

Funcionalismo, reduccionismo y antirreduccionismo en la filosofía de la mente (I)

Soraya Vega

Profesora

Departamento de Humanidades y Letras
Universidad Central

Introducción

El funcionalismo en la filosofía de la mente pretende dar explicación al complejo mental humano dentro de lo que se ha caracterizado como Ciencia Cognitiva. Las diferentes significaciones del término “función” que han sido comprendidas en la psicología, la fisiología, las ciencias físicas y matemáticas y las ciencias sociales han jugado parte en la dinámica de la adopción de esta perspectiva de explicación en el campo de lo mental. Una primera acepción del término se refiere a *ejecución* en tanto actividad de un “algo” que cumple con un cierto fin, de modo que toda parte de un todo (organismo biológico, sociedad, psiquis individual, etc.), se puede decir que tiene una función en cuanto tal, que permite la funcionalidad del todo. Lo que este “algo” sea, su naturaleza, no se considera de importancia para el funcionalismo, sino el cómo funciona, cómo actúa. Una segunda acepción del término se refiere a *relación*, como concepción de orden proveniente de las ciencias físicas y matemáticas, que no entra en contradicción con la primera acepción funcional, y en la cual se considera fundamental la relación de las diferentes partes para la eficiencia del todo.

Estas consideraciones muy generales del punto de vista teórico del funcionalismo, se

concretan en el dominio de lo mental en la problemática mente-cuerpo, de modo que la concepción de la independencia funcional (ejecución, relación) de la cuestión estructural (naturaleza), propende por la autonomía de la psicología que separa las descripciones y explicaciones mentales de las físicas o fisiológicas. En este mismo sentido, el funcionalismo niega su compatibilidad con la teoría de la identidad en tanto Tipo, es decir, por ejemplo, con la tesis fisicalista de que “los estados mentales son estados cerebrales”, considerando tal teoría como un reduccionismo. El funcionalismo no pretende ser, con ello, una teoría ontológica, en tanto no responde al qué son los estados mentales.

Todas las versiones del funcionalismo sostienen que los estados mentales se designan en términos de sus relaciones causales entre ellos mismos (relaciones internas), los inputs y outputs. El cómo se especifiquen estas relaciones, así como los inputs, outputs, y estados mentales, es la respuesta que diferencia a las diversas versiones del funcionalismo.

Teniendo esto en cuenta, también se ubicarán las diferentes versiones del funcionalismo de acuerdo con la división hecha por N. Block entre las que se explican en términos de una máquina de Turing o funcionalismo de máquinas, que Block llama

psicofuncionalismo, y las que no, que llama Funcionalismo (con F mayúscula). Dentro del psicofuncionalismo se describirán dos versiones: la primera versión del funcionalismo desarrollada por Putnam o teoría de la identidad de estado funcional (FSIT), y el funcionalismo computacional o de inteligencia artificial propuesto por Fodor y Block.

En la primera parte de este trabajo se expondrá en qué consiste cada una de las versiones del funcionalismo y qué tipo de objeciones se han argumentado en contra de tales versiones.

En la segunda parte se expondrán las principales polémicas que han entablado reduccionistas (teorías de la identidad de Tipo) y antirreduccionistas entre los que se cuentan los funcionalistas y algunas versiones del fisicalismo. Lo que está allí en juego es la autonomía de lo mental frente a un reduccionismo simplista.

PRIMERA PARTE: EL FUNCIONALISMO

1. Psicofuncionalismo

Los representantes del psicofuncionalismo en sus diferentes versiones son H. Putnam, Fodor, Harman, Dennett y Block.

El psicofuncionalismo considera las identidades funcionales como parte de una psicología empírica, identificando el estado mental *S* con el correlato funcional Ramsey de *S* con respecto a una teoría psicológica científica, que considera los análisis funcionales como hipótesis científicas sustantivas, de un modo que las relaciones causales entre los estados mentales, los inputs y outputs, sean plausibles como comprobaciones empíricas de tales hipótesis.

Si bien tanto Funcionalistas como Psicofuncionalistas insisten en la especificación física, no mental, de inputs y outputs, los Funcionalistas requieren de una especificación siempre externa al agente, mien-

tras que los psicofuncionalistas tienen la posibilidad de especificar inputs y outputs en términos de parámetros internos tales como señales en las neuronas (o impulsos neuronales) como input o output.

1.1. Funcionalismo de Putnam (FSIT)

Putnam toma como punto de partida la crítica a la teoría de la identidad como Tipo y a su análisis reductivo, para plantear la teoría funcionalista.

Su hipótesis argumentativa plantea que “el dolor o el estado de dolor es un estado funcional de un organismo completo”; tal hipótesis funcional se apoya en ciertas nociones técnicas, a saber: la máquina de Turing, el autómata probabilístico y la noción de descripción de un sistema; nociones que a su vez permiten especificar la hipótesis en cuatro numerales relacionados con ellas.

Una máquina de Turing se presenta como modelo de un organismo o sistema; ésta consiste en: una cinta de longitud potencialmente infinita en la que existe una secuencia de cuadrados, cada uno de los cuales contiene o un cero (0) o un uno (1), una unidad de ejecución que puede estar en uno de los estados internos finitos (por ejemplo, *s1* y *s2*) de la unidad. Las actividades de dicha unidad se determinan por un grupo de reglas finitas que se especifican en la tabla de máquina y que son de la forma: si la unidad está en el estado *s1* y el input es 0 emite el output 1 y pasa al estado *s2*. Esta definición de una máquina de Turing especifica un autómata determinista.

Un autómata probabilístico se define similarmente, sólo que las transiciones entre los estados y los outputs dados como respuestas permiten probabilidades que toman valores entre 0 y 1, y que dependen de las iniciales condiciones del conjunto

de inputs, el estado interno, y las reglas de acción del autómata.

Una descripción de un sistema S , es cualquiera en la que el sistema posea diferentes estados S_1, S_2, \dots, S_n , los cuales están relacionados entre sí, con los outputs motores, y los inputs sensoriales, mediante las transiciones de probabilidad dadas en las reglas de una tabla de máquina. Esta tabla de máquina es la organización funcional del sistema S , cuyo conocimiento no envuelve el conocimiento de las realizaciones físicas de S .

Así, Putnam especifica su hipótesis de acuerdo con las anteriores nociones en los siguientes cuatro numerales:

- 1) Todo organismo capaz de sentir dolor es un autómata probabilístico.
- 2) Todo organismo capaz de sentir dolor posee al menos una descripción de un cierto género (una organización funcional).
- 3) Ningún organismo capaz de sentir dolor posee una descomposición en partes, las cuales separadamente poseen descripciones del género referido en 2).
- 4) Para toda descripción del género referido en 2), existe un subgrupo de inputs sensoriales tales que un organismo con dicha descripción se encuentra en estado de dolor cuando y sólo cuando algunos de sus inputs sensoriales están en el subgrupo. Para este funcionalismo los estados mentales son equivalentes a estados funcionales de la tabla de máquina que representa la organización funcional del sistema mental.

1.1.1. Objeciones al funcionalismo de Putnam

El mismo Putnam considera las hipótesis anteriores contentivas de una cierta vaguedad, que puede expresarse especialmente en dos cuestiones. La

primera concierne a qué debemos conocer del género de organización funcional que deba tener un organismo capaz de sentir dolor; a este respecto se requiere adicionalmente de tal descripción que sea una función de preferencia, que sea algo que parezca una lógica inductiva, es decir, una máquina que aprenda de la experiencia y que la máquina posea órganos sensoriales que den la señal de daños al cuerpo de la máquina, peligros que transmiten un subgrupo de inputs (aquellos referidos en 4)). La segunda cuestión se refiere a qué rasgos distinguen el subgrupo de inputs sensoriales referidos en 4). Esta cuestión requiere que los inputs en el subgrupo distinguido tienen un alto disvalor sobre la función de preferencia de la máquina.

Las objeciones que presentan Block y Fodor a este funcionalismo tocan más directamente a su constitución. El funcionalismo de Putnam es caracterizado como una teoría de tipos de identidad de estado funcional (FSIT); en este sentido afirma que cada tipo de estado psicológico es idéntico a un (tipo de) estado de la tabla de máquina, la cual ofrece la descripción de un organismo como modelo de sus operaciones mentales. Entonces, el funcionalismo afirma dos aspectos más que la simple afirmación de los organismos que satisfacen predicados psicológicos, se describen mediante un autómata probabilístico: 1. Existe una descripción única mejor (función de preferencia) de un organismo, tal que; 2. Cada estado psicológico es de un tipo idéntico a un estado, relativo a tal descripción de la tabla de estados del autómata Fodor y Block dan un conjunto de seis argumentos que presentan dificultades a la explicación funcionalista, y que son decisivos en contra de ésta, en cuanto a la

Puesto que al psicofuncionalismo pertenece una psicología empírica, los sistemas funcionales son una realización de la psicología humana. En efecto, la psicología cognitiva no explica los qualia, pero tampoco los explica la psicología de la sensación, como fisiología, psicofísica o conjunto heterogéneo de estudios descriptivos.

posibilidad de poner en correspondencia (de identidad) los estados psicológicos con los estados de una tabla de máquina. Tales argumentos son: 1. La FSIT no tiene una forma plausible de distinguir entre estados disposicionales (creencias, deseos, inclinaciones) y estados ocurrentes (sensaciones, pensamientos, sentimientos), sin abandonar su principio fundamental de correspondencia. En efecto, si los estados de una tabla de máquina son del género de estados ocurrentes, la respuesta de que o bien para cada estado disposicional hay una tabla de estados tal que un organismo está en tal estado disposicional si y sólo si su tabla de estados contiene el estado ocurrente de dicho estado disposicional, o bien las condiciones necesarias y suficientes para tener estados disposicionales pueden estar dadas en términos de alguna propiedad abstracta de la organización de la tabla de máquina, entonces a un estado disposicional no le corresponde nunca un estado en la tabla de máquina. 2. El funcionalismo afirma que la conducta es producto de interacciones entre estados psicológicos. Pero es muy probable que existan empíricamente dos géneros de interacciones conductualmente eficaces de las cuales la FSIT provee el modelo sólo para una de ellas. El primer género (explicado por la FSIT) se refiere a la conducta como producto de una *serie* de estados psicológicos ocurrentes como *interacciones*

secuenciales. El segundo género (no explicado por la FSIT) en que la conducta puede ser el resultado de *interacciones simultáneas* entre estados mentales, cuyo modelo sería el de un grupo de autómatas intercomunicándose en *paralelo*. Esto significa que la mejor descripción no puede ser la de un autómata singular con sus estados correspondientes. 3. Si en lugar de una correspondencia biunívoca entre estados mentales y estados de una tabla de máquina, se comprende una equivalencia funcional, de modo que los estados de una tabla de máquina son idénticos si y sólo si ellos están relacionados idénticamente con otros estados de una tabla de máquina y sus mecanismos de estados de entradas (inputs) y salidas (outputs); los estados psicológicos son pues de idéntico tipo si comparten aquellas propiedades que deben ser especificadas para individuar un estado de una tabla de máquina. Este criterio de tipos de identidad no tiene en cuenta aquello que caracteriza un tipo de estado psicológico, es decir su carácter “cualitativo”; de manera que no podría discernir entre un estado y otro de cualidad invertida. Los funcionalistas presentan dos defensas ante este contraargumento: la primera, es más bien una hipótesis, y es que cualquier diferencia de cualidad introduce también una diferencia de organización funcional; la segunda, que las diferencias entre propiedades cualita-

tivas no determinan diferencias en la organización funcional y pueden por ello ser ignoradas. Esto último sigue siendo problemático, pues resultaría indiferente que un organismo esté en estado de dolor, o mejor, que el dolor sentido resulte distinto en diferentes organismos, o todavía más que un organismo esté en estado de dolor aunque no sienta dolor. Lo que esta consideración muestra es que los estados psicológicos no pueden ser definidos funcionalmente en todo, no pueden ser puestos en correspondencia con cualquier propiedad definible sobre autómatas abstractos. 4. Según el FSIT, dos estados de dos autómatas deterministas son distintos si ellos difieren o en sus salidas asociadas o en su estado sucesor asociado. Análogamente, dos estados de dos autómatas probabilistas difieren si ellos difieren en su rango de salidas asociado o en su rango de estados sucesores asociado, o en las distribuciones de probabilidad asociado a cada uno de esos rangos. Para el caso de un autómata determinista, x y y son tipos distintos de estados de una tabla de máquina si los estados sucesores inmediatos de x y y son de distinto tipo. Pero, los estados sucesores inmediatos de x y y son de distinto tipo si sus estados sucesores inmediatos son de distinto tipo, y así sucesivamente; entonces si se asume el paso computacional de un estado a otro se concluye algo absurdo, y es que dos autómatas que tienen algunos de sus estados en común no podrán tener ninguno de sus estados en común. Los criterios de identificación se muestran como insuficientemente abstractos, pues no pueden ser usados como tipos de identidad de estados psicológicos. 5. El grupo de estados de una tabla de máquina pueden ser especificados por definición por una *lista* de ellos, mientras que los estados menta-

les son empíricamente *productivos*, es decir, especificados por una axiomatización finita, a partir de la cual puedan ser ellos generados. En contra de esto se podría argumentar que las leyes psicológicas limitan el conjunto de proposiciones creíbles a un conjunto finito, pero hay situaciones aritméticas que son imposibles de creer, y que son consecuencia del carácter de la memoria de la gente, y no consecuencia de su representación mental de la aritmética. Ello muestra que los estados psicológicos de un organismo no pueden ponerse en correspondencia con los estados de un autómata. 6. Si el criterio de identidad de la FSIT puede ser efectuado, necesariamente tiene que fallar en la representación de propiedades esenciales de los estados psicológicos. Es clara la existencia de similitudes estructurales entre algunos estados psicológicos, pero, representando los estados psicológicos como una lista, tal modelo no tiene recursos conceptuales para especificar similitudes estructurales, tales como la noción de “es un constituyente de ..”. Pues resulta que las similitudes estructurales se explican de la misma manera que se explica la productividad, es decir, como generados, y no como una lista.

De otra parte, es importante aclarar que la afirmación de la FSIT de que los organismos tienen una única mejor descripción como autómata probabilista, no es obviamente cierta ni implica condiciones sobre el tipo de identidad de estados psicológicos.

1.2. Funcionalismo computacional o de inteligencia artificial

Precisamente a raíz de las dificultades que enfrenta el funcionalismo de Putnam como teoría de tipos de identidad, y que han expuesto Fodor y Block en lo tocante

a su criterio de identidad de estados mentales que pone en correspondencia tipos de estados psicológicos con los estados de una tabla de máquina en su mejor descripción, estos autores han propuesto un funcionalismo con un criterio de identidad diferente para estados mentales.

Fodor y Block proponen una condición de identidad entre estados mentales y estados computacionales de un autómata probabilista, entendiéndose por estado computacional cualquier estado de la máquina caracterizado en términos de sus inputs, outputs y/o los otros estados de la tabla de máquina. Es condición para esta correspondencia que los estados psicológicos sean computables.

Desde esta perspectiva ya no se conciben tipos de identidad, pues un estado mental no se concibe como un género (tipo psicológico), sino como un estado que en sí mismo es *relacional*, o mejor, *operacional*.

Este punto de vista resulta compatible con la productividad de estados mentales de la que se habló en el numeral anterior, pues la cognición se concibe como una actividad en la que se procesan símbolos o se llevan a cabo operaciones sobre símbolos codificados en ella. Esto funciona precisamente como un sistema formal o de axiomatización finita, en el que los símbolos se operan de acuerdo con un conjunto finito de reglas.

Así que sólo puede establecerse una equivalencia entre sistemas si tienen a la base el mismo conjunto de reglas y representaciones, con lo que éstas deben precisarse. Ejecutar la misma operación es seguir un mismo algoritmo, es decir, una misma secuencia de pasos, en los que cada cual es una operación primitiva previamente especificada.

La *inteligencia artificial* considera que el objetivo de este funcionalismo es diseñar máquinas que realicen funciones cognitivas, sin importar si las realizan del mismo modo que los humanos, mientras que la *simulación cognitiva* se interesa en que lo hagan de la misma manera.

2. Funcionalismo

El funcionalismo ha sido desarrollado por Lewis, Armstrong, Smart, Shoemaker, y es según Block, el heredero del conductismo lógico.

El funcionalismo considera las identidades funcionales como parte de una psicología a priori que es, por lo demás, una psicología del sentido común, de modo que el funcionalismo identifica, en términos de la oración Ramsey de una teoría psicológica, el estado mental S con el correlato funcional Ramsey de S, con respecto a una teoría psicológica del sentido común; tal teoría nos da las significaciones de los términos mentales o actitudes proposicionales, que especifican las relaciones causales que determinan la conducta, es decir, relaciones causales entre estados mentales, estímulos o inputs y respuestas motrices o outputs que se presentan como generalizaciones de la forma: “quien quiera que esté en el estado w y reciba el input x emite el output y, y pasa al estado z.”. Así, los análisis funcionales se consideran como análisis de los significados de los términos mentales.

Los funcionalistas requieren de especificaciones de los inputs y outputs que sean externamente observables.

Esta clase de funcionalismo no se presenta como una posición crítica a la teoría de la identidad como tipo, y en este sentido va dirigida la crítica de Lewis a Putnam, que se verá más adelante.

3. Objeciones al funcionalismo en general: liberalismo y chauvinismo

La principal crítica al funcionalismo, se inscribe en la crítica general hecha por N. Block en la que muestra una sin salida del funcionalismo, en sus diferentes versiones, de los puntos de vista liberal y chauvinista en la filosofía de la mente, puesto que, o bien atribuye estados mentales a sistemas a los que no deberían atribuirse, o bien niega que ciertos sistemas tengan estados mentales, cuando en verdad deberían adjudicárseles.

Block sostiene que el funcionalismo en general es liberal y en este sentido se le acusa de igual manera que el conductismo filosófico, en tanto este último atribuye estados mentales a todo sistema que tenga conductas o disposiciones conductuales apropiadas. Sin embargo el funcionalismo puede ser modificado de modo que evite el liberalismo, pero entonces cae en el chauvinismo, del cual ha sido acusado también el fisicalismo, pues éste, basado en la identidad de estados mentales con estados cerebrales excluye a criaturas sin cerebro que no obstante tienen mente.

¿De qué manera el funcionalismo cae en el liberalismo? Block presenta ciertos dispositivos ante los cuales todas las versiones del funcionalismo caen en el liberalismo. Tales dispositivos son simulaciones funcionales humanas, basadas en la idea de que las funciones cognitivas son realizadas por un homunculus (un hombrechillo) que tenemos en la cabeza, de modo que sus actividades mecánicas en el análisis funcional pueden ser emuladas finalmente por una máquina que se constituye en un sistema funcionalmente equivalente a nosotros, y descrito también por la máquina de Turing. En ello se basa la construcción mental de un robot de cabeza homuncular funcionalmente

equivalente a nosotros pero físicamente diferente y el sistema Chino en el que cada ciudadano ejecuta una cuádrupla de la tabla de máquina siendo también funcionalmente equivalente a nosotros.

Pues bien, las relaciones de equivalencia funcional son relativas a la especificación de los inputs y outputs independientemente de si es una descripción funcionalista o psicofuncionalista; luego, para una descripción dada tendremos una especificación de inputs y outputs en tales dispositivos de manera que se cumpla la equivalencia funcional. Con ello se generaliza el argumento a todo el funcionalismo.

Según la equivalencia funcional tendríamos que atribuir mentalidad a tales dispositivos con lo que el funcionalismo cae en el liberalismo, puesto que existe una duda prima facie, de que esos dispositivos tengan estados mentales cualitativos, es decir el funcionalismo concibe sistemas con qualia ausentes a los que atribuye mentalidad. Puesto que las condiciones de equivalencia funcional del funcionalismo especifican inputs y outputs con respecto a una psicología a priori, resulta demasiado liberal, ya que tiene que sostener que tales sistemas equivalentes tienen estados mentales y cualitativos, o sea que experimentan el mismo qualia que un humano —v. gr. sienten realmente dolor—, con lo cual el funcionalismo llega a una conclusión absurda a la que no ofrece una razón para creer en ella.

El psicofuncionalismo evita ser demasiado liberal en su equivalencia funcional, al ser ella una condición necesaria pero no suficiente de atribución de estados mentales. Ya que este funcionalismo no puede dar cuenta de la naturaleza de los qualia, lo que puede ser más obvio en la condición de los qualia

invertidos, en los que nuestra experiencia afectiva puede alterarse o desaparecer sin cambio alguno en nuestra organización funcional, lo cual muestra que las relaciones causales del análisis funcional no captan dichos qualia. La simulación funcional comparte con nosotros nuestros estados mentales no-cualitativos, pero no comparte los cualitativos.

Esto significa que, puesto que al psicofuncionalismo pertenece una psicología empírica, los sistemas funcionales son una realización de la psicología humana. En efecto, la psicología cognitiva no explica los qualia, pero tampoco los explica la psicología de la sensación, como fisiología, psicofísica o conjunto heterogéneo de estudios descriptivos, de los que sólo la psicofísica podría decirse que se ocupa de los qualia, pero ésta sólo se ocupa del aspecto funcional de la sensación. De todo ello surge la duda de que los qualia no pertenecen al dominio de la psicología. Por otra parte, se considera un estado mental como un estado compuesto de un qualia y un estado funcional, o como un quale que desempeña un rol funcional.

¿De qué manera cae aquí el psicofuncionalismo en un chauvinismo? Resulta que los sistemas de simulación funcional han de caer en el dominio de las teorías de la psicología humana a las que se refiere el psicofuncionalismo, luego las teorías psicológicas que son verdaderas para nosotros deben ser verdaderas para ellos.

Pero no hay una razón suficiente para creer que la condición de lo mental sea la equivalencia funcional con nosotros puesto que se puede concebir la existencia de diversidad de procesos psicológicos funcionalmente diferentes, de los que indicamos sólo un tipo; pensemos por ejemplo en marcianos cuyo comportamiento es diferente, no podemos suponer con base en ello que no piensan, ni desean, ni creen. Así, el psicofuncionalismo cae en el chauvinismo del cual ha acusado a la teoría de la identidad como Tipo.



Bibliografía

- BLOCK, N. Troubles with Functionalism, in: Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980.
- BLOCK, N. and FODOR, J. What Psychological States are Not, in: Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980.
- CHURCHLAND, P. S. Reductionism and Antireductionism in Functionalist Theories of Mind, in: Churchland P. S. *Neurophilosophy*, M.I.T., Cambridge, 1989.
- DAVIDSON, Mental Events, in: Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980. And in: Rabossi, E. (Comp.) *Filosofía de la Mente y Ciencia Cognitiva*. Ed. Paidós, Barcelona, 1995.

- FODOR, J. A. Making Mind Matter More, in: Fodor, J., *Philosophical Topics*, Vol 17, 1989.
- KIM, J. The Myth of Non-Reductive Materialism, in: Kim, J., *Supervenience and Mind*, C.U.P. (Cambridge), 1993
- KIM, J. Physicalism and the Multiple Realizability of Mental States, Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980.
- LEWIS, D. Review of Putnam, in: Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980.
- PUTNAM, H. The Nature of Mental States, in: Block, N. (Ed.) *Readings in the Philosophy of Psychology*. 2 Vol. – Harvard U. Press, 1980.