

## BASE EMPIRICA Y TEORIA FUNCIONALISTA EN LAS CIENCIAS COGNITIVAS

Pascual F. Martínez-Freire  
Universidad de Málaga

### Resumen

Después de destacar el carácter interdisciplinar de las ciencias cognitivas, se considera el concepto de sistema de procesamiento de información como el supuesto básico común a las ciencias cognitivas. Defiendo que el funcionalismo constituye una hipótesis filosófica acerca de la naturaleza de los procesos mentales, aneja de manera natural a la noción de sistema de procesamiento de información. Para ello trazo una breve historia del funcionalismo señalando las confusiones introducidas por Ned Block así como el abandono del funcionalismo por Putnam. Finalmente me refiero a los tres niveles de descripción en las ciencias cognitivas, que nos llevan desde teorías funcionalistas a bases empíricas.

*Palabras clave:* funcionalismo, ciencias cognitivas, sistema de procesamiento de información, Hilary Putnam, Jerry Fodor, Ned Block, David Marr, Zenon Pylyshyn.

### Abstract

After emphasizing the interdisciplinary characteristic of cognitive sciences, the concept of information processing system is considered as the basic assumption common to cognitive sciences. It is my claim that functionalism is a philosophical hypothesis about the nature of mental states, attached in a natural way to the notion of information processing system. Hence I sketch a brief history of functionalism, pointing out the misunderstandings introduced by Ned Block and also the abandonment of functionalism by Putnam. Lastly I refer to the three levels of description in cognitive sciences, that lead us from functionalist theories to empirical bases.

*Key words:* functionalism, cognitive sciences, information processing system, Hilary Putnam, Jerry Fodor, Ned Block, David Marr, Zenon Pylyshyn.

### 1. Las ciencias cognitivas como campo interdisciplinar

Aunque es una realidad bastante bien conocida, me interesa destacar que el conjunto de las diversas ciencias cognitivas constituyen un campo de investigación interdisciplinar. Ahora bien, tal como he señalado en mi trabajo “Las

ciencias cognitivas como puente entre campos científicos” (2001), cabe establecer tres tipos (al menos) de relaciones interdisciplinarias entre las ciencias. En primer lugar, es posible establecer relaciones de hecho motivadas por el encomiable y laudable afán de ampliar los propios conocimientos, tal como sucede cuando el matemático se interesa por la historia política; se trata entonces de las relaciones interdisciplinarias de inspiración enciclopédica. En segundo lugar, existen relaciones entre las ciencias exigidas para perfeccionar los propios conocimientos, tal como sucede cuando el psicólogo estudia estadística o solicita la ayuda de un estadístico; tenemos ahora las relaciones interdisciplinarias de colaboración. Y en tercer lugar, se dan relaciones entre las ciencias que son constitutivas de un campo de investigación científico con personalidad propia; nos encontramos aquí con relaciones interdisciplinarias sustanciales.

Pues bien, las relaciones interdisciplinarias establecidas entre las diferentes ciencias cognitivas son del tercer tipo, esto es, las ciencias cognitivas constituyen y conforman un campo de investigación interdisciplinar. Al respecto, hay una cuestión de denominación que no es irrelevante. En un principio este campo de investigación se denominó “ciencia cognitiva” (en singular), tal como aparece en el título de la revista *Cognitive Science* (fundada en 1977) o en el título de la sociedad correspondiente *Cognitive Science Society* (constituida en 1979), denominaciones que aún perduran. Pero posteriormente se ha ido abriendo camino la denominación “ciencias cognitivas” (en plural), tal como aparece en el *Vocabulaire de sciences cognitives*, editado por Olivier Houdé y otros en 1998, o en *The MIT Encyclopedia of the Cognitive Sciences*, editada por Robert A. Wilson y Frank C. Keil en 1999.

Personalmente, como ya se habrá advertido, prefiero la denominación en plural, ya que atiende al hecho de la diversidad de las ciencias que componen este campo de investigación, aunque, al mismo tiempo, entiendo que tiene perfecta justificación emplear una sola denominación para las ciencias en cuestión, puesto que todas ellas conforman un mismo campo de investigación, con un supuesto básico común (la tesis de que el conocimiento es procesamiento de información) y una hipótesis filosófica común (la teoría funcionalista, adecuadamente entendida, de los procesos mentales).

Las ciencias cognitivas que constituyen el núcleo de la investigación son la psicología (de orientación cognitiva, no conductista) y la ciencia de la inteligencia artificial. Dicho con mayor extensión, el núcleo de la investigación atiende tanto a los sistemas cognitivos humanos y animales, estudio propio de la psicología, como a los sistemas cognitivos artificiales (computadores y robots debidamente programados), estudio característico de la inteligencia artificial.

Ambas ciencias son claramente diversas en sus detalles. La psicología ha sido considerada tradicionalmente como una ciencia humana, y sólo con el conductismo (con razón por cierto) pasó a ocuparse también de la conducta

animal. En todo caso, la base empírica de la psicología actual tiene que ver con la biología y la neurociencia. En cambio, la base empírica de la inteligencia artificial es ajena, en principio, tanto a la biología como a la neurociencia. Expresado en términos simples, el soporte empírico principal de los sistemas cognitivos humanos y animales son neuronas, redes y mapas neuronales, mientras que el soporte empírico principal de los sistemas cognitivos artificiales son circuitos electrónicos. O dicho más brevemente, mientras que los sistemas cognitivos humanos y animales son seres vivos, en cambio los sistemas cognitivos artificiales no son seres vivos. Además la inteligencia artificial es un conocimiento híbrido en el que son relevantes la lógica, la matemática, la física y la ingeniería. Sin la lógica y la matemática no es posible la inteligencia artificial, pero estas disciplinas son consideradas tradicionalmente ciencias formales. Tampoco habría inteligencia artificial sin la física (sin electrónica no hay informática), que es considerada una ciencia natural. Pero también la inteligencia artificial es una ingeniería, puesto que sin la construcción real de los computadores y robots inteligentes la inteligencia artificial se queda en mera especulación (lo cual ocurre con frecuencia), pero la ingeniería no es propiamente una ciencia sino tecnología asociada. En suma, dentro del núcleo del programa de investigación cognitivo nos encontramos con una ciencia tradicionalmente humana (la psicología) junto a una ciencia híbrida de ciencia formal, ciencia natural y tecnología (la inteligencia artificial).

Existen otras ciencias implicadas en el campo de la investigación cognitiva que, aunque no agotan sus intereses en el estudio de los sistemas cognitivos en cuanto tales, constituyen poderosos recursos o instrumentos para abordar el fenómeno de la cognición. Tales disciplinas son, fundamentalmente, la lógica para la inteligencia artificial, la neurociencia cognitiva y la lingüística. Y podemos añadir la antropología y la pedagogía cognitivas.

El carácter interdisciplinar sustancial de las ciencias cognitivas ya fue advertido por Daniel Bobrow y Allan Collins, quienes, según creo, fueron los primeros en utilizar y divulgar la denominación de “ciencia cognitiva” en la obra colectiva (con diecisiete autores) editada por ellos en 1975 titulada *Representation and Understanding. Studies in Cognitive Science*. En efecto, en el prefacio dicen: “Este libro contiene estudios en un nuevo campo que llamamos *ciencia cognitiva*. La ciencia cognitiva incluye elementos de psicología, ciencia de la computación, lingüística, filosofía y educación, pero es más que la intersección de estas disciplinas. Su integración ha producido un nuevo conjunto de instrumentos para ocuparse de un amplio campo de cuestiones. En los años recientes, las interacciones entre los que han trabajado en estos campos han conducido a excitantes nuevos desarrollos en nuestra comprensión de los sistemas inteligentes y al desarrollo de una ciencia de la cognición. El grupo de trabajadores se ha dedicado a problemas que no parecían resolubles desde dentro de una disciplina aislada”.

## 2. El supuesto básico común a las ciencias cognitivas

En un libro de título prometedor (*What Is Cognitive Science*, 1993), Barbara Von Eckardt intenta presentar lo que denomina un “marco de compromisos compartidos” interdisciplinar que caracteriza a las ciencias cognitivas. Sin embargo, por razones de manejabilidad, esta autora restringe su atención a lo que considera el componente central del programa de investigación de la ciencia cognitiva, a saber, el estudio de la cognición típica y normal del adulto humano, que abrevia ANTCOG (*adult's normal, typical cognition*). Pero con esta restricción Von Eckardt pierde de vista el propio campo de la investigación cognitiva, ya que se entiende que se limita a la psicología humana, dejando de lado la psicología animal, y, lo que aún es peor, dejando también de lado la inteligencia artificial. Con todo, cuando se refiere a la estrategia explicativa de la ciencia cognitiva, formula una asunción de enlace combinada (combinando los aspectos computacional y representacional) que dice que las capacidades cognitivas humanas consisten, en gran medida, en un sistema de capacidades de procesamiento de información. En este punto se advierte, por un lado, la errónea restricción a las capacidades humanas, pero, por otro lado, la adecuada referencia a la noción de procesamiento de información.

En efecto, puede y debe considerarse que el concepto adecuado de sistema de procesamiento de información es el supuesto básico común a las ciencias cognitivas.

Allen Newell (1927-1992) y Herbert Simon (1916-2001), dos de los fundadores de las ciencias cognitivas, publicaron en 1972 su obra conjunta *Human Problem Solving*, donde ya aparece una caracterización de un sistema de procesamiento de información. En el epílogo de este libro se declara que el solucionador de problemas humano y el computador programado son ambas especies del género sistema de procesamiento de información, aunque los autores se han centrado en el estudio de la conducta humana.

Por otra parte, en la comunicación de Newell y Simon titulada “Computer Science as Empirical Inquiry: Symbols and Search” (1976) aparece formulada la denominada “hipótesis del sistema de símbolos físico”, que resulta equivalente a la tesis del sistema de procesamiento de información. Tal hipótesis sostiene que un sistema de símbolos físico tiene los medios necesarios y suficientes para la acción inteligente general. Con ello se pretende defender que cualquier sistema que exhiba inteligencia general resultará ser un sistema de símbolos físico y que, conversamente, cualquier sistema de símbolos físico de suficiente tamaño puede organizarse para exhibir inteligencia general. A su vez, se entiende por acción inteligente general el mismo alcance de inteligencia que vemos en la acción humana, esto es, una conducta apropiada a los fines del sistema y adaptativa a las exigencias del entorno.

En este punto, me parece importante señalar (como ya hice en mi trabajo citado de 2001) que la utilización reiterada del término “exhibir” por Newell y

Simon puede ser entendida como definiendo una tesis filosófica neutra o intermedia en la cuestión de si los computadores son inteligentes. Entre la tesis fuerte de que los computadores son realmente inteligentes y la tesis débil de que los computadores simulan simplemente ser inteligentes, cabe la tesis neutra (o previa) de que los computadores exhiben de algún modo inteligencia.

Ahora bien, para una detallada y clara descripción de los sistemas de símbolos físicos debemos dirigirnos al trabajo de Allen Newell titulado justamente "Physical Symbol Systems" (1981). En este lugar se establece que un sistema de símbolos físico consta de una memoria, de un conjunto de operadores (se describen diez), de un control, de una entrada y de una salida. Asimismo un sistema de símbolos es una máquina que existe en un medio que consta de objetos distribuidos en un espacio de ubicaciones. Las entradas (podemos precisar) son informaciones sobre los objetos, y las salidas son informaciones sobre la modificación o creación de objetos. La conducta externa del sistema de símbolos consiste en las salidas que produce como función de las entradas; pero el estado interno del sistema consiste en el estado de su memoria y el estado de su control, mientras que su conducta interna consiste en la variación de ese estado interno a lo largo del tiempo.

Tal como he hecho en mi trabajo "¿En qué sentido la psicología cognitiva puede ser computacional?" (2000), tenemos que introducir algunos comentarios si queremos precisar el alcance exacto de esta hipótesis de los sistemas de símbolos físicos. En primer lugar, esta hipótesis es una formulación más detallada de la tesis habitualmente citada de los sistemas de procesamiento de información. Para advertir este punto, basta con comparar la figura 2.1 de *Human Problem Solving* (1972), donde se presenta la estructura general de un sistema de procesamiento de información, con la figura 2 de "Physical Symbol Systems" (1981), donde se presenta la estructura de un sistema de símbolos paradigmático, ya que las coincidencias son obvias. En segundo lugar, se trata de una hipótesis, es decir, se postula que cualquier sistema de símbolos, tal como queda descrito, está dotado de inteligencia, y que cualquier entidad inteligente es un sistema de símbolos. En consecuencia, queda abierta a la discusión y a la investigación empírica tal identificación entre sistemas de símbolos y entidades inteligentes.

Pero conviene analizar con mayor detalle esta noción de sistema de símbolos. Para empezar, y frente a ciertas confusiones y acusaciones infundadas, en el concepto de sistema de símbolos los aspectos semánticos y de interpretación están explícitos. En efecto, el control del sistema interpreta continuamente cualquier expresión, interpreta tanto los datos como los programas, y dentro de éstos tanto los símbolos como sus roles. Los operadores descritos por Newell, en su trabajo de 1981, son diez, de los que cinco están asociados a la unidad de entrada (receptores) y otros cinco están asociados a la unidad de salida (motores). Los cinco primeros son entrar, leer, escribir, copiar y asignar, mientras que los otros cinco son hacer, salir-si, continuar-si, citar y actuar.

Entre unos y otros se sitúa la unidad de control, en relación permanente con la unidad de memoria. Ahora bien, el operador de asignar desempeña un papel muy importante, puesto que establece una relación básica entre un símbolo y la entidad que designa; más aún, el concepto de designación o representación es, para Newell, el concepto más fundamental para un sistema de símbolos. Podemos comentar que el sistema de símbolos se concibe como un mecanismo típicamente representacional.

Por otra parte, resulta claro que la hipótesis del sistema de símbolos físico se aplica por igual a la psicología cognitiva y a la inteligencia artificial. En efecto, un sistema de símbolos es una descripción abstracta tanto de un sujeto humano o animal como de un computador. En todos estos casos cabe hablar de unidad de entrada (receptores), unidad de control con sus operadores, memoria, y unidad de salida (motores). Y también en todos estos casos cabe hablar por igual de sujetos procesadores de información.

### 3. El funcionalismo en las ciencias cognitivas

El punto de vista que deseo defender y aclarar debidamente en este trabajo es que el funcionalismo constituye una hipótesis filosófica acerca de la naturaleza de los procesos mentales, aneja de manera natural a la hipótesis del sistema de símbolos físicos (o del sistema de procesamiento de información) y que permite consolidar la unificación de las diversas ciencias cognitivas, ya unificadas por la hipótesis empírica del sistema de símbolos. Asimismo las diferentes bases empíricas de cada ciencia cognitiva dan cuenta de su diversidad.

Como caracterización inicial del funcionalismo, me permitiré reproducir lo que escribí en mi artículo "El desafío de las emociones a las ciencias cognitivas" (2000). Por funcionalismo se entiende la doctrina filosófica de que los procesos mentales no son conducta externa públicamente observable, sino que son funciones mediadoras entre entradas sensoriales y salidas motoras, que se producen en el interior del sistema y son causa de la conducta externa. Esto supone que los procesos mentales pueden ser caracterizados según la función que desempeñan, con independencia del soporte físico del sistema. Es decir, si se trata, por ejemplo, de demostrar un teorema, lo relevante es el conjunto de operaciones y los datos de que disponemos, siendo en principio indiferente que la demostración sea ejecutada en un computador o en un cerebro humano. Tal como veremos, esta caracterización sigue algunas propuestas de Hilary Putnam.

Por otra parte, el funcionalismo puede tomarse en sentido restringido o en sentido amplio. En el primer sentido, y en consonancia con la hipótesis del sistema de símbolos *físico*, los soportes han de ser en todo caso de carácter físico. Pero en sentido amplio los soportes pueden ser cualesquiera, incluido un

soporte espiritual. De todas maneras, para el funcionalismo (amplio o restringido) los procesos mentales no son necesariamente procesos cerebrales, sino que son procesos que se relacionan entre sí mediante relaciones causales ejecutadas sobre tipos variados de soportes.

Esta última consideración nos lleva a poner de relieve que el funcionalismo, frente a lo que pudiera pensarse, no constituye una solución al problema mente-cerebro. Para empezar, resulta muy conveniente tener en cuenta que el problema mente-cerebro sólo tiene sentido, tal como he señalado en mi libro *La nueva filosofía de la mente* (1995), si se admite algún tipo de dualismo entre procesos mentales y procesos cerebrales. En efecto, dentro del campo de las ciencias cognitivas, el problema mente-cerebro consiste en discutir si los procesos mentales (tal como los estudia la psicología cognitiva o la inteligencia artificial) se identifican o no con los procesos cerebrales (tal como los estudia la neurociencia cognitiva), y en caso de no identificarse cuáles son las relaciones entre unos y otros. Si se rechaza la existencia de procesos mentales (o si se rechazara la existencia de procesos cerebrales) entonces no hay problema mente-cerebro. Tal es lo que ocurre con el conductismo (ya sea filosófico o psicológico), puesto que para esta doctrina no existen los procesos mentales, sino únicamente conducta públicamente observable. Tampoco existe el problema mente-cerebro para cualquier materialismo *realmente* eliminativo, esto es, para aquella doctrina que sostuviera que los procesos mentales son simples y meras ficciones del lenguaje cotidiano que no corresponden sino estrictamente a procesos cerebrales. Y finalmente, y es lo que me interesa poner de relieve, tampoco el funcionalismo aporta una solución al problema mente-cerebro, ya que defiende que los procesos mentales pueden ser caracterizados con independencia de su eventual soporte físico o de los órganos específicos que los produzcan. Por ello, en mi libro citado, he considerado que el funcionalismo es una especie de “tiempo muerto” en el problema mente-cerebro.

#### 4. Breve historia del funcionalismo cognitivo

Creo que el filósofo Ned Block es el principal responsable de la confusión que con frecuencia puede advertirse acerca del concepto de funcionalismo, aunque también Hilary Putnam, con su abandono parcial del funcionalismo, que él inventó, ha contribuido a tal confusión.

En efecto, Ned Block editó en 1980 y en 1981 dos volúmenes sucesivos muy influyentes, por la categoría e interés de los trabajos recopilados, titulados *Readings in Philosophy of Psychology*. En el primer volumen, en su tercera parte, Block reúne bajo el rótulo de “funcionalismo” estudios muy diversos incluyendo, entre otros, junto a Putnam a David Armstrong y a David Lewis (1941-2001). Asimismo Block justifica estas inclusiones en su “Introduction: What Is Functionalism” que abre esta tercera parte, así como en su “Troubles with Functionalism”, que la cierra.

En el primero de estos trabajos, Ned Block distingue tres tipos de funcionalismo: el análisis funcional, el funcionalismo de computación-representación y el funcionalismo metafísico. El primero propone explicaciones de tipo funcional que se apoyan en la descomposición de un sistema en sus partes componentes; como ejemplo de análisis funcional señala el artículo de Robert Cummins que incluye. A su vez, el funcionalismo de computación-representación considera que los estados psicológicos representan de modo sistemático el mundo mediante un lenguaje de pensamiento y que los procesos psicológicos son computaciones que envuelven esas representaciones; como ejemplos Block señala los artículos incluidos en la primera parte del volumen segundo titulada “representación mental”. Finalmente, el funcionalismo metafísico caracteriza los estados mentales en términos de sus roles causales, en particular en términos de sus relaciones causales con las estimulaciones sensoriales, las salidas conductuales y otros estados mentales; como ejemplos pone la mayoría de los trabajos que siguen, metiendo en el mismo saco el funcionalismo auténtico de Putnam y el materialismo de Armstrong y Lewis.

A su vez, en “Troubles with Functionalism”, publicado originalmente en 1978, Ned Block distingue entre funcionalistas a priori y funcionalistas empíricos. Los primeros, entre los que se encuentran Armstrong y Lewis, tienden a considerar los análisis funcionales como análisis de los significados de los términos mentales. Los segundos, entre los que se encuentran Putnam y Fodor (que yo estimo los auténticos funcionalistas o funcionalistas cognitivos), consideran los análisis funcionales como hipótesis científicas sustantivas. Para mayor complicación (y confusión), Block decide denominar ambos tipos con la palabra “funcionalismo” (con f minúscula), mientras que el primer tipo se denominará “Funcionalismo” (con f mayúscula) y el segundo Psicofuncionalismo (con p mayúscula).

Finalmente, en el artículo de Block titulado “Functionalism (2)”, publicado en *A Companion to the Philosophy of Mind* (1994), distingue entre funcionalismo conceptual y psicofuncionalismo. Para el funcionalismo conceptual (el Funcionalismo en la distinción anterior), las definiciones funcionales, dentro de una teoría psicológica popular del sentido común, se dirigen a la captura de nuestros conceptos mentales ordinarios. Para el psicofuncionalismo (el anterior Psicofuncionalismo), las definiciones funcionales, dentro de una teoría psicológica empírica (científica), se establecen para fijar las extensiones de los términos mentales.

Personalmente creo que Block se equivocó en dos aspectos. Primero, al ampliar el concepto de funcionalismo de Putnam (que ahora veremos) para incluir a los materialistas Lewis y Armstrong. Segundo, al entender que el Funcionalismo se dice sobre todo de estos materialistas. Ambos equívocos son completamente contrarios al punto de vista, que defiende, de que el funcionalismo constituye la hipótesis filosófica sobre la naturaleza de los procesos mentales que unifica las ciencias cognitivas y que es aneja, de modo natural,

a la hipótesis empírica del sistema de símbolos. Por ello, y si es necesario evitar más malentendidos, podemos hablar de funcionalismo cognitivo.

Veamos pues el funcionalismo en Hilary Putnam. Me centraré en tres trabajos suyos. En primer lugar, su artículo de 1960 “Minds and Machines”, donde aparecen los antecedentes de su concepción. En segundo lugar, su estudio “Psychological Predicates” de 1967 (reimpreso con el nuevo título de “The Nature of Mental States”), donde se establece el funcionalismo. Y finalmente su nuevo estudio “Why Functionalism Failed”, publicado en 1992 (reimpreso con el nuevo título de “Why Functionalism Didn’t Work”, y que utiliza el capítulo quinto, con este mismo título, de su libro *Representation and Reality* de 1988), donde abandona parcialmente el funcionalismo.

En “Minds and Machines” Putnam analiza la noción de máquina de Turing, es decir, la caracterización general de un computador presentada por el matemático y lógico británico Alan Turing (1912-1954). Cuando una máquina de Turing es descrita mediante una “tabla de máquina” aparece como algo que posee una cinta (con cuadros que tienen símbolos o no), un mecanismo de impresión, un mecanismo para “examinar”, y un conjunto finito de estados, que Putnam denomina “estados lógicos”. Para nuestro autor, la “descripción lógica” de una máquina de Turing no incluye ninguna especificación de la *naturaleza física* de esos “estados”, ni desde luego de la naturaleza física de la totalidad de la máquina. En cambio, en cuanto una máquina de Turing es realizada físicamente, ocurre que desde el punto de vista del lógico sólo tiene los “estados lógicos”, pero desde el punto de vista del ingeniero tiene un número casi infinito de “estados estructurales”. Por otra parte, Putnam señala que, al igual que hay dos posibles descripciones de la conducta de una máquina de Turing (el anteproyecto estructural del ingeniero y la “tabla de máquina” del lógico), también hay dos posibles descripciones de la psicología humana. En efecto, por un lado tenemos la aproximación conductista, que pretende proporcionar una descripción fisicalista completa de la conducta humana, pero, por otro lado, también sería posible buscar una descripción más abstracta de los procesos mentales humanos en términos de “estados mentales” (sin especificar la realización física, si la hay) e “impresiones” (comparables a los símbolos de la cinta de máquina).

En suma, Putnam presenta en este trabajo una *analogía* entre los estados lógicos de una máquina de Turing y los estados mentales de un ser humano, por una parte, y entre los estados estructurales de una máquina de Turing y los estados físicos de un ser humano, por otra parte. Con ello, cabe añadir, deja abierta la posibilidad de definir los estados mentales en términos abstractos, de tal modo que su caracterización pueda aplicarse por igual a mentes humanas, animales o mecánicas, como requiere una adecuada unificación filosófica de las ciencias cognitivas. Pero en este artículo de 1960 Putnam no ofrece todavía una descripción funcional de los estados mentales, con lo que no me parece correcto hablar propiamente de funcionalismo.

En cambio, en “Psychological Predicates” (1967) Hilary Putnam caracteriza los procesos mentales como estados funcionales. Nuestro autor centra su discusión acerca de la naturaleza de los procesos mentales en la cuestión de si el dolor es un estado cerebral, con la intención de generalizar la respuesta obtenida a todos los estados mentales. En principio, según Putnam, son posibles dos respuestas a tal cuestión. Por un lado, cabe sostener que decir que “los dolores son estados cerebrales” es perfectamente significativo, esto es, no envuelve extensión de uso. Dicho de otro modo, podemos comentar, tal enunciado es significativo si no amplía el uso del término “dolores”, con lo que entonces la identificación entre dolores y estados cerebrales queda justificada. Pero, por otro lado, cabe defender que decir que “los dolores son estados cerebrales” no es significativo, esto es, envuelve un cambio de significado o una extensión de uso. Dicho de otra manera, podemos comentar, tal enunciado no es significativo porque cambia el significado o amplía el uso del término “dolores”, con lo que entonces la identificación entre dolores y estados cerebrales no está justificada. Putnam declara que su posición propia no está expresada por ninguna de ambas respuestas, ya que las nociones de cambio de significado y de extensión de uso están tan mal definidas que no es posible sostener ninguna de ambas respuestas. En efecto, según nuestro autor, ni el lingüista ni el hombre de la calle ni el filósofo poseen actualmente una noción de cambio de significado que resulte aplicable a casos como el presente (casos de identificación de términos, como “dolores” y “estados cerebrales”). En estas circunstancias, la respuesta a la cuestión de si el dolor es un estado cerebral debe permitir enunciados de la forma “el dolor es A”, donde “el dolor” y “A” no son sinónimos en sentido alguno (esto, podemos añadir, no se determinan recurriendo a la noción de significado), sino que tales enunciados resultan o no aceptables recurriendo a fundamentos empíricos y metodológicos. Tal es el camino que sigue Putnam.

Nuestro autor sostiene que el dolor (y, en general, los procesos mentales) no es un estado cerebral, es decir, no es un estado físico-químico del cerebro, y tampoco un estado del sistema nervioso entero, sino que es otro *tipo* de estado. La hipótesis propuesta (en lugar de la identidad entre procesos mentales y procesos cerebrales) es que el dolor, o el estado de tener dolor, es un estado funcional de un organismo tomado en su totalidad. Pero para aclarar tal hipótesis Putnam utiliza la noción de autómeta probabilístico. A su vez, un autómeta probabilístico es una máquina de Turing con la peculiaridad de que las transiciones entre los estados internos de la máquina pueden responder a probabilidades variadas en lugar de ser deterministas. Por supuesto, un autómeta probabilístico tiene entradas sensoriales y asimismo salidas motoras; la tabla de máquina (esto es, podemos aclarar, el conjunto de especificaciones de su curso de acción) establece, para cualquier combinación posible de un estado interno y un conjunto completo de entradas sensoriales, una instrucción que determina la probabilidad del siguiente estado así como también las pro-

babilidades de las salidas motoras. Creo que debe comentarse que la idea del autómata probabilístico consiste en diseñar un organismo con actividad flexible; es decir, dadas ciertas entradas sensoriales y dado un estado interno del organismo, las salidas motoras (si las hay) y el nuevo estado interno no están determinados de modo fijo, sino que responden a ciertas probabilidades.

Ahora bien, sigue Putnam, un sistema dado empíricamente puede ser al mismo tiempo la realización física de muchos diferentes autómatas probabilísticos. Es decir, cabe añadir, que podemos para un mismo sistema u organismo diseñar muchos diversos autómatas probabilísticos. Por ello, nuestro autor se ve obligado a introducir la noción clave de descripción de un sistema. En efecto, la descripción de un sistema  $S$  es cualquier enunciado verdadero que establece que  $S$  posee distintos estados  $s_1, s_2, \dots, s_n$ , que están relacionados entre sí y con las entradas sensoriales así como con las salidas motoras mediante las probabilidades de transición dadas en cierta tabla de máquina. Pues bien, tal tabla de máquina, mencionada en la descripción del sistema  $S$ , se denominará la “organización funcional de  $S$ ” relativa a tal descripción. Y a su vez, un estado cualquiera  $s_i$ , tal que el sistema  $S$  está en ese estado  $s_i$  en un tiempo dado, se denominará el “estado total de  $S$ ” (en ese tiempo) relativo a esa descripción. Putnam pone de relieve que conocer el estado total de un sistema relativo a una descripción envuelve conocer mucho acerca de cómo se comportará probablemente el sistema, dadas varias combinaciones de entradas sensoriales, pero no envuelve conocer la realización física de tal estado total; en particular no envuelve conocer si tal estado es un estado físico-químico del cerebro.

Después del análisis de estas nociones técnicas previas, volvemos a la hipótesis de Putnam de que el dolor (y en general cualquier proceso mental) es un estado funcional de un organismo tomado en su totalidad. Tal hipótesis, siguiendo a nuestro autor, puede desarrollarse en los cuatro puntos siguientes. Primero, todos los organismos capaces de sentir dolor son autómatas probabilísticos. Segundo, cualquier organismo capaz de sentir dolor posee un tipo apropiado de organización funcional. Tercero, ningún organismo capaz de sentir dolor posee una descomposición en partes que separadamente posean la citada organización funcional (con ello se indica, podemos precisar, que el dolor es un estado funcional del organismo entero). Y cuarto, para cualquier organización funcional, existe un subconjunto de entradas sensoriales tal que un organismo con la citada organización funcional tiene dolor cuando y sólo cuando algunas de sus entradas sensoriales están en tal subconjunto (es decir, cabe añadir, que un proceso mental no es producido sino por entradas determinadas).

Putnam dice además que la organización funcional debe incluir algo parecido a una función de preferencia (y con ello, podemos suponer, la posibilidad de seleccionar entradas sensoriales y salidas motoras, ya que el autómata es probabilístico), asimismo debe incluir algo parecido a una lógica inductiva, es

decir, la “máquina” (tal como dice Putnam) debe ser capaz de aprender de la experiencia, y finalmente la máquina debe disponer de órganos sensoriales.

Aunque tal hipótesis, que identifica procesos mentales y estados funcionales, pueda parecer vaga, es para nuestro autor mucho menos vaga que la actual hipótesis materialista, que identifica procesos mentales y estados físico-químicos del cerebro. De hecho, la hipótesis funcionalista es para Putnam más plausible, porque permite más investigación matemática y empírica, y ya que conduce a cuestiones y predicciones fructíferas. Por otro lado, la hipótesis funcionalista, según declaración expresa de Hilary Putnam, no es incompatible con el dualismo, puesto que un sistema, si lo hubiera, compuesto de cuerpo y alma puede ser un autómatas probabilístico. Finalmente, y contra la tesis conductista, la conducta de la máquina dependerá no sólo de las entradas sensoriales sino también de su estado total, incluyendo valores, creencias, etc.

Para concluir la presentación del funcionalismo de Putnam, veremos algunos puntos de su trabajo “Why Functionalism Failed” (1992). Nuestro autor aclara, en el inicio, que la analogía computacional (esto es, la analogía entre la mente humana y un computador digital) fue una reacción contra la idea de que nuestra materia es más importante que nuestra función. A su vez, el “funcionalismo” insistía en que, en principio, un robot avanzado, un ser humano, una criatura con química de silicio, y, si hay espíritus incorpóreos, un espíritu incorpóreo podrían funcionar de la misma manera cuando se describen al nivel de abstracción relevante, y que es erróneo pensar que la esencia de nuestras mentes es nuestro “hardware”. Putnam declara que no abandona esto y que le parece tan verdadero como importante. Sin embargo, nuestro autor entiende aquí que los argumentos a favor del punto de vista computacional pueden ser generalizados y extendidos para mostrar que una identificación simple entre los estados mentales y los estados funcionales tampoco es correcta. En suma, creo que puede entenderse que Hilary Putnam sigue pensando que el funcionalismo es básicamente adecuado, pero piensa que algunos argumentos crean serias dificultades al funcionalismo. Por ello, podemos hablar de abandono parcial por Putnam del funcionalismo.

La principal dificultad que encuentra nuestro autor en su posición funcionalista es que es incompatible con la explicación del significado que él mismo estableció en “The Meaning of ‘Meaning’” (1975). Según tal explicación, el contenido de nuestras creencias y deseos no está determinado por propiedades individualistas del hablante, sino que depende de hechos sociales relativos a la comunidad lingüística del hablante y de hechos relativos al ambiente físico del hablante. Por ejemplo, son los expertos (frecuentemente ajenos a nosotros) quienes saben si un árbol es un olmo o si un metal es oro auténtico. La referencia (y el “significado”) dependen del ambiente no-humano al igual que dependen de la sociedad. Entonces, sigue Putnam, las actitudes proposicionales (esto es, los estados mentales) no son estados del cerebro humano y del sistema nervioso considerados aisladamente del ambiente social y no-

humano. Y *a fortiori* no son estados funcionales, es decir, estados definibles en términos de parámetros que entrarían en una descripción de software del organismo. En conclusión, siguiendo este argumento, el funcionalismo, construido como la tesis de que las actitudes proposicionales son estados computacionales del cerebro, no puede ser correcto.

Creo que son convenientes dos observaciones. En primer lugar, Putnam pretende presentar dificultades a la hipótesis general funcionalista (cualquier proceso mental es un estado funcional) argumentando en contra de uno de sus casos especiales, a saber, que los procesos mentales humanos son estados funcionales; de ahí que se refiera en particular a las actitudes proposicionales humanas y a los estados del cerebro humano. Por supuesto, esta línea de argumentación es lógicamente correcta. Pero, situados en el caso especial humano y considerando el soporte cerebral de los procesos mentales humanos, creo que Putnam se equivoca al sostener que los significados no están “en la cabeza” (como declara en su trabajo de 1975) y, con ello, su argumento no pone en dificultades al funcionalismo. Lo que parece olvidar nuestro autor es que (al menos) la mayoría de los procesos mentales son representaciones. Ahora bien, nuestras representaciones están “en la cabeza”, ya que son elaboradas por nuestros cerebros. Y además, y este es el punto serio del asunto, no sólo tenemos representaciones elaboradas por nosotros mismos sino también representaciones de lo que dice el experto, de los informes de los medios de comunicación culturales y, por supuesto, del ambiente no-humano. En suma, una posición funcionalista solidaria de la hipótesis del sistema de símbolos (que, como señalé antes, es un mecanismo típicamente representacional) sobrevive al argumento en contra de Putnam.

Para terminar este apartado me referiré a la contribución de Jerry Fodor al funcionalismo cognitivo, contenida en su libro *Psychological Explanation* (1968). En efecto, un año después de que Hilary Putnam publicase su célebre artículo, Fodor presenta sus tesis funcionalistas al distinguir entre microanálisis y análisis funcional. Aquí se advierte, de manera clara y sencilla, una idea básica en la hipótesis funcionalista, a saber, la distinción entre soporte y función.

Para Fodor, el reduccionismo de las teorías psicológicas a las neurológicas es básicamente erróneo, aunque es positivo relacionar los constructos psicológicos con los constructos neurológicos. El materialismo reduccionista se caracteriza por practicar el microanálisis en cuanto opuesto al análisis funcional. Efectivamente en el microanálisis o análisis reduccionista la pregunta usual es “¿de qué consta X?”, buscando una respuesta que consista en la especificación de la microestructura de X. Por ejemplo, son cuestiones de microanálisis las preguntas “¿de qué consta el agua?”, o bien “¿de qué consta un rayo de luz?”, siendo sus respuestas respectivas en microanálisis “de dos átomos de hidrógeno unidos a uno de oxígeno” y “de una corriente de electrones”. Contrariamente, el análisis funcional pregunta acerca del papel que juega

una parte de un mecanismo en las actividades características de tal mecanismo considerado como un todo. Por ejemplo, una cuestión de análisis funcional es la pregunta “¿qué hace el árbol de levas?” y su respuesta es “abre las válvulas, permitiendo la entrada de combustible en el cilindro, para que al hacer explosión se mueva el pistón”. Por tanto, dice Fodor, mientras que el progreso del microanálisis depende del desarrollo de poderosos instrumentos de observación y de métodos muy precisos de disección, en cambio el análisis funcional requiere una estimación de los tipos de actividad característicos de un determinado mecanismo así como de la contribución de cada una de las partes al funcionamiento del todo.

Fodor admite y defiende establecer relaciones entre los constructos psicológicos y los constructos neurológicos, pero entiende que se podrán evitar muchas confusiones si estas relaciones no las concebimos como expresión de un microanálisis, sino como una forma de atribuir determinadas funciones psicológicas a los correspondientes sistemas neurológicos. Por ejemplo, añade nuestro autor, para algunos filósofos y psicólogos es posible seguir la trayectoria de una entrada a través de los sistemas neurológicos aferentes, centrales y eferentes sin que nos encontremos con nada parecido a motivos, impulsos, estrategias y demás invenciones de las teorías psicológicas. Pero esto es tan erróneo como afirmar que en la descripción completa de un motor de combustión interna nunca aparecerá un elevador de válvulas. En ambos casos se confundirían los términos propios del microanálisis con los términos propios del análisis funcional.

Comentando estos puntos de vista, tal como he hecho en mi libro *La nueva filosofía de la mente* (1995), la identificación entre procesos mentales y procesos cerebrales, propia del materialismo reductivo, sería análoga a la identificación entre elevadores de válvulas (que se caracterizan funcionalmente) y árboles de levas (que se caracterizan refiriéndonos a su estructura física). Sin embargo, no todo elevador de válvulas es un árbol de levas, ya que existen otros mecanismos que cumplen la función de elevar válvulas. Análogamente, podemos añadir, no todo proceso mental, que se caracteriza por el rol causal o función que desempeña, es un proceso cerebral, puesto que, además del cerebro, existen otros mecanismos que cumplen funciones mentales, como, por ejemplo, un computador al establecer la demostración de un teorema.

## 5. Niveles de descripción en las ciencias cognitivas

El funcionalismo cognitivo implica que, para comprender de modo adecuado un sistema de símbolos o de procesamiento de información, es preciso describirlo a varios niveles. Al menos al nivel de la estructura o soporte del sistema y al nivel de sus funciones o roles causales. Ahora bien, el científico cognitivo inglés David Marr (1945-1980) distingue tres niveles de descripción, en su obra *Vision*, publicada en 1982 tras su muerte prematura.

Estos niveles señalados por Marr son el de la teoría computacional, el de la representación y algoritmo, y el de la implementación en hardware. En primer lugar, el nivel superior consiste en la teoría computacional abstracta del mecanismo; se trata de especificar el objetivo de la computación, determinar su carácter apropiado y la lógica de la estrategia mediante la cual se lleva a cabo. Dicho simplemente, en el nivel de la computación se trata de precisar la tarea o función que intentamos estudiar, en términos del “qué” de la computación. En segundo lugar, el nivel del centro consiste en la elección de la representación para las entradas y salidas así como del algoritmo que se empleará para transformar unas en otras. Ahora, en el nivel del algoritmo se trata de precisar el “cómo” de la computación. Y finalmente, el nivel inferior consiste en pormenorizar los detalles de cómo se realizan físicamente las representaciones y el algoritmo. Aquí, en el nivel de la implementación se trata, según David Marr, por así decir, de la arquitectura del computador detallada.

Para nuestro autor, aunque los tres niveles son necesarios para una comprensión completa de un mecanismo de procesamiento de información, están relacionados entre sí de una manera no rígida, de tal modo que algunos fenómenos pueden ser explicados en un nivel o dos únicamente. Con todo, para Marr, el nivel más importante, desde el punto de vista del procesamiento de la información, es el nivel de la computación, aunque los algoritmos y los mecanismos (soportes) son empíricamente más accesibles. En concreto, un algoritmo se comprenderá más fácilmente si comprendemos la naturaleza del problema a resolver que examinando el hardware en el que está incorporado. En su libro, David Marr aplica esta metodología de los tres niveles al estudio de la percepción visual humana.

Otro autor que insiste en la distinción de niveles de descripción en las ciencias cognitivas es el científico cognitivo canadiense Zenon Pylyshyn. En efecto, desde su libro *Computation and Cognition* (1984) y recientemente en su artículo “What’s in Your Mind” (1999), distingue también tres niveles. Estos niveles son el físico o biológico, el simbólico o sintáctico, y el semántico o nivel de conocimiento. Según Pylyshyn, tal como dice en su artículo de 1999, esta hipótesis de los tres niveles defiende que hay generalizaciones que existen en cada uno de ellos. Así hay patrones de conducta que sólo pueden explicarse recurriendo a la biología, como por ejemplo por qué las reacciones de las personas son más lentas tras ingerir alcohol, o por qué se irritan cuando se les deja sin dormir. Otras veces, en cambio, algunos patrones de conducta sólo pueden explicarse recurriendo a lo que la gente quiere y a lo que la gente cree, lo cual corresponde al nivel semántico o de conocimiento. Y finalmente nuestro conocimiento se incorpora en un sistema de códigos, que es propio del nivel simbólico o sintáctico.

Pylyshyn añade una ilustración sencilla, empleando una comparación con una calculadora electromecánica. Podemos preguntar, en su caso, por qué la impresión de los números es débil o irregular (leyéndose con dificultad), y

entonces la respuesta precisará una explicación de nivel físico, refiriéndonos a aspectos eléctricos y mecánicos. Pero también podemos preguntar por qué la calculadora emplea más tiempo en calcular senos y cosenos que en calcular sumas, y entonces la explicación requerirá describir el método o algoritmo empleado por la calculadora, moviéndonos al nivel sintáctico. Finalmente hay cuestiones acerca de las relaciones entre lo que hace la calculadora y el mundo de la matemática abstracta. Al decir que la calculadora computa senos y cosenos decimos más que cuando señalamos el algoritmo empleado; según Pylyshyn, pretendemos que el algoritmo en cuestión computa representaciones de números que corresponden a ciertas funciones abstractas definidas matemáticamente; pero cuando hablamos de las cosas que representa la máquina nos encontramos al nivel de descripción semántico.

Si comparamos las tesis de Marr con las de Pylyshyn, advertimos puntos claros de coincidencia y algunas diferencias. Resulta evidente que el nivel de implementación de Marr coincide con el nivel físico o biológico de Pylyshyn. Se trata en ambos casos del soporte físico de la computación, que puede tener elementos y configuración diferentes según las situaciones; por ejemplo, un tipo u otro de cerebro, un tipo u otro de circuitería electrónica. En cambio, no parece que haya una estricta coincidencia entre el nivel de computación de Marr y el nivel semántico de Pylyshyn, ni tampoco entre el nivel de algoritmo del uno y el nivel sintáctico del otro. La diferencia, a mi entender, consiste en que el nivel de algoritmo de Marr incluye ya las representaciones, mientras que el nivel sintáctico de Pylyshyn no es aún propiamente representacional. Por otra parte, el nivel de computación de Marr atiende a la descripción de las tareas del sistema, determinando sus objetivos y estrategias; pero a esto no parece corresponder el nivel semántico de Pylyshyn. En realidad creo que la separación entre nivel sintáctico y nivel semántico es forzada, ya que no existe una distinción tajante entre sintaxis y semántica, tal como he señalado en mi libro *La nueva filosofía de la mente* (1995) al criticar el argumento de John Searle de la habitación china.

De todas formas, lo que parece claro es que, además de introducir un nivel de descripción de las tareas del sistema, la distinción clave está entre el soporte de los procesos y los procesos mismos entendidos como representaciones (con sus aspectos sintáctico y semántico). Y aquí podemos encontrar una confusión en el concepto mismo de función, que me interesa mucho señalar. En efecto, una cosa es la función de un sistema, como por ejemplo demostrar o percibir, y otra cosa son las representaciones (los procesos mentales) que son funciones mediadoras entre las entradas y salidas del sistema. La función desempeñada por el sistema es algo que podemos apreciar como conducta externa, mientras que la organización funcional del sistema constituye un conjunto de estados internos relacionados entre sí y con sus entradas y salidas de una manera precisa. Dicho de otro modo, los procesos mentales entendidos como estados funcionales pueden tener una u otra función, ingenieril o

biológica. Por ello, debemos inclinarnos a las tesis de Marr y distinguir tres niveles: descripción de tareas del mecanismo, descripción funcional del mecanismo y descripción del soporte del mecanismo.

Finalmente, las ciencias cognitivas pueden emprender estudios generales sobre tareas y habilidades cognitivas y sus procesos mentales correspondientes. Por ejemplo, sobre percepción visual o resolución de problemas. Si tenemos en cuenta el nivel de implementación o soporte, podemos precisar los estudios anteriores y ocuparnos de visión humana o de visión artificial, de resolución humana de problemas o de razonadores mecánicos. Llegados a la base empírica las ciencias cognitivas se nos muestran en su rica diversidad.

## Referencias bibliográficas

- BLOCK, Ned (ed.), *Readings in Philosophy of Psychology*, vol. 1 y vol. 2, Harvard University Press, Cambridge (Mass.), 1980 y 1981.
- BLOCK, Ned, "Funcionalismo (2)", Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell, Oxford, 1994, pp. 323-332.
- BOBROW, Daniel y COLLINS, Allan (eds.), *Representation and Understanding. Studies in Cognitive Science*, Academic Press, Orlando, 1975.
- FODOR, Jerry A., *Psychological Explanation. An Introduction to the Philosophy of Psychology*, Random House, Nueva York, 1968.
- HOUDÉ, Olivier y otros (eds.), *Vocabulaire de Sciences Cognitives*, Presses Universitaires de France, Paris, 1998.
- MARR, David, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, Nueva York, 1982.
- MARTINEZ-FREIRE, Pascual F., *La nueva filosofía de la mente*, Gedisa, Barcelona, 1995.
- MARTINEZ-FREIRE, Pascual F., "¿En qué sentido la psicología cognitiva puede ser computacional?", Mary Sol de Mora y otros (eds.), *Actas del III Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*, Universidad del País Vasco, San Sebastián, 2000, pp. 685-693.
- MARTINEZ-FREIRE, Pascual F., "El desafío de las emociones a las ciencias cognitivas", P. F. Martínez-Freire y otros (eds.), *Emociones. Themata*, 25, 2000, 55-66.
- MARTINEZ-FREIRE, Pascual F., "Las ciencias cognitivas como puente entre campos científicos", Antonio Blanch (ed.), *La nueva alianza de las ciencias y la filosofía*, Universidad Pontificia de Comillas, Madrid, 2001, 73-98.
- NEWELL, Allen, "Sistemas de símbolos físicos", Donald A. Norman (ed.), *Perspectivas de la ciencia cognitiva* (trad. Nuria Sebastián), Paidós, Barcelona, 1987, 51-107.

- NEWELL, Allen y SIMON, Herbert A., *Human Problem Solving*, Prentice-Hall, Englewood Cliffs (N.J.), 1972.
- NEWELL, Allen y SIMON, Herbert A. "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the Association for Computing Machinery*, 19 (1976), pp. 113-126.
- PUTNAM, Hilary, "Minds and Machines", Sidney Hook (ed.), *Dimensions of Mind*, New York University Press, New York, 1960.
- PUTNAM, Hilary, "Psychological Predicates", W. H. Capitan y D. D. Merrill (eds.), *Art, Mind and Religion*, University of Pittsburgh Press, Pittsburgh, 1967. (Reimpreso posteriormente con el título "The Nature of Mental States").
- PUTNAM, Hilary, "The Meaning of 'Meaning'", K. Gunderson (ed.), *Language, Mind and Knowledge*, University of Minnesota Press, Minneapolis, 1975.
- PUTNAM, Hilary, *Representation and Reality*, The MIT Press, Cambridge (Mass.), 1988.
- PUTNAM, Hilary, "Why Functionalism Failed", John Earman (ed.), *Inference, Explanation, and Other Philosophical Frustrations*, University of California Press, Berkeley, 1992. (Reimpreso posteriormente con el título "Why Functionalism Didn't Work").
- PYLYSHYN, Zenon W., *Computation and Cognition. Toward a Foundation for Cognitive Science*, The MIT Press, Cambridge (Mass.), 1984.
- PYLYSHYN, Zenon W., "What's in Your Mind", Ernest Lepore y Zenon Pylyshyn (eds.), *What is Cognitive Science*, Blackwell, Oxford, 1999.
- VON ECKARDT, Barbara, *What Is Cognitive Science*, The MIT Press, Cambridge (Mass.), 1993.
- WILSON, Robert A. y KEIL, Frank C. (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, Cambridge (Mass.), 1999.