchor the physical-behavioral definitions of individual properties."

argumentos que muestran la incapacidad de las máquinas de Turing de recuperar varios rasgos que las mentes exhiben.

El primer argumento que Block y Fodor presentan consiste en el hecho que las máquinas de Turing no parecen ser capaces de recuperar la distinción que establece la psicología mentalista tradicional entre los estados mentales disposicionales (e.g., creencias, deseos y temores) y los estados mentales ocurrentes (e.g., sensaciones, pensamientos y emociones). Desde un punto de vista intuitivo, la distinción parece ser plausible, por lo que debería ser recuperada por la propuesta de Putnam, si la versión de funcionalismo que sostiene es en verdad una buena teoría de la mente. Sin embargo, resulta sencillo ver que, si bien los estados mentales ocurrentes no plantean problema alguno en tanto pueden ser definidos con base en las relaciones causales que mantienen con ciertos estímulos y ciertas conductas, la situación inversa no es verdadera: los estados mentales disposicionales no pueden ser definidos con base en ciertos estímulos y ciertas conductas. En efecto, una cierta sensación de miedo puede ser considerada como un estado de tabla de máquina en la medida que se encuentra definida por una clase de estímulos (e.g., todas las instancias en que un desconocido me apunta con una pistola para robarme) y una clase de conductas (e.g., todas las instancias en que palidezco y me pongo a temblar) pero mi creencia de que hay una caja de galletas frente a mí no puede serlo. Esto se explica en la medida que, aun cuando mi creencia de que hay una caja de galletas frente a mí se da generalmente en presencia de un estímulo perteneciente a una clase compuesta por todas las situaciones en que hay una caja de galletas frente a mí, dicha creencia no se encuentra definida por una clase específica de conductas, sino por muchas, en la medida que los estados mentales disposicionales son *multi-track* como señaló Ryle.[134] Ahora bien, algunas personas han sugerido que esta objeción se puede evitar transformando los estados mentales disposicionales en estados mentales ocurrentes y, para justificar esta postura, señalan que al predicado disposicional "habla francés" corresponde el predicado ocurrente "esta hablando francés" en una relación uno a uno. Lamentablemente, esta estrategia no funciona para todos los estados disposicionales como podemos apreciar en el siguiente pasaje:

[134] Cf. supra, nota 100.

> No tenemos "está creyendo que P" correspondiendo a "cree que P"; no tenemos "está deseando una paleta" correspondiendo a "desea una paleta"; no tenemos "esta prefiriendo X a Y" correspondiendo a "prefiere X a Y".[135]

Esta situación demuestra que no es posible establecer una correspondencia uno a uno entre estados mentales y estados de tabla de máquina en la medida que los estados mentales disposicionales no tienen correspondencia uno a uno con los estados de tabla de máquina como la tienen los estados mentales ocurrentes. Esto arroja una seria duda sobre la plausibilidad del funcionalismo (y en particular del funcionalismo de máquina de Turing) como teoría de la mente.

El segundo argumento que Block y Fodor presentan consiste en lo siguiente: una de las intuiciones más plausibles e inmediatas sobre la naturaleza de la mente es el hecho que la conducta puede ser el resultado de interacciones entre estados mentales simultáneos. Por ejemplo, es muy probable que el hecho que tome una galleta de una caja en un momento dado se deba tanto al deseo que tengo de comer una galleta como a la creencia de que hay galletas en la caja en ese mismo momento. Sin embargo, de acuerdo con Block y Fodor, el funcionalismo no puede explicar esto en tanto que:

> El funcionalismo no provee ninguna maquinaria conceptual para representar este estado de cosas. En efecto, el funcionalismo puede dar cuenta de interacciones secuenciales entre estados psicológicos, pero no de interacciones simultáneas. El funcionalismo también no puede explicar el hecho de que un organismo pueda estar en más de un estado psicológico ocurrente al mismo tiempo, puesto que un autómata probabilista sólo puede estar en un estado de tabla de máquina en un momento dado.[136]

[135] Block y Fodor, 1972 in Block, 1980, p. 242: "We have no 'is believing that P' corresponding to 'believes that P'; we have no 'is desiring a lollipop' corresponding to 'desires a lollipop'; we have no 'is preferring X to Y' corresponding to 'prefers X to Y'."

[136] *Ibid.*, p. 243: "FSIT provides no conceptual machinery for representing this state of affairs. In effect, FSIT can provide for the representation of sequential interactions between psychological states, but not for simultaneous interactions FSIT even fails to account for the fact that an organism can be in more than one occurent psychological state at a time, since a probabilistic automaton can be in only one machine table state at a time "

Ante esta objeción, Putnam reconoce un año más tarde en *Philosophy and our Mental Life* (1973) que la hipótesis según la cual un ser humano *es* una maquina de Turing (y que los estados psicológicos de un ser humano *son* estados internos de una máquina de Turing) es incorrecta, señalando "que se encontraba todavía demasiado en el puño de la perspectiva reduccionista."[137] Uno de los problemas más graves del funcionalismo de máquina de Turing reside en el hecho de que la teoría propone que los estados mentales de un individuo se encuentran pareados uno a uno con sus estados de tabla de máquina. De ser cierto esto, no sería posible que una persona tuviese dos (o más) estados mentales al mismo tiempo, lo cual es absurdo Es importante notar que una de las estrategias para hacer frente a esta objeción consiste en suponer que el cerebro no es una máquina de Turing, sino un conjunto de máquinas de Turing intercomunicadas entre sí, de tal manera que una persona pueda tener dos o más estados mentales al mismo tiempo. Esta es la línea de investigación seguida por Fodor cuando propone la existencia de diversos módulos especializados en tareas determinadas en el cerebro en *The Modularity of Mind* (1983).[138]

El tercer argumento propuesto por Block y Fodor para rechazar el funcionalismo de máquina de Turing consiste en el siguiente razonamiento. Por un lado es evidente que el conjunto de estados que constituye la tabla de máquina de una máquina de Turing (o de un autómata probabilista) es una lista, por lo cual pueden ser exhaustivamente enumerados por medios finitos. En cambio, por otro lado, Block y Fodor señalan que:

> El conjunto de estados mentales de al menos algunos organismos (principalmente personas) es, desde un punto de vista empírico, productivo. En particular, eliminando limitaciones teóricamente irrelevantes impuestas por la memoria y la mortalidad, hay infinitos tipos de estados psicológicos nomológicamente posibles de una persona dada.[139]

[137] Putnam, 1973, in Putnam, *op. cit.*, p. 298: "(. .) I was too much in the grip of the reductionist outlook."
[138] De hecho, la propuesta de Fodor es híbrida en la medida que sólo los sistemas perceptuales son modulares, pero no así el sistema central que tiene acceso a toda la información.
[139] Block y Fodor, 1972 en Block, *op. cit.*, p. 246: "(...) the set of mental states of at least some organisms (namely, persons) is, in point of empirical fact, productive. In particular, abstracting from theoretically irrelevant limitations

Así pues, para aquellos que quieren establecer una relación uno a uno entre estados mentales y estados de tabla de máquina según la propuesta de Putnam, se impone el problema de que los estados de tabla de máquina de cualquier máquina de Turing son finitos mientras que el número de estados mentales de una persona es, en el mejor de los casos, numerable por medio de una axiomatización finita (i.e., equivalente a $\aleph_0$). En este punto, Block y Fodor hacen una precisión de gran importancia: los estados de tabla de máquina de un autómata probabilista son finitos, pero los estados computacionales de este mismo autómata pueden ser infinitos en tanto pueden ser definidos recursivamente.[140] Así pues, si bien es imposible establecer una correspondencia uno a uno entre estados mentales y estados de tabla de máquina, acaso sea posible establecer esta misma correspondencia entre estados mentales y estados computacionales. Esta línea de investigación también es seguida por Fodor en obras posteriores como, por ejemplo, *The Elm and the Expert* (1995).

El cuarto argumento propuesto por Block y Fodor para rechazar la propuesta funcionalista de Putnam consiste en el razonamiento siguiente. Aun cuando aceptemos la posibilidad de una correspondencia uno a uno entre estados psicológicos y estados de tabla de máquina, Block y Fodor sostienen que esto no basta para recuperar ciertas propiedades de los estados psicológicos como las similitudes que ciertos estados mentales (e.g., creer que $P \wedge Q$) presentan con respecto a otros estados mentales (e.g., creer que P). Sobre este punto, los dos autores señalan lo siguiente:

> Lo que se necesita decir es que creer que P es de alguna manera un constituyente de creer que $P \wedge Q$, pero el modelo de estado de tabla de máquina no tiene recursos conceptuales para decir

imposed by memory and mortality, there are infinitely many type-distinct, nomologically possible states of any given person."

[140] La diferencia entre estados de tabla de máquina y estados computacionales establecida por Block y Fodor radica en lo siguiente (*Ibid.*, loc. cit.): "By the former [the machine table states] we will mean states specified by columns in its machine table. By the latter [the computational states] we mean any state of the machine which is characterizable in terms of inputs, outputs, and/ or machine table states. In this usage, the predicates 'has just run through a computation involving three hundred seventy-two machine tables states' or 'has just proved Fermat's last theorem' all designate possible computational states of machines." Como podemos ver, los estados computacionales incluyen los estados de tabla de máquina así como los estados que son definibles recursivamente a partir de los primeros.

esto. En particular, la noción de "es constituyente de" no está definida para estados de tabla de máquina.[141]

Lo que sostiene esta objeción es que el funcionalismo de máquina de Turing no tiene la capacidad de dar cuenta de los vínculos semánticos que existen entre los estados mentales. De todas las objeciones que hemos presentado contra la propuesta de Putnam, acaso esta es la más seria puesto que, si el funcionalismo de máquina de Turing no puede dar cuenta de los vínculos semánticos entre los estados mentales, es muy probable que no pueda dar cuenta tampoco de la intencionalidad, lo cual constituiría una razón sólida para desechar la propuesta en tanto que teoría de la mente.

Es importante señalar un punto acerca de esta última objeción a la propuesta de Putnam: nuestro lector sin duda habrá observado que no hemos hablado previamente de contenido en el curso de este capítulo hasta la presentación de la cuarta objeción a la propuesta de Putnam. Esto se debe al hecho de que tanto Lewis y Putnam pensaban que el papel causal o la posición en una tabla de máquina de un estado mental bastaban para definirlo completamente, i.e., para definir todas sus características, incluyendo un eventual contenido. Es menester notar que tanto el papel causal (en el caso de Lewis) así como la posición en una tabla de máquina (en el caso de Putnam) de un estado mental no son en último lugar sino estructuras abstractas definidas por las relaciones que mantienen con ciertas clases o tipos de estados (estímulos, conductas u otros "estados mentales"), por lo que no tienen en principio ningún contenido: son meros "esqueletos lógicos". Ambos autores parecen sostener implícitamente que el contenido sólo surge en el momento en el cual la estructura abstracta de un cierto estado mental (definida por el papel causal o la posición de un estado mental en una tabla de máquina) es realizada por un sistema dado que instancia las clases o tipos de estados relacionados con la estructura abstracta del estado mental. Por lo tanto, ambas propuestas parecen compartir la idea según la cual la estructura abstracta (o, en otros términos, la sintaxis) de un estado mental, aun cuando no tiene contenido por sí misma, es aquello que determina en última

[141] *Ibid.*, p 247-248: "What needs to be said is that believing that P is somehow a constituent of believing that P∧Q; but the machine table state model has no conceptual resources for saying that. In particular, the notion 'is a constituent of' is not defined for table machine states."

instancia el contenido de la realización de un estado mental. Esta tesis, muy popular en la década 1960-1970 y todavía sostenida por varios autores en la actualidad,[142] ha sido criticada por Searle (1980) a través de un brillante argumento conocido como el "cuarto chino" a través del cual prueba que la mera estructura abstracta de un estado mental (i.e., su sintaxis) es insuficiente para generar contenido: sólo sirve para realizar una serie de computaciones. Ahora bien, es importante notar que, según Searle, "los procesos computacionales responden solamente a las formas de los símbolos; sus significados, o aquello que representan, son computacionalmente irrelevantes."[143] Es por ello que la estructura abstracta de un estado mental definida por las relaciones causales (o por la posición en la tabla de máquina) que mantiene con respecto a otro tipo de estados no basta para determinar el contenido del estado mental. Nuestro lector habrá sin duda constatado que el argumento de Searle constituye un medio para rechazar que la estructura abstracta de un estado mental baste para determinar su contenido. Sin embargo, este rechazo se hace *desde afuera*. Para que sea efectivo, el rechazo debe de darse *desde adentro*, i.e., debe de asumirse la tesis implícita de Lewis y Putnam para mostrar ulteriormente que se derivan de ella consecuencias absurdas o inaceptables Por lo tanto, asumiremos por mor del argumento que la estructura abstracta de un estado mental definida por el papel causal o por la posición en la máquina de Turing basta para determinar el contenido del estado mental. Ahora bien, como las características esenciales de un estado mental M (entre las cuales se encuentra su contenido) están determinadas por la red de relaciones causales que mantiene con otros estados mentales, y las características esenciales de cada uno de esos estados mentales están determinadas a su vez por la red de relaciones causales que mantiene con otros estados mentales, de ello se deduce que no se puede decir que dos sujetos estén en un mismo estado mental M a menos que compartan las mismas relaciones causales que tiene M con otros estados mentales (i.e., a menos que realicen la misma máquina de Turing), lo cual nos lleva a la siguiente conclusión:

[142] Dretske, 1981, p. 76: "Perhaps the intentionality of our cognitive attitudes (the way they have a unique content), a feature that some philosophers take to be distinctive of the mental, is a manifestation of their underlying information theoretic structure."
[143] Kim, 1998, p. 100: "Computational processes respond only to the shapes of symbols; their meanings, or what they represent, are computationally irrelevant."

> Para que dos sujetos estén en el mismo estado mental, ambos deben realizar la misma máquina de Turing. Pero si realizan la misma máquina de Turing, su psicología completa debe ser idéntica. Esto es, en el funcionalismo de máquina, la psicología completa de dos sujetos debe ser idéntica si es que van a compartir un sólo estado psicológico.[144]

Esta conclusión es intuitivamente absurda puesto que el hecho que mi hermana y yo queramos ir al cine a ver la misma película no implica que el resto de nuestra psicología sea idéntica Así pues, aun cuando el carácter holista de los estados mentales que presuponen tanto la propuesta de Lewis como la propuesta de Putnam presenta una ventaja que consiste en la posibilidad de recuperar la noción de estados mentales como causas o consecuencias de otros estados mentales sin caer en un círculo vicioso o en un regreso al infinito, tiene también una seria desventaja: es suficiente que dos individuos compartan el mismo estado psicológico en un momento dado (i.e., que compartan un estado con el mismo contenido) para sus psicologías tengan que ser idénticas, lo constituye una conclusión intuitivamente absurda. Esta desventaja arroja una duda sobre la validez de las tesis funcionalistas con carácter holista que sostienen tanto Lewis como Putnam.

Hemos visto que las propuestas de Lewis y de Putnam tienen un punto de contacto muy importante que consiste en que ambas comparten una concepción holista de los estados mentales que les permite intentar explicar los estados mentales por las relaciones causales que mantienen con otros estados mentales sin caer en un círculo vicioso o en un regreso *ad inifinitum*, y que las hace vulnerables ante la objeción de que es necesario que dos compartan su psicología completa si es que comparten en un momento dado un mismo estado mental. Sin embargo, la propuesta de Putnam se distingue de la propuesta de Lewis en varios otros puntos como, por ejemplo, en el hecho que Putnam declare que es menester aceptar que las máquinas de Turing que modelan las mentes de los seres humanos deben de disponer de órganos que detecten fallas internas:

---

[144] *Ibid.*, p. 92: "For any two subjects to be in the same mental state, they must realize the same Turing machine. But if they realize the same Turing machine, their total psychology must be identical. That is, on machine functionalism, two subject's total psychology must be identical if they are to share a single psychological state."

> Una máquina de Turing físicamente realizada puede no tener manera de determinar su propio estado estructural, de la misma manera en que un ser humano puede no tener ninguna manera de determinar la condición de su apéndice en un momento dado Sin embargo, es en extremo conveniente otorgar a una máquina 'órganos de los sentidos' electrónicos que le permitan escanearse a sí misma y detectar disfuncionamientos menores.[145]

¿Por qué es importante postular esta capacidad de revisión y alteración del programa para Putnam, capacidad que es completamente ajena tanto al concepto original propuesto por Turing como al modelo causal-funcional de la mente de Lewis? Esto se explica en la medida en que los seres humanos son el producto de una larga evolución biológica regida por la selección natural. En el curso de la evolución, las condiciones del entorno se ven continuamente alteradas por una gran variedad de eventos regulares tanto a corto plazo (cambio de estaciones, mareas, sucesión de la noche al día, etc) como a largo plazo (glaciaciones, terremotos, erupciones volcánicas, desecamiento, aparición de nuevos predadores, enrarecimiento del alimento, etc.), por lo cual era importante que los seres vivientes pudiesen adaptarse para sobrevivir. En el caso de seres complejos que pueden ser modelados por máquinas de Turing, esta adaptación requería una revisión continua de los estados internos para determinar si no había errores que pudiesen conducir a la muerte. Ahora bien, ¿en qué consiste exactamente el proceso de auto-revisión de los estados mentales? Acaso una de las mejores descripciones de este proceso sea la que Dennett denomina "prueba-y-error" (expresión que retoma de Bennett) y que consiste en "el tipo de consideración que uno hace cuando imagina los posibles resultados del comportamiento propio antes de actuar."[146] La supervivencia de una criatura que es capaz de revisar sus estados internos para determinar si sus acciones son apropiadas y de cambiar su conducta si es necesario es mucho más plausible que la de una criatura

[145] Putnam, 1960 in Putnam, *op. cit*, p. 371: "A physically realized Turing machine may have no way of ascertaining its own structural state, just as a human being may have no way of ascertaining the condition of his appendix at a given time. However, it is extremely convenient to give a machine electronic 'sense organs' which enable it to scan itself and to detect minor malfunctions."

[146] Dennett, 1969, p. 150: "(...) the sort of pondering one does when one imagines various outcomes to one's behavior before acting "

que es incapaz de ello y cuyos estados internos se encuentran inalterablemente determinados por ciertos tipos de estímulos, de conductas y de estados mentales antecedentes o consecutivos, sin posibilidad alguna de cambio. Este rasgo de la propuesta funcionalista de Putnam es muy importante en tanto que constituye la base a partir de la cual teorías funcionalistas que insisten en rasgos como la intencionalidad teleológica de las acciones y la plasticidad de la mente (como la teoría de Dennett que veremos más delante) se han desarrollado.

Tras haber expuesto algunas de las similitudes y de las divergencias que existen entre la propuesta de Lewis y la propuesta de Putnam, podemos hacer un balance general del funcionalismo de máquina de Turing. Al hacer este balance, podemos constatar que, aun cuando la teoría parece poder dar cuenta de los vínculos causales entre mente y cuerpo, es vulnerable también respecto a varias líneas de crítica (en particular, parece ser incapaz de recuperar la intencionalidad de los estados mentales), por lo cual existe al menos una duda *prima facie* respecto a su plausibilidad en tanto que teoría de la mente. A continuación, veremos cómo Dennett intentó aportar una respuesta a las limitaciones de la propuesta de Putnam.

III.1.3 La propuesta de Dennett[147]

La teoría que Dennett propone para suplir las limitaciones de las propuestas anteriores tiene un rasgo muy especial que conviene destacar. En principio, Dennett parece defender una versión de funcionalismo en tanto sostiene que una estructura funcional es "cualquier fracción de materia (e.g., cableado, plomería, cuerdas y poleas) de la cual puede decirse que –a causa de las leyes de la naturaleza– opera en cierta manera cuando es operada en cierta manera"[148], lo cual podemos traducir de la siguiente forma: una estructura funcional es cualquier fracción de materia que, al recibir ciertos estímulos, produce ciertas

[147] La propuesta de Dennett que será examinada en esta sección corresponde a las tesis que expone en (1969), por lo que algunos lectores más familiarizados con sus textos más recientes pueden sentirse algo desconcertados con la exposición y el análisis que aparecen aquí. Es indudable que la posición de Dennett ha cambiado considerablemente con el transcurso de los años; en particular Dennett presenta en (1987) una posición bastante distinta de la que aparece aquí. Sin embargo, la posición expuesta en (1987) también cuenta con varios problemas y objeciones en contra que no será examinadas aquí por razones de espacio.

[148] Dennett, 1969, p. 48: "[A functional structure is] any bit of matter (e.g., wiring, plumbing, ropes and pulleys) that can be counted on -because of the laws of nature- to operate in a certain way when operated in a certain way."

respuestas. Dennett acepta abiertamente como Lewis y Putnam que los estados mentales son estructuras funcionales en este sentido. Sin embargo, lo que distingue a Dennett de sus correligionarios consiste en el hecho de que, mientras Putnam y Lewis aceptan desde el primer momento que los estados mentales son funciones causales entre estímulos y conductas, y que es su rasgo definitorio, Dennett es mucho más cauto y declara que esta aseveración debe ser examinada con cuidado. La razón por la cual Dennett aconseja cautela es que, según él, muchas gentes tienden a unir en un sólo problema los vínculos causales entre mente y cuerpo así como la dimensión intencional de los estados mentales. Si bien es muy probable que esta tesis sea verdadera en última instancia según Dennett, no por ello se debe asumir como verdadera desde un primer momento como lo hacen Lewis y Putnam que parecen aceptar implícitamente que la tesis de la función causal de los estados mentales basta por sí sola para dar cuenta de todos los rasgos de los estados mentales (en particular de la intencionalidad y del contenido fenoménico o cualitativo de éstos). En contraposición a Lewis y Putnam, Dennett sugiere que el problema de la intencionalidad debe de ser tratado y resuelto separadamente para después abordar el problema de la relación causal mente-cuerpo. El problema de la intencionalidad radica en lo siguiente. Como hemos señalado previamente, la gente tiende a explicar y/ o predecir la conducta de los demás con base en oraciones del tipo: "Como creía que iba a llover y no quería mojarse, tomó un paraguas antes de salir de casa". Ahora bien, ¿en qué consiste la relación que se establece entre las dos partes de la oración (los estados mentales y la descripción de la conducta)? La mayoría de la gente (seguida en ello por los psicólogos de tendencias conductistas y por los filósofos materialistas) sostiene que la relación es causal. Ahora bien, para evitar el problema que llevó al dualismo cartesiano al colapso, se encuentran obligados a sostener que la creencia y el deseo son objetos materiales, de tal manera que pueden estar en una relación causal con otro objeto material (la conducta) sin necesidad de realizar "saltos ontológicos". Una explicación de este tipo cuenta con la ventaja de que, a primera vista, parece llenar el primer requisito que hemos establecido para que algo sea una buena teoría de la mente: dar cuenta de los vínculos causales entre la mente y el cuerpo. Sin embargo, existe un serio problema: la mayoría de la gente también sostiene que toda relación causal entre dos objetos particulares

debe ser una instancia de una ley causal general (de otra manera, la relación entre esos dos objetos sería un mero hecho fortuito, sin ningún valor científico).[149] Esto tiene la siguiente consecuencia: si hay dos objetos que se puedan articular por medio de una relación causal en una circunstancia determinada, entonces estos mismos objetos deben poder ser articulados por medio de la misma relación causal en todas las circunstancias posibles (o, por lo menos, en un número significativo de ellas para que tenga sentido hablar de la existencia de una ley causal que los vincule). Sin embargo, Dennett muestra la imposibilidad de cumplir este requerimiento en el siguiente ejemplo:

> Una actriz puede querer gritar, gritar en verdad, y sin embargo no gritar intencionalmente, porque aunque quería gritar en ese momento, gritó realmente porque estaba asustada. Así pues, el "porque" de las explicaciones intencionales resiste admirablemente ser tratado como un "porque" causal; debemos explicar la acción intencional X de A diciendo que A hizo X porque intentaba realizar X, y no puede darse de esta intención la caracterización independiente que necesita para ser una verdadera causa.[150]

Esta objeción de Dennett muestra que no puede darse un tratamiento causal (y, por lo tanto, tampoco un tratamiento nomológico) de la intencionalidad (al menos en principio) y que los vínculos causales entre mente y cuerpo y la intencionalidad de los estados mentales son dos características de la mente que deben ser tratadas a parte. ¿Cómo podemos explicar la dimensión intencional de los estados mentales para poder ulteriormente explicar el problema mente-cuerpo en términos funcionales? La hipótesis central de Dennett apunta a buscar el origen de la intencionalidad en el proceso de evolución de los organismos biológicos por medio de la selección natural. Esta hipótesis se encuentra justificada según Dennett por el

---

[149] Armstrong, *op. cit.*, p. 84: "But in speaking of the sequence as a causal sequence, we imply that there is some description that fall under a law."

[150] Dennett, *op. cit.*, pp. 36-37: "An actress can intend to scream, actually scream, and yet not intentionally scream, for though she did intend to scream at the time, she actually screamed because she was genuinely frightened. So the 'because' of intentional explanations steadfastly resists treatment as a causal 'because'; we must explain A's intentional action X by saying that A did X because he intended to do X, and this intention cannot be given the independent characterization it needs to be a proper cause."

hecho que "la descripción intencional supone la adecuación ambiental de las conexiones antecedente-consecuente; la selección natural garantiza, en el largo plazo, la adecuación ambiental de aquello que produce."[151] A continuación, examinaré brevemente cómo Dennett pretende resolver el problema de la intencionalidad para después determinar con exactitud el tipo de funcionalismo que propone así como las ventajas y las desventajas que tiene.

Según Dennett, el contenido de los estados mentales presupone la adecuación ambiental de las conexiones antecedente-consecuente; esto quiere decir que el contenido de un estado mental como el de dolor se encuentra determinado por el hecho de que aquellos grupos de estímulos que tienden a causarlo y aquellas conductas que tienden a ser sus efectos aseguran la supervivencia del organismo. De acuerdo con él, si un estado mental de dolor tiene un contenido especial, es porque los estímulos que tienden a producirlo (e.g., quemaduras, descargas eléctricas, cortaduras, etc.) son sentidos por el organismo como nocivos para su integridad y porque las conductas que tiende a producir (e.g., retirar la parte dañada de aquello que produce el estímulo, buscar aliviar la parte dañada, evitar en lo sucesivo cualquier tipo de proximidad con el origen del estímulo, etc.) tienden a restablecer o mantener la integridad del organismo, asegurando así su supervivencia. Ahora bien, es importante observar que la naturaleza no creó desde un principio organismos que estuvieran adaptados para sobrevivir y multiplicarse durante millones de años, adaptándose a las condiciones siempre cambiantes. Como bien señala Dennett, "la Madre Naturaleza (el proceso de selección natural) es singularmente miope y carente de objetivos."[152] En vez de una evolución con objetivos bien establecidos a partir de ciertos organismos con las características adecuadas, la Madre Naturaleza (para retomar la expresión de Dennett) produce ciega y continuamente un número muy grande de "experimentos" biológicos en los cuales se encuentran presentes asociaciones distintas de estímulos con varias conductas determinadas. Estas diversas asociaciones estímulo-conducta presentes en distintos tipos de organismos no contienen ni producen por sí solas una dimensión intencional; según Dennett, ésta

[151] *Ibid.*, p. 41: "Intentional description presupposes the environmental appropriateness of antecedent-consequent connections; natural selection guarantees, over the long run, the environmental appropriateness of what it produces."
[152] Dennett, 1991, p. 175: "Mother Nature (the process of natural selection) in famously myopic and lacking in goals."

sólo surge en los organismos a través de la interacción de estas asociaciones estímulo-conductas impresas en el sistema nervioso de los organismos con las características del entorno:

> El contenido, si es que existe, de un estado neuronal depende de dos factores: su origen normal en la estimulación, y cualesquiera efectos posteriores eferentes apropiados que tiene; y para determinar estos factores, se debe hacer una evaluación que vaya más allá de una descripción extensional de la estimulación y de la repuesta motora.[153]

Este concepto de contenido plantea la siguiente pregunta: si bien es cierto que el proceso de selección natural tiende a privilegiar en el largo plazo la supervivencia de los organismos cuyos sistemas nerviosos tienen impresas asociaciones estímulo-conducta que corresponden a las características del entorno (y, por lo tanto, que los contenidos intencionales de los estados mentales de los organismos producto del proceso de selección natural tienden a ser "correctos"), ¿qué ocurre cuando la naturaleza produce organismos cuyas asociaciones estímulo-conducta no son adecuadas para asegurar su supervivencia en el largo plazo (i.e., organismos cuyos estados mentales tienen contenidos intencionales "incorrectos")? Es muy natural suponer que, aun cuando estos organismos aparezcan ocasionalmente en la historia de la evolución, están destinados a desaparecer, ¿pero cómo podemos dar cuenta de ellos mientras existen? ¿Podemos decir que tienen contenidos mentales aunque éstos sean "incorrectos"? Dennett rechaza tajantemente esta hipótesis declarando lo siguiente:

> Uno sólo puede atribuir contenido a un suceso, un estado o una estructura neuronal cuando es un vínculo demostrablemente adecuado en una cadena entre lo aferente y lo eferente. El

[153] Dennett, 1969, p. 76: "The content, if any, of a neural state, event or structure depends on two factors: its normal source in stimulation, and whatever appropriate further efferent effects it has; and to determine these factors one must make an assessment that goes beyond an extensional description of stimulation and response locomotion "

contenido que uno atribuye a un suceso, un estado o una estructura no es entonces un rasgo adicional que uno descubra en él (...).[154]

Esta característica que Dennett atribuye a la intencionalidad plantea un serio problema en el contexto del funcionalismo que no puede ser resuelta, como veremos más adelante. Por el momento, consideremos solamente el concepto de contenido de un estado mental definido como un eslabón en la cadena de las conexiones aferentes y eferentes. ¿Cuáles son las implicaciones de esta noción de contenido sobre nuestro concepto de psicología según Dennett? De acuerdo con él, este concepto parece enterrar para siempre toda tentativa de hacer de la psicología una "ciencia autónoma de la intención." Si esta opción pudiera ser llevada a cabo, ello implicaría hacer de la psicología una ciencia totalmente aparte, donde las leyes causales no tienen cabida o aplicación y deben ser sustituidas por leyes intencionales. Sin embargo, es evidente que las leyes intencionales no tienen el carácter explicativo o sistematizador que las leyes causales tienen en el ámbito de las demás ciencias (como la física o la biología) en tanto que "el hecho que un evento (intencionalmente caracterizado) sea seguido de manera apropiada por otro no es ni siquiera contingente, por lo que no es sujeto de una explicación."[155] Dada esta noción teleológica de contenido, la única alternativa que tenemos es considerar que se puede dar cuenta de la intencionalidad de los estados mentales por medios únicamente materiales. Sin embargo, aun cuando la única opción para dar cuenta de la intencionalidad sea a través de medios materiales, es importante notar que hay por lo menos dos maneras de abordar el problema de la intencionalidad para Dennett en el marco de la alternativa materialista. Una manera es la posición periférica que consiste en "caracterizar los sucesos conductuales y la estimulación extensionalmente desde el principio, y llegar a leyes extensionales relacionando estos elementos."[156] Como la posición periférica corresponde claramente a las propuestas de tipo conductista, Dennett rechaza seguirla puesto que ninguna versión de conductismo plantea resolver el

[154] *Ibid.*, p. 78: "One can only ascribe content to a neural event, state or structure when it is in a link in a demonstrably appropriate chain between the afferent and the efferent. The content one ascribes to an event, state or structure is not, then, an extra feature that one discovers in it (...)"
[155] *Ibid.*, p. 38: "The fact that one event (as intentionally characterized) is followed in an appropriate way by another is not even contingent, and hence not subject to explanation."
[156] *Ibid.*, p. 41: "[While the peripheralist hopes to] characterize behavioral events and stimulation extensionally form the beginning, and arrive at extensional laws relating these (...)"

problema de la intencionalidad (acaso la única excepción sea la propuesta de Ryle que sí toma en cuenta las acciones, pero que no puede dar cuenta ellas a causa de los supuestos de trabajo a partir de los cuales se desarrolla). La otra manera consiste en la estrategia centralista que, de acuerdo con Dennett, consiste en lo siguiente:

> El centralista hace su caracterización inicial intencional, describiendo los eventos que deben ser relacionados de manera cuasi-nomológica usando expresiones intencionales ordinarias, semi-ordinarias o completamente artificiales. Posteriormente, espera que una base física adecuada pueda ser encontrada entre los estados y sucesos internos del organismo de tal manera que sean posibles 'reducciones' de las oraciones intencionales de la teoría a oraciones extensionales de la teoría.[157]

Como ni la "ciencia autónoma de la intención" ni la posición periférica son alternativas viables para Dennett, resulta sencillo constatar que se inclina más bien hacia una posición centralista en la cual, si bien las descripciones de los estados mentales se pueden dar intencionalmente en un principio, deben de poder ser reducidas en última instancia a descripciones extensionales (i.e., a descripciones que involucren única y exclusivamente objetos físicos). Así pues, en el ámbito de la teoría funcionalista que Dennett propone, los contenidos de los estados mentales se encuentran completamente determinados por las relaciones casuales adecuadas que mantienen con ciertos estímulos y con ciertas conductas.[158] De este concepto particular de contenido, se derivan algunas consecuencias bastante particulares para la determinación de las relaciones causales entre mente y cuerpo. Hemos señalado en varias ocasiones que la mayoría de la gente tiende a dar generalmente una interpretación causal a oraciones del tipo "Como creía que iba a llover y no quería mojarse, tomó un paraguas al salir de casa." Al igual que la mayoría de los

[157] *Loc. cit*: "The centralist makes his initial characterization Intentional, describing the events to be related in law-like ways using either ordinary, or semi-ordinary, or even entirely artificial Intentional expressions. He then hopes that an adequate physical basis can be found among the internal states and events of the organism so that 'reductions' of Intentional sentences of the theory to extensional sentences of the theory is possible."

[158] Esta conclusión a la que llega Dennett debe indudablemente mucho a las tesis conductistas, hecho que le ha sido reprochado en numerosas ocasiones por sus críticos que le reprochan el ser un "conductista" o "neo-conductista".

materialistas, Dennett rechaza que pueda haber una relación causal entre los dos elementos de las oraciones de este tipo (i.e., entre los estados mentales y las conductas) en la medida en que piensa que sólo puede haber una relación causal entre los dos elementos si caen bajo el dominio de una ley causal general (en este punto, sigue los pasos de Davidson). Por ello, Dennett no admite que se pueda dar una interpretación del estado mental como causa de la conducta: lo más que se puede hacer, en el contexto de la intencionalidad pura, es decir que el estado mental constituye una *razón* para la conducta. Creer que va a llover constituye sólo una razón (entre muchas otras razones) para exhibir la conducta que consiste en tomar un paraguas al salir de casa. Esto quiere decir que, si una persona P cree que va a llover, puede darse el caso que no tome un paraguas al salir de casa o puede darse el caso que, si lo toma, lo tome por una razón completamente distinta de la creencia que tiene que va a llover (e.g., puede querer regalárselo a un amigo). De acuerdo con Dennett, si un suceso S es una razón de otro suceso T, S no constituye una condición necesaria para que T ocurra: sólo puede ser una condición suficiente (una entre muchas otras). Ahora bien, si aceptamos la tesis según la cual los estados mentales pueden ser razones de las conductas, pero no causas (porque, de ser esto el caso, cada instancia de un cierto estado mental implicaría la ocurrencia de una conducta particular), ¿cómo se puede recuperar la noción de vínculos causales entre mente y cuerpo en el contexto de la teoría funcionalista de Dennett? La respuesta presentada por Dennett a esta pregunta consiste en señalar que sólo puede haber relaciones causales entre la conducta y estados funcionales realizados físicamente y que debemos tener la esperanza de que la neurología progrese de tal manera que, algún día, la intencionalidad de los estados pueda ser reducida a algo con bases materiales.[159]

Por todo lo anterior, la propuesta de Dennett sólo puede ser considerada como un candidato plausible para el papel de una buena teoría de la mente (en principio, al menos) con la condición de que aceptemos (1) que algún día podrá darse eventualmente cuenta de la dimensión intencional de los estados mentales apelando a la adecuación de las asociaciones estímulo-conducta con las características del ambiente

[159] *Ibid.*, p. 161: "The regularity with which the receipt of sad news is followed by crying suggests that there is a causal relation between the two, and neurologists may someday provide detailed confirmation of this hypothesis." Cf. también Dennett, 1987, p. 14: "If we knew more about physiological psychology, we could in principle determine whether or not you believe there is milk on the fridge, even if you were determined to be silent or disingenuous on the topic."

generada por el proceso de selección natural en el largo plazo y (2) que el contenido podrá eventualmente ser reducido en última instancia a algo material de tal manera que se pueda explicar la relación entre mente y cuerpo por medio de vínculos causales.[160] Ahora bien, para que la propuesta de Dennett pueda ser una buena teoría de la mente, también debe de poder dar cuenta del carácter fenoménico de los estados mentales. Sobre este punto, es interesante notar que Dennett adopta en *Content and Consciousness* una posición bastante distinta a la que adopta respecto a la intencionalidad de los estados mentales. Hemos señalado que una de las dos hipótesis sobre las cuales reposa la propuesta de Dennett consiste en la esperanza de que la neurobiología progrese en el porvenir a tal grado que haga posible una reducción de la dimensión intencional de los estados mentales a ciertos estados físicos del cerebro Sin embargo, esta perspectiva reduccionista no se mantiene en el caso de las características fenoménicas (e.g., la "dolorosidad" de una sensación de dolor) por la siguiente razón:

> Un análisis de nuestra manera ordinaria de hablar acerca de los dolores muestra que ningún suceso o proceso que exhiba las características de los "fenómenos mentales" putativos del dolor puede ser descubierto en el cerebro, porque el discurso acerca de los dolores es esencialmente no-mecánico, y los sucesos y procesos del cerebro son esencialmente mecánicos.[161]

Como las características fenoménicas de los estados mentales no son reducibles en principio a estados físicos del cerebro, Dennett sugiere que deben ser consideradas como elementos pertenecientes al nivel personal en el cual ninguna explicación puede ser mecánica mientras que los sucesos y los procesos del cerebro pertenecen al nivel sub-personal donde las explicaciones son por naturaleza mecánicas. Ambos

[160] Como nuestro lector puede sin duda apreciar, el razonamiento de Dennett depende en última instancia de dos "wishful thoughts" a través de los cuales nos pide que tengamos fe en el progreso de la neurobiología que habrá de vindicar su propuesta en un porvenir no muy lejano. Dejo a la consideración de mi lector la pregunta sobre si estos "buenos deseos" pueden constituir una base sólida para construir un buen argumento filosófico.

[161] Dennett, 1969, p. 91: "An analysis of our ordinary way of speaking about pains shows that no events or processes could be discovered in the brain that would exhibit the characteristics of the putative 'mental phenomena' of pain, because talk of pains is essentially non-mechanical, and the events and processes of the brain are essentially mechanical."

niveles de explicación deben ser separados en la medida en que, según Dennett, los términos que denotan los elementos del nivel de explicación sub-personal refieren a sucesos y procesos concretos y analizables por medio de las herramientas de la ciencia mientras que los términos que denotan los elementos del nivel de explicación personal refieren a entidades vagas y no-analizables. Este hecho sugiere que "es cierto en cierto sentido que no hay relación entre los dolores y los impulsos neuronales porque no hay dolores; el término 'dolor' no refiere."[162] La posición original de Dennett que consiste en separar las características fenoménicas de los estados mentales tanto de los vínculos causales mente-cuerpo como de la dimensión intencional de estos mismos estados mentales ha cambiado en el transcurso de los años, haciéndose más radical, hasta llegar a la tesis básica actual según la cual, como las características fenoménicas no son reducibles a ningún soporte material ni son analizables en términos relacionales, se debe sencillamente asumir que no existen.[163] De acuerdo con los tres criterios que hemos establecido para que algo pueda ser una buena teoría de la mente, la tesis actual de Dennett sobre la naturaleza de los rasgos fenoménicos de los estados mentales constituye una razón suficiente para descartarla automáticamente. Sin embargo, esta estrategia no es demasiado interesante puesto que constituye una refutación de la propuesta de Dennett *desde afuera*, i.e., desde una base que Dennett rechaza terminantemente. Para refutar el funcionalismo de Dennett de manera inteligente, debemos refutarlo *desde dentro*, i.e., desde sus propios supuestos. El propósito de la siguiente sección consistirá en mostrar que la propuesta de Dennett (así como las otras teorías funcionalistas de la mente en general) no pueden dar cuenta de la intencionalidad de los estados mentales como sostienen que pueden hacerlo en tanto que hay muchos problemas de por medio. Además, mostraremos que esta imposibilidad no obedece simplemente a cuestiones de hecho, sino que obedece a cuestiones de principio.

---

[162] *Ibid.*, p. 96: "(...) it is in one sense true that there is no relation between pains and neural impulses, because there are no pains; pain does not refer."

[163] Como podemos constatar, el argumento que parece hallarse detrás de la posición que Dennett asume sobre los qualia es (puesto muy crudamente) "Si no lo veo y no lo puedo analizar, entonces seguramente no debe de existir." Una vez más, dejo a la atenta consideración de mi lector determinar si esto constituye un buen argumento para deshacerse de los qualia.

III.2 El funcionalismo y la intencionalidad de los estados mentales

En la sección anterior, hemos visto que una de las dos hipótesis básicas del pensamiento de Dennett consiste en la creencia de que puede darse cuenta de la intencionalidad de los estados mentales apelando a la adecuación de las asociaciones estímulo-conducta con las características del ambiente generada por el proceso de selección natural en el largo plazo.[164] Dicho de otra manera, nuestros estados mentales tienen contenido en tanto nuestros cerebros son el producto de un largo proceso de selección natural, i.e., el producto de una historia. Ahora bien, algunos autores señalan que es posible imaginar un sistema que no sea el producto del largo proceso de selección natural y que, a pesar de ello, tenga exactamente el mismo conjunto de asociaciones estímulo-conducta (i.e., la misma organización funcional) que tiene una persona que es el producto actual del proceso de selección natural.[165] Aunque estemos obligados a aceptar que el desarrollo espontáneo de un sistema como el que hemos descrito no es demasiado plausible, es claro que Dennett (o cualquier otro funcionalista que quiera sentar las bases del contenido en el proceso evolutivo) debe admitir que existe por lo menos la posibilidad de que, en algún momento, se llegue a desarrollar un sistema con estas características. Este argumento muestra claramente que el proceso de selección natural no es una base necesaria y suficiente para explicar la existencia de contenido en los estados mentales de organismos como los seres humanos, puesto que hay una posibilidad (en extremo reducida, pero no nula) de que, algún día, los elementos que componen el fango de un pantano se reagrupen espontáneamente para formar un hombre del pantano ("Swampman") hecho de lodo y material orgánico en descomposición que reproduzca exactamente la organización funcional que tiene un hombre normal, producto del proceso de selección natural.

Ahora bien, aun cuando Dennett estuviese dispuesto a flexibilizar su postura para admitir la posibilidad de los hombres de pantano, existe otro serio problema en su propuesta que se deriva del hecho de que, según él, sólo tiene sentido hablar de contenido en un sistema cuando existe una adecuación de la

[164] Nuestro lector seguramente habrá seguramente constatado que esta noción funcionalista de contenido de Dennett es idéntica a la noción de aprendizaje del conductista, por lo cual sostener que el funcionalismo no es sino una versión particular del conductismo no es una tesis tan descabellada como algunas personas piensan

[165] Braddon-Mitchell y Jackson, *op. cit.*, p. 265

organización funcional del sistema con las condiciones del entorno. ¿Qué ocurre entonces con aquellos sistemas cuyas organizaciones funcionales no corresponden a las características del entorno? ¿Es plausible sostener que estos sistemas carecen de contenido? Esto parece poco probable, pero resulta sencillo constatar que aún menos probable es otra conclusión que se deriva de la misma idea en conjunción con otra: si asumimos que sólo existe contenido en los estados mentales de un sistema dado cuando existe una adecuación de la organización funcional del sistema con el entorno (como Dennett nos invita a hacerlo) y si asumimos que el contenido de los estados mentales es correcto cuando existe también el mismo tipo de adecuación (lo cual es una hipótesis muy plausible), resulta sencillo constatar que en la propuesta de Dennett (1969) *no puede existir el contenido erróneo*. De hecho, este problema no se limita sólo a la propuesta de Dennett: es un problema común para la totalidad de propuestas funcionalistas que buscan dar cuenta del contenido de los estados mentales a través de las relaciones causales que éstos mantienen con estímulos, conductas y otros estados mentales, como Fodor destaca:

> Las teorías causales tienen problemas para distinguir las condiciones para la *representación* de las condiciones para la *verdad*. El problema es intrínseco; las condiciones que las teorías causales imponen a la representación son tales que, cuando son satisfechas, la representación errónea no puede, por el mismo hecho, ocurrir.[166]

Muchos funcionalistas contemporáneos han intentado dar cuenta del problema del contenido erróneo a través de estrategias variadas. Acaso la más conocida y empleada consiste en declarar que el contenido de nuestras creencias o de nuestros deseos es erróneo cuando hay algo en el entorno que es anormal o cuando hay algo dañado en los órganos de los sentidos o en los mecanismos cognitivos internos responsables de la formación de las creencias o de los deseos:

---

[166] Fodor, 1990, p. 34: "Causal theories have trouble distinguishing the conditions for *representation* from the conditions for *truth*; This trouble is intrinsic; the conditions that causal theories impose on representation are such that, when they're satisfied, misrepresentation cannot, by that very fact, occur."

> Ahí donde las creencias son falsas (...) esperamos también algún tipo de explicación para la desviación de la norma: ya sea una anormalidad en el entorno, como en las ilusiones ópticas u otros tipos de evidencia que confunda, o una anormalidad en los mecanismos internos de formación de creencias.[167]

El problema con este acercamiento al problema del contenido erróneo consiste en que presupone, como bien destaca Fodor, la existencia de situaciones *normales* en las que el contenido de los estados mentales es correcto. Según Fodor, la normalidad de estas situaciones tiende a ser explicada por los funcionalistas "apelando a la teleología (natural), e.g., a una cierta noción más o menos darwiniana/histórica de los mecanismos biológicos *que hacen aquello para lo que fueron seleccionados*."[168] Según los funcionalistas, cuando disponemos de una historia que nos permite dar cuenta de un mecanismo cognitivo de un sistema intencional en concordancia con el proceso de selección natural, entonces es necesario que haya un cierto conjunto de circunstancias C (externas o internas al sistema) tales que el mecanismo sirva de mediador causal entre ciertos estados del entorno E y ciertos estados mentales M cuando C ocurre (*ceteris paribus*) y que la posesión de este mecanismo otorgue al sistema una ventaja decisiva sobre otros en el proceso de selección natural precisamente porque actúa como mediador causal entre E y M cuando C ocurre (*ceteris paribus*). Partiendo de esto, se puede presentar una noción de normalidad basada en C con las siguientes características:

> Identificamos situaciones "normales" como aquellas donde C ocurre; y decimos que si instancias de un estado mental del tipo S son causadas por instancias de P en tales situaciones, entonces, las instancias del estado mental S significan (expresan la propiedad P). Como las situaciones en las que C no ocurre son *ipso facto* anormales para la instanciación de

[167] Stalnaker, 1984, p. 19: "Where beliefs are false (. .) we also expect some explanation for the deviation of the norm: either an abnormality in the environment, as in optical illusions or other kinds of misleading evidence, or an abnormality in the internal belief-forming mechanisms."

[168] Fodor, *op. cit.*, p. 64: "[Normality should somehow be cashed out] by appeal to (natural) teleology; e.g ,to some more-or-less Darwinian/historical notion of biological mechanisms *doing what they were selected for* "

> S, y como se supone que es solamente en circunstancias normales que la causación es constitutiva del contenido, las instancias de S que aparecen cuando no ocurre C tienen la posibilidad de ser causadas por cualquier cosa. En particular, tienen la posibilidad de ser falsas.[169]

Con base en esta noción de normalidad, algunos funcionalistas como Israel (1987) sostienen que pueden explicar el problema del contenido erróneo a través del siguiente ejemplo. Si observamos a una rana cazar moscas y tener éxito la mayoría de las veces en su empeño, podemos sostener que, cada vez que ve una mosca, esta rana tiene un cierto estado mental cuyo contenido es que hay una mosca cerca. Los casos en que la rana persigue un señuelo volador de plástico similar a una mosca son explicados como casos de instanciaciones de un estado mental con un contenido erróneo en la medida en que no ocurre C sino C' (que es muy parecido a C). En tanto que no ocurre C, el mecanismo de causación no es relevante para el establecimiento del contenido en estos casos, por lo que el contenido puede ser erróneo. Al examinar minuciosamente esta estrategia para dar cuenta del contenido erróneo, Fodor destaca (muy atinadamente, creo yo) que la estrategia en cuestión no resuelve el problema del contenido erróneo, sino que solamente lo replantea de manera distinta, por la siguiente razón:

> Aunque se pueda describir el mecanismo teleológico de la mordida guiada de la rana como quiere Israel -en circunstancias normales, reacciona respecto a las moscas, por lo que su función es reaccionar respecto a las moscas, por lo que su contenido intencional es acerca de las moscas- no hay nada que nos detenga para contar la historia de manera distinta. En la versión alternativa, aquello para lo que el mecanismo neuronal está diseñado para responder son pequeñas cosas negras del entorno. (...) La moral es que (...) a Darwin no le importa

[169] *Ibid.*, pp. 69-70: "We identify 'Normal' situations as the ones in which it's the case that C; and we say that if mental states tokens of type S are caused by P-instantiations in such situations, then tokens of mental state S mean (express the property) P. Since situations where it isn't the case that C are *ipso facto* not normal for the tokening of S, and since it's only in normal circumstances that causation is supposed to be constitutive of content, S-tokens that transpire when it isn't the case that C are free to be caused by anything they like. In particular, they are free to be false."

> cómo se describe los objetos intencionales del cerramiento de la boca de la rana. Todo lo que importa para la selección es cuántas moscas la rana logra ingerir como consecuencia de sus mecanismos guiados para cerrar la boca, y este número resulta ser exactamente el mismo cuando uno describe la función de los mecanismos de cerramiento de la boca con respecto a un mundo poblado por moscas que son, de hecho, puntos negros del entorno, o con respecto a un mundo poblado por puntos negros del entorno que son, de hecho, moscas.[170]

Así pues, si queremos establecer una teoría teleológica del contenido (y esta tendencia es muy común entre los funcionalistas contemporáneos),[171] el problema principal al que nos vemos confrontados consiste en que dicha teoría no puede dar cuenta del problema del contenido erróneo (y tampoco del contenido correcto que es "mosca") puesto que es posible que tanto el contenido correcto como el contenido erróneo tengan exactamente el mismo valor o la misma utilidad con respecto a la supervivencia de los organismos en los cuales están presentes, por lo cual resulta imposible distinguirlos desde la perspectiva del proceso de selección natural. ¿Cómo podemos entonces dar cuenta del contenido erróneo? Una respuesta posible es la siguiente: si no podemos dar cuenta del problema de contenido erróneo por medio de las relaciones causales actuales que, según los funcionalistas, definen la naturaleza de un estado mental, acaso sea posible dar cuenta del problema por medio de relaciones causales contrafácticas (e.g., sustituyendo oraciones del tipo "Dado C, la presencia de un estado del entorno E causa la instanciación de un estado mental M" por oraciones del tipo "Dado C, si tuviésemos la presencia de un estado del entorno E, esto causaría una instanciación del estado mental M"). A primera vista, la ventaja clave de esta estrategia

---

[170] *Ibid.*, p. 71-72: "Though you can describe the teleology of the frog's snap guidance mechanism the way that Israel want you to -in Normal circumstances, it resonates to flies; so its function is to resonate to flies; so its intentional content is about flies- there is precisely nothing to stop you from telling the story in quite a different way. On the alternative account, what the neural mechanism is designed to respond to is little ambient black things. (...) The Moral (...) is that (...) Darwin doesn't care how you describe the intentional objects of the frog snaps. All that matters for selection is how many flies the frog manages to ingest in consequence of its snapping, and this number comes out exactly the same whether one describes the function of the snap-guidance mechanism with respect to a world that is populated by flies that are, de facto, ambient black dots, or with respect to a world that is populated by ambient black dots that are, de facto, flies."

[171] Además de Dennett (1987), los otros funcionalistas que buscan naturalizar la noción de contenido por medio de la teleología evolucionista son Millikan (1984), Papineau (1988), Israel (1987), Stalnaker (1984) y Sober (1985).

consiste en el hecho de que podemos dar cuenta de los casos de contenido erróneo en los estados mentales en la medida que, si consideramos la relación causal contrafáctica, es posible que ocurra una instanciación de M que no sea causada por E, sino por otra cosa (lo que constituye un error). Sin embargo, Fodor señala que esta estrategia no puede tener éxito a causa de la siguiente razón:

> Emplear contrafácticos para definir la función (y, por lo tanto, el contenido) sería renunciar a una solución darwiniana al problema de la disyunción [que es otra manera de designar el problema del contenido erróneo] en tanto que la utilidad que aumenta sólo en los entornos contrafácticos no produce ventajas actuales en la selección. (...) Consideremos, por ejemplo, los peces de colores brillantes que, según la leyenda popular, se hallan en los fondos carentes de sol de los océanos. No estoy seguro de la explicación evolutiva que esto tiene, pero algo es seguro: no puede consistir en que los peces sean de colores porque sería ventajoso para ellos si su entorno estuviera iluminado.[172]

Así pues, es claro que la introducción de condicionales contrafácticos para intentar definir el *contenido* de los estados mentales no es relevante, aunque si lo es para definir la *función* de los mismos de acuerdo con Fodor. Esta distinción es muy importante puesto que la mayoría de los funcionalistas tienden a creer que, una vez que ha sido establecida la función de un estado mental, también queda establecido el contenido del mismo estado de manera automática.[173] Sin embargo, esto no es el caso por la razón que habremos de exponer a continuación. En el caso en que logramos determinar la función de un estado mental, una de las cosas que determinamos también al mismo tiempo es su etiología, i.e., las causas que tienden a producir este mismo estado mental. Ahora bien, de acuerdo con Fodor, "el corazón de la teoría

[172] *Ibid.*, p. 75-76: "Going counterfactual to define function (and hence content) would be to give up on a Darwinian solution to the disjunction problem since utility that accrues only in counterfactual environments doesn't produce actual selectional advantages. (...) Consider, for example, the brightly colored fish that, according to the natural legend, are found in sunless ocean deeps. I don't know what the evolutionary explanation is supposed to be, but one thing is for certain: it can't be that the fish are colored because for them to be so would be advantageous if their environment were lit up."

[173] Dennett, 1987, p 321: "You can't have realism about meanings without realism about functions."

teleológica es la idea que, en 'circunstancias normales', las instancias de un símbolo sólo pueden tener un tipo de causa, es decir, el tipo de causa que fija el significado (Normalmente, sólo las vacas causan 'vacas')."[174] Esta tesis aparece con recurrencia en los autores funcionalistas contemporáneos puesto que la gran mayoría de ellos buscan dar cuenta del contenido por medios teleológicos. Sin embargo, esta tesis es errónea como bien lo muestra Fodor a través de la siguiente situación:

> Supongamos que, al no tener nada mejor que hacer, ocupo mi tiempo pensando acerca de las ranas. Y supongamos que, en el curso de esta meditación, por un proceso de asociación tan natural como pueda ser, mis pensamientos acerca de las ranas me llevan a pensamientos acerca de las moscas. El resultado es una instancia del tipo de estado mental *acerca del concepto MOSCA*, que es, seguramente, causado de una manera completamente natural (la teleología del funcionamiento mental puede evitar el *error*, pero no puede evitar el *pensar*). Pero no es una instancia de un estado intencional que haya sido causado por aquello que *significa*. Lo que provocó que pensara acerca de las moscas fue pensar acerca de las ranas, pero el efecto de esta causa era un pensamiento acerca de las moscas a pesar de ello.[175]

Así pues, es evidente por todo lo anterior que la tesis según la cual el contenido de los estados mentales puede ser explicado apelando a explicaciones de tipo teleológico es claramente errónea. Ahora bien, es muy importante delimitar el alcance de los argumentos que hemos presentado para no caer en errores: los argumentos de Fodor muestran la imposibilidad de naturalizar la intencionalidad vía la teleología, pero no muestran que sea imposible naturalizar la intencionalidad. De hecho, Fodor ha destacado en numerosas

[174] Fodor, *op. cit.*, p. 90: "The heart of the teleological theory is the idea that 'in Normal circumstances' the tokens of a symbol can only have one kind of cause -viz., the kind of cause that fixes meaning (Normally, only cows cause 'cows')."

[175] *Ibid.*, p. 80: "Suppose, having nothing better to do, I while away my time thinking about frogs. And suppose that, in the course of this meditation, by a natural process of association as it might be, my thoughts about frogs lead me to thoughts about flies. The result is a token of the mental state type *entertaining the concept FLY*, which is, surely, caused in a perfectly Normal way (the teleology of mental functioning may abstract from *error*, but surely it doesn't abstract form *thinking*). But it is not an instance of an intentional state that was caused by what it means. What caused me to think about flies was thinking about frogs; but the effect of this cause was a thought about flies for all that."

ocasiones que uno de los supuestos centrales alrededor del cual se articula su propuesta consiste en que la tesis (a probar) que psicología intencional debe poder ser explicada vía explicaciones naturalistas:

> Los psicólogos no tienen derecho a asumir que hay estados intencionales a menos que puedan proveer, o que al menos puedan prever el proveer, o que al menos no puedan prever ninguna razón de principio por la cual alguien no pudiera proveer condiciones naturalistas suficientes para que algo esté en un estado intencional.[176]

Hemos visto en las líneas anteriores las razones por las que las explicaciones naturalistas del contenido no pueden ser de tipo teleológico. Pero Fodor no se contenta con mostrar esto, sino que muestra que estas explicaciones tampoco pueden ser de tipo funcional, al menos en el sentido que los funcionalistas clásicos como Putnam atribuían al término. Como el funcionalismo clásico sostiene que los estados mentales se encuentran definidos por el hecho de que son funciones causales que mapean ciertos estados (estímulos o estados mentales) a otros estados (conductas y otros estados mentales consecutivos), podemos apreciar que inevitablemente nos conduce a un tipo de holismo semántico en el cual el contenido de cada estado mental está establecido por todas las relaciones causales que mantiene con otros estados mentales cuyo contenido también está establecido de la misma manera. Fodor se opone a cualquier tipo de holismo semántico derivado del funcionalismo con base en el argumento presentado por Block y él mismo en *What Psychological States are Not* (1972) según el cual, si dos sistemas intencionales comparten al menos un estado mental particular en el mismo instante y el holismo semántico es verdadero, entonces dichos sistemas comparten exactamente la misma psicología en su totalidad. Así pues, la propuesta que Fodor presenta en contraposición al holismo semántico derivado del funcionalismo consiste en el hecho que "el contenido de un pensamiento depende de sus relaciones *externas*, en la manera en que el

[176] Fodor, 1994, p. 5: "Psychologists have no right to assume that there are intentional states unless they can provide, or anyhow foresee providing, or anyhow foresee no principled reason why someone couldn't provide naturalistic sufficient conditions for something to in an intentional state"

pensamiento está relacionado con el mundo, *pero no en la manera en la que está relacionado con otros pensamientos.*"[177]

Si el contenido de un estado mental depende solamente de las relaciones que mantiene el estado mental con el mundo (por lo cual no es determinable empleando una estrategia holista), es claro que la semántica que requerimos debe ser atomista, lo que quiere decir en otros términos que "si todo lo que importa para saber si su pensamiento es sobre perros es saber si está causalmente conectado con perros, entonces sería posible *prima facie* tener el pensamiento de perro aún si no se tuvieran pensamientos sobre ninguna otra cosa."[178] Ahora bien, si aquello que requerimos es una semántica atomista, ¿cuáles son los candidatos más plausibles para ocupar este papel? En varias ocasiones, Fodor ha señalado que el candidato más plausible para ocupar el papel de una semántica atomista por medio de la cual se pueda naturalizar la dimensión intencional de los estados mentales es una semántica informacional ("informational semantics"). La semántica informacional se encuentra basada en la idea según la cual las instancias del símbolo "perro" tienden a ser producidas en presencia de entidades que instancian la propiedad de la "perreidad" (i.e., los distintos perros con los cuales nos topamos) en la medida que las distintas instancias del símbolo "perro" contienen información sobre los perros. Es muy importante notar que, de acuerdo con Fodor, el hecho de que un símbolo o un estado mental tenga contenido se encuentra íntimamente ligado al hecho de que ese símbolo o ese estado mental contiene una cierta información, pero ambos hechos no son equivalentes (esto es llamado por Fodor dependencia asimétrica). En caso de serlo, ello implicaría que cada objeto que contiene información también tiene contenido –una tesis que es denominada por Fodor *pansemanticismo*.

---

[177] *Ibid.*, p. 4: "The content of a thought depends on its *external* relations; on the way the world is related to the world, *not on the way it is related to other thoughts.*" Ahora bien, es importante notar que para Fodor sólo el contenido de un estado mental es atómico. Esto no implica que la estructura de la mente sea atómica igualmente, sino todo lo contrario. Como Fodor (1983) sostiene que la mente esta compuesta por distintos módulos con distintas funciones y actividades que dependen unas de otras (nótese bien que son las funciones de los módulos las que son interdependientes, no los contenidos) el modelo de la mente que propone consiste en un atomismo de contenido y en un holismo metodológico. Puesta en otros términos, la propuesta de Fodor consiste en sostener que el origen del contenido intencional es atomista mientras que su implementación es holista en tanto es computacional.

[178] *Ibid.*, p. 6: "If all that matters to whether your thought is about dogs is how it is causally connected to dogs, then, *prima facie*, it would be possible for you to have *dog* thoughts even if you didn't have thoughts about anything else."

Fodor presta una gran atención al rechazo del pansemanticismo y a la justificación de la distinción entre información y contenido a través del siguiente argumento:

> La intuición de que "significa" es unívoco –y que significa *contiene información sobre*– en "'humo' significa *humo*" y "humo significa fuego" se encuentra cerca del corazón de la semántica informacional. Pero esto no puede ser correcto. Si lo fuese, entonces (como "contiene información sobre" es transitivo) se seguiría que "humo" significa *fuego*, lo que no es el caso.[179]

Si el hecho que un símbolo o un estado mental contenga información es una condición necesaria, pero no suficiente para que el símbolo o el estado mental signifique algo, entonces la pregunta que se plantea automáticamente es la siguiente: ¿cuál es la condición suficiente para que un cierto objeto que contiene información tenga también contenido? Sobre este respecto, la propuesta de Fodor consiste en estipular la existencia de lo que denomina la "robustez" del contenido además de la dependencia asimétrica. La "robustez" del contenido consiste en el hecho de que, aun cuando las instancias de "gato" pueden ser producidas por todo tipo de objetos (gatos, perros, ratas gigantes, autómatas con forma de gato, etc.), cada instancia *significa* gato. Nótese que la noción de robustez del contenido permite la posibilidad del contenido erróneo en la medida que acepta la posibilidad de que haya objetos que no son gatos, pero que tienden a producir instancias de "gato" que significan *gato*. Ahora bien, Fodor señala que la robustez del contenido sólo tiene sentido en la medida en que postulamos que hay por lo menos algunos gatos que tienden a producir instancias de "gato", i.e., que "las instancias falsas son metafísicamente dependientes

[179] Fodor, 1990, pp. 92-93: "The intuition that "means" is univocal –and means *carries information about*– in "'smoke' means *smoke*" and "smoke means fire" is close to the heart of information based-semantics. But this can't be right. If it were, then (since "carries information about" is transitive") it would follow that "smoke" means *fire*; which it doesn't."

de las verdaderas."[180] La razón principal para introducir esta condición es que, de otra manera (i.e., si no hubiese por lo menos algún gato que produjese de vez en cuando una instancia de "gato"), no habría mucho provecho en el hecho de ser una criatura pensante puesto que todos nuestros pensamientos que contienen representaciones serían erróneos. Recapitulando lo que hemos señalado hasta ahora, los tres criterios básicos que Fodor establece como necesarios para elaborar una teoría del contenido correcta (i.e., una teoría que pueda dar cuenta del contenido erróneo y del hecho que "X" significa X) son los que enunciamos a continuación:

1°) "X causa 'X'" es una ley.

2°) Algunos "X" son causados realmente por X.

3°) Para todos los Y distintos de X, si los Y en tanto que Y causan realmente "X", entonces el hecho que Y cause "X" es asimétricamente dependiente del hecho X cause "X".

El rasgo más curioso de esta teoría del contenido consiste en que no apela tanto a las intuiciones del funcionalismo como a aquellas del conductismo en tanto sólo considera como elementos relevantes para determinar el contenido de los estados mentales los estímulos y las conductas. El contenido de *vaca* de un pensamiento de vaca que tenemos depende entonces solamente de las vacas con las que nos encontramos en el mundo así como de las instancias de "vaca" que pronunciamos observarlas. El conductismo tiene la ventaja de permitir el desarrollo de una semántica atomista, lo cual resulta imposible en el contexto del funcionalismo clásico (Lewis, Putnam y Dennett) que sólo permite el desarrollo de teorías de la mente con semánticas holistas. Así pues, un rasgo muy interesante de la propuesta de Fodor en el cual se aprecia una divergencia notable con la ortodoxia funcionalista consiste en que, si bien los funcionalistas sostienen que la estructura básica que articula la mente es Relaciones Causales → Función → Contenido, Fodor

[180] *Ibid.*, p. 91: "False tokens are metaphysically dependent on true ones." Es interesente notar que esta idea también aparece en los funcionalistas ortodoxos como Dennett. Cf. Dennett, 1987, p. 18: "The false beliefs that are reaped grow in a culture medium of true beliefs."

sostiene que la estructura básica es en realidad Relaciones Causales → Contenido → Función donde el vínculo entre relaciones causales y contenido se explica por medio del atomismo semántico y el vínculo entre contenido y función por medio del holismo metodológico:

> No se puede derivar el contenido intencional de un estado mental de su función biológica (al menos no si la explicación de su función biológica está basada en la historia de su selección). Pero podría ser aconsejable intentar ir por el otro camino: dada una explicación independiente, no-teleológica, y naturalista del contenido (...) se puede intentar construir la función de un estado mental con base en lo que representa. Por ejemplo, la función de la creencia que P es representar el mundo como siendo el caso que P en (ciertas) ocasiones cuando es el caso que P.[181]

Así pues, si bien el término "funcionalismo" pueda aplicarse a la propuesta de Fodor sin cometer un error, esto debe de hacerse con mucho cuidado en tanto el funcionalismo de Fodor es muy distinto del funcionalismo ortodoxo que otros autores como Dennett sostienen en tanto que la función de un estado mental no es derivable directamente de la red de relaciones causales que mantiene, sino del contenido del mismo. Asumiendo que la propuesta de Fodor es correcta (y que es posible, no dar cuenta del contenido en el contexto del funcionalismo, pero sí al menos reconciliar contenido y función definiendo ésta a partir del primero), se plantea otra pregunta: ¿basta esto para hacer una buena teoría de la mente? Aun cuando se asuma que la teoría del contenido de Fodor es correcta y satisfactoria (aseveración que no deja de suscitar varias interrogantes y dudas)[182], Fodor reconoce que "aun cuando sea verdad que la

[181] *Ibid.*, p. 129: "You cannot derive the intentional content of a mental state from its biological function (not, at least, if your account of its biological function is grounded on its selectional history.) But it might be well-advised to try going the other way 'round: given an independent, non-teleological, naturalistic account of content (...) you might try construing the function of a mental state by reference to what it represents. For example, the function of the belief that P is to represent the world as being such that P on (certain) occasions when it's the case that P."

[182] La limitación principal de la semántica de Fodor consiste en el hecho siguiente que él mismo reconoce (Fodor, 1994, p. 25): "The coinstantiation of broad content with its computational implementers is reliable and explicable,

intencionalidad es igual a la información más la robusteza, no se tendría que seguir de ello que la información más la robusteza es suficiente para tener mente."[183] A través de esta observación, podemos notar que Fodor parece seguir los lineamientos que hemos establecido en la introducción del presente trabajo: además de las relaciones causales y de la intencionalidad de los estados mentales, una teoría correcta y satisfactoria de la mente debe poder dar cuenta de los rasgos cualitativos de los estados mentales. En la siguiente sección, habremos de mostrar que el funcionalismo (en cualquiera de sus versiones posibles) no puede dar cuenta de estos rasgos cualitativos, por lo cual está condenado a ser una teoría incompleta de la mente.

III.3 El funcionalismo y las propiedades fenoménicas de los estados mentales

La presente sección constituye sin duda el punto más delicado que hayamos abordado hasta ahora en la medida en que las propiedades fenoménicas (o qualia) de ciertos estados mentales tienen un estatuto por demás especial en comparación a otras propiedades de los estados mentales (como la intencionalidad). Un punto que conviene sin duda precisar consiste en que no todos los estados mentales posibles tienen qualia, sino sólo aquellos que denominamos "experiencias" (e.g., sensaciones puras como el dolor o el hambre, colores y ciertos sentimientos). Es importante señalar que, si bien estas "experiencias" tienen un quale particular, en general carecen de contenido intencional –un dolor en general no es acerca de nada– mientras que los estados mentales con contenido intencional como las actitudes proposicionales carecen de qualia –mi creencia de que los Yankees van a ganar la próxima Serie Mundial no tiene ninguna característica fenoménica. Esta disyunción entre estados mentales intencionales y estados mentales con qualia (cuyos elementos parecen ser mutuamente excluyentes) ha sugerido a algunos autores como Block (1990) y Davies (1996) que lo que conocemos como *qualia* de los estados mentales es un tipo de

but metaphysically contingent; that they coinstantiate depends on some very general facts of the world, not on the metaphysical constitution of content as such."

[183] *Ibid.*, p 130: "Even if it's true that intentionality equals information plus robustness, it wouldn't have to follow that information plus robustness is sufficient for mentality."

contenido (habitualmente denominado *contenido perceptual* o *contenido fenoménico*) distinto del contenido intencional.[184] Este contenido tiene la particularidad de poseer una dimensión informativa (esto es claro en la medida que un dolor generalmente nos dice algo acerca del mundo, e.g., que ciertos tejidos de cierta parte de nuestro cuerpo están dañados) pero no es robusto puesto que sólo las instancias de "dolor" son producidas por instancias de dolor o, poniendo esto en otros términos, no hay posibilidad de error en el caso del dolor como lo hay en el caso de los estados mentales que tienen contenido intencional. Si veo un hombre beber un vaso que contiene un líquido transparente que parece agua, generalmente creo que está bebiendo agua, pero siempre existe la posibilidad de que el contenido de mi creencia sea errónea (e.g , el líquido en el vaso puede ser vodka). En cambio, cuando siento un dolor, no puede haber error alguno sobre el contenido de mi estado mental en la medida que un estado mental no puede instanciar la propiedad de la "dolorosidad" (i.e., no puede ser doloroso) independientemente del dolor. Varios materialistas han intentado demostrar que el concepto de indubitabilidad de los qualia es erróneo en tanto esta basada en una concepción absurda y contradictoria de la conciencia.

Tradicionalmente, los partidarios de los qualia han sostenido que éstos son indubitables en la medida que se encuentran *directamente presentes ante la conciencia*: no se puede padecer dolor o sentir hambre sin estar consciente de que se padece dolor o se siente hambre. Es menester notar que resulta imposible argumentar este hecho: los qualia son conscientes por la sencilla razón que no podemos concebir que no lo sean. Los intentos sucesivos de elaborar un argumento que sostenga el carácter consciente de los qualia han desembocado en una serie de *impasses* teóricas como la noción cartesiana de "idea clara y distinta" que no hacen sino oscurecer el problema.[185] Aprovechando la carencia de un argumento, los materialistas han señalado que la idea del carácter consciente de los qualia implica que una experiencia particular y la

[184] Es menester observar que, si bien la denominación de los qualia como "contenido fenoménico" es aceptada por autores como Block o Davies, hay otro autores como Stroud (2000) que rechazan que los qualia puedan ser contenido en la medida que el único contenido es el intencional. Para ellos, los qualia son sólo un cierto modo de presentación del pensamiento. En este trabajo, seguiré la línea de Block y Davies en la medida que pienso que los qualia son en verdad un tipo de contenido en tanto que no sólo proporcionan una cierta información acerca del estado de cosas del mundo, sino que esta información es comunicable.

[185] Este juicio se basa parcialmente en una observación de Wilson que admite que incluso ella es "incapaz de defender con algún detalle la noción de percepción clara y distinta." (Wilson, 1990, p. 61)

conciencia que tenemos de ella son la misma entidad. Los materialistas han buscado demostrar que esta conclusión sencillamente no se sostiene (lo cual implicaría, por *modus tollens*, que la idea del carácter consciente de los qualia tampoco se sostiene) a través de una serie de argumentos de los cuales el que presentamos a continuación constituye un caso paradigmático:

> Consideremos el análogo mecánico de nuestros estados mentales: el escaneo por un mecanismo de sus propios estados internos. Es claro que la operación de escaneo y la situación escaneada deben ser "existencias distintas". Una máquina puede escanearse a sí misma sólo en el mismo sentido en que un hombre puede comerse a sí mismo. Debe de existir una distinción absoluta entre el comedor y lo comido: por ejemplo, entre la mano y la boca. De igual manera, debe de haber una distinción absoluta entre el escáner y lo escaneado.[186]

Este argumento por analogía es muy sugestivo pero, como el propio Armstrong lo reconoce, está muy lejos de ser concluyente. En las líneas que siguen, intentaremos mostrar que este tipo de argumentaciones hechas por los materialistas para demostrar la distinción entre la sensación y la conciencia que tenemos de ella son erróneas. Es importante observar que, detrás de la noción de conciencia como escaneo sostenida por Armstrong y otros materialistas como Rosenthal (que sostiene que "el hecho que un estado mental sea consciente consiste en que uno tenga un pensamiento de que uno está en ese mismo estado mental")[187], se esconde un cierto tipo de contenido intencional.

---

[186] Armstrong, *op. cit.*, pp. 106-107: "Let us consider the mechanical analogue of awareness of our own mental states: the scanning by a mechanism of its own internal states. It is clear that the operation of scanning and the situation scanned must be 'distinct existences'. A machine can scan itself only in the same sense that a man can eat himself. There must remain a distinction between the eater and the eaten: mouth and hand, say. Equally, there must be an absolute distinction between the scanner and the scanned."

[187] Rosenthal, 1991, p. 31: "A mental state's being conscious consists in one's having a thought that one is in that very mental state."

Si admitimos el principio según el cual los qualia son conscientes, de ello se deriva que no sólo los qualia son siempre indubitables, sino que no existe la posibilidad de ignorarlos: siempre que ocurren, somos conscientes de que ocurren.[188] Esta tesis es correcta, pero no así la interpretación que los materialistas hacen de ella. Según ellos, el hecho de que siempre seamos conscientes de los qualia de nuestros estados mentales no es otra cosa más que la aplicación de la regla de auto-intimación ("self-intimation") que se enuncia de la siguiente manera:

Regla de auto-intimación: P (donde P es una oración que expresa un estado mental con qualia)

$\therefore$ P lógicamente implica que A cree P

Según los materialistas, si pronuncio la oración "Tengo un dolor de muelas" y soy sincero, entonces esto implica lógicamente que creo que tengo un dolor de muelas. En la medida que si una persona cree algo, es evidente que debe ser consciente de ese algo (e.g., no puedo creer que los Yankees van a ganar la próxima Serie Mundial y no tener conciencia de que creo que van a ganar la próxima Serie Mundial), la regla de auto-intimación parece constituir una expresión correcta de la tesis según la cual los qualia son conscientes. Sin embargo, esto no es el caso por la siguiente razón: como asumimos que P es una oración con contenido fenoménico y el contenido fenoménico es no-intencional (i.e., no puede ser algo distinto del sujeto que lo experimenta), no podemos considerar P como si tuviera contenido intencional (en tanto que el contenido intencional presupone la existencia de una distinción clara entre el objeto intencional y la actitud proposicional dirigida sobre el objeto intencional), y esto es precisamente lo que los materialistas hacen al hacer de ella el objeto de una actitud proposicional por medio de la regla de auto-intimación. En otros términos, no tiene sentido decir, desde la perspectiva de la primera persona, que creo que tiego un

[188] Burge ha argüido en (1997) que pueden existir propiedades fenoménicas de las que no somos conscientes. Esta aseveración es errónea en la medida en que se apoya en una distinción errónea entre la "fenomenalidad" ("what-it-is-likeness") y la conciencia fenomenal ("what it is currently like for the individual). El problema de esta distinción es que la "fenomenalidad" no puede ser una propiedad abstracta, sino que siempre debe ser instanciada. El dolor de dientes no puede ser dolor de dientes sino es dolor de dientes *de alguien o para alguien*.

dolor o que no lo tengo: sólo se puede decir que siento dolor (en cuyo caso tengo dolor) o que no lo siento (en cuyo caso no lo tengo).[189] Así pues, en tanto que los materialistas tienden a interpretar "ser consciente de" como "creer que" o "pensar que" en el caso de los qualia, es evidente cometen un serio error en tanto que dan un tratamiento intencional a algo que manifiestamente no lo es. De lo que hemos señalado anteriormente, se deduce que los qualia parecen tener una cierta autonomía con respecto al contenido intencional que debemos respetar, so pena de caer en contradicciones.

Es menester notar que la intuición que hemos mencionado en las líneas anteriores sobre la naturaleza de las propiedades fenoménicas choca con el proyecto básico del pensamiento materialista (del cual las dos principales versiones del conductismo, la teoría de la identidad mente-cerebro y el funcionalismo en general constituyen distintas instancias) que pretende naturalizar la mente en su totalidad. En principio, para poder ser llevado a cabo, este ambicioso proyecto de naturalización de la mente presupone que todas las propiedades o rasgos de la mente (incluyendo, por supuesto, las relaciones causales entre mente y cuerpo, la intencionalidad de los estados mentales y las propiedades fenoménicas de éstos) deben poder ser naturalizados, i.e., deben poder ser definidos con base en o reducidos a entidades o propiedades físicas. Sin embargo, esto implica que estas propiedades o rasgos de la mente deben de ser intencionales. En efecto, si una propiedad es física, entonces puede ser un objeto intencional como podemos apreciar a través del siguiente ejemplo: la propiedad de la "gatidad" (que es una propiedad física en tanto que es realizada por ciertos objetos físicos) es un objeto intencional en tanto que un gato es una entidad distinta de la idea que podemos tener de él. ¿Cómo hacer entonces para salir de esta *impasse* que surge de los intentos de caracterización de las propiedades fenoménicas? Los funcionalistas (así como otros materialistas) han propuesto varias alternativas para resolver el problema acerca de las propiedades fenoménicas. En las líneas que siguen, enlistaremos las distintas alternativas que han sido propuestas

[189]Algunos podrían objetar que, en el caso de oraciones como "Juan cree que María tiene dolor de muelas", hay un contenido intencional tanto como un contenido fenoménico. Lamentablemente, este no es el caso en la medida que, para que tengamos contenido fenoménico, debe de haber única y exclusivamente acceso de primera persona. Como en la oración que hemos mencionado, hay acceso de tercera persona, sólo tiene contenido intencional en ella.

hasta ahora, mostrando en cada ocasión las razones por las cuales la alternativa propuesta es errónea o insuficiente.

Una de las alternativas más populares actualmente entre los miembros de la comunidad filosófica para naturalizar la mente consiste en sostener que todas las propiedades y los rasgos de la mente pueden ser naturalizados, incluyendo los qualia. Esta corriente de pensamiento sostiene que, aun cuando no existe en la actualidad una relación clara que permita ligar los qualia a estados físicos del cerebro, eventualmente la investigación y el progreso en las diversas áreas de la neurociencia nos permitirá llevar esto a cabo como señalan los Churchland al declarar que "la naturaleza de los qualia específicos será revelada por la neurofisiología, la neuroquímica y la neurofísica."[190] Asumir que este proyecto puede ser llevado a cabo (y no sólo ser llevado a cabo, sino ser llevado a cabo por el funcionalismo) presupone que los qualia pueden ser reducidos a estados o propiedades físicas. Sin embargo, esta presuposición implica, según Kim, que "los fenómenos que deben ser reducidos [i.e., los qualia] son conceptualizados relacionalmente o extrínsecamente, en términos de sus relaciones causales/ nomológicas con otros fenómenos [físicos], y no intrínsicamente, en términos de su carácter cualitativo interno o de su estructura composicional."[191] Esta presuposición choca directamente con la tesis según la cual los qualia son propiedades intrínsecas no-relacionales. Ahora bien, ¿existen acaso buenas razones para suponer que los qualia son propiedades relacionales? Varios experimentos recientes en el campo de la neurociencia han sugerido (mas no demostrado) que los qualia dependen estrechamente de ciertos sucesos eléctricos del cerebro (por lo que es plausible que sean propiedades relacionales). Por ejemplo, se ha mostrado en Gray *et al.* (1987) que dos neuronas pertenecientes al cortex visual primario de un gato y separadas por 7 milímetros (una distancia considerable considerando el tamaño de las neuronas) que son sensibles al mismo estímulo (una barra moviéndose en el campo visual del gato) producen impulsos eléctricos sincronizados en un rango de

[190] Churchland, 1981 en Churchland, 1989, p. 31 "The nature of specific qualia will be revealed by neurophysiology, neurochemistry, and neurophysics."

[191] Kim, *op. cit*, p 175-176· "The phenomena to be reduced [i.e., the qualia] are conceptualized relationally or extrinsically, in terms of their causal/nomological relationships to other [physical] phenomena, and not intrinsically, in terms of their qualitative character or compositional structure."

frecuencias de 35 a 75 Hertz (aunque por lo general se considera el valor promedio de 40 Hz). Haciendo uso de estos resultados experimentales, Crick y Koch (1990) han propuesto una hipótesis interesante según la cual la conciencia (y, en especial, la conciencia de una sensación de rojo o, hablando en términos más precisos, de un quale rojo) puede ser explicada con base en las oscilaciones neuronales de 40 Hz:

> Sugerimos que una de las funciones de la conciencia es presentar el resultado de varias computaciones subyacentes y que esto implica un mecanismo de atención que temporalmente liga las neuronas relevantes sincronizando sus picos en oscilaciones de 40 Hz. Estas oscilaciones no codifican ellas mismas ninguna información adicional, excepto en la medida que unifican parte de la información existente en una percepción coherente.[192]

Además de los neurobiólogos, algunos miembros de la comunidad filosófica como Flanagan (1992) han expresado un cierto optimismo con respecto a la hipótesis de que la naturaleza de los qualia pueda ser explicada por medio de este tipo de experimentos. Sin embargo, aun cuando supongamos que algún día la neurociencia podrá proporcionar una lista exhaustiva de las correlaciones causales entre los qualia y las propiedades físicas o eléctricas de los estados neuronales, esto no resuelve el problema de la naturaleza de la mente, sino sólo lo plantea de manera más acuciante como bien señala Kim al escribir lo siguiente:

> Los neurocientíficos podrán algún día proporcionarnos una lista exhaustiva de correlaciones [causales] entre qualia y estados cerebrales, y esto incrementaría nuestro conocimiento del cerebro y de las maneras particulares en las que nuestra vida consciente depende de aquello que ocurre en el cerebro. Pero estas correlaciones son exactamente aquello que hace surgir

[192] Crick y Koch, 1990 en Block *et al.*, 1997, p. 288: "We suggest that one of the functions of consciousness is to present the result of several underlying computations and that this involves an attentional mechanism that temporarily binds the relevant neurons together by synchronizing their spikes in 40 Hz oscillations. These oscillations do not themselves encode additional information, except in so far as they join together some of the existing information into a coherent precept."

> los problemas filosóficos; lo que necesitamos es una explicación de porqué este particular sistema de correlaciones, entre una miríada de otros sistemas posibles, se sostiene en nuestro mundo.[193]

Así pues, podemos apreciar que la estrategia propuesta por Crick y Koch para naturalizar los qualia es sin duda satisfactoria para los científicos, pero no lo es para los filósofos en la medida que, mientras que los científicos se contentan con mostrar que las correlaciones son un hecho, los filósofos deben exigir que se den las condiciones necesarias y suficientes para que este hecho sea metafísicamente necesario. Pero es sencillo constatar que dichas condiciones necesarias y suficientes no pueden ser dadas en tanto que esto requeriría que se demostrásemos que es metafísicamente necesario que los qualia estén relacionados con las oscilaciones neuronales de 40 Hz en todos los mundos posibles donde hay qualia y cerebros con oscilaciones neuronales de 40 Hz. Es por ello que la propuesta de Crick y Koch para naturalizar los qualia (así como todas las propuestas similares) dista mucho de ser idónea.

Hemos visto anteriormente que la condición para que los qualia puedan ser reducidos a sucesos físicos es que sean conceptualizados relacionalmente o extrínsecamente, en términos de sus relaciones causales/ nomológicas con otros sucesos físicos como estímulos o conductas. Algunos autores han sugerido que, si bien la conceptualización relacional o extrínseca de los qualia es indispensable para poder llevar a cabo el proyecto de naturalización, acaso la causalidad sea un criterio demasiado estricto para poder establecer las relaciones entre los qualia y los otros sucesos físicos. Para llevar a cabo el proyecto de naturalización, se ha buscado entonces establecer la conceptualización extrínseca de los qualia con base en otro vínculo relacional distinto de la causalidad, y el candidato más plausible para ocupar el lugar de este vínculo es la

[193] Kim, *op. cit.*, p. 177: "Neuroscientists may someday deliver to us an exhaustive list of [causal] qualia-brain state correlations, and this would add to our knowledge of the brain and the particular ways in which our conscious life depends on what goes on in the brain. But these correlations are exactly what gives rise to philosophical puzzles; what we need is an explanation of why this particular system of correlations, out of the myriad of possible ones, holds in our world."

relación de superveniencia.[194] Hemos hablado previamente de este concepto en el capítulo anterior, pero sin duda conviene recordar brevemente a nuestro lector en que consiste. Para ello, emplearemos la definición de superveniencia presentada por Davidson que, al describir los rasgos de su propuesta (el monismo anómalo), destaca lo siguiente:

> Aunque la posición que describo niega que haya leyes psicofísicas, es consistente con el punto de vista de que las características mentales dependen en cierto sentido de, o supervienen, en las características físicas. Tal superveniencia podría tomarse en el sentido que no puede haber dos sucesos iguales en todos sus aspectos físicos pero diferentes en algún estado mental sin que se altere en algún aspecto físico.[195]

Al hacer el análisis de la noción de superveniencia propuesta por Davidson, Kim señala que existen tres componentes básicos que estructuran la noción: covariancia, dependencia e irreductibilidad.[196] Entre estos tres componentes, el más importante sin duda es la covariancia en tanto que, según la interpretación que tengamos de él, tendremos una noción distinta de superveniencia. Kim ha señalado que existen al menos dos conceptos básicos de covariancia que pueden ser formuladas de la siguiente manera:

Covariancia débil: Necesariamente, si alguna cosa tiene la propiedad F en A, existe una propiedad G en B tal que la cosa tiene G, y todo lo que tiene G tiene F.

---

[194] Mi lector puede plantearse sin duda la siguiente pregunta: si hemos presentado previamente un razonamiento que demuestra la vaguedad y la confusión de la noción de superveniencia al ser introducida en argumentos filosóficos (Cf. supra pp. 85-87), ¿por qué debemos reintroducir la noción en este punto para intentar explicar la naturaleza de los qualia? La razón que nos motiva a hacer esto es que muchos autores no aceptan el radicalismo de Schiffer, y sí consideran que la superveniencia puede ser una alternativa. Como la crítica de Schiffer es llevada a cabo desde una perspectiva externa, acaso no sea lo más idóneo para rechazar la superveniencia como alternativa. Lo que habremos de hacer en este punto es rechazar la superveniencia desde dentro, i.e., asumirla en primera instancia como opción para mostrar ulteriormente que es insuficiente para demostrar concluyentemente aquello que se quiere demostrar.

[195] Davidson, 1995, pp. 271-272

[196] Kim, 1993, p. 140.

Covariancia fuerte: Necesariamente, si alguna cosa tiene la propiedad F en A, existe una propiedad G en B tal que la cosa tiene G, y *necesariamente* todo lo que tiene G tiene F

Estos dos conceptos básicos de covariancia constituyen las bases sobre las cuales posteriormente se puede construir los dos conceptos básicos de superveniencia que pueden ser formuladas de la siguiente manera:

Superveniencia débil: A superviene débilmente en B si y sólo si necesariamente para cada propiedad F en A, si un objeto X tiene F, entonces existe una propiedad G en B tal que X tiene G, y si cualquier Y tiene G también tiene F.

Superveniencia fuerte: A superviene fuertemente en B cuando, necesariamente, para cada X y cada propiedad F en A, si X tiene F, entonces existe una propiedad G en B tal que X tiene G, y *necesariamente* si cualquier Y tiene G, entonces también tiene F.

En las líneas que siguen, habremos sólo de examinar el concepto de superveniencia débil. El concepto de superveniencia fuerte no será estudiado puesto que, como el concepto de superveniencia fuerte implica el concepto de superveniencia débil, basta con demostrar que la tesis de la superveniencia débil de los qualia en los estados físicos del cerebro es insuficiente para poder concluir, por *modus tollens*, que la tesis de la superveniencia fuerte es también insuficiente. Por un lado, es importante observar que han sido llevados a cabo varios experimentos que sugieren que en verdad los qualia supervienen en los estados físicos del cerebro. Se ha observado que existen varias sustancias que, al ser administradas a un individuo, provocan cambios en la estructura química de su cerebro, y que estos cambios químicos parecen alterar los qualia (e.g., la ciencia médica ha constatado que algunos alcaloides suprimen las sensaciones de dolor y que el litio tiene notables propiedades antipsicóticas). Sin embargo, estas observaciones sólo sugieren que existe una superveniencia contingente de los qualia en los estados físicos del cerebro, pero no *muestran* que esta relación de superveniencia exista como verdad relevante para los filósofos (i.e., que sea

metafísicamente necesaria). Es interesante observar que, en aquellas situaciones donde el vínculo de causalidad entre qualia y estados neuronales es remplazado por un vínculo de superveniencia débil, aparece exactamente el mismo tipo de problema que encontramos en las situaciones donde presuponemos un vínculo causal –un problema que es descrito por Kim en los siguientes términos:

> La investigación en neurofisiología quizás nos dirá más acerca de la base biológica de las experiencias fenoménicas, pero es difícil ver cómo podría esto ser relevante como evidencia para la superveniencia de los qualia. En el mejor de los casos, tales descubrimientos nos dirán más acerca de las correlaciones nomológicas entre qualia y estados neuronales subyacentes, pero la pregunta tiene que ver con el hecho si estas correlaciones son *metafísicamente necesarias* –si hay mundos posibles en que las correlaciones fallen.[197]

Intuitivamente, la condición central para que la superveniencia de los qualia en los estados neuronales sea metafísicamente necesaria consiste en que la existencia de un estado neuronal sea condición necesaria y suficiente para la existencia del qualia que se supone superviene en él. Sin embargo, hay varios autores como Kripke que han señalado que esto no es verdadero:

> ¿Qué pasa con el caso de la estimulación de las fibras C? Parecería que para crear este fenómeno, Dios necesita crear solamente seres con fibras C capaces de tener el tipo adecuado de estimulación física; que los seres sean o no conscientes es algo aquí sin importancia. Parecería, sin embargo, que para hacer que la estimulación de la fibra C corresponda al dolor, o sea sentida como dolor, Dios tiene que hacer algo además de la mera creación de la fibra C;

[197] Kim, 1998, p. 171: "Research in neurophysiology will perhaps tell us more about the biological basis of phenomenal experiences, but it is difficult to see how this could be relevant as evidence for qualia supervenience. At best such discoveries will tell us more about the lawful correlations between qualia and underlying neural states, but the question has to do with whether these correlations are *metaphysically necessary* –whether there are possible worlds in which the correlations fail."

> tiene que hacer que las criaturas sientan la estimulación de la fibra C como *dolor*, y no como una cosquilla, o como calor, o como nada, tal y como aparentemente también habría estado en su poder.[198]

Si el razonamiento de Kripke es correcto (y existen muy buenas razones para suponer que es correcto puesto que, como el dolor no puede ser idéntico a una fibra C excitada, Dios tiene que hacer algo para que el dolor supervenga en la fibra C excitada), entonces se presentan dos tipos de objeciones que demuestran de manera patente la imposibilidad de dar cuenta de los qualia en el contexto del funcionalismo por medio de la superveniencia. El primer tipo de objeciones corresponde a los mundos posibles que son idénticos al mundo actual en todos los aspectos físicos relevantes, pero en los cuales los vínculos de superveniencia entre qualia y estados neuronales son distintos de lo que son en el mundo actual. El paradigma de este tipo de mundos posibles son los mundos donde existen "espectros invertidos". Una de las más depuradas exposiciones del argumento de los "qualia invertidos" aparece en el artículo *Inverted Earth* (1990) donde Block argumenta que, por medio de él, se puede mostrar que el funcionalismo es incapaz de dar cuenta de los qualia siguiendo la siguiente estrategia:

> Describiré un caso de dos personas/niveles cuyas experiencias son cualitativamente las mismas, pero intencionalmente y funcionalmente invertidas. Si tengo razón sobre este caso, la distinción entre el contenido intencional y el contenido cualitativo de la experiencia será justificada, y la teoría funcionalista del contenido cualitativo refutada.[199]

---

[198] Kripke, *op. cit.*, p. 149

[199] Block, 1990 en Block *et al.*, 1997, p. 682: "I will describe a case of two persons/stages whose experiences are qualitatively the same but intentionally and functionally inverted. If I am right about this case, the distinction between the intentional and qualitative content of experience is vindicated, and the functionalist theory of qualitative content is refuted."

La situación que Block plantea para mostrar que no puede darse cuenta de los qualia funcionalmente es la siguiente: Block imagina a dos gemelos A y B que han sido criados juntos en la Tierra (el hecho que los individuos sean gemelos es necesario para que podamos hablar de una identidad neuronal entre A y B en los aspectos relevantes). Tras haber llegado a la edad adulta, el individuo B es raptado un día mientras duerme por un grupo de científicos locos extraterrestres y transportado a una velocidad hiperlumínica a su planeta que llamaremos Tierra Invertida. La Tierra Invertida tiene la particularidad de ser idéntica a la Tierra en todos los aspectos relevantes (e.g., ambos planetas tienen exactamente el mismo tamaño, la misma distribución de masas de agua y tierra seca, la misma historia evolutiva y social así como el mismo número de hispanohablantes distribuidos a través de las mismas extensiones de tierra), con dos pequeñas salvedades. En primer lugar, en la Tierra Invertida todo tiene el color complementario que tiene la Tierra (el pasto es rojo, las berenjenas son amarillas, las zanahorias y las naranjas azules, etc.) y, en segundo lugar, el vocabulario de los habitantes de la Tierra Invertida también está invertido, i e., cuando se le pregunta a un habitante hispanohablante de la Tierra Invertida de qué color (azul) son las zanahorias, responde (diciendo la verdad) "Naranja". Los dos rasgos que hemos mencionado hacen que los habitantes de la Tierra Invertida se encuentren no sólo funcionalmente invertidos con respecto a los habitantes de la Tierra (lo cual resulta sencillo constatar en la medida que los estímulos y las respuestas están invertidos), sino también intencionalmente invertidos (en efecto, cuando un habitante de la Tierra Invertida dice "El naranja de la zanahorias es muy pálido", su pensamiento es en realidad acerca de *azul*). Cuando el gemelo B llega a la Tierra Invertida, los científicos locos proceden a insertar en sus pupilas unos lentes inversores mientras aún está dormido que hacen que los objetos de la Tierra Invertida parezcan exactamente como los de la Tierra de tal manera que, cuando B despierta, piense que se encuentra todavía en la Tierra.[200] Asumiendo que esto ocurre, Block señala que los qualia de B en la Tierra Invertida son exactamente los mismos que los que tenía previamente en la Tierra gracias a los lentes inversores. En lo que concierne el

[200] La situación propuesta por Block requiere asumir que, además de la inserción de los lentes, el cuerpo de B ha sido recubierto por una pintura del color complementario que normalmente tiene. Esto es necesario para que, si B tiene la tentación de mirarse en un espejo o de mirar directamente alguna parte de su cuerpo, no se dé cuenta de que ha sido llevado a la Tierra Gemela.

contenido intencional, Block señala que, en un primer momento, el contenido intencional de las creencias y de los pensamientos de B está invertido con respecto a lo que era Tierra (e.g., si B dice al mirar el cielo de la Tierra Invertida "¡Qué azul más hermoso!", su pensamiento de *azul* se encuentra dirigido sobre algo que en realidad es *naranja*, por lo cual el contenido intencional de su pensamiento es erróneo), pero que, con el transcurso del tiempo, tendríamos una adaptación progresiva de B de tal manera que los contenidos intencionales de sus estados mentales terminarían por ser los mismos de que los de los habitantes nativos de la Tierra Invertida. Según Block, la mera posibilidad de la Tierra Invertida basta para rechazar la teoría funcionalista del contenido cualitativo (y, al mismo tiempo, las tentativas de explicar exhaustivamente la naturaleza de la mente en un marco funcionalista) por la siguiente razón:

> Si el estado mental M = el estado funcional F, entonces cada instancia de M debe ser una instancia de F. Pero en el caso de la Tierra Invertida descrito más arriba, dos instancias del mismo estado cualitativo –el suyo y el de su gemelo– tienen distintos papeles funcionales. Y el estado cualitativo que usted comparte con su gemelo no puede ser idéntico a su estado funcional y al estado funcional de su gemelo, en tanto esto sería un caso de $Q=F_1$, $Q=F_2$ y $\neg(F_1=F_2)$, lo que contraviene a la transitividad de la identidad.[201]

El segundo tipo de objeciones que se desprende del razonamiento de Kripke respecto a la posibilidad de dar cuenta de los qualia en el contexto del funcionalismo corresponde a los mundos posibles que son idénticos al mundo actual en todos los aspectos físicos relevantes, pero en los cuales no hay vínculos de superveniencia entre qualia y estados neuronales puesto que los seres humanos en esos mundos posibles, aun cuando son funcionalmente idénticos a nosotros, no tienen qualia. En estos mundos posibles, aunque

[201] *Ibid.*, p. 684: "If mental state M = functional state F, then any instance of M must also be an instance of F. But in the Inverted Earth case just described, two instances of the same qualitative state –yours and Twin's– have different functional roles. And the qualitative state that you shae with Twin cannot be identical with to both your functional state and Twin's functional state, since that would be a case of $Q=F_1$, $Q=F_2$ and $\neg(F_1=F_2)$, which contravenes the transitivity of identity."

las personas que pronuncian la oración "Tengo un dolor de muelas" se comportan exactamente como nosotros cuando tenemos un dolor de muelas, en realidad no experimentan nada: son *zombis*. Las voces de varios funcionalistas (así como de otros materialistas) se han alzado para rechazar la posibilidad de los zombis a través del siguiente argumento:

> La condición de zombi se convierte en una posibilidad sólo bajo una perspectiva acorde con el epifenomenalismo. Si sostenemos que la conciencia tiene poderes causales, entonces la ausencia de conciencia en mi gemelo zombi, que es idéntico a mí en cualquier otro respecto, implicaría una diferencia. Pero, por estipulación, no hay ninguna diferencia entre las personas y sus gemelos zombis, excepto en el hecho que éstos últimos carecen de conciencia. Por lo tanto, negar el epifenomenalismo sería bloquear la posibilidad de la condición de zombi.[202]

Tras haber presentado el argumento, es menester destacar un punto: para demostrar que la posibilidad de zombis no es viable, no basta con demostrar que el epifenomenalismo no ocurre en el mundo actual –de hecho, existen razones empíricas de sobra para sostener que los qualia tienen poderes causales con respecto a la materia en el mundo actual, y coincido plenamente con los que las juzgan correctas,[203] debe mostrarse que el epifenomenalismo es metafísicamente imposible para excluir definitivamente la posibilidad de que existan zombis en otros mundos posibles. Lamentablemente (para los funcionalistas),

[202] Güzeldere, 1995 en Block *et al.*, *op cit.*, p. 41: "Zombiehood becomes a possibility only under a certain view that accords with epiphenomenalism. If we maintain that consciousness has causal powers, then the abscence of consciousness in my zombie twin, which is identical to me in every respect, would make some difference. But by stipulation, there is no difference whatsoever between persons and their zombie twins except the fact that the later lack consciousness. Hence, denying epiphenomenalism would also block the possibility of zombiehood."

[203] Esta aseveración sin duda puede desconcertar a ciertos lectores que pueden pensar que hay una contradicción en mis razonamientos. En efecto, ¿cómo puedo sostener por un lado que los qualia son propiedades no relacionales y, por otro lado, admitir que tienen en este mundo poderes causales? La respuesta a esta aparente paradoja consiste en el hecho que los poderes causales no se encuentran situados en el mismo nivel que las propiedades no-relacionales de los qualia. Mientras que los poderes causales de los estados mentales con qualia son algo contingente (podemos imaginar mundos posibles con entes con estados mentales con qualia, pero sin poderes causales) de lo cual podemos prescindir sin atentar contra la esencia de los qualia, las propiedades cualitativas y no-relacionales son algo necesario para que los qualia sean precisamente lo que son. Así pues, una imagen de los estados mentales con qualia en la cual se admite que tienen propiedades no-relacionales necesarias y propiedades relacionales contingentes no es absurda.

este proyecto no es algo que pueda ser llevado a cabo. En efecto, podemos imaginar un mundo físico posible en el cual sólo hay cerebros en cubetas: a causa de este hecho, estos cerebros no pueden tener ninguna acción sobre el mundo (por lo cual la conciencia no tiene en este mundo poderes causales) pero esto no excluye la posibilidad que estos cerebros tengan qualia (podemos concebir que estos cerebros padezcan dolor o tengan sensaciones de colores sin ninguna contradicción). Así pues, si existe al menos un mundo posible el cual los estados mentales son epifenomenales y en el cual algunos de esos estados mentales tienen rasgos cualitativos, de ello podemos deducir que la existencia de los zombis no es metafísicamente imposible como los funcionalistas afirman (y esto es todo lo que nos basta).

Además de Güzeldere (que es funcionalista), otros materialistas han intentado demostrar que tanto la objeción de los qualia invertidos como la objeción de los qualia ausentes no tienen sentido en la medida que se pueden presentar algunas objeciones muy sugerentes en contra de ambas intuiciones. En el caso de los qualia ausentes, Chalmers ha presentado en *The Conscious Mind* (1996) una situación hipotética (que denomina situación de los "qualia evanescentes") que consiste en lo siguiente: imagina a un hombre que experimenta una sensación con un quale particular (e.g., una sensación de dolor) y en el cual las neuronas de su cerebro son sustituidas una por una por chips de silicón que cumplen exactamente el mismo papel que cumplen las neuronas. Considerando esta situación, la primera pregunta que Chalmers se plantea es la siguiente: cuando remplazamos las neuronas progresivamente por chips de silicón, ¿acaso los qualia desaparecen? Si respondemos afirmativamente (como quizás estén tentados de hacerlo ciertos autores como Block para los cuales la existencia misma de la mente así como la presencia de los qualia dependen de ciertas estructuras neuronales precisas),[204] entonces sólo tenemos dos alternativas: suponer que en el curso del proceso de sustitución de las neuronas por los chips los qualia desaparecen bruscamente o que, al contrario, se desvanecen poco a poco. Después de mostrar que las dos alternativas son improbables en el mundo actual, Chalmers se ve obligado a admitir que, a pesar de que son muy improbables, "los qualia

[204] Block, 1978 en Block, 1980, p. 277: "It is a highly plausible assumption that (...) mentality depends crucially on psychological and neurological processes and structures"

evanescentes son lógicamente posibles."[205] Ahora bien, para demostrar que la imposibilidad de los qualia ausentes, tendríamos que mostrar que los qualia evanescentes son metafísicamente imposibles, pero como no podemos llevar esto a cabo (según el reconocimiento del propio Chalmers), sus razones en contra de los qualia ausentes no son conclusivas.

En el caso de los qualia invertidos, las razones que presenta Chalmers adolecen del mismo problema que tiene el argumento de los qualia evanescentes: sólo muestran que la situación de inversión de qualia en dos individuos que comparten la misma organización funcional y las mismas estructuras neuronales es altamente improbable, pero no metafísicamente imposible, que es lo que realmente importa para que el argumento sea concluyente. Por lo tanto, aun cuando admito que los casos de qualia invertidos y de qualia ausentes son muy poco probables en el mundo actual, los funcionalistas no pueden demostrar que estos casos son imposibles en todos los mundos posibles, lo cual basta para demostrar que el funcionalismo no es una buena teoría de la mente en tanto no es posible naturalizar los qualia.

Ahora bien, como existen buenas razones para pensar que no se puede naturalizar los qualia en el marco del funcionalismo en tanto no se encuentran inmersos en la red de relaciones funcionales/causales que caracterizan a los estados funcionales ni supervienen en éstos o en los estados neuronales, ciertos funcionalistas han señalado lo siguiente: como los intentos sucesivos de naturalizar los qualia a través de la causalidad y de la superveniencia no han sido exitosos, acaso esto se deba al hecho de que no se puede naturalizar *aquello que no existe*. Dennett es, sin duda alguna, el mejor representante de esta corriente de pensamiento eliminativista que sostiene que los qualia no existen. En *Quining Qualia* (1988), Dennett ha presentado una serie de intuiciones o razones que buscan demostrar que los qualia no existen en tanto que las propiedades que generalmente se les atribuyen (indubitabilidad, atomicidad, inefabilidad y el hecho que sean conscientes) son erróneas y no corresponden a nada real. No es mi propósito examinar y discutir cada una de las razones de Dennett (lo cual sería una tarea que llevaría demasiado tiempo y espacio); en vez de ello, me contentaré con citar un pasaje de Wittgenstein que sintetiza admirablemente las dudas que

[205] Chalmers, 1996, p. 257: "Fading qualia are logically possible."

Dennett (y los demás materialistas que adoptan una postura eliminativista respecto a los qualia) expone en su artículo:

> Supongamos que cada uno tuviera una caja y dentro hubiera algo que llamamos "escarabajo". Nadie puede mirar en la caja del otro; y cada uno dice que él sabe lo que es un escarabajo sólo por la vista de su escarabajo. Aquí podría muy bien ser que cada uno tuviese una cosa distinta en su caja (...) La cosa que hay en la caja no pertenece en absoluto al juego del lenguaje; ni siquiera como un algo: pues la caja podría incluso estar vacía. [206]

El punto que Wittgenstein quiere mostrar a través de esta situación es que no es claro cómo puede haber un discurso público sobre qualia si no existe manera de comprobar empíricamente que el dolor de muelas que siento es realmente el mismo que el dolor de muelas que otra persona dice tener, aunque los términos que ambos empleamos sean los mismos. Cuando Dennett niega la existencia de los qualia sosteniendo que no tienen las propiedades que habitualmente les atribuimos (acceso restringido desde la perspectiva de la primera persona, inefabilidad, indubitabilidad, etc.), es importante entender que hace esto porque en el fondo resulta desconcertante para él que no puedan ser analizados desde una perspectiva pública y empírica, que sería acorde con la comprobabilidad intersubjetiva de los resultados que exige la ciencia para poder ser objetiva. Como podemos constatar, Dennett al igual que Wittgenstein tiene en el fondo la misma noción conductista según la cual una cosa debe tener un carácter empírico y públicamente comprobable para que podamos afirmar que existe en verdad. He señalado anteriormente que este tipo de argumentos no resultan muy convincentes en tanto que presentan un criterio demasiado verificacionista de la existencia: tanto para Wittgenstein como para Dennett, si una cosa tiene criterios de verificación empíricos e intersubjetivos, entonces esta cosa debe de existir. Como los qualia no tienen esos criterios de verificación empíricos e intersubjetivos, entonces Dennett concluye que seguramente no existen. Ahora

[206] Wittgenstein, 1986, p. 245 (§293)

bien, este argumento dista mucho de ser concluyente.[207] Para demostrar que los qualia no existen, no basta con señalar que no tienen criterios de verificación empíricos e intersubjetivos: se debe mostrar por medio de un argumento formal que son metafísicamente imposibles. Como los qualiafóbicos no han podido llevar esto a cabo todavía, los qualiafilos (entre los cuales me incluyo) podemos sostener sin demasiadas dificultades la existencia de los qualia. Además, es importante señalar que esta existencia de los qualia no es meramente gratuita: los qualia tienen una gran utilidad en la medida que nos permiten establecer una taxonomía de las experiencias subjetivas que, de otra manera, serían todas indistinguibles entre sí. De igual manera que el contenido intencional nos permite diferenciar estados mentales distintos como mi creencia de que los Yankees van a ganar la Serie Mundial de este año y mi creencia de que hay helado de vainilla en el congelador, los qualia nos permiten distinguir distintos tipos de dolores para los cuales tenemos una serie de calificativos fenoménicos (leve, agudo, punzante, agónico, etc.) y distintos tipos de experiencias subjetivas alternas como las emociones (aprehensión, miedo, terror, etc.). Sin los qualia, es evidente que no podríamos hacer este tipo de distinciones entre distintos tipos de experiencias subjetivas que hacemos cotidianamente (e.g., entre el miedo y el terror o entre la irritación y la ira). Así pues, es claro que la eliminación de los qualia que propone Dennett mutilaría gravemente el concepto de mente en tanto tendríamos que renunciar a la posibilidad de hacer distinciones entre familias o clases de experiencias subjetivas. En conclusión, es evidente que cualquier tentativa de naturalización de la mente que ignore a los qualia pretendiendo que no existen sólo porque no pueden ser naturalizados a través de la red de relaciones funcionales/causales o de la superveniencia está condenada al fracaso en la medida que los qualia son elementos que no podemos eliminar arbitrariamente si queremos desarrollar una teoría satisfactoria de la mente. Esta eventual teoría que se encuentra todavía por desarrollar debe poder dar cuenta de los qualia así como de las relaciones causales entre la mente y el cuerpo y de la intencionalidad.

[207] El mismo Dennett está consciente de esta limitación cuando declara lo siguiente (Dennett, 1988 en Block *et al.*, *op. cit.*, p. 621): "Rigorous arguments only work on well defined materials, and since my goal is to destroy the faith in the pre-theoretical or 'intuitive' concept [of qualia], the right tools for my task are intuition pumps, not formal arguments."

Conclusiones

A lo largo del presente trabajo, hemos visto que las principales teorías materialistas de la mente que se han desarrollado a lo largo del siglo XX parecen ser, sino erróneas, por lo menos bastante incompletas en tanto no pueden dar cuenta de las tres condiciones mínimas que hemos establecido en la introducción para que una teoría dada pueda ser considerada una buena teoría de la mente: dar cuenta de los vínculos causales entre la mente y el cuerpo, dar cuenta de la intencionalidad que tienen ciertos estados mentales y dar cuenta del carácter cualitativo o intrínseco que tienen otros estados mentales.

En el primer capítulo, hemos visto que las dos principales versiones de conductismo no pueden dar cuenta de los vínculos causales entre la mente y el cuerpo en la medida que, para ciertos conductistas, los estados mentales son cosas que ni siquiera existen mientras que para otros, como son idénticos a la conductas, no puede ser causa de las conductas en tanto esto violaría el principio de Hume sobre la causalidad. Asimismo, hemos visto que el conductismo no puede dar cuenta de la intencionalidad de los estados mentales en la medida que las meras conductas no bastan para recuperar la dimensión intencional, como vimos a través del ejemplo de los ratones en los laberintos. Incluso el intento de Ryle de dar cuenta de la intencionalidad por medio de la aceptación de las *acciones* en el ámbito de las conductas constituye una prueba fehaciente de las limitaciones del conductismo en la medida en que dicha aceptación traiciona los principios del conductismo ortodoxo que sólo reconoce como conductas las respuestas fisiológicas y los meros movimientos corporales. Como la intencionalidad de las acciones no es una característica directamente observable en los meros movimientos corporales según los conductistas puristas, sino algo que inferimos o atribuimos posteriormente, aquellos que sostienen que las *acciones* son conductas como Ryle se alejan de la ortodoxia conductista para convertirse en teóricos de la intención, por lo cual su pertenencia al conductismo queda en entredicho. Finalmente, el problema de las cualidades intrínsecas es algo que los conductistas ni siquiera se plantean en la medida que, en tanto no son directa y públicamente observables, es algo que no requiere ser explicado puesto que se asume que dichas cualidades no existen. Además de estos tres problemas, el conductismo cuenta con otras limitaciones que hacen poco plausible

que pueda ser una buena teoría de la mente. Sin embargo, es precisamente a causa de todos estos errores y limitaciones que el conductismo tuvo (y sigue teniendo) un gran valor en tanto que señaló a los filósofos posteriores los derroteros que no es conveniente seguir.

En el segundo capítulo, hemos visto que la teoría de la identidad constituye un notable progreso con respecto al conductismo en tanto que, al establecer una identidad entre los estados mentales y los estados neuronales, permite explicar la existencia de vínculos causales entre mente y cuerpo sin tener que violar el principio de clausura física del mundo. Sin embargo, el establecimiento de una identidad entre estados mentales y estados neuronales es totalmente incompatible con el sostenimiento del principio de identidad de los indiscernibles, que constituye el criterio de identidad más plausible e intuitivo que tenemos en la actualidad. Si bien algunos autores como Davidson han intentado escapar a esta incompatibilidad entre la tesis básica de la teoría de la identidad y el principio de identidad de los indiscernibles arguyendo que la identidad que sostienen es, no una identidad de tipo, sino una identidad de instancia (i.e., una identidad que se da entre un objeto y el mismo, aun cuando las diversas propiedades que le atribuimos en distintas descripciones no coincidan entre sí), esta estrategia nos lleva en última instancia a un epifenomenalismo de propiedades. Así pues, el monismo anómalo de Davidson (así como los otros fisicalismos basados en identidades de instancia) es insostenible en tanto que teoría de la mente en la medida que una buena teoría debe explicar porque las propiedades mentales (i.e., las propiedades que hacen que un estado mental sea precisamente mental) tienen un papel causal respecto a ciertas propiedades físicas.

Además de las múltiples objeciones basadas en la incompatibilidad entre la tesis básica de la teoría de la identidad y el principio de identidad de los indiscernibles, hemos visto que existen diversos argumentos que demuestran la escasa plausibilidad de considerar la teoría de la identidad como una buena teoría de la mente. Por ejemplo, algunos autores como Putnam han argüido que es posible imaginar dos organismos que comparten el mismo estado neuronal en el mismo momento, pero que tienen dimensiones intencionales distintas, lo cual basta para demostrar que la teoría de la identidad no puede dar cuenta de la dimensión intencional de los estados mentales. En el caso de las cualidades intrínsecas, los teóricos de la identidad pretendieron dar cuenta de ellas por medio de los análisis tópico neutrales. Sin embargo, hemos

mostrado que este método es erróneo en la medida que, detrás de cada análisis tópico neutral, se encuentra escondida una petición de principio, por lo que la teoría de la identidad tampoco puede dar cuenta de las características intrínsecas de los estados mentales. El hecho de que la teoría de la identidad no pueda explicar ni la intencionalidad de los estados mentales ni su carácter cualitativo e intrínseco constituye una razón suficiente para poder descartarla como candidato a ocupar el papel de una buena teoría de la mente, pero todavía cuenta con otros problemas. Por ejemplo, los teóricos de la identidad sostienen habitualmente que la identidad entre estados mentales y estados neuronales que buscan establecer es una identidad contingente. Sin embargo, Kripke ha mostrado que el concepto mismo de identidad contingente es una aberración en la medida que todas las identidades son necesarias (i.e., verdaderas en todos los mundos posibles). Asumiendo la validez de los argumentos de Kripke a favor de la necesidad de la identidad, se torna entonces imposible para los teóricos de la identidad poder dar cuenta de la realizabilidad múltiple de los estados mentales (i.e., del hecho que puedan ser realizados por estructuras materiales distintas de los cerebros humanos en este o en otros mundos posibles) que había sido un *desideratum* del materialismo clásico, por lo cual la teoría de la identidad se encuentra condenada a ser una propuesta irremediablemente antropocéntrica y chauvinista.

En el tercer capítulo, hemos analizado las distintas versiones de funcionalismo que se desarrollaron en la década 1960-1970 como resultado de la decadencia progresiva de la teoría de la identidad, buscando determinar cuáles son sus puntos comunes y sus divergencias. Hemos visto que la propuesta funcionalista de Lewis constituye un notable avance respecto al chauvinismo inherente a la teoría de la identidad en la medida en que los estados mentales son considerados como clases definidas por su papel causal respecto a otras clases de estímulos o de respuestas motoras y realizables por clases de orden menor en la jerarquía de los universales como las clases biológicas o electrónicas. Sin embargo, hemos visto que esta definición abstracta de los estados mentales nos condena en última instancia a una fragmentación de la psicología en la medida que no existe una ley causal universal bajo la cual puedan caer las realizaciones de un estado mental, puesto que cada realización, en tanto es una clase de orden menor, tiene poderes causales distintos a los de otras clases. Así pues, el principal problema del funcionalismo de Lewis consiste en que su teoría

de la mente parece no poder dar cuenta de los vínculos causales entre mente y cuerpo de manera uniforme y general, sino que tenemos un conjunto de propiedades o poderes causales distintos para cada instancia de un estado mental. Por lo tanto, en tanto no cumple el primer requisito que hemos establecido, resulta complicado pensar que la propuesta de Lewis pueda constituir una buena teoría de la mente.

Tras haber analizado la propuesta de Lewis, hemos podido observar que la propuesta funcionalista de Putnam, que se encuentra articulada alrededor del supuesto de que una máquina de Turing es un buen modelo de la mente, recupera muchas ventajas de la propuesta de Lewis, pero también supera varias limitaciones de ésta. Por ejemplo, es muy evidente que la propuesta de Putnam es tan sensible como la de Lewis a la necesidad de dar cuenta de la realizabilidad múltiple de los estados mentales. En la medida en que los estados de tabla de máquina son entidades abstractas definidas por su posición en una tabla de máquina (i.e., por el papel causal que juegan entre cierto tipo de estímulos y cierto tipo de respuestas), la identidad entre estados mentales y estados de tabla de máquina permite que los estados mentales puedan ser instanciados por entidades que mapean ciertos símbolos a otros símbolos (neuronas biológicas, chips de silicón, etc.) sin importar su constitución física o química. Por otro lado, como los símbolos mapeados por las máquinas de Turing tienen poderes causales, resulta sencillo ver que la propuesta de Putnam constituye una buena teoría de la mente, al menos en cuanto a la necesidad de dar cuenta de los vínculos causales entre mente y cuerpo. Sin embargo, el problema principal que aqueja a la propuesta funcionalista de Putnam consiste en que es incapaz de dar cuenta de los vínculos semánticos que hay entre los estados mentales, i.e., de la intencionalidad de los estados mentales como lo señalan Block y Fodor. En la medida que la teoría de Putnam no cumple con el segundo requisito que hemos establecido, resulta complicado pensar que puede ser una buena teoría de la mente.

Al exponer la propuesta funcionalista de Dennett, hemos podido constatar que Dennett tiene un doble objetivo: por un lado, pretende recuperar los rasgos positivos de las propuestas funcionalistas de Lewis y Putnam mientras que, por otro lado, pretende presentar una teoría de la mente que pueda dar cuenta de la intencionalidad de los estado mentales. La propuesta de Dennett para explicar la dimensión intencional de los estados mentales consiste en mostrar que el contenido intencional surge en los organismos cuando hay

una adecuación de las conexiones nerviosas aferentes y eferentes con el entorno. Esta adecuación según Dennett no se encuentra dada en la mayoría de los organismos de manera natural, al menos en un inicio. En el inicio de la historia evolutiva, sólo algunos organismos disponen del precableado nervioso adecuado (i.e., un precableado que corresponde a las condiciones del entorno) que les permite tener una dimensión intencional. Según Dennett, el proceso de selección natural se encarga en el largo plazo de que sólo los organismos que disponen de contenido intencional, i.e., de un precableado nervioso adecuado sobrevivan mientras que los otros tienden a ser eliminados. Esta explicación teleológica del contenido tiene el grave incoveniente, como hemos visto, de que no permite la existencia del contenido erróneo. Es por ello que otras alternativas para naturalizar la intencionalidad, como la propuesta de Fodor, han sido propuestas. La propuesta de Fodor para dar cuenta de la intencionlidad dentro de un contexto funcionalista consiste en asumir que el origen del contenido intencional de los estados mentales es atómico (i.e., se da en virtud de las relaciones que tienen con el mundo) mientras que la implementación del contenido intencional es holista y computacional (i.e., se da en virtud de las relaciones que los estados mentales que tienen entre sí). Si aceptamos la propuesta de Fodor (aun cuando cuenta con ciertas limitaciones), es evidente que tenemos a nuestra disposición una teoría materialista que puede dar cuenta, al menos parcialmente, de los vínculos causales entre mente y cuerpo así como de la intencionalidad de los estados mentales. Sin embargo, la propuesta de Fodor no puede dar cuenta del contenido intrínseco de los estados mentales y, por lo que hemos visto en la última sección del tercer capítulo, ninguna de las múltiples propuestas funcionalistas contemporáneas parece poder dar cuenta de las propiedades fenómenicas.

Así pues, aun si cada versión de materialismo que hemos revisado cuenta con objeciones en contra y limitaciones propias, hemos podido constatar que existe un común denominador entre ellas: la aparente imposibilidad de dar cuenta del carácter intrínseco de ciertos estados mentales. Incluso el funcionalismo, que es actualmente la propuesta materialista más popular entre los miembros de la comunidad filosófica en la medida que parece poder dar cuenta de los vínculos causales entre la mente y el cuerpo y (si aceptamos la semántica atomista de Fodor) de la intencionalidad de ciertos estados mentales, parece ser incapaz, como lo hemos mostrado en nuestro trabajo, de dar cuenta de las propiedades cualitativas o

intrínsecas de los estados mentales no-intencionales. No parece haber manera de naturalizar los qualia ni por medio de la red de relaciones funcionales/causales, ni por medio de la postulación de un vínculo de superveniencia (en la medida que se requiere que este vínculo sea metafísicamente necesario); tampoco parece ser posible ignorarlos o eliminarlos como ciertos autores sugieren, so pena de mutilar gravemente el concepto de mente. Por lo tanto, es bastante plausible que debamos desarrollar una teoría alternativa de la mente distinta del funcionalismo que pueda llenar los tres requisitos básicos que hemos establecido.

Tras haber descartado sucesivamente el conductismo, la teoría de la identidad, el monismo anómalo, el materialismo eliminativista y, finalmente, el funcionalismo en sus distintas versiones como alternativas para ocupar el papel de una buena teoría de la mente, ¿qué podemos hacer entonces? No me he atrevido a presentar una solución a la pregunta planteada inicialmente ("¿Qué teoría puede ser una buena teoría de la mente?") a lo largo de mi trabajo, y no lo haré tampoco ahora en la medida que considero que el debate todavía sigue abierto. Como hemos podido constatar a lo largo de mi exposición, ningún filósofo ha conseguido hasta ahora aportar un argumento concluyente y definitivo que apoye una teoría particular en detrimento de las demás. Sin embargo, aun si no puedo presentar en el estado actual de las cosas una buena teoría de la mente, lo que puedo hacer es señalar las posibles direcciones de investigación que se desprenden de mi trabajo. Considerando todo lo que he dicho anteriormente, parece que sólo dispongo de dos direcciones de investigación posibles para poder eventualmente presentar una respuesta a mi pregunta pregunta inicial:

(1) Se puede sostener que por lo menos uno de los tres requisitos básicos que he establecido para que algo sea una teoría satisfactoria de la mente es prescindible con respecto a los demás.

(2) Se puede sostener que los tres requisitos son necesarios e irreducibles unos con respecto a otros

En el caso de la primera opción, es menester notar que las sospechas se centran de manera automática en el tercer requisito que hemos establecido, a saber, el hecho de que una buena teoría de la mente debe de dar cuenta del carácter fenoménico de los estados mentales. Casi nadie se arriesgaría a decir

actualmente que los otros dos requisitos (causalidad mente-cuerpo e intencionalidad) son prescindibles en una buena teoría de la mente. En cambio, la mayoría de los filósofos contemporáneos sospechan que hay algo erróneo acerca del requisito de los qualia en la medida que, si bien existen ciertos criterios empíricos y públicos que nos permiten sostener la plausibilidad de la existencia de los vínculos causales entre la mente y el cuerpo o la plausibilidad de la existencia de la dimensión intencional de la mente, parece no haber ningún criterio empírico que nos permita sostener la existencia de los qualia en tanto pertenecen al ámbito subjetivo. Por lo tanto, es posible que me haya equivocado al establecerlo como requisito básico e independiente de los otros dos requisitos que una teoría satisfactoria de la mente debe cumplir. Si este fuese el caso, entonces o bien los qualia deben poder ser eventualmente reducidos a ciertas propiedades físicas o bien deben poder ser radicalmente eliminados. En este caso, el funcionalismo en alguna de sus versiones volvería a presentarse como la alternativa más inmediata para ocupar el papel de una buena teoría de la mente. Sin embargo, para poder sostener que el requisito anteriormente mencionado es o bien prescindible o bien reducible a los otros dos, es menester presentar argumentos sólidos y no meras razones o intuiciones (como han hecho los filósofos hasta ahora) que demuestren de manera concluyente que los qualia son o bien prescindibles o bien reducibles a propiedades físicas en una buena teoría de la mente, sin lesionar el concepto mismo de mente. Ahora bien, en el caso que sea imposible elaborar sólidos argumentos que demuestren el carácter prescindible o reducible de los qualia, nos veremos obligados a admitir que el requisito de que una teoría debe de dar cuenta del carácter cualitativo de los estados mentales para ser una buena teoría de la mente es necesario e imprescindible. Si aceptamos (2), i.e., si aceptamos que los tres requisitos que hemos establecido son entonces necesarios e irreductibles los unos con respecto a los otros, y que una buena teoría de la mente debe poder dar cuenta de todos ellos, entonces se derivan dos posibles líneas de investigación. Podemos intentar desarrollar otras alternativas teóricas, distintas de las que hemos visto en este trabajo, que cumplan con los tres requisitos básicos, o bien podemos sostener como McGinn en *The Mysterious Flame* (2000) que los tres requisitos son necesarios e irreducibles los unos a los otros, pero que por sus características hacen imposible obtener una buena teoría de la mente, por lo que una eventual solución de la naturaleza de la mente se encuentra más

allá de las capacidades humanas por principio. En trabajos ulteriores, examinaré las líneas de investigación aquí mencionadas para determinar cuál de ellas es correcta y puede aportar una solución al problema que consiste en presentar una buena teoría de la mente (aun si la solución consiste precisamente en que no hay solución).

Bibliografía:

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Baier, K. (1964) 'The Place of a Pain', *Philosophical Quarterly*, 14, pp. 138-150

Bergson, H. (1889) *Essai sur les données immédiates de la conscience*, Thèse de doctorat, Paris.

Block, N. y Fodor, J. (1972) 'What Psychological States are not.' *Philosophical Review*, 81, pp. 159-181 [Reimpreso en Block, 1980, pp. 237-250]

Block, N. (1978) 'Troubles with Functionalism' in C. W. Savage, (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 9, Minneapolis: University of Minnesota Press [Reimpreso en Block, 1980, vol. I, pp. 268-305]

Block, N. (ed.) (1980) *Readings in Philosophy of Psychology*, vol. I, London: Methuen.

Block, N. (1990) "Inverted Earth" in J. Tomberlin, (ed.), *Philosophical Perspectives*, vol. 4, Northridge: Ridgeview Publishin Co. [Reimpreso en Block *et al.*, 1997, pp. 677-694]

Block, N. (1995) "On a Confusion about a Function of Consciousness", *Behavioral and Brain Sciences*, 18, pp. 227-247

Block, N., Flanagan, O. y Güzeldere, G. (1997) (eds.) *The Nature of Consciousness*, Cambridge, MA: MIT Press

Braddon-Mitchell, D. y Jackson, F. (1996) *Philosophy of Mind and Cognition*, London: Blackwell.

Burge, T. (1997) "Two kinds of Consciousness" in Block et al., 1997, pp. 427-434.

Campbell, K. (1970) *Body and Mind*, London: The MacMillan Press. [Traducción al español de Susana Marín (1987), *Cuerpo y Mente*, México: Instituto de Investigaciones Filosóficas]

Carnap, R. (1928) *Die logische Aufbau der Welt*, Leipzig: Felix Meiner. [Traducción al español de Laura Mues (1988), *La construcción lógica del mundo*, México: Instituto de Investigaciones Filosóficas]

Carnap, R. (1935) *Philosophy and Logical Syntax*, London: Kegan Paul, Trench, Tubner & Co. [Traducción al español de César Molina, segunda edición (1998), *Filosofía y sintaxis lógica*, Cuadernos, 12, México: Instituto de Investigaciones Filosóficas]

Chalmers, D. (1996) *The Conscious Mind*, Cambridge, MA: MIT Press.

Chomsky, N. (1959) "Review of *Verbal Behavior* by B. F. Skinner", *Language*, 35, pp. 26-58 [Reimpreso en Block, 1980]

Churchland, P. M. (1981) 'Functionalism, Qualia and Intentionality', *Philosophical Topics*, 12, pp. 121-145 [Reimpreso en Churchland, 1989, pp. 23-46]

Churchland, P. M. (1985) 'Reduction, Qualia and the Direct Introspection of Brain States', *The Journal of Philosophy*, **82**, pp. 8-28 [Reimpreso en Churchland, 1989a, pp. 47-66]

Churchland, P. M. (1988) *Matter and Consciousness* (Edición revisada), Cambridge, MA: MIT Press.

Churchland, P M. (1989) *A Neurocomputational Perspective*, Cambridge, MA: MIT Press.

Coder, D. (1973) 'The Fundamental Error of Central State Materialism', *American Philosophical Quarterly*, **10**, pp. 289-298

Coffa, J. A. (1991) *The Semantic Tradition from Kant to Carnap. To the Vienna Station*, Cambridge: Cambridge University Press.

Crick, F. y Koch (1990) "Towards a Neurobiological Theory of Consciousness" in *Seminars in the Neurosciences*, **2**, pp. 263-275 [Block *et al.*, 1990, p. 277-292]

Davidson, D. (1967) 'Causal Relations', *The Journal of Philosophy*, 64, pp. 691-703 [Traducido al español como *Relaciones causales* y reimpreso en Davidson, 1995, pp. 189-206]

Davidson, D. (1969) 'The Individuation of Events' in N. Rescher (ed.) *Essays in Honor of Carl Hempel*, Holland: D. Reidel [Traducido al español como *La individuación de los sucesos* y reimpreso en Davidson, 1995, pp. 207-229]

Davidson, D. (1970) 'Mental Events' in L. Foster y J. W. Swanson (eds.) *Experience and Theory*, London y Amherst: Duckworth y The University of Massachusetts Press [Traducido al español como *Eventos mentales* y reimpreso en Davidson, 1995, pp. 263-287]

Davidson, D. (1973) 'The Material Mind' in P. Suppes, L. Henkin, G. C. Moisil y A. Joja (eds.) *Proceedings of the Fourth International Congress of Logic, Methodology and Philosophy of Science*, Holland: North-Holland Publishing Co. [Traducido al español como *La mente material* y reimpreso en Davidson, 1995, pp. 309-326]

Davidson, (1974) "Psychology as Philosophy" in S. C. Brown (ed.) *Philosophy of Psychology*, London: The Macmillan Press & Barnes, Noble, Inc [Traducido al español como *La psicología como filosofía* y reimpreso en Davidson, 1995, pp. 290-307]

Davidson, D. (1980) *Essays on Actions and Events*, New York: Oxford University Press. [Traducción al español coordinada por Olbeth Hansberg (1995), *Ensayos sobre acciones y sucesos*, México y Barcelona: Instituto de Investigaciones Filosóficas y Crítica]

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, New York: Oxford University Press.

Davies, L. H. (1996) "Extrernalism and Experience" in A. Clark, J. Ezquero y J. M. Larrazabal (eds.) *Philosophy and Cognitive Science: Categories, Consciousness and Reading*, Dordrecht: Kluwer Academic Publishers. [Reimpreso en Block *et al.*, 1997, pp. 309-328]

Dennett, D. C. (1969) *Content and Consciousness*, London: Routledge & Kegan Paul.

Dennett, D. C. (1978a) 'Why you can't make a computer that feels pain.' *Synthese*, **38** [Reimpreso en Dennett, 1978b, pp. 190-229]

Dennett, D. C. (1978b) *Brainstorms*, Vermont: Bradford Books

Dennett, D. C. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.

Dennett, D. C. (1988) 'Quining Qualia' in Marcel y Bisiach, 1988, pp. 42-77

Dennett, D. C. (1991) *Consciousness explained*, Boston: Little, Brown & Co.

Descartes, R. (1964-1974) "Méditations métaphysiques" in *Oeuvres*, vol. IX, Paris: CNRS y Vrin.

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.

Feigl, H. y Brodbeck, M. (1953) *Readings in the Philosophy of Science*, New York: Appleton-Century-Crofts.

Feigl, H. (1960) 'Mind-Body, not a Pseudo-Problem" in S. Hook (ed.), *Dimensions of Mind*, New York New York University Press [Reimpreso en Feigl, 1981]

Feigl, H. (1980) *Inquiries and Provocations* (ed. Robert S. Cohen), Boston. Dordrecht Reidel.

Feyerabend, P. K. (1963a) 'Materialism and the Mind-Body Problem.' *Review of Metaphysics*, 17, pp. 49-66

Feyerabend, P. K. (1963b) 'Mental Events and the Brain.' *The Journal of Philosophy*, 60, pp. 295-296 [Reimpreso en Lycan, 1990, pp. 204-205]

Flanagan, O. (1992) *Consciousness reconsidered*, Cambridge, MA: MIT Press.

Fodor, J. (1974) 'Special Sciences, or the Disunity of Science as a Working Hypothesis', *Synthese*, 28, pp. 97-115 [Reimpreso en Block, 1980, pp. 120-133]

Fodor, J. (1975) *The Language of Thought*, Cambridge, MA: Harvard University Press.

Fodor, J. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press.

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA. MIT Press.

Fodor, J. (1990) *A Theory of Content*, Cambridge, MA: MIT Press.

Fodor, J. (1994) *The Elm and the Expert*, Cambridge, MA: MIT Press.

Frege (1892) "Über Sinn und Bedeutung", *Zeitschrift fur Philosophie und philosophische Kritik*, 100, pp. 25-50 [Traducido por P. Geach y M.Black (eds.), *Translations of the Philosophical Writings of Gottlob Frege*, Bail Blackwell: Oxford]

Gray, C. M. König, P., Engel, A. y Singer, W. (1989) "Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties", *Nature*, 338, pp. 334-337

Güzeldere, G. (1995) "Introduction" in Block *et al.*, 1997, p. 1-67.

Haugeland, J. (1998) *Having Thought: Essays in the Metaphysics of Mind*, Cambridge, MA: Harvard University Press.

Hempel, C. G. (1935) "The Logical Analysis of Psychology" in H. Feigl y W. Sellars (eds.) *Readings in Philosophical Analysis*, New York: Appleton Century Crofts [Reimpreso en Block, 1980]

Horgan, T. (1989) "Mental Quausation", *Philosophical Perspecctives*, 3, pp. 47-76.

Hornsby, J. (1997) *Simple Mindedness: A Defense of Naive Naturalism in the Philosophy of Mind*, Cambridge, MA: MIT Press.

Jackson, F. (1986) 'What Mary didn't know', *The Journal of Philosophy*, 83, pp. 291-295

Kim, J. (1993) *Supervenience and Mind*, New York: Cambridge University Press.

Kim, J. (1998) *Philosophy of Mind*, Boulder, CO: Westview Press.

Kripke, S. (1981) *Naming and Necessity*, Oxford: Basil Blackwell [Traducción al español de Margarita Valdés, segunda edición (1995), *El nombrar y la necesidad*, México: Instituto de Investigaciones Filosóficas]

Lepore, E. y Loewer, B. (1987) "Mind Matters", *The Journal of Philosophy*, 84, pp. 630-642

Lewis, D. (1966) "An Argument for the Identity Theory." *The Journal of Philosophy*, 63, pp. 17-25 [Reimpreso en Lewis, 1983]

Lewis, D. (1969) "Review of *Mind, Art and Religion*", *The Journal of Philosophy*, 66, pp. 23-55 [Reimpreso en Block, 1980]

Lewis, D. (1970) "How to define theoretical terms", *The Journal of Philosophy*, 67, pp. 249-258 [Reimpreso en Lewis, 1983]

Lewis, D. (1980) "Mad Pain and Martian Pain"in Block, 1980, pp. 216-222

Lewis, D. (1983) *Philosophical Papers*, vol. I, New York: Oxford University Press.

Lycan, W. (1987) *Consciousness*, Cambridge, MA: MIT Press.

Lycan, W. (1990) *Mind and Cognition: A Reader*, Oxford: Basil Blackwell.

Marcel, A. J. y Bisiach, E. (1988) *Consciousness in contemporary science*, New York: Oxford University Press.

McGinn, C. (2000) *The Mysterious Flame: Conscious Minds in a Material World*, New York: Basic Books.

Millikan, R. (1984) *Language, Thought, and other Biological Categories*, Cambridge, MA: MIT Press.

Mendelson, E. (1987) *Introduction to Mathematical Logic*, Belmont, CA: Wadsworth & Brooks.

Nagel, T. (1974) 'What is it like to be a bat?', *Philosophical Review*, 83, pp. 435-450 [Reimpreso en Block, 1980, vol. I, pp. 159-168]

Papineau, D. (1988) *Reality and Representation*, Oxford: Basil Blackwell.

Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47, pp. 44-50 [Reimpreso en Lycan, 1990, pp. 29-36]

Putnam, H. (1960) "Minds and Machines" in S. Hook (ed.) *Dimensions of Mind*, New York, New York University Press [Reimpreso en Putnam, 1975, pp. 362-385]

Putnam, H (1963) 'Brains and Behavior' in R. Butler (ed.), *Analytical Philosophy, Second Series*, Oxford· Basil Blackwell [Reimpreso en Putnam, 1975, pp. 325-341]

Putnam, H. (1967a) 'The Mental Life of some Machines' in H.N. Castañeda (ed.) *Intentionality, Minds and Perception*, Wayne State University Press. [Reimpreso en Putnam, 1975, pp. 408-428]

Putnam, H. (1967b) "Psychological predicates" in W. H. Capitan y D. D. Merrill (eds.), *Art, Mind and Religion*, Pittsburgh: University of Pittsburgh Press. [Reimpreso en Putnam, 1975, pp. 429-440]

Putnam, H. (1969) "Logical Positivism and the Philosophy of Mind" in P. Achinstein y S Barker (eds.), *The Legacy of Logical Positivism*: Johns Hopkins Press [Reimpreso en Putnam, 1975, pp. 441-451]

Putnam, H. (1973) "Philosophy and our Mental Life" in Putnam, 1975, pp. 291-303

Putnam, H. (1975) *Mind, Language and Reality. Philosophical Papers*, Vol. 2, New York: Cambridge University Press.

Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press [Traducción al español de Gabriela Venturiera (1990), *Representación y realidad*, Barcelona: Gedisa]

Quine, W. V. O. (1952) "On mental entities" *Proceedings of the American Academy of Arts and Sciences*, 80 [Reimpreso en Quine, 1976, pp. 221-227]

Quine, W. V. O. (1961) *From a Logical Point of View*, New York : Harper & Row

Quine, W. V. O. (1976) *The Ways of Paradox and other Essays*, Cambridge, MA: Harvard University Press.

Rosenthal, D. M. "Two Concepts of Consciousness", *Philosophical Studies*, 94, pp. 329-359

Rosenthal, D. M. (ed.) (1991) *The Nature of Mind*, New York: Oxford University Press.

Rorty, R. (1965) "Mind-Body Identity, Privacy and Categories" *Review of Metaphysics*, 19, pp. 25-54

Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.

Russell, B. (1905) "On Denoting", *Mind*, XIV, pp. 479-393. [Reimpreso en R. C. Marsh (ed.), (1956) *Logic and Knowledge*, London: Allen Unwin]

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

Schiffer, S. (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press

Searle, J. (1983) *Intentionality*, New York: Cambridge University Press.

Skinner, B. F. (1953) *Science and Human Behavior*, New York: Appleton-Century-Crofts.

Smart, J. J. C. (1959) 'Sensations and Brain Processes.' *Philosophical Review*, 68, pp. 141-156

Smart, J. J. C. (1963) 'Materialism.' *The Journal of Philosophy*, 60, pp. 651-662

Sober, E. (1985) 'Panglossian Functionalism and the Philosophy of Mind', *Synthese*, 64, pp. 165-193

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.

Stich, S. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.

Stroud, B. (2000) *The Quest for Reality: Subjectivism and the Metaphysics of Colour*, New York: Oxford University Press.

Taylor, D. M (1965) 'The Location of Pain', *Philosophical Quarterly*, pp. 53-62

Villanueva, E. (2000) "La mente es una estructura causal: el funcionalismo teórico", *Teorema*, XIX, pp. 27-44

Wilson, M. D. (1978) *Descartes*, London: Routledge & Kegan Paul. [Traducción al español de José Antonio Robles (1990), *Descartes*, México: UNAM-IIF]

Wittgensteing, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge and Kegan Paul [Traducción al español de Jacobo Muñoz e Isidoro Reguera (1973), *Tractatus Logico-Philosophicus*, Madrid: Alianza Editorial]

Wittgenstein, L. (1958) *Philosophical Investigations* (G. E. M. Anscombe y R. Rhees eds.), London: Basil Blackwell [Traducción al español por Alfonso García Suárez y Ulises Moulines, (1988), *Investigaciones Filosóficas*, México y España: Instituto de Investigaciones Filosóficas y Crítica]