ALGUNAS TENDENCIAS EN LA FILOSOFIA DE LA MENTE

Juan J. Botero C.
Universidad Nacional de Colombia

Como se sabe, la "filosofía de la mente"* no gozó durante algunos años de muy buen prestigio en los círculos académicos. Tanto filósofos como psicólogos llegaron a considerar al concepto mismo de "mente" sospechoso: para algunos, de introducir entidades fantasmagóricas o al menos superfluas en lo que se suponía debería ser un estudio científico; para otros, por el contrario, de restarle prestancia y status filosófico al ámbito de lo "trascendental" y lo *a priori* por excelencia.

Ahora bien, durante los últimos años han ocurrido en el campo de la filosofía de la mente cambios tan drásticos que algunos de sus temas tradicionales más importantes nos son casi irreconocibles, al tiempo que empiezan a surgir nuevas preocupaciones y problemas sin antecedentes claros en la tradición filosófica.

Mi intención es pasar revista a algunas de las tendencias contemporáneas más importantes en este ámbito con el fin de hacer un poco de claridad acerca de los temas y problemas que constituyen el núcleo de los debates más significativos en el área. Como lo indica el título, sólo se consideran algunas tendencias dejando de lado otras, sin que esta selección signifique valoración alguna que pretenda favorecer a las elegidas.

---

* Quiero retomar la nota de Eduardo Rabossi, traductor de G. Ryle (véase nota 2), para hacer la siguiente advertencia: *Mind* y *Mental* se utilizan en la tradición anglosajona para abarcar la totalidad de la vida consciente, incluyendo lo volitivo y emocional (intenciones, deseos, dolores, etc.) y no solamente lo abarcado por "mente" y "mental" en español, que parece limitarse a lo intelectivo. El sentido con el que se usa aquí es el anglosajón, por razones inherentes al contenido del ensayo.

Comenzaré con un brevísimo esbozo de los antecedentes filosóficos que explican la orientación fundamentalmente antidualista de las tendencias consideradas. Posteriormente presentaré las dos o tres opciones propuestas como respuestas al dualismo y me centraré definitivamente en la forma como se presentan más comúnmente en la actualidad los trabajos y debates que tienen como tema central a la mente considerada como objeto de indagación filosófica.

Antecedentes

El punto de referencia común a las teorías modernas y contemporáneas de la mente es el dualismo cartesiano. En términos generales se conoce como "dualismo" a la teoría que sostiene que solamente existen dos tipos de substancias: una espiritual, o mental, y una material, o física. En su forma cartesiana, y con respecto al hombre, el dualismo sostiene que estamos compuestos de dos substancias distintas: una substancia corporal, la *res extensa*, y una espiritual, la *res cogitans*. Pero somos *esencialmente* seres espirituales y sólo de manera contingente seres con un cuerpo.

En efecto, en el ejercicio de la duda metódica Descartes llega a la conclusión de que, aunque estamos formados de un alma y un cuerpo, como lo piensa el sentido común prefilosófico anterior a la duda, la existencia del cuerpo, como la de cualquier objeto físico, es dudable, mientras que la de la mente que piensa y duda no lo es. Esta evidencia de que su mente que duda no puede no existir lleva a Descartes a concluir que esta mente es su atributo esencial. La consecuencia más problemática de esto es que para ser lo que es, una persona solamente necesita de su mente. "¿Qué clase de cosa soy?" se pregunta Descartes. "Una cosa que piensa", responde. Es decir: mente, entendimiento, razón.

Vale la pena subrayar que la tesis ontológica se desprende de consideraciones epistemológicas o, si se quiere, metodológicas: si se puede dudar de la existencia del cuerpo, pero no de la mente, entonces la existencia del cuerpo no es algo esencial, como sí lo es la de la mente. En otro lugar he intentado mostrar la perspectiva especulativa de este tipo de argumentación, que llevó a Hegel a afirmar que "toda la filosofía moderna gira en torno al *cogito*" y a catalogar a Descartes como el "fundador" de la filosofía moderna[1]. Aquí solamente quiero hacer notar lo siguiente: el carácter

1 J. J. Botero, "Alma y cuerpo en la sexta meditación metafísica de

"dudable" del cuerpo es una tesis lógica y no psicológica. ¿Es imaginable que no tengamos un cuerpo? Quién sabe. Tal vez sí. O tal vez no. Pero no es contradictorio suponer que no lo tenemos. O dicho de otro modo: es posible avanzar un argumento válido que niegue la existencia del cuerpo. Pero negar la existencia de la mente es autocontradictorio.

La tesis dualista es una consecuencia lógica (por la "ley de Leibniz") de la tesis ontológica acerca del carácter esencial de la mente e inesencial del cuerpo, así como de las demás diferencias entre los dos: mente y cuerpo son dos substancias distintas.

Ahora bien: si el dualismo mente-cuerpo es verdadero, es necesario ofrecer una explicación de la relación entre estas dos substancias, en particular las relaciones causales entre ellas. Pero no parece existir una explicación semejante que sea válida, y este es el principal problema del dualismo. El mismo Descartes pensaba que no era posible hacer plenamente inteligible la idea de la unión del alma y el cuerpo. En estas condiciones, para muchos filósofos no habrá otra salida que renunciar al dualismo. Otros tal vez sostendrán que nos encontramos aquí con un problema insuperable y, por esa misma razón, auténticamente filosófico. Otros, basados en el mismo argumento, sostendrán precisamente lo contrario: que no se trata de un problema filosófico genuino.

La "disolución" del problema y el conductismo lógico

Podemos decir que la nueva época en la filosofía de la mente arranca con la publicación en 1949 de *The Concept of Mind*[2], de Gilbert Ryle, un inteligente y entretenido alegato en contra fundamentalmente del dualismo cartesiano mente-cuerpo, pero en general en contra de cualquier teoría de la mente, exceptuando la que se expone en ese libro, admitiendo con fuertes reservas por parte del propio autor que lo que allí se expone sea una teoría de la mente.

Descartes", *Universitas Philosophica*, Bogotá, Año 3, No. 6, junio de 1986. Véase G.W.F. Hegel, *Enciclopedia de las Ciencias Filosóficas*, parágrafo 64.

2 Barnes & Noble, Inc. (N.Y). Traducción española: *El concepto de lo mental*, Paidós (Buenos Aires).

En efecto, en esta obra Ryle sostiene que la filosofía de la mente ha reposado en un error colosal, y que es justamente este error el que ha conducido al surgimiento de este campo de investigación. Pero no se trata de un error acerca de alguna concepción particular de la mente o de la materia, sino de un error de otro tipo, más fundamental, que ha hecho que todos los esfuerzos por responder a los planteamientos que surgen en la "filosofía de la mente" conduzcan de manera ineluctable al sinsentido.

Antes de Ryle, y en especial como consecuencia de la tesis dualista, había habido diversas teorías de la mente: los temibles "ismos", tales como idealismo (toda realidad es mental), materialismo (toda realidad es física), monismo neutral (sólo hay una realidad, pero no es ni mental ni física: es neutra), epifenomenismo (la mente como un subproducto —epifenómeno— de procesos nerviosos), interaccionismo (influencia causal de doble vía entre la mente y el cuerpo), etc. Estas teorías se enfrentaban entre sí, pero el campo de confrontaciones estaba formado por numerosos presupuestos compartidos acerca de la naturaleza del problema que lo definía: el "problema mente-cuerpo". Ryle sugirió que en realidad no existía tal problema, pues de hecho no lo había en nuestra vida cotidiana, sino solamente una confusión cuyo origen era claramente filosófico pero cuya naturaleza era lingüística: solamente parece haber un "problema de la mente" cuando la reflexión filosófica conduce a un uso —y abuso— poco juicioso del lenguaje ordinario que usamos en la vida diaria para referirnos a hechos acerca de la mentalidad que nos son familiares, los cuales constituyen los datos para cualquier investigación científica de ella.

Un análisis cuidadoso del habla común acerca de la mente disiparía las confusiones, disolvería los problemas, y haría así ociosos tanto al dualismo como a su negación, el monismo, y entre las variedades de monismo, tanto al idealismo como a su opuesto, el materialismo, es decir, todos los puntos de vista metafísicos rivales que habían sido el producto principal de este campo de estudio.

Ryle da el nombre de "dogma del fantasma en la máquina" al dualismo cartesiano: se piensa en la mente como en una entidad que habita en el interior de un cuerpo y desde allí realiza diversas operaciones sobre este cuerpo y sobre el mundo en general. Una consecuencia de este punto de vista es que tenemos un acceso privilegiado, directo, a nuestras operaciones mentales, por lo cual el conocimiento que tenemos de ellas es cierto y evidente. La opinión de Ryle es que el dualismo es completamente falso, pero no por algún error de detalle, sino falso por principio. El error sobre el que reposa es lo que él llama un "error de categoría", que consiste en

concebir los hechos que tienen que ver con la vida mental como si pertenecieran a un determinado tipo lógico, o categoría, cuando en realidad pertenecen a otro muy diferente.

Un "error de categoría" surge por desconocimiento de la manera como determinadas palabras funcionan en el lenguaje común. El error no consiste en ignorar el sentido de tal o cual término, sino en utilizarlo con un sentido que no tiene, desconociendo el que le confiere su uso en el habla cotidiana. Y esto ocurre generalmente en el ámbito filosófico, pues es allí en donde las palabras son sacadas de su ámbito normal de utilización para construir teorías. Los "errores de categoría" se corrigen restaurando el ámbito normal de uso de las palabras y observando lo que ocurre con ellas.

En el caso de los conceptos que tienen que ver con la vida mental, psicológica, de las personas, una observación cuidadosa de su uso en situaciones normales lleva a la conclusión de que no están destinadas a designar ningún objeto, ninguna "cosa" en particular, sino principalmente *disposiciones* a asumir o producir determinados comportamientos, y en unos pocos casos ciertas ocurrencias de fenómenos psicológicos. En su análisis de la palabra "creer", o "creencia", Ryle muestra que la utilización común de esta palabra no la vincula a un "hecho" que ocurra en algún lugar, sino a una disposición a asumir ciertos comportamientos dictados por el contenido de la creencia (por ejemplo, creer que el suelo está resbaloso es estar dispuesto o tener la disposición a caminar con cuidado, advertirle a alguien que el piso está resbaloso, discutir con alguien que lo niegue, etc.).

Al aplicar esta clase de "análisis disposicional" a los términos mentales se llega a la conclusión de que cuando las personas hablan de "lo mental", o de "la mente", no se están refiriendo a ningún lugar en particular, ni a ninguna cosa en especial. Se refieren más bien a determinadas habilidades, conductas, inclinaciones a hacer ciertas cosas o a evitar otras, etc. Por consiguiente, toda la teoría dualista y sus contradictorias metafísicas pueden disolverse simplemente evitando utilizar las palabras con un sentido que no tienen y extraer conclusiones teóricas de este desconocimiento del uso normal.

El trabajo de Ryle fue el principal dentro de lo que se llamó la "filosofía del lenguaje común" en la filosofía de la mente, y su influencia debería medirse, no por un censo de los conversos a ella sino por el hecho de que por más de una década después de la aparición de *The Concept of Mind* las teorías de la mente se volvieron tan pasadas de moda que prácticamente se extinguieron. Se sostenía en esta época que las teorías habían sido creadas

por quienes no habían sido capaces de ver que los problemas —al menos los que los filósofos podían abordar con sus herramientas— surgían de presupuestos ingenuos y errados acerca de la manera como la relación mente-palabras funciona en el lenguaje. La construcción de teorías fue reemplazada por la actividad más modesta y cauta del "análisis conceptual": el escrutinio delicado y persistente, aunque informal y asistemático, de las formas de expresión del lenguaje común. Esto ha producido, en cuanto a lo mejor, en los trabajos de Ryle, Austin, Wittgenstein y Anscombe, algunos planteamientos profundos que orientaron y a veces continúan orientando el pensamiento corriente acerca del tema de la mente. Y en cuanto a lo peor, el resultado han sido montañas de trivialidades. En el medio hubo un gran número de trabajos inteligentes y útiles, cuyo interés era raras veces el de resolver problemas y casi nunca el de avanzar teorías generales, sino más bien, y de manera característica, el de alertar a los incautos acerca de la existencia de más problemas y complicaciones de los que aparecían a primera vista.

La estrategia básica era el "giro lingüístico" o, según la expresión de Quine, el "ascenso semántico": cuando el hablar acerca de las cosas (en este caso la mente, las sensaciones, pensamientos y cosas semejantes) produce cierta perplejidad, casi siempre es una gran ayuda cambiar de enfoque y hablar acerca de la manera como se habla de esas cosas, acerca de "lo que uno diría habitualmente" en diversas circunstancias, o, si no se está muy comprometido con el lenguaje común, acerca de lo que se debería decir en diversas circunstancias. Aunque la táctica es buena, contrariamente a la creencia generalizada de entonces, resultó que no todos los problemas de la filosofía de la mente se evaporaron bajo el análisis lingüístico.

Pero hay más: aunque se suponía que el "giro lingüístico" era una defensa en contra de la tendencia a exponer o defender cualquier "ismo", tampoco permitía operar completamente inocente de presupuestos, los cuales en última instancia siempre implican de alguna manera una teoría. De modo característico, es claro que en filósofos como el mismo Ryle y Anscombe había una teoría tácita, el "conductismo lógico".

El "conductismo lógico" es, en general, un punto de vista según el cual la verdad de las adscripciones de estados y eventos mentales está implicada por la verdad de varias oraciones referidas exclusivamente a comportamientos. Por consiguiente, los enunciados que contienen términos destinados a hacer referencia a estados mentales deben reducirse a enunciados equivalentes que no los contienen pues han sido reemplazados por términos que se refieren a comportamientos observables. El objetivo, pues, es reduccionista:

eliminar del lenguaje de las explicaciones psicológicas toda referencia a causas mentales. Los análisis disposicionales propuestos por Ryle buscan cumplir esta función. Tomemos, por ejemplo, la oración

(1) "Juan tiene sed".

Un análisis disposicional permitiría traducirla a una oración hipotética ("contrafáctica") del tipo

(2) "Si Juan tuviera agua a su disposición, la bebería".

Pero es obvio que este tipo de traducciones no resuelve el problema para el que han sido propuestas. Pues para admitir que la segunda oración es verdadera hay que empezar por admitir otras, y éstas contienen términos que hacen referencia a estados mentales. Habría que admitir, en nuestro ejemplo, al menos estas dos:

(3) "Juan cree que el agua calma la sed".

(4) "Juan desea calmar la sed".

No podemos admitir (2) a menos que admitamos la creencia y el deseo incluidos en (3) y (4). Si Juan ignora que el agua calma la sed, o si prefiere morirse de sed, no podemos considerar equivalentes (1) y (2).

Pero no todos los filósofos que apelan al análisis del lenguaje común están comprometidos en filosofía de la mente con el conductismo lógico. Hay excepciones, como por ejemplo P.F. Strawson, quien aparentemente sigue más bien alguna forma un tanto críptica de la *teoría del doble aspecto*: la *persona* no debe analizarse en una mente más un cuerpo, pues ella es, hablando más propiamente, un tipo de entidad unitaria caracterizable de manera esencial, aunque no exhaustiva, como sujeto de atributos tanto materiales, o físicos, como psicológicos, o mentales. Strawson quiere ante todo oponerse a la teoría de la identidad, que se verá más adelante, admitiendo que un evento en el que se encuentren involucrados estados mentales permite tanto descripciones físicobiológicas como descripciones "biográficas" en términos de lo que se ha denominado (un tanto peyorativamente) la "psicología popular" (*folk psychology*), esto es, que apela a términos utilizados comúnmente cuando se ofrecen explicaciones o recuentos de acciones o comportamientos humanos. El problema surge cuando se quiere tender a toda costa el puente y postular, por ejemplo, que todo estado mental tiene una base física, o es idéntico a un estado físico. Según Strawson,

hay que dejar las cosas como están y abstenerse de postular identidades pues en este caso simplemente no es posible ofrecer una explicación unificada. Las dos "historias", la físicobiológica y la "biográfica", no están *in pari materia*.

Esta teoría, cuyo antecedente en la filosofía moderna se encuentra en Spinoza, está emparentada estrechamente con el llamado *monismo neutral*: en lugar de admitir la existencia de dos substancias, la una mental y la otra física, o de solamente admitir una de las dos, se postula la existencia de una que no es ni física ni mental. Esta es la posición de B. Russell y de algunos otros filósofos que, por lo demás, no comparten para nada la metodología del análisis lingüístico como estrategia filosófica.

Independientemente del concepto que se tenga acerca de la vigencia o no del "análisis del lenguaje común" en la filosofía actual, lo cierto es que la vía del análisis lingüístico realmente destruyó la manera tradicional de elaborar una filosofía de la mente. Las teorías tradicionales generalmente procedían a base de generalizaciones apriorísticas que no respondían a nada, o tal vez a análisis conceptuales inconscientes y poco claros, mezclados con introspecciones casuales y observaciones acerca de las experiencias de la gente normal, promoviendo toda esta mezcla al status de verdades metafísicas sobre las esencias de entidades mentales. Pero resultó claro que, si debiera haber teorías de la mente, éstas ya no podrían ser elaboradas con los antiguos métodos artesanales. De modo que el filósofo de la mente se encontraba frente a una triple alternativa: abandonar la filosofía y proseguir su trabajo con teorías empíricas en el campo de la psicología o las ciencias del cerebro; abandonar la teorización y contentarse con los modestos esclarecimientos y las recomendaciones profilácticas del análisis puramente lingüístico; o convertirse en una especie de meta-teórico, un crítico conceptual de las teorías propuestas por los científicos de las ciencias pertinentes. Es esta última concepción, que concibe a la filosofía de la mente como una especie de rama de la filosofía de la ciencia, la que domina los mejores trabajos en este campo en los últimos años. Su diferencia más sobresaliente con relación a las teorías tradicionales y al enfoque del lenguaje ordinario es su interés por las teorías y los datos provenientes de la psicología, las neurociencias, la inteligencia artificial y la lingüística.

La "teoría de la identidad"

La primera alternativa seria al conductismo lógico derivado del análisis del lenguaje común fue la llamada "teoría de la identidad": las mentes son

cerebros, y los contenidos de las mentes —dolores, pensamientos, sensaciones, etc.— *son* (idénticos a) diversos eventos, procesos y estados del cerebro[3].

Al comienzo, los trabajos que proponían este enfoque estaban claramente encaminados a completar el cuadro fisicalista con una teoría fisicalista de la mente que se opusiera tanto a las desviaciones del conductismo lógico como a los excesos metafísicos de la alternativa dualista. El objetivo era una teoría diseñada conceptualmente por la filosofía pero cuyos detalles fueran rellenados por la ciencia y cuya ontología no admitiera más que entidades portadoras de credenciales científicas. Al final, la teoría de la identidad tuvo el curioso resultado de dividir a la gente en dos bandos: aquellos a quienes les parecía obviamente verdadera y aquellos a quienes les parecía obviamente falsa. Aunque se ha escrito y polemizado mucho alrededor de ella, podemos resumir las dificultades que encuentra en tres grandes áreas: las que se derivan de la ley de Leibniz, los problemas de generalización, y los abstractos enigmas lógicos que plantea la relación de identidad.

Problemas con la "ley de Leibniz"

La llamada "ley de Leibniz" hace que la identidad sea una relación mucho más fuerte que la simple similaridad o equivalencia. Según esta ley, de "x es idéntico a y" se concluye que cualquier verdad de la cosa denotada por "x" es verdad también de la cosa denotada por "y", y viceversa. El principio es indiscutible pues en el caso de cualquier identidad verdadera, "x" y "y" denotarán la misma cosa, y cualquier cosa que sea verdad de esta cosa es verdadera de ella, como quiera que la llamemos. Supongamos ahora que algún pensamiento mío es ingenioso, o profundo, u obsceno. Según la teoría de la identidad mente-cerebro habrá que sostener que algún evento o proceso cerebral (aquel que es idéntico a mi pensamiento) es ingenioso, o profundo, u obsceno, y al menos a primera vista no parece que los procesos o eventos cerebrales sean la clase de cosas que pueden ser ingeniosas, profundas u obscenas, como tampoco podrían ser, por ejemplo, la raíz cuadrada de 7, o capitalistas, o partidarios de la perestroika.

---

3 Véase, p. ej., J.J. Smart, "Sensations and Brain Processes", *The Philosophical Review*, Vol. 68, 1959; *Philosophy and Scientific Realism*, London, 1963, así como D. Davidson *Actions and Events*, London, Oxford U. Press, 1980.

Otros problemas, quizás más fuertes[4], que surgen a raíz de la ley de Leibniz tienen que ver con la dificultad para identificar las llamadas "post-imágenes" (la persistencia de impresiones retinianas después de que el estímulo ha sido retirado) con eventos cerebrales[5]: si un lamparazo de un "flash" de fotografía, por ejemplo, deja en mí una post-imagen redonda y anaranjada, ¿habrá que suponer que en el cerebro existe algún evento, o estado cerebral redondo y anaranjado que es idéntico a mi post-imagen? Nada que no sea al menos redondo y anaranjado puede ser idéntico a ella. Por consiguiente, o bien efectivamente existe en mi cerebro ese evento o proceso cerebral redondo y anaranjado (cosa muy difícil de sostener); o bien mi post-imagen no es en realidad redonda y anaranjada (¿sólo "me lo parece"?); o quizás la post-imagen es algo (mental) diferente de cualquier evento cerebral (y por consiguiente la teoría de la identidad es falsa).

El "materialismo eliminativo"

Este caso reviste una especial importancia pues él ha sido a menudo el pretexto para proponer otra forma de teoría materialista, la llamada "teoría eliminacionista". Frente a este problema, algunos materialistas han adoptado un punto de vista que puede resumirse así: quizás, después de todo, *no existen* post-imágenes redondas y anaranjadas, ni sensaciones, ni cosas semejantes. Lo que hay que cuestionar es la semántica normal de los términos utilizados para hablar de la mente y de los "eventos" mentales. Se habla, de manera casual e inadvertida, usando términos como "dolor", "imagen", "creencia", etc., como si se tratara de términos referenciales que designan sin ambigüedades cosas o hechos reales que ocurren en la mente (cualquier cosa que ésta pueda ser). Pero quizás esta forma común de hablar encarna alguna teoría mítica fosilizada y una vez que la ciencia pueda mostrarnos lo que realmente sucede en nuestras "mentes" dejaremos de buscarles referentes a esos términos. El vocabulario mentalista, pues, está condenado a desaparecer, a ser eliminado y reemplazado por un vocabulario fisicalista.

---

4 El anterior podría responderse haciendo una distinción entre el pensamiento como *evento* y el pensamiento como *contenido* y argumentando que rasgos como la obscenidad son rasgos del contenido, y no del evento, o algo similar, pero aquí encontraríamos de nuevo el problema de individualizar el evento por medio del "contenido".

5 Véase W.G. Lycan, "Materialism and Leibniz Law", en *The Monist*, Vol. 56, 1972, págs. 276-287.

Este "materialismo eliminativo", o "desaparicionista", como también se le conoce, tiene en realidad la forma de una hipótesis[6]: se conjetura que llegará el día en que se reformulen los enunciados que utilizan términos mentalistas como enunciados de la forma: "lo que se acostumbraba a llamar 'un dolor' no es más que...", en donde los puntos suspensivos se reemplazan por una expresión fisicalista, por ejemplo una proveniente de la neurofisiología. Hay que notar que aquí también se establece una identidad, pero metalingüística —de ahí las comillas en "un dolor"—, con lo cual se establece que el término no se usa referencialmente, pues no tiene referente, sino que se lo está mencionando. En este contexto, Richard Rorty[7] ha propuesto incluso una comparación entre nuestro uso actual de términos mentalistas (como "sensación", p. ej.) y el uso de términos como "demonios" y "brujas" en culturas diferentes a la nuestra que mantienen creencias que incluyen la existencia de entidades que son designadas por esos términos. Las sensaciones serían a la psicofisiología, por ejemplo, lo que los "demonios" son a las ciencias físicas. Hablamos como si realmente hubiera "post-imágenes", y así parece que fuera. Pero también hablamos de que el sol se oculta, y aunque así parece que ocurriera, la ciencia puede hacer que esta expresión, como la primera, se vuelva metafórica. No es que la ciencia, pues, descubra las identidades que corresponderían a los términos mentalistas, sino que estos términos deberán desaparecer una vez que un cuadro científico más sofisticado sustituya al que poseemos actualmente.

Problemas de generalización: identidad de tipos o de casos

El segundo problema que surge de la teoría de la identidad es el de la

---

6 Hay tres variantes prominentes del "materialismo eliminativo", sostenidas respectivamente por P. K. Feyerabend ("Mental Events and the Brain", *The Journal of Philosophy*, LX, 11 May 1963), W.V.O. Quine ("On Mental Entities", in *The Ways of Paradox*, Random House, N.Y. 1966) y R. Rorty ("Mind-Body Identity, Privacy and Categories", *Review of Metaphysics*, XIX, 1, 1965). La razón exacta para eliminar las entidades mentales, y su significado epistemológico, metafísico y semántico son el núcleo de la disputa entre estas tendencias, pero todos están de acuerdo en rechazar la postulación de una identidad estricta entre lo físico y lo mental. La versión más contemporánea del eliminacionismo radical puede ser la de "los Churchland", Paul y Patricia. Véase Paul Churchland, "Eliminative Materialism and Propositional Attitudes", *The Journal of Philosophy*, LXXVIII, 1981, y Patricia Smith Churchland, *Neurophilosophy: Toward an Unified Science of the Mind-Brain*, The MIT Press, 1986.

7 *Op. cit.*

generalización. Este problema surge porque el papel normal de una afirmación de identidad es permitir la generalización: si esta nube es idéntica a un conglomerado de goticas de agua, entonces tal vez también lo son todas las nubes; si este gene es una molécula de DNA, entonces la hipótesis que se quisiera someter a verificación es que todos los genes lo son. Pero no es nada claro que la teoría de la identidad mente-cerebro nos permita ninguna generalización, más allá de la afirmación englobante de que todo ítem mental es un evento cerebral.

Supongamos que María piensa en π al mediodía. Diríamos entonces, de acuerdo con la teoría de la identidad, que el pensamiento de María es idéntico a su proceso o evento mental *p* al mediodía, habiendo definido los rasgos físicos F, G, H... correspondientes. Pero no es ni siquiera remotamente plausible suponer que todo pensamiento acerca de π es un proceso mental con los rasgos F, G, H..., aunque sólo sea porque no hay ninguna razón para suponer que criaturas inteligentes de cualquier otra parte del universo tengan que compartir necesariamente con María su neurofisiología, o incluso su bioquímica, para poder pensar en π[8]. No es plausible ni siquiera que todo pensamiento humano en π, o incluso que todo pensamiento de María en π sea idéntico a un proceso o evento cerebral que caiga dentro de una clase especificable únicamente en términos de las características físicas de sus miembros. Parece que un tema clave de una teoría de la mente es el de decirnos qué es lo que hace que un pensamiento sea un pensamiento y qué es lo que hace que un pensamiento acerca de π sea un pensamiento acerca de π , pero no parece que los rasgos generales que estamos buscando sean rasgos físicos.

Una respuesta de la teoría de la identidad podría consistir en debilitar sus pretensiones. Podría decir que lo único que se necesita para evitar el dualismo es que cada evento mental particular (cada "caso") sea algún evento cerebral (es decir, que ningún evento mental sea un evento no-físico o no-cerebral). Así, distinguiríamos entre una teoría de la identidad referida a "casos" y una referida a "tipos", y abandonaríamos esta última. Quizás no haya nadie que hoy en día suponga que los tipos de hechos mentales puedan ser distinguidos directamente por rasgos puramente físicos, pero casi nadie supone tampoco que esto sea un objetivo razonable del fisicalismo.

Pero también hay una respuesta fuerte que va más allá del reconoci-

---

8 Cfr. H. Putnam, "The Nature of Mental States", en *Mind, Language and Reality. Philosophical Papers*, Vol. II. Cambridge U. Press, 1975.

miento de una teoría de la identidad de los "casos". Sostiene que las marcas distintivas buscadas de los *tipos* de hechos mentales pueden ser definidas en términos de los estados lógicos de la máquina abstracta de Turing "realizada" por el sistema nervioso de un ser humano, o en términos de los roles *funcionales* cumplidos por ellos. De estas formulaciones, en apariencia diferentes entre sí, surge sin embargo un punto de vista ampliamente compartido, que ha sido llamado *funcionalismo*: los estados mentales son estados funcionales, es decir, estados que se individualizan por el rol funcional que cumplen en la totalidad del sistema. Por ejemplo, decir que una creencia o un dolor particular es un particular estado funcional es decir que cualquier otra cosa, independientemente de su composición, forma, química o cualquier otra característica física, que cumpla el mismo rol funcional en un sistema funcionalmente equivalente sería la misma creencia, o dolor, y que nada que cumpla el mismo rol funcional podría no ser esa creencia o dolor. El funcionalismo se ha convertido en una de las teorías dominantes en la filosofía de la mente contemporánea. Por eso hablaré más extensamente de ella un poco más adelante.

Identidad y necesidad

El tercer problema que hace surgir la teoría de la identidad tiene que ver con la lógica de la relación misma de identidad. Inicialmente se suponía que el concepto de identidad era transparente y nada problemático. Pero a medida que se lo fue explorando comenzaron a multiplicarse las dificultades, hasta que el problema de la identidad tomó vida propia como tema lógico y metafísico, alejado de las preocupaciones propias de la filosofía de la mente.

En realidad, los filósofos fisicalistas partidarios del "desaparicionismo" habían dejado de preocuparse por cualquier noción de identidad. Más que una teoría de la identidad, proponían una alternativa fisicalista al dualismo. Y, por otra parte, como lo hizo notar Putnam en su temprana exposición del funcionalismo,[9] la cuestión de identificar un estado lógico de una máquina

9 Aunque el funcionalismo, tal como aquí se le describe, fue una propuesta inicial de Hilary Putnam, en los últimos años este filósofo ha hecho público su repudio a esta teoría. Véase, para la propuesta funcionalista: "Minds and Machines", in Sidney Hook (ed.), *Dimensions of Mind*, New York University Press, 1960. Y para el "revisionismo": *Representation and Reality*, MIT Press, 1988 (hay traducción española en Gedisa, 1990: *Representación y realidad: Un balance crítico del funcionalismo*).

(realizada) de Turing con su realización concreta en el "hardware" sería más bien una ociosidad metafísica. Cuando las dos tácticas, el "materialismo eliminativo" o "desaparicionismo", y el funcionalismo, convergen, las identidades que se pueden reconocer tienen que ver más bien con curiosas y abstractas entidades[10]. Llegados a este punto, la inicial motivación para proclamar identidades con el fin de salvarnos de la fantasmagoría del dualismo ha desaparecido.

De ahí que cuando S.Kripke en 1971 incluyó una brillante "refutación" de la teoría de la identidad mente-cerebro como un subproducto ilustrativo de su nueva versión, extraordinariamente influyente, de la necesidad, la referencia y la identidad, el asunto pudo haber aparecido como un anacronismo. El problema que enfrenta Kripke se deriva del tipo de formulaciones con las que se enuncian las tesis materialistas. Un enunciado de la teoría de la identidad tendría esta forma:

(5) "El dolor es la estimulación de las fibras C del cerebro".

Se podría argumentar contra esta tesis que en su formulación se está cometiendo una especie de "error de categoría": el dolor es un estado privado, mientras que un estado cerebral es una entidad observable y localizable espacialmente. Yo no me puedo equivocar cuando digo que tengo un dolor, pero sí puedo sentirlo sin tener la más mínima noción de neurofisiología. En esta crítica se está suponiendo que, en cuanto declara-

---

10 El argumento se basa en la exposición misma del funcionalismo, que identifica, en todos sus aspectos lógicos y metodológicos, el problema mente-cuerpo con el problema estado lógico-estado estructural de una máquina de Turing. En este último caso, según Putnam, nadie toma en serio la "identidad" o "no-identidad" de los dos estados, pues no tiene ninguna importancia cuál pueda ser la respuesta. Si, por ejemplo, se llegara a la muy improbable conclusión de que los estados lógicos de las máquinas de Turing son diferentes de sus estados estructurales, eso sería un descubrimiento puramente *verbal*, y lo mismo valdría para el caso humano. Y si se piensa que el "problema" tiene algún interés extra, por ejemplo demostrar que existen "almas", tendría que aceptarse alguna de estas tres conclusiones: a) que ningún argumento usado por algún filósofo ha arrojado jamás la más leve claridad sobre el asunto; o b) que algún argumento filosófico en favor del mecanismo es correcto; o c) que algún argumento dualista muestra en efecto que *tanto* los seres humanos *como* las máquinas de Turing tienen almas!

ción de una identidad, el enunciado (5) es "analítico": sólo así sería posible la identidad.

Ahora bien: la respuesta de los partidarios de la tesis materialista será que esta tesis, así como (5), que no es más que una formulación particular de ella, no es un enunciado analítico basado en la sinonimia. Dirán que tiene el mismo estatuto que los siguientes enunciados:

(6) "El calor es la energía cinética promedio de las moléculas".

(7) "El agua es H_2O"

Para los físicos y químicos, (6) y (7) no son enunciados de sinonimia, es decir, enunciados "analíticos". Si lo fueran, tendríamos derecho a decir también que allí hay "errores de categoría". Pero es claro que (6) y (7) son enunciados científicos perfectamente refutables. Lo que los partidarios de la teoría de la identidad sostienen es que (5) es un enunciado análogo a (6) y (7). Este es el punto de partida de Kripke para su crítica.

Los argumentos de Kripke dependen de algunas innovaciones técnicas. Un "designador rígido" es una expresión que designa a la misma entidad "en todos los mundos posibles". Así, "Benjamín Franklin" es un designador rígido, mientras que "el inventor de los lentes bifocales" no lo es, aunque en este mundo ambos designen al mismo individuo. Ahora bien: Kripke sostiene convincentemente que todas las auténticas afirmaciones de identidad se componen de términos que son designadores rígidos. De ahí que cuando son verdaderas lo son necesariamente. Los enunciados (6) y (7), por ejemplo, si son verdaderos, expresan verdades necesarias. La ilusión de que son enunciados contingentes proviene de la asimilación comúnmente aceptada de los términos *necesario, a priori, analítico*, entre sí, y de sus opuestos: *contingente, a posteriori, sintético*. Los enunciados (6) y (7), y en general las identidades teóricas, pueden ser perfectamente a la vez enunciados necesarios y *a posteriori*, pues la necesidad es una característica relativa a como son las cosas, mientras que el hecho de ser *a posteriori* es una característica epistemológica, que se refiere a la manera como es conocida una verdad. Los enunciados teóricos, pues, en cuanto enunciados de identidad que involucran dos "designadores rígidos", son ejemplos de necesidad *a posteriori*. Y lo que sea válido lógicamente para ellos lo será también para los enunciados de la teoría de la identidad mente-cerebro como (5). Por consiguiente, lo que se afirma en este tipo de enunciados, y en general en la teoría de la identidad mente-cerebro, tendría que ser una identidad necesaria.

Ahora bien: a diferencia de lo que ocurre con los enunciados (6) y (7), en el caso de (5) es posible distinguir entre el dolor y la estimulación de las fibras C (cualquier cosa que ésta sea). Es decir: es posible que (hay un mundo posible en donde) haya estimulación de fibras C sin que haya la sensación de dolor. En el caso del calor, la apariencia de contingencia de la identidad calor/movimiento molecular se debe a que podemos imaginar una situación en la que se tenga *la sensación de calor* sin que haya movimiento molecular, pero la sensación de calor *no es* el calor. El conocimiento de la identidad es *a posteriori*, y en él puede intervenir la sensación de calor como un intermediario; pero el calor *es* movimiento molecular. En el caso del dolor no hay una "ilusión de contingencia" semejante, pues la sensación de dolor *sí es* el dolor. Por consiguiente, si es posible que haya estimulación de fibras C sin dolor (la contingencia no es "ilusoria") entonces el dolor no es *necesariamente* idéntico a la estimulación de fibras C. Si aceptamos, pues, que establecemos la identidad estado mental/estado cerebral por medio de designadores rígidos, y si aceptamos que todo enunciado de identidad cuyos términos son designadores rígidos es una identidad necesaria, tenemos que concluir que la teoría de la identidad es falsa.

A menos que se acepte la llamada "teoría de la identidad contingente", es decir, que una persona *pueda* existir sin su cuerpo (o su equivalente en términos mente/cerebro). Pero este tipo de identidad no le sirve a quienes sostienen esta teoría, pues ella equivaldría a concederle la razón a Descartes. La argumentación de Kripke es sutil y difícil, pero podemos tratar de aclararla brevemente.

Recordemos que Descartes sostenía que la mente era una propiedad *esencial* mientras que el cuerpo era sólo una propiedad contingente. Como la mente puede existir sin el cuerpo, una persona (una mente) es distinta de su cuerpo. El argumento puede resumirse así:

> Sea "*A*" un *nombre* (un designador rígido) del cuerpo de Descartes. Entonces, Descartes argumenta que ya que él pudo haber existido aun cuando *A* no existiese, (Descartes *A*), por lo tanto, Descartes *A*.[11]

Ahora bien: decir que Descartes pudo haber existido sin que *A* existiese

---

11 Cfr. S. Kripke, "Identity and Necessity", en Stephen P. Schwartz (ed.), *Naming, Necessity and Natural Kinds*. Cornell University Press, 1977, pág. 101 n.

es lo mismo que decir que *A* puede existir sin que Descartes exista. De hecho, esto puede mostrarse sin necesidad de argumentos modales: si se acepta que Descartes deja de existir cuando muere, es verdad que *A* puede existir cuando Descartes ya no existe (*A* sigue existiendo como cadáver), y por consiguiente *A* tiene una propiedad (existir en determinado momento) que no tiene Descartes. O sea que se puede llegar a la misma conclusión partiendo de la premisa de que el cuerpo (*A*) pudo haber existido sin la mente[12]. Algunos partidarios de la tesis de la identidad mente-cuerpo, dice Kripke, aceptan la premisa cartesiana, es decir, aceptan la contingencia de la identidad mente-cuerpo, pero no la conclusión, pues acusan al argumento de contener una falacia modal. La argumentación de Kripke destruye esta pretensión al hacer notar que "*A*" es rígido y que, por consiguiente, el argumento es correcto: si la premisa es verdadera la conclusión también lo es. Para sostener la tesis de la identidad mente-cuerpo sus partidarios deberían admitir el argumento cartesiano pero rechazar la premisa (la contingencia de la identificación) a pesar de que ésta parezca plausible. Y esto no parece ni trivial ni viable.

Llamemos "*A*" a una sensación de dolor particular y "*B*" al estado cerebral correspondiente. Parece lógicamente posible, como se dijo, que *B* podría existir sin la presencia de *A*, es decir, que por ejemplo Juan podría estar en ese determinado estado cerebral sin sentir ningún dolor. Pero esto no sería posible pues la identidad de *A* y *B*, si es verdadera, es una identidad necesaria, dado que "*A*" y "*B*" son designadores rígidos. Quienes sostienen la teoría de la identidad mente-cerebro quisieran que la identidad entre un estado cerebral y su correspondiente estado mental pudiera analizarse en términos tales que lo mental (no-físico) resultara siendo una *propiedad contingente* del estado cerebral, es decir, que lo que una oración como (5) declara es que el ser un estado mental es solamente una propiedad contingente de un determinado estado cerebral. Pero esto es incluso un absurdo autoevidente: equivaldría a declarar que el dolor que ahora siento, es decir, este estado mental específico, podría existir sin que yo tenga este estado mental.

---

12 Dado que los textos a que se hace referencia son transcripciones de conferencias realizadas sin notas, no es completamente claro que Kripke suscriba estos argumentos, los cuales atañen a una teoría de la identidad persona-cuerpo y aparentemente podrían ser refutados postulando que dentro de esta teoría la propiedad de estar vivo es una propiedad contingente de las cosas que la poseen.

En síntesis: si se sostiene la identidad "$A = B$", entonces la identidad entre A y B es una identidad necesaria y cualquier propiedad esencial del uno debe ser también una propiedad esencial del otro. Quien quiera sostener la tesis de la identidad no puede simplemente aceptar la intuición cartesiana de que A puede existir sin B, de que B puede existir sin A, de que la presencia correlativa de algo que tiene propiedades mentales es solamente un hecho contingente para B y que la presencia correlativa de propiedades físicas es solamente un hecho contingente para A. Tiene que explicar estas intuiciones y mostrar su carácter ilusorio[13].

Hay que decir que los argumentos de Kripke no han encontrado mucho eco. El principal obstáculo con el que se enfrentan es quizás que, aun si se acepta su teoría de la identidad y la necesidad, la postulación de los términos pertinentes como "designadores rígidos" parece muy vulnerable a un escrutinio minucioso. Por ello no se puede afirmar que la argumentación de Kripke en contra de la teoría de la identidad haya conducido a una revitalización del dualismo.

El funcionalismo

El funcionalismo puede ser visto como una variante de la teoría de la identidad, tal como quedó indicado más arriba, aunque estrictamente no tiene por qué admitir ningún compromiso ontológico. De hecho, fue propuesto por Hilary Putnam como alternativa a las *impasses* a las que conduce tal teoría, en especial las que ponen en tela de juicio las oraciones de identidad sobre bases semánticas. Para ello Putnam se apoyó en una analogía entre la relación mente/cerebro y la relación estado lógico/estado estructural de una máquina de Turing[14].

Se puede sostener que los estados mentales son a los estados cerebrales lo que los estados funcionales, o lógicos, de una máquina de Turing son a sus

---

13 Cfr. S. Kripke, "Naming and Necessity", en Davidson & Harman (eds.), *Semantics of Natural Language*, Reidel (Dordrecht), 1972, pág. 334 ss.

14 Una "máquina de Turing" es un dispositivo matemático "ideado" para darle precisión a la noción de "procedimiento efectivo", la cual es básica en la teoría de la computabilidad. Como tal es solamente un "ente ideal", pero no es completamente fantasioso. Las máquinas computadoras actualmente existentes no son más que realizaciones de la máquina de Turing. Véase Putnam, "Minds and Machines", *op. cit.* (nota 9).

estados estructurales. Los estados funcionales se toman como estados "internos" de un organismo que (en relación con las "entradas", "salidas" y otros estados) satisfacen una determinada teoría formal, la cual es vista como un "programa" para ese organismo. Esta noción de "estado funcional" es a su vez una noción de "rol". El mismo estado funcional S_p puede ser "realizado" en diferentes organismos de diferentes maneras físicas, y cada uno de estos estados físicos cumple el mismo papel en la determinación del resultado (*output*) que caracteriza a S_p. Como lo ha subrayado Putnam en numerosas oportunidades, es perfectamente concebible que S_p pueda ser realizado por un estado *no*físico, por ejemplo de un organismo inmaterial. Un marciano muy bien podría ser una criatura puramente electromagnética que tenga periódicamente estados del tipo S_p realizados sólo por campos; o un fantasma podría en determinado momento encontrarse en un estado del tipo S_p de su núcleo ectoplásmico. A diferencia de un materialista eliminacionista, el funcionalista plantea una identidad estricta; sinembargo, también es compatible con un dualismo cartesiano, pues no hay contradicción en la noción de una causa inmaterial. En sentido estricto, la tesis funcionalista no implica necesariamente ningún compromiso ontológico. Aunque un funcionalista *materialista* sostendrá que de hecho S_p siempre se encuentra realizado por algún tipo de estado físico[15].

El funcionalismo tiene la ventaja de proporcionar el cimiento conceptual para el trabajo actual en psicología cognitiva, psicolingüística, y en las modelizaciones de la inteligencia artificial. En estas disciplinas, una estrategia de investigación puede caracterizarse en términos de (una versión de) la distinción de Chomsky entre competencia y actuación: dada la especificación de una determinada clase de competencia, digamos discriminativa, o lingüística, la tarea consiste en diseñar un modelo de actuación, a menudo un programa modelo de computador, que posea esa competencia, y de ser posible, que incluya también pretensiones de realidad psicológica. Esta especie de estrategia investigativa permite explorar dificultades y problemas altamente abstractos sin tener que preocuparse por la mecánica o la bioquímica de sus "realizaciones" concretas en la cabeza, sin abandonar por ello la restricción fisicalista fundamental de que los sistemas descritos funcionalmente sean de alguna manera realizables. En su nivel más

15 En la actualidad los funcionalistas no sostienen que los "procesos mentales" sean idénticos a procesos establecidos en el formalismo de la máquina de Turing, sino en general en un formalismo computacional. Pero el modelo que estuvo en el punto de partida es el de la máquina de Turing y es éste el que le otorga todo su sentido.

abstracto y general, este tipo de investigación se mezcla con investigaciones en epistemología y filosofía de la mente, y es precisamente en este terreno de contacto en donde se realiza en la actualidad el trabajo más interesante y prometedor.

Quien con mayor claridad y algo de audacia ha expuesto los principios metodológicos del funcionalismo en conexión con el estudio de la mente (en particular con la psicología) ha sido J. Fodor. En 1975 H. Putnam acuñó la expresión "solipsismo metodológico" para describir, y de paso desacreditar, la posición defendida por Fodor[16]. Pero éste la asumió y la utilizó para exponer su punto de vista en un artículo titulado "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology"[17]. Allí Fodor distingue entre lo que llama la "concepción representacional de la mente" y lo que llama también "la concepción computacional de la mente". La primera es una versión débil del funcionalismo y es la menos discutible: afirma simplemente que las teorías psicológicas no pueden tener menos de dos grados de libertad. Si Pedro cree que *p*, entonces no teme que *p*, ni desea que *p*, ni tampoco cree que *q* (a menos que *p* implique *q* y que Pedro lo sepa): los dos parámetros de las teorías psicológicas son la actitud (estado mental) expresada por el verbo de actitud proposicional y la proposición expresada por la frase subordinada. Este es el programa mínimo de las ciencias cognitivas, pero no es aún el solipsismo metodológico. Este deberá estar de acuerdo con la concepción computacional de la mente, según la cual los procesos mentales tienen acceso solamente a propiedades formales, no-semánticas, de las representaciones mentales según las cuales se les define, y con el "principio de formalidad" contenido en esta concepción. Así, pues, lo que afirma el solipsismo metodológico es que los procesos mentales se aplican a las representaciones solamente en virtud de la sintaxis de estas representaciones. Y con esto se define el programa "fuerte" de las ciencias cognitivas.

La concepción computacional, con su principio de formalidad, es mani-

---

16 Aunque no refiriéndose a Fodor explícitamente, sino a algunos "filósofos tradicionales" que sostienen que ningún estado psicológico (mental) propiamente dicho presupone la existencia de ningún individio diferente de aquél a quien se le adscribe dicho estado. Cfr. "The Meaning of "Meaning"", en *Mind, Languge and Reality. Philosophical Papers, Vol. II*, Cambridge University Press, 1975, pág. 220.

17 Publicado inicialmente en *The Behavioral and Brain Sciences*, 3 (1980), págs. 63-73.

fiestamente mucho más restrictiva que la concepción representacional de la mente: afirma que los procesos mentales pertinentes para las teorías psicológicas son *sintácticos*, es decir, que no pueden ser semánticos, o al menos que su semántica no es psicológicamente pertinente. En principio, esto parece plantear una contradicción con lo sostenido por el mismo Fodor en *The Language of Thought*[18], en donde sostiene, entre otras cosas, que aprender una lengua es aprender una definición de la verdad para esa lengua. Hay que tener en cuenta que el sostenimiento de la teoría computacional de la mente y su condición de formalidad hacen de la analogía entre las computaciones de la mente y las de un computador algo muy fuerte, y que en este sentido, puesto que manifiestamente los computadores "no tienen una semántica", habrá que admitir que la mente humana tampoco tiene estas propiedades, al menos no en cuanto objeto de estudio de la psicología: las relaciones causales entre el mundo y la mente no es un asunto de la psicología funcionalista, sino, a lo sumo, de alguna teoría naturalista de la percepción.

La teoría computacional de la mente y el principio de formalidad implica que la psicología deba hacer taxonomías también puramente sintácticas de los estados mentales. Descritos en estos términos, los estados mentales conservan la opacidad propia de las expresiones de actitud proposicional en la cual se expresan comúnmente. Esto quiere decir que, por ejemplo, Pedro y Pablo estarán en el mismo estado mental cuando cada uno se crea enfermo, es decir, que individualizados sólo formalmente, ambos estados son el mismo tipo de estado mental, pues al decirse la misma frase: "yo estoy enfermo", ambos enuncian mentalmente el mismo objeto sintáctico. El hecho de que cuando Pedro cree que él está enfermo el contenido de su creencia es que *él* (Pedro) está enfermo, y en cambio cuando Pablo lo cree lo que cree es que *él* (Pablo) está enfermo, es decir, que las dos creencias difieren en cuanto a sus referentes y, por ello mismo, en cuanto a sus valores veritativos, no es pertinente y no logra hacer diferir los dos tipos de creencia, pues la referencia, y *a fortiori* las condiciones de verdad son propiedades semánticas. Así como los computadores ejecutan cálculos en su lenguaje-máquina obedeciendo instrucciones puramente sintácticas, así mismo la mente (según la concepción computacional y la condición de formalidad) pasa de estado computacional a estado computacional, y estos estados sólo son definidos en términos sintácticos. Como afirma el propio Fodor, la

---

18 New York, 1975. Hay traducción española: *El lenguaje del pensamiento*, Alianza, 1984.

verdad, la referencia y el resto de nociones semánticas no son categorías psicológicas.

A diferencia de la concepción de Fodor, que parece demasiado fuerte, la de D. Dennett en *Intentional Systems*[19] parece demasiado débil. Dennett parece encontrar aún demasiado reduccionista el funcionalismo que adopta literalmente el punto de vista de la mente como una "máquina de Turing", pues éste definiría una creencia, por ejemplo, más o menos en los siguientes términos:

(1) (Para todo x) (x cree que la nieve es blanca = x "realiza" una máquina de Turing *k* en el estado lógico Y).

En este caso lo que es objeto de identificación son los tipos de estados mentales y de estados de la máquina. Dennett propone que se debilite una condición como (1) dándole más bien la forma siguiente:

(2) (Para todo x) (x cree que la nieve es blanca = se le puede atribuir a x, con el fin de poder efectuar predicciones referentes a ella, la creencia de que la nieve es blanca).

Aquí aparentemente sólo se dice que una creencia es una creencia si y solamente si es una creencia. Pero es que Dennett no propone una estrategia de reducción de la creencia o de los demás estados mentales. Según su punto de vista, la construcción de una teoría psicológica supone que se admitan las actitudes proposicionales en términos de actitudes *intencionales*, lo cual, entre otras cosas, coloca al lenguaje de la psicología en el mismo nivel de descripción que el lenguaje común. De hecho, la reducción llegará, pero en un nivel ulterior, una vez hayamos definido lo que Dennett llama un *sistema intencional.*

Un sistema intencional posee esencialmente un estatuto heurístico: es intencional todo sistema que permita predecir el comportamiento de un organismo, siempre que reúna tres factores: un programa o una cierta función, cierta composición física y cierta racionalidad. La instancia de racionalidad es la instancia intencional y tiene la forma de actitudes que se le atribuyen al sistema, por ejemplo a una máquina de jugar ajedrez: se dirá

---

19 D. Dennett, "Intentional systems", en *The Journal of Philosophy*, vol. 68 (1971), págs. 87-106.

que la máquina *espera* a que el contrincante juegue, que *cree* que si el adversario coloca su rey en E4 está amenazando al alfil y, sobre todo, que la máquina *desea* realmente ganar el juego. Ahora bien, estas atribuciones son necesariamente "sub-determinadas", en el sentido que le da Quine a este término, por los datos empíricos. En efecto, atribuir a un sistema una creencia solamente tiene sentido si se le atribuyen también un conjunto de creencias consecuentes y de reglas lógicas. Pero este conjunto tiene que ser necesariamente difuso, pues en cuanto actitudes proposicionales las creencias no cumplen la llamada Ley de Leibniz y, por consiguiente, el conjunto puede presentar inconsistencias. Así mismo, es imposible determinar a priori la relación entre reglas y representaciones de reglas en el interior de un conjunto de creencias, pues de la constatación que un sistema funciona según determinadas regularidades intencionales no se puede inferir que estas reglas son necesariamente creencias y, por consiguiente, que hacen parte del plan que rige el comportamiento del sistema.

Al descuidar esta distinción, Fodor cree poder afirmar que la inteligencia que se manifiesta en el comportamiento implica necesariamente la existencia de un sistema interno de representación, dotado de las propiedades lógicas de una lengua, el llamado "lenguaje del pensamiento", o "Mentalés". Pero si aceptamos lo que dice Dennett, el punto de vista intencional no es más que una estrategia metodológica y como tal debe ser eliminado una vez que la investigación haya avanzado lo suficiente en la comprensión del sistema. Con otras palabras, para Dennett la intencionalidad solamente tiene existencia teórica: no decimos que el sistema, por ejemplo la máquina de jugar ajedrez, se encuentre realmente en estado de espera, o que tenga la creencia tal, o que realmente quiera ganar el partido. De hecho, las investigaciones en inteligencia artificial proceden de este modo: se desea construir programas que sean "simulaciones" o modelos de determinadas actividades humanas y para ellos se presupone la racionalidad del sistema, o simplemente se le atribuye tal racionalidad. Como dice Dennett, se hace un "préstamo" sobre la inteligencia del sistema, préstamo que habrá que reembolsar más tarde para no considerar más que la manera como funciona la presunción de racionalidad. Bien sea que el sistema *crea* las verdades de la lógica o que no las crea, lo que debemos suponer es que *las sigue*.

Los demás componentes de este principio de racionalidad son el carácter normativo de las creencias (si se cree que *p*, entonces se cree que *p* es verdad) y un "principio de caridad" según el cual la mayor parte de las creencias que se atribuyen a un sistema deben ser verdaderas para que su comportamiento pueda ser interpretado. Cada una de estas suposiciones de racionalidad puede ser desechada en uno u otro momento, pero el uso de una noción como

"sistema intencional" requiere que en un determinado punto cesemos de verificar su racionalidad, y que la admitamos sin más.

La importancia que Dennett atribuye a su teoría de los sistemas intencionales puede evaluarse por las palabras del mismo Dennett, quien se pregunta: "¿existe algún tesoro mental que no pueda ser comprado con dinero intencional?". Con otras palabras: la estrategia de investigación en filosofía de la mente que permite abrigar las mejores esperanzas con respecto a su desarrollo es la estrategia intencional.

Existen variantes del funcionalismo en diversas aplicaciones a la psicología, la inteligencia artificial y en general a lo que se conoce como "ciencia cognitiva". Recientemente algunas propuestas en inteligencia artificial (llamadas *conexionistas*) plantean abandonar toda noción de "representación" o "programa" y hablan en su lugar de "redes neuronales", con lo cual se acercan a una reconciliación con las teorías biológicas. Llegados a este punto es difícil catalogar a esta empresa como una variante más del funcionalismo.

Algunas críticas al funcionalismo

Voy a mencionar dos críticas filosóficas que han sido dirigidas al funcionalismo, sin ahondar mucho en ellas y sin exponer completamente las teorías de la mente a las que responden. Será solamente una indicación de tendencias.

La primera es la del filósofo norteamericano J.R. Searle, quien ha propuesto una teoría de la mente basada en el concepto de *intencionalidad*[20]. Searle se confiesa un "realista", en el sentido de admitir que realmente existen cosas tales como creencias, deseos, intenciones y demás estados mentales a los cuales hace referencia normalmente el habla común con estos términos. Y expone una teoría según la cual lo que caracteriza a estos estados mentales no es su naturaleza ontológica (si se trata de configuraciones

---

20 La crítica aprareció inicialmente en "Minds, Brains and Programs", y en "Intrinsic Intentionality", en *Behavioral and Brain Sciences*, 3, 1980, págs. 450-456. También en *Mind, Brain and Science (Reith Conferences 1984)*. (Hay traducción: *Mentes, cerebros y ciencia*, Madrid, ed. Cátedra). La teoría de la mente ha sido expuesta en numerosos artículos y recogida sistemáticamente en *Intentionality: An Essay in the Philosophy of Mind*, Cambridge U. Press, 1983.

neuronales, o imágenes mentales, o modificaciones de un "yo", etc.) sino un rasgo lógico esencial, que él denomina "intencionalidad", y que consiste en estar referidos a objetos o estados de cosas en el mundo. Su teoría no está pues, en principio, comprometida con el materialismo de la teoría de la identidad, dada esta neutralidad ontológica, y más bien parece cercana al funcionalismo. No obstante, el carácter intencional de los estados mentales no permite esta identificación.

Que los estados mentales sean intencionales significa que poseen un contenido representacional, y que es este contenido el que hace que estén "dirigidos" hacia objetos o estados de cosas en el mundo. Con otras palabras, los estados mentales son "intencionales" en el mismo sentido en que el lenguaje posee una "semántica": en ambos casos, el contenido representacional determina ciertas "condiciones de satisfacción" que establecen la relación con el mundo. Ahora bien: este contenido de las representaciones mentales, que les confiere su carácter intencional, es precisamente lo que hace falta en las "representaciones" postuladas por Fodor y otros funcionalistas. Y este es el punto central del ataque.

Si el funcionalismo "fuerte" estuviera en lo correcto, tendríamos que atribuirle intencionalidad a los programas que procesan las máquinas de computación, por ejemplo. Pero las operaciones de estas máquinas son cálculos que se efectúan sobre elementos que han sido especificados sólo formalmente (y este rasgo de formalidad le es inherente y esencial al funcionalismo "fuerte", como se vio) y que no tienen *per se* ninguna relación con el mundo. Por ejemplo, el punto de vista funcionalista-cognitivista que discute Searle supone que los principios formales especificados en el programa y las operaciones computacionales de la máquina bastan para comprender una lengua. No obstante Searle muestra por medio de un sencillo "experimento de pensamiento" cómo nosotros mismos podemos realizar toda clase de operaciones, siguiendo instrucciones precisas, sobre símbolos exclusivamente formales (del modo como lo hace una máquina) sin por ello haber comprendido en lo más mínimo lo que esos símbolos significan. La comprensión de una lengua, como toda actividad intencional, no se reduce a operar sobre símbolos formales.

Para Searle, el modelo programa/máquina como explicación de la relación mente/cerebro no sirve sencillamente porque el programa, a diferencia de la mente, no posee intencionalidad. Las manipulaciones de símbolos que ocurren cuando una máquina "corre" un programa no poseen en sí mismas ninguna intencionalidad, y en esta medida no tienen *sentido*. No son ni siquiera manipulaciones de *símbolos*, pues los tales "símbolos",

carentes de significación, no simbolizan nada: solamente poseen sintaxis, no semántica. La intencionalidad que parecen exhibir algunos modelos de programas corridos en un computador se encuentra solamente en la mente de quienes programan y de quienes utilizan el programa, es decir, de quienes introducen el "input" y quienes interpretan el "output".

Los estados intencionales (mentales) no son formales en el sentido en que lo son los programas o estados lógicos de un computador. Los estados intencionales están descritos y caracterizados en términos de sus contenidos, y se los define como un determinado contenido mental con determinadas condiciones de satisfacción, dirección de adecuación entre él y el mundo, etc., es decir, en términos de sus características semánticas. Un estado intencional no tiene ni siquiera una forma "sintáctica" pues a la misma creencia, por ejemplo, podemos darle un número indefinido de expresiones sintácticas diferentes en sistemas lingüísticos diferentes.

En síntesis, el problema fundamental del funcionalismo, según Searle, radica en su tesis principal: la "neutralidad ontológica" postulada de los estados mentales obliga a no tomar de ellos más que sus rasgos estrictamente formales; pero al hacerlo se está excluyendo la intencionalidad de lo mental; y sin intencionalidad no hay representación.

Ahora bien: ¿por qué la "neutralidad ontológica" excluye la intencionalidad? Es cierto que ésta no puede venir de la sola realización de un programa formal. Por consiguiente, termina sosteniendo Searle, no se pueden disociar las operaciones mentales de las operaciones del cerebro tal como se disocian los programas de las operaciones de las máquinas. La intencionalidad es una propiedad física, real, del cerebro, un fenómeno biológico que deben explicar las neurociencias, puesto que él es producido por "poderes causales del cerebro". La intencionalidad depende tanto de la bioquímica de sus orígenes como, por ejemplo, la "lactancia" y la "fotosíntesis" dependen causalmente de la bioquímica de sus orígenes. Como filosofía de la mente, la teoría no puede decir más y se remite a lo que algún día puedan decir las ciencias del cerebro al respecto.

La segunda crítica al funcionalismo como filosofía de la mente tiene un interés particular por provenir de su propio "fundador", el también filósofo norteamericano Hilary Putnam[21]. Este lo propuso para otorgarle un estatuto

---

21 Véase *Representación y realidad, op. cit.*, (nota 9).

científico a la psicología anticonductista basada en creencias y deseos identificándola con la psicología computacional que se deriva del planteamiento funcionalista (el cerebro como un computador y los estados mentales como el *software* de ese computador). Su ataque va dirigido ahora directamente contra el centro de su propuesta: no es posible, dice, individualizar los conceptos y las creencias aislándolos de su entorno social y humano como se aísla el *software* o programa de un computador de la máquina que lo computa. Si bien el funcionalismo acertó en su momento al señalar que, por ejemplo, un pensamiento en π no necesita tener una definición en términos físico-químicos que comparta cualquier pensamiento en π, también tendríamos que admitir que, contrariamente a la pretensión funcionalista, los distintos "estados computacionales" en los que uno puede encontrarse mientras piensa en π no necesitan tener "algo en común" definible en términos computacionales. La razón básica del error funcionalista estriba en su pretensión de buscar una explicación no intencional de lo intencional, es decir, en su pretensión reduccionista de "explicar" las actitudes proposicionales apelando a algo "básico" —los "procesos mentales" caracterizados computacionalmente— que concuerde con nuestro sistema de "metafísica científica".

Putnam aplica la técnica del "*Jiu-Jitsu*" y utiliza la fuerza del adversario en su contra: los mismos argumentos esgrimidos para mostrar la incorrección de la tesis que identifica los estados mentales con estados físico-químicos se generalizan para mostrar también la incorrección de la identificación de esos estados con estados caracterizados computacionalmente: no solamente se puede atribuir el mismo estado mental a sistemas que no poseen la misma constitución física (contra la teoría de la identidad), sino que también se puede, en principio, atribuir el mismo estado mental a sistemas que no poseen la misma estructura computacional (contra el funcionalismo). El ataque, pues, no se dirige a negar que puedan existir estados computacionales, sino a negar que puedan encontrarse equivalencias, por ejemplo entre las estructuras de todos los sistemas que tienen una creencia determinada, que puedan definirse en términos computacionales. El funcionalismo, como el materialismo de la teoría de la identidad, es falso por su pretensión reductiva.

Como Searle, Putnam es un "realista" y admite que hay estados mentales ("intencionales") y que no se los puede desechar, como hacen los "eliminacionistas" (Quine, etc.), calificándolos como "psicología popular". Pero niega que pueda encontrarse alguna propiedad científicamente descriptible que sea común a todos los casos de algún fenómeno intencional particular. Entre estos fenómenos intencionales incluye no solamente a los

estados mentales como creencias, deseos, etc., sino a los fenómenos básicos de la significación y la referencia y, en cierta medida, también a la verdad. Aunque "realista", pues, Putnam se opone tanto a los funcionalistas "fuertes" como a Searle y a todo aquel que sostenga en general que existen "representaciones mentales" descriptibles en términos físicos o computacionales y que dan cuenta de todos nuestros conceptos y de nuestro funcionamiento mental.

La equivocación del mentalismo como postura general radica en pretender explicar la significación y la referencia, y *a fortiori* los estados mentales como la creencia y el deseo, con recurso a "representaciones mentales", debido a que esta postura choca con dos ideas claves, la una sostenida por Quine y Davidson, la otra propuesta por el mismo Putnam.

La primera es el llamado "holismo del significado": las significaciones no se pueden individualizar separadamente, pues cada una remite a la totalidad de teorías, creencias y conocimientos generales compartidos por una comunidad. Así como Quine sostuvo en su momento contra los empiristas lógicos que no era posible la verificación de enunciados individuales sino que las teorías se verificaban "como un todo", Putnam sostiene que los conceptos significan en medio de la totalidad de conocimientos y no de manera individual. No es posible entonces aislar una representación que pueda figurar como la "significación" de un término o un concepto dado.

La segunda idea clave atañe a la determinación de la referencia. Putnam ha librado una larga batalla contra la idea mentalista en semántica según la cual "la intensión (significación) determina la extensión (referente)". Contra esta postura, cuyo más ilustre forjador sería el propio Aristóteles, pero cuyo impulso en semántica filosófica y lógica proviene de Frege, Putnam ha argumentado que si ella debiera ser cierta, entonces "los significados no están en la cabeza": la referencia no se determina por ninguna representación mental del objeto referido. Su argumento consiste en mostrar que en la fijación de la referencia los factores determinantes son el entorno, es decir la particular configuración del mundo en el que se encuentran los objetos referidos, y la cooperación social y cultural que forma el tejido en el cual se dan los intercambios lingüísticos. Es posible, por ejemplo, encontrar dos individuos con idénticas representaciones mentales asociadas a términos de sus respectivos lenguajes pero refiriéndose a objetos distintos. Y la razón es que todos los términos del lenguaje poseen en alguna medida un componente indexical que los vincula de manera esencial a una situación y a un contexto determinados. La referencia, pues, no puede fijarse por una representación mental, independiente de esta situación y de este contexto.

Es necesario, en consecuencia, separar las cuestiones relativas a la fijación de la referencia de las expresiones del lenguaje de las cuestiones relativas a sus contenidos conceptuales.

De estas dos ideas surge la concepción de la generalidad del fenómeno de la *interpretación*, señalado por Quine y por Davidson, pero también por filósofos como Gadamer. Si el "holismo del significado" es correcto, y si la determinación de la referencia depende del entorno natural y social y de la labor de cooperación lingüística, entonces los discursos tienen que ser interpretados como totalidades. Esto quiere decir que para interpretar un lenguaje es necesario, en general, tener alguna idea de las teorías y los métodos de inferencia que comparten los miembros de la comunidad que habla ese lenguaje. Ahora bien, si esto es un problema insuperable para el mentalismo en general, lo es aún más para la forma que toma esta postura en el "funcionalismo fuerte" de Fodor y de la mayoría de los teóricos de la inteligencia artificial: la noción misma de una psicología computacional implica que todas las representaciones deben ser descritas en términos sintácticos o en términos de procedimientos, o bien por una combinación de ambos factores. Todo lo que se necesita, pues, es que la situación pueda ser descrita en algún lenguaje regimentado. Pero esto requeriría, por ejemplo para saber solamente a qué se refiere un término usado en una situación y en un entorno determinado, una descripción completa de la situación, la historia y el entorno en términos de algún conjunto estandarizado de parámetros físicos y computacionales, lo cual enfrenta el insuperable problema de que ese "entorno" deberá incluir *todo el universo*. Para saber si dos términos son correferenciales, o sinónimos, habría que examinar todas las teorías posibles y anticipar de alguna forma los procesos de fijación de creencias de que depende la comprensión de la teoría a la cual pertenece el término.

La interpretación de un discurso requiere que se lo pueda continuar. Un algoritmo que interpretara un discurso arbitrario, dice Putnam, tendría que ser lo suficientemente "inteligente" para examinar todos los posibles discursos racionales, semirracionales y poco racionales pero aún inteligibles que pudiesen construir de modo físicamente posible todas las criaturas físicamente posibles. Pero es posible examinar, y establecer reglas de interpretación para ellos, el razonamiento y las creencias de todos los seres y todas las sociedades humanas posibles? Si lo fuera, es decir, si existiera una *teoría* sobre todos los discursos humanos, concluye Putnam, "sólo un dios podría escribirla". La "psicología computacional" tiene como obstáculo esencial "la infinitud abierta e inagotable de los esquemas con-

ceptuales que [...] deben ser interpretados". Y esto es lo que defiene sus límites.

Porque la cuestión parece ser nuevamente de límites. Como lo señala el mismo Putnam, el enfoque computacional ha querido introducir un paradigma totalmente nuevo acerca de lo que debería ser una explicación realista científica de la mente y la intencionalidad. Este paradigma vendría a reemplazar al viejo (pero aún sostenido, aunque con más convicción que pruebas, por filósofos como Searle) paradigma realista científico que pretendía explicar la mente en términos de la "mecánica del cerebro" o de una psicología de asociación de ideas, contra el cual habían escrito ya filósofos como Brentano y Husserl y, antes que ellos, Kant, quien había observado en la *Crítica de la Razón Pura* que probablemente no nos sea posible dar una explicación del *esquematismo* en términos de la ciencia natural. Ahora bien, así como el viejo paradigma encontró sus límites, según Putnam lo mismo ocurre con el nuevo paradigma computacional. Y en este plan de alcanzar límites no es casual que el argumento final sea de tipo gödeliano.

Gödel demostró que no es posible formalizar completamente la capacidad matemática porque hace parte de esta misma capacidad el trascender aquello que formaliza. Del mismo modo, una formalización completa de la interpretación es un proyecto tan utópico como una "formalización" completa de la fijación de creencias, pues saber cómo se *usan* las palabras implica saber cómo se fijan las creencias que contienen esas palabras, y la fijación de las creencias es holística. El proyecto de formalización de la razón fracasa porque es uno de los rasgos esenciales de la razón el trascender aquello que ella misma examina.