

El advenimiento de la metáfora mente-computador* Del alma al *software* (3)

Felipe de Brigard**

Teoría de identidad de instancias

Existe, por otra parte, una versión más débil de la teoría de identidad que, en lugar de exigir una equivalencia entre tipos de estados mentales y tipos de estados físicos o cerebrales, sostiene que:

[Enunciado 9] Para cada instancia o ejemplar (*token*) particular de un estado/suceso mental deberá haber algún estado/suceso o estado neurofisiológico al que sea idéntico (tesis de la identidad de instancias).

Esto implica, por otra parte, que una persona puede estar en el mismo estado mental en dos ocasiones distintas pero, en cada una, en un *tipo* de estado cerebral diferente, y viceversa.

Una de las versiones más controvertidas de la *teoría de identidad de instancias* la propuso hace unos años Donald Davidson. Él sostenía que los materialistas se enfrentan a una aparente contradicción a la hora de res-

.....

* En el artículo pasado consideramos, en primera instancia, los argumentos en contra de la postura conductista. Dicho análisis nos llevó a los orígenes del reduccionismo fiscalista en la teoría de identidad de tipos. Sin embargo, tras considerar sus más importantes objeciones, llegamos a una sin salida, en la que el materialismo eliminativo parecía ser la solución. Ahora la historia se bifurca. Lo que sigue a continuación no ha de verse necesariamente como una crítica al materialismo eliminativo, pues es bien sabido que una rama importante —y muy comedida— de la filosofía de la mente y de las ciencias cognitivas sigue siendo partidaria de esta corriente. La teoría de identidad de instancias, con la que comienza esta última parte, puede verse como una alternativa a la ya mencionada teoría de identidad de tipos, más que como una propuesta contra el materialismo eliminativo (aunque oponerlas no resulta en lo absoluto descabellado). Al igual que en el artículo pasado, la numeración de las tesis y los subtítulos mantiene coherencia con los anteriores.

** Pontificia Universidad Javeriana. Facultad de Filosofía.

ponder la pregunta: “¿cómo puede reconciliarse el hecho de que los estados mentales se resistan a ser capturados por la red nomológica de la teoría física, con el hecho de que ellos juegan un papel causal en el mundo físico?” (1). Y la contradicción surge –en su opinión– de la supuesta imposibilidad de reconciliar tres principios aceptados por los materialistas: (1) el *principio de interacción causal*, que sostiene que por lo menos algunos sucesos mentales interactúan causalmente con sucesos físicos; (2) el *principio del carácter nomológico de la causalidad*, que dice que donde hay causalidad hay una ley, esto es, los sucesos relacionados como causa y efecto caen bajo leyes deterministas estrictas, y (3) el *principio de la anomalía de lo mental*, que propone que no hay leyes deterministas estrictas basándose en las cuales puedan predecirse y explicarse los sucesos mentales.

Así, para evitar la pretendida inconsistencia que acarrearía aceptarlos sin más, Davidson propuso una nueva postura –conocida como *monismo anómalo*– que, a pesar de afirmar que un mismo evento puede ser a la vez físico y mental (de ahí que sea monismo), niega la existencia de leyes que relacionen la descripción mental con la descripción física (por lo cual es anómalo). De ahí que asevere: “El monismo anómalo se parece al materialismo al afirmar que todos los eventos son

físicos, pero rechaza la tesis, usualmente considerada esencial para el materialismo, de que los fenómenos mentales pueden ser dados en explicaciones puramente físicas” (1).

La razón de este rechazo, dice él, es que si bien podríamos llegar a desarrollar generalizaciones que ligaran eventos descritos en términos mentales con eventos descritos físicamente, estas generalizaciones no tendrían el carácter de una ley, es decir, no serían legaliformes; pues los predicados que se unirían en ese enunciado pertenecen a vocabularios distintos. Lo máximo que éstas podrían hacer por nosotros sería proveernos de “sabidurías prácticas”, que asumiríamos como toscamente ciertas y en virtud de las cuales formularíamos predicciones en términos, si acaso, probabilísticos. Para que pudiéramos formular leyes psicofísicas adecuadas, los predicados mentales, por un lado, y los predicados físicos, por otro, que se vincularan en éstas, deberían pertenecer a un mismo marco lingüístico cerrado, “pero no es plausible que los conceptos mentales puedan proveernos de un marco tal, simplemente porque lo mental no constituye, en principio, un sistema cerrado” (1). Así, el monismo anómalo se convierte en una teoría materialista de la mente que se opone al reduccionismo, pero no negándolo desde un punto de vista ontológico, sino desde uno nomológico.

Una de las ventajas que, se supone, presenta esta postura frente a la vieja teoría de identidad de tipos es que resiste las críticas modales de Kripke a la identidad. McGinn (2) desarrolló con cierto detalle este argumento a favor de la teoría del monismo anómalo. De acuerdo con su exposición, la teoría de identidad de instancias permite que “ellos [los ejemplares de estados mentales] puedan ser (y ser esencialmente) de un tipo tal que otros ejemplares de ese tipo fallen en tener propiedades que ellos, *qua* ejemplares, necesariamente tienen” (2). De esta forma, cuando las propiedades que vinculan los dos lados de la identidad son esenciales a los eventos físicos y a los eventos mentales como ejemplares y no como tipos, la relación puede entenderse como necesaria para estos casos en particular, y se evita así la tentación de entender también esta identidad como contingente.

No obstante, sostener la postura de la identidad de instancias nos acerca a una difícil conclusión: si es verdad que sólo hay eventos físicos en el mundo, de los cuales algunos de ellos son a la vez mentales, pero no es posible formular leyes psicofísicas estrictas, entonces la posibilidad de llevar a cabo una psicología científica –esto es, con leyes científicas– se pondría peligrosamente en tela de juicio. Un análisis de los eventos mentales en términos únicamente de sus ejemplares

nos impediría, por ejemplo, dar razones de por qué, si es que no compartimos un tipo de estado mental, podemos, a pesar de ello, coincidir en tener una determinada creencia o esperanza. Al parecer, dicen sus objetores, la única manera en que podemos establecer leyes psicofísicas estrictas es reconciliando tipos de estados mentales y tipos de estados físicos pero, como vimos anteriormente, esto no se puede hacer.

Por ello –entre otras razones– surgió una postura que no sólo permite formular leyes de este estilo sino que, además, resulta compatible con la teoría de identidad de instancias. En lugar de promulgar correlaciones entre tipos de estados físicos y tipos de estados mentales, en virtud de alguna propiedad física compartida por todos los ejemplares de dicha clase, lo hace en virtud de una propiedad funcional. Coincide con la teoría de identidad de instancias al afirmar que cada ejemplar de estado mental es idéntico a un estado físico particular (que, como veremos, no tiene que ser necesariamente cerebral), pero se separa de ella –y, en general, de las teorías fisicalistas– al proponer que las descripciones de los estados mentales deben darse en términos funcionales. Esta teoría se conoce como funcionalismo, y terminaremos este breve paso por la historia de la filosofía de la mente con un análisis de sus principales características.

Funcionalismo

El funcionalismo constituye una de las más importantes teorías filosóficas sobre la mente de los últimos tiempos. Aunque existen algunas variantes de la teoría, todas ellas coinciden en la idea primigenia de que:

[Enunciado 10] La característica esencial o definitoria de un estado mental es el conjunto de las interacciones causales mutuas que mantiene con (1) los efectos del ambiente sobre el cuerpo (o estímulos), (2) otros estados mentales y (3) el comportamiento corporal (tesis funcionalista).

Probablemente, como anota Block (3), el funcionalismo en su versión original tiene dos fuentes principales: por una parte, la opinión expuesta por Smart (4), quien, preocupado por el hecho de que los estados mentales parecían tener propiedades tanto físicas como irreductiblemente mentales, propone para ellos un análisis conceptual “temáticamente neutral”, cuyo objetivo era formular descripciones de los estados mentales que no se comprometieran ni con la física ni con la mente. Esta idea promovió una línea investigativa en la que el lenguaje utilizado para describir los estados mentales carecía de un compromiso ontológico, lo cual, como veremos, es una de las características más importantes del funcionalismo.

La otra fuente se remonta a 1960, cuando Putnam (5) propuso comparar los estados mentales de una persona con los estados de una tabla de una máquina de Turing, valiéndose de la noción de equivalencia funcional. En términos muy generales, una máquina de Turing puede describirse como un dispositivo abstracto para resolver funciones matemáticas de forma totalmente mecánica, en virtud de una serie de reglas determinadas que especifican el modo en que la máquina ha de comportarse según sus datos de entrada, sus datos de salida y otros estados en que pudiese encontrarse. Al conjunto de reglas que gobiernan las operaciones de la máquina se le llama *tabla* de la máquina. Así, Putnam propuso entender a los seres con estados mentales como si fueran funcionalmente isomórficos o equivalentes a una determinada tabla de máquina de Turing, de forma que pudiera darse una correspondencia uno-a-uno entre sus estados mentales y los estados de dicha tabla. La posibilidad de esta equivalencia funcional lo llevó, en consecuencia, a suponer que los seres con estados mentales podían considerarse *autómatas probabilísticos* (aunque, estrictamente, la versión de máquina que Turing propuso era determinística), en el sentido en que la tabla de máquina que gobernaría el comportamiento global del sistema estaría definida según unas ciertas probabilidades de emitir tal o cual

output (o dato de salida), y ciertas otras probabilidades de pasar a tal o cual estado subsiguiente, en virtud de un específico *input* (o dato de entrada).

Aunque este modo de pensar puede ser similar al conductismo, dado que nuestros criterios para identificar si un organismo está o no, digamos, en estado de ‘dolor’ siguen siendo comportamentales, tiene una importante diferencia con aquél, al sostener que una caracterización adecuada de los estados mentales involucra una ineliminable referencia a otros estados mentales internos con los cuales está relacionado. Además, se opone también al reduccionismo fisicalista de la teoría de identidad de tipos, al argumentar que no es en virtud de una propiedad física, sino más bien de una propiedad funcional, esto es, un determinado rol causal dentro de la organización funcional del organismo (6) hace que un específico estado mental sea de tal o cual tipo. De ahí que Putnam afirme que los predicados psicológicos son independientes de los predicados físicos, puesto que no es posible llegar a una identificación entre los unos y los otros:

...si podemos encontrar aunque sea sólo un predicado psicológico que pueda claramente ser aplicado tanto a un mamífero como a un pulpo (digamos “hambriento”) pero cuyos correlatos físico-químicos sean diferentes en los

dos casos, la teoría del estado-cerebro [reduccionismo fisicalista de la teoría de identidad de tipos] ha colapsado. (6)

Es en este sentido –en el que debe entenderse la independencia ontológica del funcionalismo–, lo importante para ‘lo mental’ no es la materia con que una criatura está hecha, sino la estructura de las actividades internas, es decir, el conjunto de operaciones o computaciones que dicha materia soporta. También debe quedar claro que el hecho de que el funcionalismo se oponga a la idea de reducir la mente a la física no significa que sea incompatible con ella, así como tampoco resultaría incompatible con una teoría dualista, dado que podría ser posible que un sistema con cuerpo y alma fuera un autómata probabilístico (6). Lo que busca, en cambio, al caracterizar los estados mentales como funcionales, es que los intereses de la psicología se establezcan en una forma que abstraiga la actividad mental de su constitución física. Así, veamos qué tendría que decir el funcionalismo con respecto a cada una de las tesis que arrojó el análisis de las secciones anteriores.

En primer lugar, puede considerarse que, en un sentido, coincide con la tesis de la teoría de identidad de instancias (enunciado 9), al aceptar que, en el caso de los seres humanos, para cada ejemplar

particular de estado mental debe haber un estado cerebral que le corresponda. No obstante, se distancia un poco de ésta cuando afirma su independencia ontológica con respecto a un tipo particular de material que insta a determinada función, dado que es con su descripción funcional –y no con el tipo de sustancia que la soporta– con lo que ha de identificarse un estado mental. En segundo lugar, se opone radicalmente a la tesis del materialismo eliminativo (enunciado 8), al procurar establecer un nivel de descripción (el funcional), que no tiene por qué coincidir con el de descripción neurológico o físico, dado que las categorías que se establecen en el primero están lejos –al parecer– de coincidir con las que se establecen en el segundo. Además, como veremos más adelante, el funcionalismo podría incluso convertirse en una estrategia teórica para preservar el lenguaje de la psicología popular, lejos de eliminarlo.

Por otra parte, debe resultar también evidente que el funcionalismo se opone a la tesis de la identidad de tipos (enunciado 7), al sostener que no es en virtud de una propiedad física, sino más bien de una funcional, que los tipos de estados mentales se correlacionan con tipos de estados físicos, que, dada su noción de independencia ontológica, pueden o no ser cerebrales. Similarmente, se opondría también a la tesis del conductismo

(enunciado 6), al sostener que, además de los estímulos y las respuestas, debe haber una ineludible referencia a estados internos del organismo en la caracterización de un estado mental. Consecuentemente, tampoco podría aliarse con la tesis del positivismo fisicalista (enunciado 5), porque sostiene que un enunciado psicológico podría ser significativo, aun cuando no pudiese ser reducido al lenguaje de la física. Así tampoco tomaría partido por la tesis del monismo radical (enunciado 4), ya que al no comprometerse con un tipo particular de ontología, establece un nivel de explicación para la psicología que nada tiene que ver con si se es monista o dualista. Razones de más para suponer, finalmente, que el funcionalismo sí aceptaría la tesis de la irreductibilidad de lo mental (enunciado 3), pero por motivos radicalmente distintos a los de Descartes, es decir, no porque existan dos sustancias distintas en el mundo, sino porque las categorías que usamos en el lenguaje de la psicología no pueden reconciliarse con las de la física, cuando sí lo hacen con nuestras caracterizaciones funcionales. Mas, ¿qué ocurre con la tesis de la autonomía de la psicología (enunciado 2) y con la tesis de la introspección (enunciado 1)? Para poder descubrir la posición que el funcionalismo adoptaría respecto de ellas tenemos antes que estudiar algunas de las más relevantes objeciones que se le formularían.

Una de ellas es esgrimida por Block (7), quien al considerar al funcionalismo como una teoría 'intermedia' entre el conductismo y el reduccionismo fisicalista, lo convierte en heredero de importantes problemas de una y otra postura. Él sostiene que, al igual que el conductismo, el funcionalismo apela a criterios comportamentales para caracterizar los fenómenos mentales; pero, a diferencia de éste, interpreta los estados mentales como estados internos y les otorga un papel causal en la producción de la conducta.

Y al apoyar los procesos mentales como procesos internos, el funcionalismo podría estar de acuerdo con la teoría de identidad, aunque diferiría de ella en que no insiste en qué tipo de estados mentales se identifican con estados del cerebro. No obstante, arguye, el funcionalismo, como el conductismo, podría ser acusado de 'liberalidad', al permitir atribuirle estados mentales a seres que parecerían tenerlos pero que en realidad no los tienen. De modo similar, podría también ser acusado de 'chauvinista', como la teoría de identidad de tipos, al no permitir la atribución de estados mentales a organismos que sí los tienen. "A qué problema sucumbirá el Funcionalismo depende, de acuerdo con Block, de la forma de funcionalismo que se adopte"(8).

Por otra parte, Block y Fodor (9) desarrollaron un par de experimen-

tos mentales que buscaban demostrar que los funcionalistas, en su definición relacional de los estados mentales, olvidan el aspecto cualitativo –el carácter *qualia*– de los estados mentales. En el primero de ellos, conocido como *experimento de los qualia invertidos*, nos piden que imaginemos a un organismo cuyos estados funcionales son idénticos a los nuestros, pero que ve invertidos los colores del espectro (por ejemplo, rojo como verde). Consideran que dado que no podemos comparar nuestros estados cualitativos y que todos (nosotros y el individuo imaginado) hacemos las mismas discriminaciones observacionales, no hay forma de mantener que nuestros espectros están invertidos, es decir, que aun cuando nuestros espectros estén invertidos permaneceremos funcionalmente isomórficos. Un segundo experimento mental, esbozado por ellos y luego reformulado individualmente por Block (7), busca que imaginemos la existencia de un ser con los mismos estados funcionales que nosotros, pero sin carácter cualitativo alguno asociado a ellos. En su reformulación, Block imagina a una especie de organismo (o un robot), cuyo procesador es toda la población de China accionando palancas y dispositivos diversos. La tabla de la máquina de este ser –el conjunto de instrucciones que son ejecutadas por los chinos– sería funcionalmente equivalente al de nosotros, no obstante, parece imposible que

los estados funcionales de dicho sistema puedan tener un carácter cualitativo.¹

Finalmente, Block y Fodor (9) se encargan también de demostrar que hay un gran número de rasgos de los fenómenos psicológicos que no pueden manejarse con el funcionalismo, por ejemplo, la distinción entre estados que ocurren efectivamente (sensaciones, pensamientos, sentimientos, etc.) y estados disposicionales (creencias, deseos, inclinaciones, etc.), pues al ser todos estados de una tabla de una máquina de Turing, dicha diferenciación resulta impracticable (9). De hecho, unos años más tarde, Block afirmaría –siguiendo esta misma línea argumentativa– que lo que le da a un sistema una organización funcional en un momento dado no es lo que *hace* en ese momento, sino también los contrafácticos, que son verdaderos de él en ese preciso instante, “lo que *hubiera* hecho (y lo que hubieran sido sus transiciones de estado) de haber tenido un input diferente o de haber estado en un estado diferente” (7).

Por otra parte, ambos filósofos anotan que el funcionalismo es una teoría que puede dar cuenta de comportamientos producto de una serie de estados psicológicos sucesivos, pero que no permitiría que en sus computaciones se presentaran estados intermedios simultáneos, cosa que evidentemente ocurre en la realidad con los seres con mentalidad. Consideran, además, que mientras los estados de una tabla de máquina pueden ser listados, los estados psicológicos, al ser productivos, son potencialmente infinitos y que, a lo máximo, el conjunto de estados mentales de una persona podría ser especificado por axiomatización finita (9). Y, finalmente, piensan que el funcionalismo no puede explicar la noción de constituyente, pues si la mente es sólo una lista de estados, no se vería la similaridad entre quien cree que P y quien cree que P & Q; es decir, no se puede mostrar que P es constituyente de P & Q (9). En este orden de ideas, afirman que “...con los estados psicológicos como con las oraciones, el hecho de que ellos sean productivos y el hecho de que ellos

¹ Este argumento contra el funcionalismo guarda un aire de familia (un fuerte aire, de hecho) con una de las más importantes críticas esbozadas por Searle a la inteligencia artificial: el argumento del Cuarto Chino. Searle (10) nos pide que imaginemos un cuarto en el que hay una persona encerrada que no sabe hablar chino, pero tiene con él una serie de cestos con símbolos en ese idioma y un instructivo para traducirlos al inglés, atendiendo sólo a su forma. La idea es que personas de afuera del cuarto introducen en éste preguntas en chino que, gracias al instructivo, el hombre traduce al inglés y emite respuestas en este último idioma. El punto que busca concluir Searle es que si bien este cuarto se comporta como si entendiera chino, en realidad no lo hace, razón por la cual no tenemos razones suficientes para suponer que algo que tenga un programa (*software*) para, digamos, pensar, piense en realidad.

exhiban estructura interna son dos aspectos del mismo fenómeno” (9).

Para superar estas dificultades proponen una reformulación de la teoría computacional de la mente, planteada por el funcionalismo, basándose igualmente en la analogía entre la mente y la máquina computadora, y que toma a la mente como un *sistema formal automático e interpretado*, es decir, como un sistema en el que símbolos discretos (esto es, representaciones mentales) se manipulan de acuerdo con un conjunto finito de reglas que operan en virtud, no de la semántica de esos símbolos, sino de su sintaxis. Esta nueva versión del funcionalismo ha sido principalmente desarrollada por Fodor, a lo largo de su trayectoria filosófica, aunque es posible encontrar en los trabajos de otros filósofos –como Stephen Stich (11)– versiones de una teoría computacional de la mente un poco distintas.

En términos generales, lo que busca una teoría computacional de la mente (o lo que debería buscar) es proveernos de una suerte de “programa para la mente” (3); de este modo, una vez contemos con la descripción del *software* humano, podríamos encontrar razones suficientes para negar que la tesis de la introspección (enunciado 1) tenga que ser un postulado irrefutable de una teoría psicológica científica. No obstante, en caso de ser cierto que

nuestra mente sea una especie de programa de computador, quizá sí resulten valiosos los avances en neurociencia, en la medida en que mostrarían algo acerca del *hardware* en que corre dicho programa. Por ello, a menos que podamos establecer criterios específicos para determinar qué corresponde al programa de la mente y qué a su *hardware*, no podríamos aceptar la tesis de la autonomía de la psicología (enunciado 2), según la cual ningún conocimiento del cerebro podrá darnos luces acerca del nuestro respecto a la mente. Pero ¿cuál podría ser la naturaleza de un *software* mental? Esto, justamente, es lo que intentaré dilucidar a continuación.

Teorías computacionales de la mente

La noción de algoritmo, y en particular la de máquina de Turing, nos ha provisto de una interesante perspectiva, según la cual una habilidad que se creía inteligente puede simplificarse en una serie de operaciones mecánicas que no requieren inteligencia para ser llevadas a cabo. Y sobre la base de dicha concepción, heredada del funcionalismo, se ha formulado una de las doctrinas más ampliamente aceptadas sobre la naturaleza de nuestro *software* mental: la *teoría computacional de la mente* (TCM en adelante), la cual sostiene, en breve, que todo proceso mental es

computacional y que la mente cuenta con una serie de reglas operativas que conducen la computación. No obstante, a pesar de que la TCM esté basada en la doctrina funcionalista, no está necesariamente comprometida con la tesis de que la mente (esto es, en nuestro caso, el cerebro) sea literalmente una máquina de Turing; la tesis afirma, sencillamente, que algunos marcos computacionales pueden explicar y replicar procesos cognitivos humanos, pero no nos obliga a aceptar ningún tipo particular de estructura (por ejemplo, Turing o Von Neumann) para dichas computaciones (12).

Ahora bien, si concebimos nuestra mente de acuerdo con los parámetros de una TCM, y reconocemos que para ello se necesita que ésta cuente, además, con un sistema de representaciones con las cuales operan las computaciones, tendríamos de un aparato teórico para explicar la manera como manejamos nuestras representaciones mentales durante nuestras actividades cognitivas.² Es decir, contaríamos con una TCM que sostendría que nuestra mente funciona “según un conjunto de reglas básicas que el cerebro efectúa, no en

virtud de que se represente cómo deben ser efectuadas, sino en virtud de su constitución física” (13), y que operarían sobre nuestras representaciones mentales, las cuales, por una parte, tienen una dimensión semántica y, por otra, una dimensión formal –que además es física, para el caso del cerebro humano–.

Los estados y los procesos mentales, así entendidos, serían simbólicos, porque se definirían sobre representaciones, y como su naturaleza es representar algo, éstas habrán de tener un contenido semántico. Asimismo, serían formales, porque apelarían a la forma (posiblemente sintáctica) de las representaciones para determinar el rol funcional que desempeñarían dentro del procesamiento cognitivo. De ahí que para definir las representaciones mentales, teniendo una dimensión formal y una semántica, estemos obligados a aceptar que la única manera como en el ámbito computacional dos instancias de representaciones mentales cuenten como distintas es si formalmente se diferencian. Además, este mismo hecho nos permite evadir el dualismo, en cuanto podemos concebir

² La idea de que nuestra mente funcione de acuerdo con los parámetros de una TCM y, además, que sea por naturaleza un sistema de representaciones parece que proviniera de dos ideas distintas. En principio se podría sostener una teoría representacional de la mente que no sea computacional o una computacional que no cuente con representaciones. No obstante, la exposición que sigo aquí se suscribe a la idea de Fodor, de que una TCM debe suponer la existencia de un sistema de representaciones en la mente para poder funcionar, y pretendo que la solidez de sus argumentos me evite el trabajo de analizar otras alternativas.

dichos objetos sintácticos operándose entre sí según su forma física, en virtud de una serie de reglas predefinidas, y no de algún tipo de 'fantasma de la máquina', que se encargará de operar las representaciones ateniendo a la semántica. En este sentido, en caso de que el enfoque computacional-representacional de la mente resultara cierto, podríamos suscribirnos a la opinión de Sterelny, quien sostiene que:

Una teoría computacional de la cognición junto con un reporte de las relaciones causales entre la mente y el mundo explican cómo podemos tener nosotros mentes representadoras. La descomposición de las capacidades complejas en simples, junto con un reporte de cómo las operaciones simples pueden ser construidas dentro de la organización física de la mente, explican cómo es que tenemos mentes computadoras.(13)

Cabe anotar, en todo caso, que aunque Putnam pensara en un principio que la mente humana podría ser como una máquina de Turing, Block y Fodor (véase sección III) creyeron que éste sólo modelo realmente no reflejaba toda la complejidad de la psicología humana. Sin embargo, resulta evidente que los avances con respecto al procesamiento de información, conseguidos principalmente por los desarrollos de la lógica simbólica y de la teorización de la máquina de Turing, le ofrecen a las TCM un enorme po-

der explicativo dentro de un marco teórico naturalista. Por una parte, porque más allá del modelo algorítmico que Turing propuso, en realidad lo que finalmente demostró es que si un proceso puede ser formalmente especificable, entonces puede ser mecanizado. Y, por otra, porque la lógica simbólica moderna ha mostrado que la mayor parte del razonamiento puede ser formalizado, de forma que si interpretamos el pensamiento racional como transacciones de representaciones, entonces es posible suponer que la mayor parte de las relaciones semánticas entre símbolos puede ser enteramente capturada por las propiedades y relaciones formales-sintácticas entre ellos; entre otras cosas porque, como Turing lo habría sugerido, las transacciones sintácticas entre símbolos se pueden hacer respetando su contenido semántico.

Por ello, desde hace ya varios años, Fodor se ha dado a la tarea de construir una versión simbólica de la teoría computacional de la mente –conocida como la *hipótesis del lenguaje del pensamiento* (LDP desde ahora)–, que busca superar las dificultades señaladas por él para el funcionalismo, dando cuenta de cómo es posible que sistemas físicos como los humanos puedan tener estados mentales que: (a) representen el mundo, (b) tengan poder causal sobre el comportamiento y (c) estén definidos dentro de un marco explicativo que, sin contra-

decir a la neurofisiología, permitan la formulación de leyes y explicaciones en el lenguaje de la psicología. Teniendo en mente estas condiciones, pretendió demostrar que la *única* teoría adecuada para explicar la naturaleza de los estados y operaciones mentales sería aquella que sostuviera la existencia de un medio de representación mental, un sistema representacional interno definido computacionalmente y con las características de un lenguaje, de manera que los pensamientos tendrían la forma de una oración (*sentence-like*). En términos del propio Fodor:

Lo que trataré de demostrar es que, prescindiendo de las propias suposiciones sobre los detalles de las teorías psicológicas de la cognición, su estructura general presupone procesos computacionales y un sistema representacional en que se realizan tales procesos [...] Una forma de describir mi punto de vista es que los organismos (o, en cualquier caso, los organismos que tengan algún tipo de comportamiento) tienen no sólo los lenguajes naturales que puedan tener, sino también un lenguaje privado en que realizan las computaciones que están en la base de sus conductas. (14)

Su teoría busca “naturalizar la psicología popular” sin eliminarla, integrándola, más bien, dentro del complejo marco de la imagen científica. Y para hacerlo empezó por demostrar la necesidad de ese sis-

tema representacional interno para poder dar cuenta de algunos procesos cognitivos (toma racional de decisiones, aprendizaje de conceptos y percepción) que, de otro modo –según él– no podrían explicarse. Así, por ejemplo, nos dice que en la elección de una conducta el organismo debe ser capaz de representarse, primero, una situación (S), luego las opciones conductuales ($B_1 \dots B_n$) y, posteriormente, computar las probabilidades y los valores que él mismo les asigna a las posibles consecuencias ($C_1 \dots C_n$) que se derivan de las conductas que se ha representado. De modo similar, presentó un modelo de aprendizaje de conceptos que resulta ser, básicamente, un proceso de formación y confirmación de hipótesis, en el que “la conducta del organismo dependerá de la relación de confirmación entre los datos y las hipótesis, de forma que las explicaciones de su conducta requieran información sobre la manera en que, en el curso del aprendizaje, se representan los datos y las hipótesis” (14). Y su idea es que la postulación y confirmación de hipótesis implica relaciones con representaciones cuyos contenidos reflejan propiedades de los objetos de la experiencia.

Sin ir más lejos, a lo que Fodor quiere llegar es a la conclusión de que procesos cognitivos, como los que ocurren al tomar una decisión o al aprender un concepto, exigen un medio en el cual las representa-

ciones puedan efectuarse, un sistema representacional interno que, como el lenguaje, presente ciertas condiciones que posibiliten dicho procedimiento. Esto es, Fodor sugiere que para poder realizar estos procesos cognitivos, el agente debe tener la posibilidad de representar el mundo como es y como podría ser, y debe también ser capaz de extraer inferencias apropiadas de esas representaciones; por eso requiere (1) un sistema representacional para formular las hipótesis, (2) una manera de ordenar las hipótesis candidatas para las pruebas, (3) un sistema para representar los datos y (4) un medio para poder aparear datos e hipótesis (13). En conclusión, la cognición necesita un sistema representacional interno, anterior al lenguaje natural (porque aprender un lenguaje es un caso particular de aprendizaje de conceptos), pero que sea como un lenguaje, para poder representar esos muchos e indefinidos estados complejos actuales y posibles de su entorno. Y supone que el mejor medio es el LDP.

¿Cómo debe ser el LDP, este particular sistema representacional interno, para poder mediar procesos como éstos? Por un lado, el LDP debe considerarse una propuesta que surge de la sugerencia de Fodor de que una TCM habrá de suponer la existencia de un sistema representacional en el cual acaezcan las computaciones mentales. El LDP,

justamente, es ese sistema de representaciones mentales, las cuales tienen la forma de oraciones y exhiben, además, dos dimensiones distintas: una semántica (pues al ser representaciones, deben significar algo) y una sintáctica o formal. Por otra parte, la TCM le permite asegurar que esas representaciones o fórmulas en el LDP puedan combinarse entre ellas en virtud de principios computacionales sensibles únicamente a la sintaxis (es decir, a la forma) de los símbolos mentales, y garantizar en todo caso su coherencia semántica, lo cual, como antes había dicho, lo habría pensado Turing.

Por otra parte, considera que el sistema de símbolos que comprenden el LDP tiene importantes paralelos semánticos y sintácticos con el lenguaje. En primer lugar, sostiene que el pensamiento, como el lenguaje, es *productivo*, lo que significa que hay, en principio, infinitas oraciones que uno puede entender, producir o captar, idea ésta que estaría amparada en la división chomskiana entre competencia y desempeño (*performance*) lingüístico. Aquí la productividad del pensamiento debe entenderse en cuanto a capacidad, de forma que uno puede tener una capacidad lingüística infinita aunque un finito desempeño, debido a nuestras humanas limitaciones de tiempo. En segundo lugar, el pensamiento, como el lenguaje, es *sistemático*, de modo que

si puedo pensar aRb, puedo también pensar bRa. Aparentemente, el argumento a favor de la sistematicidad del pensamiento se basa en el hecho –alegadamente empírico– de que la habilidad para captar ciertos pensamientos está intrínsecamente conectada con la habilidad de captar ciertos otros. Y podría incluso llegar a suponerse, por analogía con el lenguaje, que si es una ley que la habilidad para entender una oración está sistemáticamente conectada con la habilidad para entender muchas otras, entonces es una ley que la habilidad para pensar un pensamiento está sistemáticamente conectada a la habilidad para pensar muchos otros (15). Tal requerimiento legaliforme podría exponerse, en términos de Fodor, así: “La sistematicidad requiere que, en general, si un lenguaje puede expresar la proposición P y la proposición P es semánticamente cercana a la proposición Q, entonces el lenguaje puede también expresar la proposición Q” (16).

Finalmente, tanto la sistematicidad como la productividad del pensamiento quedarían garantizadas por el paralelo semántico (y, en cierto modo, también sintáctico) con otra importante característica de los lenguajes: su *composicionalidad*. Desde hace algún tiempo, la vieja idea de que uno aprende el lenguaje como si memorizara puntualmente un diccionario se ha puesto seriamente en duda, pues nos obligaría

a preguntarnos cómo puede ser posible que alguien que no ha oído nunca una oración como “existen ovejas de cuernos cuadrados en el mar de la Tranquilidad” pueda, sin embargo, entenderla –por más absurdo que resulte su significado–. La idea, en contraposición a la vieja doctrina, es que nuestro lenguaje tiene elementos atómicos, que son los que se combinan en las distintas oraciones que, aun cuando no hayamos oído, podemos entender. Quizá si viéramos todas las oraciones del lenguaje como átomos, entonces la sistematicidad en su comprensión sería un misterio; pero si reconocemos una estructura sintáctica constitutiva en éste, la sistematicidad de nuestra capacidad lingüística queda garantizada. Además, siguiendo con los paralelos lingüísticos, el LDP, en virtud de ser composicional, debe tener una sintaxis y una semántica combinatorias, de forma que los pensamientos sistemáticamente conectados también están semánticamente relacionados, por lo cual los pensamientos resultantes se compondrían de los mismos elementos semánticos.

Esto es, el mismo constituyente sintáctico haría la misma contribución semántica en otra configuración en la que apareciera. Y si el sistema representacional interno tiene una sintaxis y una semántica combinatorias, entonces debe haber un conjunto de reglas sintácticas de

formación –que estarían computacionalmente definidas– que gobiernen la construcción de expresiones bien formadas en el sistema. Tales reglas sintácticas de combinación determinan la estructura sintáctica o formal de las representaciones mentales complejas, y resolverían la manera como las operaciones mentales se llevarían a cabo, por el hecho de ser sensibles a su estructura sintáctica (formal). Este carácter combinatorio de los átomos simbólicos del LDP asegura, en consecuencia, la sistematicidad, la productividad y la composicionalidad del pensamiento (15).

En suma, podemos decir que el LDP es un sistema representacional interno que consiste en representaciones, que está físicamente realizado en el cerebro y que tiene una sintaxis y una semántica combinatorias tales que las operaciones hechas sobre las representaciones son causalmente sensibles únicamente a las propiedades sintácticas de éstas. En este sentido, el pensamiento, simplemente, consistiría en operaciones sintácticas definidas sobre tales representaciones, por lo cual no resulta absurdo afirmar que uno piensa en ‘mentales’.

No obstante la fuerza de la teoría, el modelo simbólico –como el propuesto por el LDP– no es la única TCM trabajada en el interior de

los círculos investigativos en ciencia cognitiva. Recientemente ha venido tomando fuerza una postura computacional distinta, en la que las representaciones mentales no tienen forma proposicional o lingüística (esto es, no son *language-like*) y, por lo tanto, no se comportan según los parámetros del LDP, sino que se encuentran distribuidas a lo largo de diversos patrones neuronales. Este sistema teórico se conoce técnicamente con el nombre de *modelos de redes neuronales*, pero el término *conexionismo* se ha popularizado más a la hora de hablar de dicha postura.

Contrario al modelo simbólico, que se basa en la idea de arquitectura computacional propuesta inicialmente por Turing, el modelo conexionista propone un acercamiento a la estructura de la mente a partir de la morfología misma de las neuronas. Esto hace que dicho modelo se acerque más a la realidad biológica y que así explique mejor ciertos procesos mentales, como el rápido acceso a los registros en memoria, la asignación por omisión, la degradación elegante y la generalización espontánea (17). En palabras de uno de sus principales exponentes:

El enfoque de PDP³ sugiere que debemos formular la pregunta directamente acerca del tipo de

³ PDP: procesamiento distribuido en paralelo, una de las más completas variantes de los modelos conexionistas.

computador que el cerebro puede ser, experimentar con computadores 'tipo cerebro', y luego extraer nuestra inspiración de esos sistemas computacionales. En breve, queremos reemplazar la metáfora del computador con la metáfora del cerebro para los sistemas cognitivos. [...En general] los modelos de PDP, como los cerebros, consisten en grandes redes de trabajo de simples unidades de procesamiento, que se comunican a través del paso de mensajes inhibitorios y excitatorios de una a otra [...]. (18)

Cabe anotar, en todo caso, que la idea de construir modelos abstractos para explicar el funcionamiento del cerebro basándose en la actividad neuronal no es nueva. Probablemente el primer intento serio en este campo fue realizado por McCulloch y Pitts (19), en su teoría de redes neuronales formales. En ésta se proponía un sistema expresado en lógica formal proposicional, cuyas reglas reflejaban ciertas propiedades de las neuronas, particularmente la "ley del todo o nada", y su capacidad de excitación e inhibición. Los descubrimientos posteriores en neurociencia y en inteligencia artificial invalidarían varios de los postulados de su teoría, por lo cual su aceptación pronto se vio disminuida (20).

Sin embargo, la nueva perspectiva conexionista parece superar esa brecha entre la parte neurológica y la parte funcional en la explicación

de la mente. La estrategia básica de esta nueva postura es, en términos muy generales, tomar como su unidad fundamental de procesamiento algo cercano a una neurona abstracta e imaginar que las computaciones se efectúan a través de la simple interacción entre esas unidades de procesamiento o nodos. En estos modelos se propone que tales nodos se comunican entre sí, enviando números por medio de las líneas que los conectan, y que ellos representan conexiones inhibitorias o excitatorias (17).

Se asume, por otra parte, que –contrario a los modelos simbólicos– las unidades de los sistemas conexionistas no son significativas, es decir, la información procesada en sus computaciones no es nunca una propiedad de algún nodo particular, sino que se encuentra distribuida a lo largo de todo el patrón de conexión neuronal. Además, dado que las operaciones cognitivas son sumamente rápidas, a pesar de que las neuronas son bastante más lentas que los componentes de los computadores actuales, se propone que las computaciones involucren gran cantidad de procesamientos en paralelo, y no únicamente uno secuencial (de ahí que a este tipo de sistemas se les llame *procesamiento distribuido paralelo*) (20).

Ciertamente, el modelo conexionista es mucho más complejo de lo aquí esbozado, y ahondar en sus

pormenores es una tarea que no me propongo desarrollar. Es suficiente, considero, con haber señalado el hecho de que en el ámbito computacional parece haber una interesante alternativa para el modelo simbólico, sobre la cual aún hoy se realizan sendas investigaciones.

Una última consideración metodológica

Hasta ahora hemos visto, con cierto detalle, algunas de las más influyentes concepciones de la filosofía de la mente, que se remontan hasta hace unos quince años. A mi juicio, esta discusión filosófica contemporánea en torno a la mente y a los estados mentales es muy similar a las viejas discusiones escolásticas que giraban alrededor de la naturaleza del alma. Además del hecho de que ahora contamos con algunos conceptos técnicos distintos a los de antaño, quizá la única diferencia significativa entre las dos épocas radica en que mientras los viejos pensadores se dedicaban a formular explicaciones ‘psicológicas’ que no contradijeran lo dicho en las Sagradas Escrituras, los teóricos de nuestros días se preocupan por no contrariar la ‘biblia’ que escriben los científicos. Posiblemente este afán de coherencia con las doctrinas cristianas fue lo que motivó a Descartes a postular su teoría dualista. Y nuestro posterior afecto hacia las ciencias físicas, influido en parte por

el arraigo del enfoque positivista en la investigación, empezó a conducir a los filósofos, psicólogos y, en general, a los científicos de la mente, hacia la creación de teorías basadas en supuestos ontológicos monistas. Conjuntamente, la inevitable dificultad que se seguiría de aceptar la propuesta dualista de Descartes, al momento de dar cuenta del poder causal de los estados mentales (problema que, como vimos, se ha denominado el *problema mente-cuerpo*), ha servido también para convencernos de que la mejor –si no la única– solución viable es adoptar una teoría de la mente materialista.

El problema mente-cuerpo terminó por convertirse en una especie de paradigma y, como hemos visto, durante mucho tiempo se mantuvo la idea de que la única forma de no caer en el absurdo que representaba sostener algún tipo de mentalismo era considerarse a sí mismo un materialista. Sin embargo –como ya lo había anunciado–, comparto la opinión de Searle (21), quien dice que en la base de esta concepción se oculta otra suerte de dualismo que podría llamarse *conceptual*, un punto de vista que nos invita a creer que ‘físico’ implica ‘no mental’ y ‘mental’ implica ‘no físico’. La tácita adopción de este supuesto llevó a los partidarios de casi todas las clases de reduccionismo a sostener que la aceptación de cualquier tipo de mentalismo *impli-*

caba necesariamente casarse con el dualismo, al igual que la conformidad con el monismo *implicaba necesariamente* la adopción de algún tipo de teoría materialista.

Quizá sea por la adopción del dualismo conceptual que la reiterativa pregunta filosófica original –¿qué distingue a los eventos del mundo físico de los eventos del ‘mundo psíquico’?– haya adquirido hoy la forma de otro interrogante con marcados visos materialistas: ¿por qué seguimos resistiéndonos a creer que algún día podremos construir un marco teórico para los eventos mentales que sea fácilmente ubicable dentro de nuestro amplio marco teórico de la física? Algunos sostienen –Davidson (1), por ejemplo– que nuestros eventos mentales esencialmente se distinguen de los eventos físicos, porque se resisten a caer en la red nomológica de las leyes causales con la que trabaja dicha ciencia. Pero, podríamos inquirir, ¿acaso eso es necesariamente cierto? En un sentido parece serlo, pero en otro no. Un neurólogo podría, en caso de contar con las herramientas y conocimientos suficientes, hacer un rastreo de todos los eventos neurológicos que precedieron a mi elocución “creo que está lloviendo afuera”, e incluiría desde los impulsos sinápticos que, viniendo de los conos de mi retina, recorrieron mi vía óptica hasta llegar al lóbulo occipital, hasta el recorrido de los impulsos subsiguientes

que, tras pasar por el área 44 de Brodmann, arriban a la zona prefrontal, la cual se encarga de regular el movimiento de mis órganos fonarticuladores. Y parece además cierto de suyo que en caso de que mis sinapsis hubieran sido diferentes, yo no habría pronunciado dicha frase, tal vez habría emitido otro sonido o ninguno en lo absoluto. Pero ¿es esto dar cuenta de la causación mental? Si alguien me preguntara por qué dije lo que dije, ¿sería suficiente –para efectos de una explicación psicológica– contestar con una historia como la anterior? Probablemente el sentido común nos invitaría a dar una respuesta diferente: “siento que está haciendo frío afuera” o “creo que oí gotas cayendo sobre el tejado”, etc. Mas este aspecto, evidentemente, no se considera en el rastreo neurológico.

Parece, en consecuencia, que una parte de nuestras explicaciones psicológicas todavía se resiste a la reducción neurológica; pero que, con todo, parece también ineliminable. Tal vez sólo hasta que el funcionalismo apareció en escena la verdad de la presunción del dualismo conceptual se puso en tela de juicio. La noción de *equivalencia funcional*, que trajo consigo dicha teoría, heredada de la metáfora mente-computador, permitió crear un contexto teórico para las doctrinas de la mente que quisieran seguir sosteniendo el mentalismo sin caer en el riesgoso dualismo de sus-

tancias. Sin embargo, la idea de establecer una identidad entre la actividad cognitiva y un cierto tipo de arquitectura algorítmica, que bien puede ser simbólica, conexionista o de otro tipo, resulta también problemática. A este respecto me suscribo a la opinión de Carello, Turvey, Kugler y Shaw, quienes afirman que:

La mera existencia de un algoritmo no constituye la explicación de un fenómeno. [...] Simplemente porque un algoritmo pueda ser escrito para simular una actividad dada de un organismo, no se sigue necesariamente que el organismo use tal algoritmo al efectuar la actividad en cuestión. El algoritmo es meramente una descripción de la actividad; ésta puede ser sólo una de las muchas descripciones alternativas. (22).

Esto no significa, tampoco, que yo esté apoyando el que los modelos y explicaciones funcionales/computacionales deban dejarse de lado. Éstos tienen, a pesar de las limitaciones que muchos autores les han señalado (y que no considero oportuno indicar aquí), una gran importancia. Son –digamos– posibles vínculos entre nuestros términos del lenguaje intencional y los eventos cerebrales. De hecho, el propio Fodor habría reconocido en ellos este valor:

...la relación entre los constructos neurológicos y psicológicos [...] está en perfecta consonancia con

los argumentos que defendimos anteriormente en favor de la no eliminabilidad del lenguaje mental de las teorías de la conducta [...] Tendríamos que acudir al lenguaje mental para establecer las condiciones que permitieran efectuar esas adscripciones de equivalencia funcional. (16)

Finalmente, todo lo que pasa en el mundo es físico, aunque uno puede dar explicaciones de dichos eventos desde distintas esferas. Aquello que en nuestra vida cotidiana agrupamos como creencias o temores pueden ser, descritos físicamente, procesos neuronales que no tienen más característica común que la de ser realizados por (y en) el cerebro y, por lo tanto, las descripciones neurológicas no necesariamente coincidirán con nuestras descripciones psicológicas; no obstante, los datos acerca de nuestro cerebro –de los que podría dar cuenta la neurología– pueden participar (y de hecho lo hacen) en nuestras explicaciones y predicciones psicológicas. Incluso es posible que las mejores explicaciones psicológicas de uno u otro evento, que comúnmente catalogamos como ‘mental’, tengan la estructura de una disposición conductual, sin que eso signifique, en todo caso, que deban identificarse necesariamente como tales.

Algo similar puede decirse entonces con respecto a las teorías computacionales de la mente: todo parece indicar que tampoco tenemos buenas razones para agrupar

los conceptos intencionales según tal o cual descripción funcional, por el hecho de que este tipo de explicaciones no depende de cómo es el mundo, sino de *cómo podemos construir una alternativa viable para explicarlo*. Podemos, para ciertos efectos, utilizar explicaciones funcionales de los eventos cerebrales que se asemejen a la hipótesis del LDP, mientras que en otros casos será mejor usar modelos conexionistas, como el de procesamiento distribuido paralelo (23).

Lo que quiero sugerir, en últimas, es que agrupar una serie de operaciones cerebrales dentro del *cajón* de ‘creencias’ o de ‘deseos’ es algo que hacemos en virtud de que tomamos el lenguaje natural –esto es, las viejas explicaciones de la psicología popular– como *casillero*, y no porque exista realmente un espacio funcional en el que ocurra algo así.

La hipótesis del LDP tiene el gran atractivo de procurarnos explicaciones muy similares en forma y contenido a las explicaciones de nuestro lenguaje común intencional, pero no tenemos más razón para suponer la verdad del lenguaje del pensamiento que el hecho de que su estructura “guarda un aire de familia” con la estructura del lenguaje en el que comúnmente describimos nuestra la psicología popular. Creo, incluso, que la decisión entre adoptar uno u otro mo-

delo funcional depende de nosotros. Siempre queremos que nuestras explicaciones sean las más adecuadas a la verdad y, por eso mismo, si en el ámbito computacional el modelo conexionista funciona mejor que el simbolista, apliquemos la navaja de Occam y salgamos de una buena vez de la carga de una teoría que postula entidades tan problemáticas como las de Fodor. Y si llegara a haber otro mejor, ¡hagámoslo de nuevo!

Con todo, no parece viable adoptar ninguna teoría de las analizadas en este artículo como “estructura metateórica” para la psicología. Y pareciera que la única forma en la cual la psicología puede adquirir el estatuto de ciencia es encontrando los pilotes metateóricos adecuados para poder construir, sobre ellos, las leyes de una disciplina científica. ¿Debemos entonces aceptar que la psicología no tiene, ni tendrá, más valor que el de clasificar bajo un solo término una cantidad de expresiones ‘populares’? Algunos teóricos parecerían sostener algo así, por ejemplo, Daniel Dennett (24), uno de los principales exponentes de la postura *instrumentalista*, quien sostiene un punto de vista en el que los términos que usamos para referirnos a los estados mentales apuntan simplemente a entidades ficticias, que no tienen verdadera existencia, sino que son postuladas para ejercer una labor meramente práctica o heurística. Pero ¿estaría-

mos dispuestos a renunciar a la científicidad de la psicología, arguyendo que sus expresiones no tienen más valor que el meramente práctico?

Quizá la respuesta esté un poco más atrás, en la definición misma de *ciencia*. Y tal vez la dificultad que aparece en nuestros intentos por encuadrar a la psicología dentro de un marco teórico más ‘científico’ se deba al simple hecho de que no parece haber un marco tal que pueda contener los muchos fenómenos que la mente humana carga consigo. El ser humano está definido en gran parte por su vida mental, pero con *vida mental* nos referimos a muchas cosas, más que a su (no tan) simple *vida cerebral*. Tener vida mental implica estar inmerso en una sociedad, haber vivido una serie de experiencias emocionales, tener problemas, ambiciones, desengaños. Implica también haber sufrido ciertas enfermedades, haber leído ciertos libros, saber jugar ciertos juegos, guardar secretos, desarrollar hábitos y hasta alergias. Quizá simplemente ocurra que estudiar la mente humana sea poner al ser humano, como un todo, bajo el microscopio de una *ciencia casi exacta*, y puede ser posible que el objeto de estudio –por lo pronto– venza los mecanismos científicos con que dicha ciencia cuenta.

Bibliografía

1. Davidson D. Mental events. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 107-19.
2. McGinn C. Anomalous monism and kripke's cartesian intuitions. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 156-8.
3. Block N. Introduction: what is functionalism. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 171-84.
4. Smart JJC. Sensations and brain processes. In: Chappell V C, editor. The philosophy of mind. Englewood: Prentice Hall; 1962.
5. Putnam H. Minds and machines. In: Hook S, editor. Dimensions of mind. New York: New York University Press 1960.
6. Putnam H. The nature of mental states. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 223-31.
7. Block N. Troubles with functionalism. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 268-305.
8. Bechtel W. Filosofía de la mente: una panorámica para la ciencia cognitiva. Madrid: Tecnos; 1991.
9. Block N, Fodor J. What psychological states are not. In: Block N, editor. Readings in philosophy of psychology. Cambridge: Harvard University Press; 1980. p. 237-50.
10. Searle JR. Minds, brains and programs. In: The behavioral and brain sciences; 1980.
11. Stich S. From folk psychology to cognitive science: the case against belief. Cambridge: MIT Press; 1983.
12. Chalmers DJ. A computational foundation for the study of cognition. 1995. Available from: URL: <http://www.u.arizona.edu/~chalmers/papers/computation.html>.

13. Sterelny K. The representational theory of mind: an introduction. Cambridge: Blackwell; 1990.
14. Fodor JA. El lenguaje del pensamiento. Madrid: Alianza; 1984.
15. Aydede M. The language of thought hypothesis. In: The Stanford encyclopedia of philosophy. 1988. Available from: URL: <http://www.plato.stanford.edu/entries/language-thought/>
16. Fodor JA, Lepore E. Why meaning (probably) isn't conceptual role. In: Stich S, editor. Mental representation. Cambridge: Blackwell; 1994.
17. Rumelhart DE, McClelland JL, The PDP Research Group. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT Press; 1986.
18. Rumelhart DE. Towards a microstructural account of human reasoning. In: Davis S, editor. Connectionism: theory and practice. New York: Oxford University Press; 1992.
19. McCulloch WS, Pitts WH. A logical calculus of the ideas immanent in nervous activity. In: McCulloch WS. Embodiments of mind. Cambridge: MIT Press; 1965.
20. Montañés P, De Brigard F. Neuropsicología clínica y cognoscitiva. Bogotá: Universidad Nacional de Colombia; 2001.
21. Searle JR. El redescubrimiento de la mente. Barcelona: Crítica; 1996.
22. Carello C, Turvey MT, Kugler N, Shaw RE. Inadequacies of the computer metaphor. In: Gazzaniga MS, editor. Handbook of neuroscience. New York: Plenum Press; 1984.
23. Ramos J. Simbolismo vs. conexionismo: la estructura de las representaciones. In: Botero JJ, Ramos J, Rosas A. Mentes reales: la ciencia cognitiva y la naturalización de la mente. Bogotá: Siglo del Hombre, Universidad Nacional de Colombia; 2000.
24. Dennett DC. La actitud intencional. Barcelona: Gedisa; 1988.