

## *Dismantling the Machine: Putnam's Reasons against functionalism*

Mariano RODRÍGUEZ

Recibido: 26/01/2006

Aprobado: 01/03/2006

### **Resumen**

En este artículo se examina la fundación putnamiana del así llamado “funcionalismo de la tabla de máquina”, para después, y sobre todo, pasar revista a las razones que llevaron al pensador americano a abandonar la teoría funcionalista de la mente y a impugnar el programa de naturalización de la intencionalidad. Desde aquí, y a la vista de la pobreza de los argumentos que se le han venido oponiendo a Putnam desde que consumó su ruptura con la ortodoxia funcionalista, y de lo poco convincente de los mismos, se concluye apuntando la necesidad de someter a crítica los presupuestos filosóficos del modelo computacional de la mente en psicología cognitiva como desarrollo sofisticado del materialismo moderno, en el sentido de la crítica de las ideologías.

*Palabras clave:* máquina de Turing, funcionalismo, intencionalidad, materialismo, mecanicismo, externismo semántico, metáfora computacional, ideología.

---

\* Este trabajo está basado en una conferencia organizada por el Departamento de Filosofía IV de la UCM que tuvo lugar el 17 de Noviembre de 2005, y fue realizado en el marco del Proyecto de Investigación Complutense PR/05-13341.

## Abstract

This paper examines in the first place the putnamian foundation of the so-called “table machine functionalism”, in order, mainly, to be able to review the reasons that led the American thinker to abandon the functionalist theory of mind, and also to refute all the program of intentionality naturalization. Hereafter, and in view of the fact that the arguments that have been opposed to Putnam ever since are generally poor and not much convincing, the conclusion we arrive to is pointing out the necessity of a critical consideration of the philosophical foundations of the computational model in cognitive psychology, taken as the sophisticated development of modern materialism, in the sense of the critique of ideologies.

*Keywords:* Turing machine, functionalism, intentionality, materialism, mechanism, semantic externalism, computational metaphor, ideology.

*“L’ère de l’informatique sera celle des crétins”*

Michel Henry 1987, 93

## 1. La Máquina como punto de partida

Cuando a comienzos de los sesenta Hilary Putnam se decide a ocuparse del enigma clásico de la mente y el cuerpo, lo hace guiado por la intención de mostrar que se trata sólo de un enigma *aparente*, es decir, que propiamente estaríamos ante una cuestión “nada más que” lógica y lingüística. Más específicamente: quiere dejar demostrado que la poderosa intuición según la cual la experiencia consciente (subjetiva) *no puede ser* idéntica a un proceso en el sistema nervioso habrá de ser definitivamente descartada como “hecho empírico” *porque en absoluto se presentaría en relación con un sistema computacional* “capaz de contestar preguntas acerca de su propia estructura” (MM, 362). Con el modelo de la Máquina de Turing no puede entrar para nada en consideración la pretendida naturaleza única de la experiencia subjetiva, o sea, aquello que ha constituido el quebradero de cabeza más lacerante para el Materialismo de la Identidad desaparecería de inmediato en la nueva panorámica que se abre desde la posición funcionalista. Por eso piensa Putnam que con un modelo como el que ha elegido y reivindicado podremos por fin llegar a ver a qué viene realmente a parar, cuál es el valor en efectivo que le corresponde, para emplear la metáfora del viejo pragmatismo americano, al famoso problema de la relación de la mente y el cuerpo.

Podemos pensar justificadamente que Putnam aspiraba por aquellos años a un

mecanicismo lo suficientemente sofisticado como para evitar las simplificaciones y las limitaciones del materialismo clásico<sup>1</sup> heredado de la interpretación metafísica de la física y la astronomía modernas que se sacralizó sobre todo a partir de la portentosa síntesis newtoniana. Desde luego que no vamos a aplicarnos aquí a detallar el sentido matemático de la Máquina de Turing. Tan sólo recordaremos en primer lugar que un tal dispositivo teórico puede ser concebido como una computadora universal, capaz de hacer lo que cualquier computadora posible puede hacer, todo cálculo posible. Y recordar además que Alan Turing respondió afirmativamente a la pregunta de si puede pensar una máquina. En este sentido de la modelización de la inteligencia humana, la máquina puede imprimir incluso las palabras “estoy en el estado A” cuando efectivamente se halla en ese estado, del mismo modo que Mariano puede decir que le duele la cabeza cuando la cabeza le duele.

La superioridad indiscutible del nuevo materialismo, el computacional, venía curiosamente del hecho de poner entre paréntesis la “materia” concreta de que estamos hechos nosotros o las máquinas propiamente dichas. Queda claro, desde luego, que siempre tendrá que haber algo de lo que estemos hechos, pero como este algo es en todo caso variable, debe quedar necesariamente fuera de consideración al considerar la inteligencia. Dicho de otro modo, la descripción lógica de una Máquina de Turing, o sea la especificación de las instrucciones que integran su “tabla”, no da entrada bajo ningún concepto a la precisión de la naturaleza física de los estados computacionales cuya sucesión controla el programa. Cualquier sistema que atravesase en el tiempo por una serie sucesiva de estados es una Máquina de Turing. Lo que no significa sino que una Máquina de Turing es una Máquina *abstracta*, que puede realizarse en una variedad de modos prácticamente infinita: se abre así la posibilidad tan apreciada de una funcionalización de la mente humana. El punto de vista del ingeniero es perfectamente válido e incluso decisivo, sin duda, pero no es el de la investigación de la inteligencia. Al contrario, para estudiar objetivamente

---

<sup>1</sup> Distingue el autor, efectivamente, entre un materialismo clásico o tradicional, que rechaza dándolo como falso, y un materialismo moderno para el que los estados mentales serían estados funcionales. La versión del materialismo que acepta como correcta es esta última, según la cual un ser humano sería “simplemente” un sistema físico con una determinada organización funcional compleja (*LP*, 450). Por eso no nos debemos dejar desconcertar por las numerosas declaraciones según las que el funcionalismo de la tabla de máquina no es como tal lógicamente incompatible con el dualismo cartesiano (por ejemplo: *R*, 393; *ML*, 412, 414; *NMS*, 436). Aparte de que la inspiración nuclear del funcionalismo es desde luego mecanicista y de que la tesis materialista básica es que el hombre es una máquina, debemos considerar estas afirmaciones de Putnam y de otros funcionalistas como Fodor como manifestación de que el funcionalismo constituye una hipótesis empírica revisionista, y no, en absoluto una tesis sobre el significado de los términos psicológicos en el lenguaje corriente. Por lo demás, está claro que el argumento de la múltiple realizabilidad que Putnam populariza contra el materialismo de la identidad va dirigido contra la llamada identidad de tipos y a favor del materialismo de la identidad de casos, completamente afín al “dualismo”, *sui generis* y desde luego nada cartesiano, de la metáfora computacional de la mente.

los estados y los procesos mentales trataríamos de describirlos especificando las leyes en que se suceden unos a otros y su relación con la expresión verbal. Así que la descripción de su *organización funcional*, sin entrar en la referencia a la realización material de turno, sería propiamente el punto de vista del psicólogo.

De lo que por descontado no hay duda es de que, si consideramos la Máquina de Turing, no hay enigma alguno en lo que hace a la relación entre programa y dispositivo físico, entre los estados lógicos y los estados estructurales: *por lo tanto* tampoco lo tiene que haber, desde el momento en que nos hallamos ante una analogía completa, en lo que se refiere a la relación entre el cerebro y los estados mentales de una persona<sup>2</sup>.

Tal vez nos podríamos poner a hablar del alma de los ordenadores, o si no dejar de hablar del alma de las personas. Pero evidentemente lo primero que se nos ocurre es que la analogía falla en lo fundamental, porque las personas tenemos conciencia y las máquinas desde luego que no. Por eso nosotros podemos pensar que estamos vivos y no somos robots. A esta objeción tan natural y por eso mismo de tan grueso calibre, nuestro pensador va a responder en lo esencial de la misma manera a lo largo de más de veinte años. Así, en su gran obra del 81, *Razón, verdad e historia*, dejará escrito lo siguiente: “Un argumento que a los funcionalistas nos gusta emplear, incluyéndome a mí, es el siguiente argumento antichovinista: en principio, las diferencias entre un robot y un humano (en organización funcional, en definitiva) podrían reducirse a pequeños detalles de física y de química. Podríamos incluso tener un robot que se correspondiera con nosotros hasta el nivel neuronal, podría ser de carne y hueso, aparte del cerebro. La diferencia sería que mientras que nosotros tenemos neuronas hechas de carbono, hidrógeno, proteínas y cosas por el estilo, él tendría neuronas electrónicas, pero desde el nivel neuronal hacia arriba todos los circuitos se corresponderían exactamente. Ahora bien, como no seas un ‘chovinista del hidrógeno y el carbono’ que piense que el hidrógeno y el carbono son intrínsecamente más conscientes, ¿por qué no podrías decir que este robot es una persona cuyo cerebro ocurre que tiene más metal y menos hidrógeno y carbono?” (*RTH*, 96). Si la conciencia suponemos que no queda atrapada en la organización funcional, idéntica en el robot y en mí, entonces queda automáticamente remitida a la constitución física concreta. Pero como la realización física es abierta, múltiple, la conciencia dejaría de tener ningún sentido especificable, se tornaría completa-

---

<sup>2</sup> Putnam concluirá *MM* con esta declaración tan rotunda: “La moraleja, creo yo, está perfectamente clara: ya no es posible creer que el problema mente-cuerpo es un problema teórico genuino, o que una ‘solución’ del mismo iluminaría en lo más mínimo el mundo en que vivimos. Porque está perfectamente claro que ningún adulto en sus cabales se tomaría ni por un momento en serio el problema de la ‘identidad’ o ‘no identidad’ de los estados lógicos y estructurales de una máquina—no porque la respuesta sea obvia, sino porque obviamente carece de importancia *cuál* sea la respuesta. Pero si el denominado ‘problema mente-cuerpo’ no es nada más que una diferente realización del mismo grupo de temas lógicos y lingüísticos, entonces tiene que ser justo igual de vacío y de verbal” (384).

mente absurda. Luego *de algún modo* debe estar recogida en la organización funcional...

Esta aparentemente paradójica sofisticación putnamiana del materialismo pasa por reconocer, directamente en contra del falso materialismo que, como habíamos visto, nuestro autor denomina “tradicional”, que un atributo psicológico no es lo mismo que un atributo físico, si bien, en un mundo como este en el que nos ha tocado vivir, este último *realiza* el primero, lo “implementa” (“materialmente”). Lo cual permite una especie de “libertad”, por así decir, frente a la realización física. En el sentido, por ejemplo, de que si un ser humano es un “autómata probabilístico”, entonces cualquier robot en el que corra el mismo programa será desde luego psicológicamente isomórfico con el ser humano, independientemente de aquello de que estén hechos, que en realidad no vendría al caso. Por eso se puede decir que marginar a una máquina pensante porque está hecha de metal es tan irracional e injusto como marginar a quien tiene la piel de distinto color que la nuestra (R, 407).

Humanizamos a las máquinas: es decir, con el funcionalismo se vendría a cumplir y a consumir la robotización de lo humano, el sueño de Hobbes y de La Mettrie, y ello gracias a un inteligente distanciamiento de lo groseramente “material”, que sin embargo quedará siempre en el fondo, como en reserva. Es así que en la filosofía de la mente de nuestros días se presume y escenifica la victoria definitiva del materialismo bajo la forma de fisicismo “no reduccionista”. Y en la teoría del conocimiento asistimos al despliegue de la revolución cognitiva, que arrastra los problemas de la ausencia del sujeto y el correlativo del más que enigmático sentido de la conciencia. De lo que se trata en general es de dispositivos mecánicos que reciben *inputs* (argumento de la función), los elaboran aplicándoles reglas algorítmicas, los “computan”, y emiten *outputs* verbales y conductuales (valor de la función).

Con este concepto de función llegaríamos al punto final de la historia del conocimiento, sobre todo si a lo mejor queremos ampliarla con la historia evolucionista: en último término se trata de que las máquinas asignen el valor supremo a la propia supervivencia, sólo en esta medida no sufrirán “patologías”, serán normales en suma<sup>3</sup>. (Putnam llega a establecer que las Máquinas de Turing que toma como modelo obedecen la regla de oro de la teoría económica: obra de modo que maximices la utilidad estimada). Nuestras máquinas pensantes, en suma, son agentes racionales, por eso no está muy lejos el día en que veamos avanzar hacia nosotros a los robots gritando: “¡estamos vivos!”, “¡somos conscientes!” (un robot levinasiano sin duda añadiría: “¡no me mates!”). Desde luego que será el día en que se nos vendrá encima el problema de los derechos de los robots, cuando todo dependerá, Putnam *dicit*, de una decisión, la de tomarlos como miembros de nuestra comuni-

---

<sup>3</sup> En *ML*,422, podemos leer: “Supongamos que llamamos a una función de preferencia racional *no patológica* si no asigna un peso relativamente infinito a nada excepto posiblemente a la supervivencia de la máquina misma”

dad lingüística. Será tomarla o no tomarla, y no ningún descubrimiento por nuestra parte, lo que hará o no de ellos seres vivos conscientes (R, 406).

Como dijimos, late en todo esto una idea de conocimiento muy determinada, y sin duda de conocimiento objetivo, *científico*. Ryle dejó escritas algunas cosas que sugieren que Descartes no tuvo más remedio que incurrir en el error categorial que él mismo iba a llamar “metáfora paramecánica de la mente”, porque el único sentido de “conocimiento” que había abierto la revolución científica moderna es el del conocimiento del reloj, del mecanismo. Es el formato sobre el que C. McGinn llamaba nuestra atención cuando nos hablaba de descomponer el objeto a conocer en sus partes elementales integrantes, para mostrar a renglón seguido las leyes que presiden su combinación para formar el todo del objeto en cuestión<sup>4</sup>. Ajustándose a esto sin tal vez ser muy consciente de ello, Putnam llega a decir de pasada que *el cometido de la psicología consiste en diseñar modelos mecánicos de los organismos y su comportamiento* (NMS, 435). Con este sentido preciso de “conocimiento” parece hasta inevitable que el sujeto autoconsciente se evapore y resulte muy difícil de recuperar.

## 2. Las primeras escaramuzas: contra el reduccionismo

Sin más remedio que desatender aquí los importantes desarrollos internos, por así decir, que lo constituyeron como la versión clásica de funcionalismo computacional, por ejemplo de la mano de un filósofo como Fodor<sup>5</sup>, iba a convertirse pronto el funcionalismo en la ortodoxia de la filosofía de la mente y de la psicología, en el paradigma filosófico que aspiró a fundamentar las investigaciones empíricas en la ciencia cognitiva, por mucho que desde el comienzo también se dirigieran contra él críticas de envergadura muy considerable. Desde luego, todo invitaba a disfrutar del éxito tan fulgurantemente obtenido a nivel internacional.

Pues bien, ya en un artículo publicado en 1973 en contra del reduccionismo en filosofía de la ciencia<sup>6</sup>, Putnam escribía:

En trabajos previos, he argumentado a favor de la hipótesis de que (1) un ser humano completo es una Máquina de Turing y (2) de que los estados psicológicos de un ser humano son estados de Máquina de Turing o disyunciones de estados de Máquina de

<sup>4</sup> Cfr: C. McGinn (1993).

<sup>5</sup> Para una revisión de conjunto del recorrido de Putnam y Fodor, cfr: la iluminadora obra de Luis M. Pujadas (2002), que lleva el descriptivo subtítulo “de cómo Putnam y Fodor idearon primero, y abandonaron después, el funcionalismo computacional”. Al final haremos una referencia a la conclusión crítica de la misma.

<sup>6</sup> “Reductionism and the Nature of Psychology”, aparecido originalmente en *Cognition*, 2, 131-146.

Turing. En esta sección quiero argumentar que este punto de vista estaba esencialmente equivocado, y que yo estaba entonces completamente en manos de la concepción reduccionista recién descrita (*RNP*, 433).

En un primer momento, sin duda podríamos pensar que lo que se está rechazando propiamente es el modelo concreto de la Máquina de Turing, tal vez buscando desarrollos más complejos, más realistas, del funcionalismo computacional, en el sentido maduro o “clásico”, como escribimos más arriba. Pero no estamos tratando con Fodor, el caso es completamente diferente porque en realidad tendríamos aquí un primer argumento, una primera respuesta a la pregunta de por qué Putnam, su genuino inspirador, abandonó el funcionalismo como proyecto global, y resulta ciertamente una respuesta paradójica a tenor de lo anteriormente dicho. Parece ser que en este preciso momento nuestro autor cobra conciencia de que en el fondo de la concepción funcionalista que él mismo impulsara está latiendo la característica adoración de la Física y el típico menosprecio de las ciencias de “alto nivel” que animaron a todo el programa de la Ciencia Unificada.

O sea, el único modo de disipar el innegable aire de contradicción que muestran sus palabras consiste en advertir que ahora Putnam empieza reconociendo el último poso de reduccionismo que se halla larvado en el fondo de la concepción funcionalista de los estados mentales. En definitiva, de lo que se trataba en realidad y en último término con la nueva teoría de la mente era de reducir lo psicológico a lo funcional *como única manera de conseguir la ulterior y definitiva reducción fisicalista*. Esta no se podía lograr directamente, como había quedado bien patente en el “fracaso” de los denodados esfuerzos de los materialistas de la identidad de tipos.

Nos hallamos por tanto con el funcionalismo ante el último ensayo de reducción fisicalista de lo psicológico contra el que ahora se decide a pronunciarse tajantemente el pensador americano. Y surgiría en este punto, y a decir verdad bastante inopinadamente, un elemento nuevo en todo este discurso, nuevo en el sentido de que con el se operaría un tránsito a una dimensión diferente, poco frecuentada en la tradición a la que Putnam innegablemente pertenece. En resumidas cuentas se acaba por declarar que con el reduccionismo mediato o indirecto que impulsaría al funcionalismo se incurre en el grave error de suscribir la idea de una naturaleza humana invariable, es decir, definible de una vez por todas al margen de la historia y de la cultura, y de consagrarla como universalmente válida en el núcleo mismo de la única psicología que desde ese momento se reconoce como científica, por ello mismo. En contra del naturalismo psicológico se manifiesta que “la idea de un repertorio fijo de emociones, actitudes, etc., independiente de la cultura, parece ser evidentemente cuestionable” (*RNP*, 438). Lo decisivo en este contexto es reparar en que esta denuncia putnamiana de una psicología recortada según el exclusivo patrón de las ciencias naturales se halla íntimamente vinculada con el inicial abandono del programa funcionalista.

Hay por otra parte un problema importante e interesante, de tipo más técnico, en el que vendría a sustanciarse el error naturalista señalado. Los estados de la Máquina de Turing son tales que ésta únicamente puede hallarse en uno de ellos en un momento determinado<sup>7</sup>. Además, la realidad temporal no entra aquí en consideración en absoluto, por cuanto el pasado y el futuro, o sea, psicológicamente hablando, la memoria y el aprendizaje, no vienen dados para la Máquina como adquisición de nuevos estados, sino sólo como adquisición de nueva información impresa en la cinta. Es decir, la memoria y el aprendizaje pueden hacer que la Máquina entre en un estado nuevo, pero la identidad del mismo no dependerá ni del uno ni de la otra. De manera que llegamos a la conclusión fundamental de que *un estado psicológico no puede ser en absoluto un estado de Máquina*. Lo cual quiere decir para Putnam que mi descripción como conjunto de programas y mi descripción como persona o ser humano se llevan a cabo en niveles diferentes. Aquello que podía parecer tan plausible en estados casi biológicos como el dolor, o sea, la reducción a un nudo de relaciones causales, resulta increíble por descabellado en estados tan complejos y tan mediados culturalmente como los celos, el amor o el “espíritu de competencia”, que no son en absoluto instantáneos y tienen una relación *constitutiva* con el marco social.

Para decirlo en los términos de lo que podríamos denominar “pensamiento mayor” putnamiano, si el realismo científico es lo mismo que el *imperialismo científico* (y por este nuestro autor entiende, como se pondrá bien de manifiesto en un escrito en un año posterior a *Razón, verdad e historia*, simplemente el fisicismo o materialismo tradicional, es decir, la doctrina de que todas las propiedades son propiedades físicas), entonces Putnam deja ya de ser un realista científico, por mucho que no esté de ningún modo claro que en este mismo momento se despida definitivamente del funcionalismo en filosofía de la mente. Hay observaciones bien es cierto que marginales en su gran obra de 1981 en las que se sigue declarando funcionalista, aunque tal vez sobre todo en el sentido más general de que lo verdaderamente importante para ser humano no debemos ponerlo en eso de lo que estamos hechos: aquí se localiza una “verdad” que le va a acompañar a lo largo de toda su producción futura. En este sentido podemos hablar de dualismo, pero de un dualismo, entiéndase bien, que “no es de mentes y cuerpos, sino de propiedades físicas y propiedades intencionales” (SR, 493), y que por lo tanto no lleva aparejada ninguna metafísica de interés más allá de los conceptos de superveniencia y emergentismo.

Pero hay mucho más, una vez que hemos empezado con los desacuerdos. El reconocimiento de la inadecuación radical de la Máquina de Turing para la empresa de la modelización científica de la mente humana se proyectará sin mucho tardar

<sup>7</sup> Esta sería una de las razones por las que Fodor descartó el funcionalismo de tabla de máquina, pero en el caso de Putnam esto va a significar romper con el funcionalismo en vez de desarrollarlo desde su mismo centro. Cfr: el capítulo correspondiente del libro de Carlos J. Moya (2004).



a la impugnación global de la metáfora computacional *tout court*. Así por ejemplo, en contra de la plausibilidad de los objetivos últimos de la psicología cognitiva computacional, y haciendo gala de su arrollador talento filosófico, Putnam se va a complacer en demostrar el teorema, de inequívoco aspecto gödeliano, de que “si hay una descripción computacional completa de nuestra propia competencia prescriptiva” (es decir, una descripción del modo en que *debe* funcionar nuestra mente, en el sentido lógico-deductivo o lógico-inductivo de “debe”), “entonces nosotros no podemos llegar a pensar que esta descripción es correcta cuando nuestra mente está trabajando efectivamente de acuerdo con la misma” (RR, 417). El caso es que, *por suerte o por desgracia*, siempre podemos ir más allá de nuestros propios razonamientos para comprobar si son correctos. De manera que una descripción de nuestra competencia psicológica que sea completa, y prescriptiva en el sentido recién indicado, no deja de ser una tarea imposible para nosotros, y dejarlo claro supone un verdadero jarro de agua fría para las pretensiones en verdad desmesuradas del funcionalismo computacional.

Por eso acostumbra a referirse Putnam al programa de la Inteligencia Artificial con títulos en los que se menciona la ciencia ficción, o el “mucho ruido y pocas nueces”. Porque a fin de cuentas, el objetivo declarado de la ciencia cognitiva no sería sino describir algo que es imposible de describir completamente. Está claro por una parte que hay un sentido trivial en el que cualquier sistema físico puede ser modelado como un programa de ordenador, como señalará asimismo Searle, porque bajo determinadas condiciones la conducta del sistema se puede describir con una función recursiva. Pero simular *en la práctica* la inteligencia humana, como pretende la IA cuando se nos presenta además como una disciplina fundamental del tipo de la teoría del conocimiento, también le parecerá a Putnam una empresa de todo punto imposible, desde el momento en que todo parece llevarnos a la conclusión de que no habría Programa Maestro alguno. Es decir, que lo que llamamos “inteligencia” presupone la totalidad de la “naturaleza humana”: ahí están el problema de la inducción y la necesidad de acudir al indefinible “conocimiento de trasfondo” que nos permitiría reconocer similitudes y diferencias entre casos concretos (*cf.* AI, 391-402).

### 3. Las actas de la ruptura

En el año 1988 se haría explícita y definitiva la ruptura con el funcionalismo en todas sus variedades, y justamente por el camino más filosófico: una detallada refutación de la doctrina que Putnam mismo había fundado. *Representación y realidad* se muestra pletórico de fuerza argumental, aunque tal vez por eso mismo su lectura da la impresión de ponernos ante un panorama embrollado, por así decir. Leemos

ya en la “Introducción” que los estados psicológicos no pueden ser “programas” en sentido literal, por la sencilla razón de que diferentes sistemas físicamente posibles pueden hallarse en el mismo estado mental y, sin embargo, tener distintos “programas” (*ReRe*, xiv), como quedará en evidencia en ejemplos como el de los hablantes inglés y tailandés que dicen creer que el gato está sobre la alfombra, como más abajo repetiremos. Putnam aplica así a su anterior concepción de lo mental el mismo argumento con el que ella había desactivado la pretensión central del materialismo de la identidad de tipos.

La intencionalidad, definitivamente, es irreductible a lo físico-computacional. Tampoco podemos decir, en absoluto, que es un mito, *pero ni siquiera se puede aceptar la concepción que haría de ella un fenómeno especial irreducible*. Y es que de lo que ahora se trata en realidad es de aplicar al problema de lo mental la actitud filosófica general que consiste básicamente en haber abandonado la asunción de que cada fenómeno tiene una naturaleza última y fundamental, de la que tendríamos que dar cuenta en todo caso a la manera reduccionista.

La doctrina que se busca destruir, en particular, es la teoría chomskyana que pondría representaciones semánticas, innatas y universales, en la mente-cerebro; la teoría fodoriana del lenguaje del pensamiento o *lingua mentis*, doctrinas que van dirigidas sobre todo a hacer científica a la psicología natural (*folk or commonsense psychology*), por el procedimiento de identificarla con la psicología computacional. A concepciones así Putnam las denominará en general “mentalismo”, y sería evidente para él y para casi todos que el funcionalismo presupone necesariamente el mentalismo en este preciso sentido. Pero el mentalismo no puede ser correcto porque contradice los hallazgos más definitivos de la teoría del significado. En primer lugar, el significado es holístico y en parte normativo y sensible al contexto, así que nuestros conceptos dependerían de nuestro entorno físico y social de un modo que desde luego la evolución no podía prever cuando completó la obra de nuestros cerebros hace unos 30.000 años. Los significados son en resumidas cuentas entidades tan históricas como las personas o las naciones.

Además, nuestras prácticas interpretativas recíprocas, que son lo único que puede determinar en último término lo que entendemos por sinonimia y co-referencialidad, pondrían en juego a la inteligencia que podríamos denominar general, justamente la que no tiene nada que ver con la modularidad fodoriana: es de todo punto insostenible seguir pensando que dos palabras son sinónimas cuando están asociadas con la misma *representación mental* en la mente de los hablantes que la usan. Antes bien hemos de comprender que la asociación de las representaciones con su referente es contingente y susceptible de cambiar cuando el mundo y la cultura cambian. Por eso Putnam nos aconseja desconfiar de quienes nos presentan un ámbito de representaciones puras, sin asomo de ambigüedad ni sinonimia. “Los significados no están en la cabeza” sino que la referencia vendría social o físicamente fija-

da, como nos había ilustrado ya en 1975 el experimento mental de la Tierra Gemela. “El significado es de naturaleza interactiva. El entorno mismo juega un papel a la hora de determinar a qué refieren las palabras del hablante o de la comunidad” (*ReRe*, 36).

En definitiva, y en resumen, si “el funcionalismo no funcionó”, como reflexiona ahora Putnam, fue fundamentalmente porque las actitudes proposicionales no son estados del cerebro humano que se puedan considerar al margen del entorno social y físico. *Por lo tanto, no son estados funcionales*. Y, en general, lo que definitivamente no existe, y esta es la lección esencial que debemos extraer del episodio funcionalista en filosofía de la mente, es ese “algo básico”, definible en términos no intencionales y en todo caso identificable por procedimientos científicos, a lo que habría que reducir las actitudes proposicionales para ir a dar en aquello que justamente es capaz de explicar que un campesino tailandés que cree que el “meew” está sobre la alfombra y un hablante castellano que cree que el “gato” está sobre la alfombra, se hallen ambos en la misma condición psicológica. Simplemente estamos buscando en el lugar equivocado. Una definición funcionalista cabal de la sinonimia y la co-referencialidad tendría que formalizar nuestros juicios de razonabilidad, que en todo caso tienen carácter intuitivo. Y esto es imposible porque para ello haría falta formalizar la naturaleza humana como un todo.

La teoría de la interpretación en la filosofía analítica del presente se constituye sobre todo a partir de las obras de Quine y de Davidson, y de éstas sería sin duda deudor el planteamiento específico de nuestro pensador que estamos revisando. Lo que Putnam nos está descubriendo son algunas de las razones centrales por las que resulta imposible definir, empíricamente y de modo finito, las relaciones y propiedades intencionales en términos físico-*cum*-computacionales. Jamás podremos llegar a poseer el Algoritmo Maestro para la Interpretación porque es de todo punto imposible llevar a cabo la “valoración” de todos los posibles modos de conceptualización de los seres humanos en todos sus lenguajes y todas sus culturas, y en todas sus formas de fijación de creencias. Por no decir nada, evidentemente, de las posibilidades inimaginables de los lenguajes y las culturas no humanos. Por ejemplo, sería necesario comprender los esquemas conceptuales de culturas más avanzadas y de seres más inteligentes que nosotros, lo que por definición es imposible. *No cabe una teoría general de la interpretación*, en suma.

Vistos computacionalmente, es decir, como dispositivos para ‘confirmar’ o ‘aceptar’ aseveraciones sobre la base de la estimulación sensorial, y para dar ‘respuestas motoras’, los seres humanos difieren enormemente unos de otros y de una cultura a otra. Aunque todos los seres humanos sean computadoras del mismo tipo en el momento de su nacimiento, no sucede que todos los humanos adultos tengan que atravesar la misma secuencia de estados cuando fijan una creencia que nosotros traduciríamos a nuestro lenguaje por la frase ‘Hay muchos gatos en el vecindario’ (*ReRe* 104).

Por eso el funcionalismo no funciona, porque no hace justicia a los fenómenos. Pero descubriéndolo hemos aprendido mucho (105). Lo que ha ocurrido es que lo que podríamos llamar teoría putnamiana del significado, densa y compleja, haría de raíz inviable el mentalismo, y con ello acabaría de frustrar todas las pretensiones del funcionalismo. Putnam subrayaría, como de sobra sabemos, la “teoría causal” del significado, aunque sin llegar a reconocerla como propia en sentido cabal, teoría que Burge no dudaría en aplicar directamente a la filosofía de la mente. Y a su lado, el *meaning holism* de estirpe quiniana, para culminar poniendo mucho énfasis en la normatividad del significado tal y como se transparenta en toda la complejidad de nuestras prácticas interpretativas cotidianas: ni el principio de caridad ni la mencionada “razonabilidad” se podrían caracterizar computacionalmente en absoluto<sup>8</sup>.

Como antes apuntamos, esta refutación del funcionalismo, que se nos presenta en la obra del 88 verdaderamente tumultuosa, podríamos decir, constituye el resultado de la aplicación de ese “temperamento filosófico” que Putnam bautizó con el nombre de “realismo interno” al caso concreto de la intencionalidad, que es el núcleo duro de la filosofía de la mente moderna. La reducción fisicalista y computacional de la intencionalidad, o, lo que es lo mismo, el ensayo de dar una “explicación científica” de las propiedades intencionales, todo el programa de la naturalización de la intencionalidad, sólo tendría sentido sobre el telón de fondo del realismo metafísico que Putnam rechazará con tanto ahínco. *Y porque este realismo metafísico no es filosóficamente sostenible, el proyecto mismo de naturalizar la intencionalidad está destinado a fracasar en sus propios términos de realismo científico.* Ya habíamos sugerido que por esto no resultaría tampoco viable la otra alternativa, la representada por la línea Brentano-Chisholm, que toma la intencionalidad como un “hecho primitivo” irreducible. Se trataría, en definitiva, de abandonar de una vez por todas la concepción del Ojo de Dios, para pasar a centrarnos en nuestro mundo y en nuestras prácticas humanas. “La verdad no trasciende al uso”, sería a partir de aquí el lema del filósofo.

El Algoritmo Maestro es el Ojo de Dios, la vigilancia completa de la absoluta transparencia y el determinismo, o la probabilidad controlada, de todas nuestras acciones y pensamientos. La destrucción sistemática del funcionalismo y la denuncia del realismo científico, en tanto realismo metafísico en último término sin sentido, constituyen sin ningún género de dudas operaciones profundamente filosóficas, desde el punto de vista práctico, que a fin de cuentas es el que decide, porque tienen efectos radicalmente liberadores: nos permiten eludir la inevitabilidad del marco conceptual mecanicista que nos tiene atrapados, ontológica y epistemológicamente, desde por lo menos el siglo XVII. La idea de la mente como máquina de computar hunde sus raíces en el materialismo de ese siglo y del siguiente, el que

---

<sup>8</sup> Si queremos un resumen más ligero y más claro que *ReRe*, es muy recomendable la lectura de “Why Functionalism Didn’t Work”, del año 1992.

Putnam denomina “tradicional”, y en la ciencia mecanicista que le sirvió de plataforma. “Por ejemplo, Hobbes sugirió que el pensamiento hay que denominarlo de la manera más adecuada ‘cálculo’, porque realmente consiste en la manipulación de signos de acuerdo con reglas (análogas a reglas de calcular), y La Mettrie escandalizó a su época dictaminando que el hombre es sólo una máquina (*L’Homme Machine*). Estas ideas, como era de esperar, se asociaron con el materialismo” (*RPh*, 3-4), porque para el materialismo la totalidad del ser humano es el cuerpo vivo, y el cuerpo vivo se concebía simplemente como un sistema que obedece las leyes de la física newtoniana.

Por eso, y también en la medida en que todavía casi nadie pensaba que las singularidades de la mecánica cuántica pueden tener importantes implicaciones para la filosofía de la mente (abandonar el funcionalismo no tiene por qué ser lo mismo que abandonar el materialismo), el marco conceptual mecanicista venía teniendo una potencia colosal a la hora de pensarnos a nosotros mismos como partes de la naturaleza. Definitivamente, a lo largo de centurias el humano es una máquina, aunque sólo sea porque todo el universo lo es. Putnam se apresura a matizar que al menos metafóricamente, pero ya sabemos de la fuerza determinante que han tenido las metáforas en el pensamiento científico y filosófico. La Máquina de Turing ejerció sobre nuestro filósofo tan mayúsculo atractivo porque suponía una magnífica manera de hacer del materialismo algo preciso, matemáticamente especificable, por mucho que, a decir verdad, una máquina no tenga por qué ser una Máquina de Turing, es decir, no todos los sistemas físicos posibles tienen por qué ser describibles por una función recursiva. Pero pareció a muchos mecanicistas que una aproximación lo suficientemente satisfactoria a la “trayectoria real” de un ser humano puede predecir sus *outputs*, verbales y conductuales, durante el tiempo finito en que transcurra su vida, puesto que éstos son describibles por parámetros físicos especificados sólo a un nivel de precisión macroscópico, y esa aproximación puede ser una función recursiva. De manera que lo que un humano haga y diga puede ser pronosticado por una Máquina de Turing, o sea, puede ser simulado.

Parece no ser infrecuente que las imágenes metafísicas sugeridas por las revoluciones científicas se vengán a confundir con la ciencia misma, de manera que acaban disfrutando de un enorme prestigio, y ejercen sobre todos nosotros un poder de seducción desmesurado (como es natural, nadie en su sano juicio se atrevería a desafiar a “la ciencia”). Pues así ocurrió con la imagen del hombre-máquina, nacida como *ideología* a partir de la física de Galileo y de Newton, y consumada mucho después por la revolución de los ordenadores y la ciencia cognitiva. “Una imagen nos mantuvo cautivos”, para recordar aquí el célebre estribillo wittgensteiniano con el que tiene mucho que ver lo que podríamos llamar la apostasía de Putnam.

Y no se trata sólo, o no se trata exactamente, de que los intentos de reducir la intencionalidad a lo no intencional, físico-computacional, no hayan tenido éxito

hasta el momento, es que lo intencional se nos aparece en todas partes cuando nos ponemos en marcha para describir lo no intencional. Cuando nuestro autor pasa a formularse la siguiente pregunta decisiva, convoca ante nosotros sus lectores a la mismísima “filosofía de la sospecha”:

Una cosa que me interesa en este libro [se está refiriendo a *RPh*] es por qué somos tan reacios a admitirlo [que la intencionalidad está por doquier]. ¿Qué es lo que revela sobre nuestra cultura y toda nuestra forma de pensar que sea tan difícil para nosotros admitirlo, y cómo podría ser una filosofía que comenzara a abandonar todos los sueños reduccionistas? (*RPh*, 59).

El materialismo (en filosofía de la mente) no sólo resulta insatisfactorio por lo menos en su forma actual, es que además su radical insuficiencia acabaría delatando que en realidad constituye la ideología de la sociedad dominante occidental. La confianza acrítica en la ciencia y la tecnología, la convicción ciega de que todos los problemas filosóficos y humanos en general acabarán encontrando resolución o disolución en el avance futuro y victorioso de las ciencias particulares: lo que propiamente ha pretendido Putnam es insistir en la advertencia de que, simplemente, todo esto es algo propio de ilusos carentes de reflexión que caminan sin ver muy bien por donde van.

#### 4. Esbozo de una imagen alternativa

Vamos a echar un vistazo, para concluir toda esta revisión más bien apresurada, al espléndido conjunto de conferencias y ensayos reunidos en 1999 bajo el título de *The Threefold Cord. Mind, Body and World*<sup>9</sup>, puesto que en sus páginas encontraremos la al parecer última palabra de Putnam sobre los problemas centrales de la filosofía de la mente, en primer lugar en todo lo referente a la coronación de su crítica al funcionalismo como ejemplo más potente y actual de reduccionismo materialista, pero también, en segundo término, en lo que hace a la exposición de las tesis positivas del filósofo sobre toda la cuestión de lo mental, tesis entreveradas con el último avatar de la evolución general del pensamiento putnamiano, esa posición que William James denominara “realismo natural del hombre corriente”.

Para comenzar, nos las vemos con la destrucción de la teoría computacional de la mente en su punto culminante, donde ya no va a ser posible ninguna clase de recitificación que suponga una marcha atrás. Porque ahora se insiste no en que las

---

<sup>9</sup> Esta publicación recoge en lo esencial *The Dewey Lectures* (“Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind”), de 1994, y *The Royce Lectures* (“Mind and Body”), de 1999.

expectativas levantadas por dicha teoría no puedan tener cumplimiento, sino en que, mucho peor desde luego, carecen de un sentido que se pueda especificar. El programa del funcionalismo computacional, dirigido a la identificación de nuestras diversas actitudes proposicionales con los denominados estados computacionales, no puede considerarse por tanto un programa científico, aunque a muchos les siga pareciendo lo contrario, sino que no pasaría de ser, simplemente, una *quimera*. A poner al descubierto la impostura se habría venido aplicando, como hemos ido viendo a lo largo de este trabajo, una parte nada desdeñable del esfuerzo intelectual de nuestro filósofo.

Es evidente que las propiedades formales de los estados computacionales son bien diferentes de las de los estados psicológicos propiamente tales. Por ejemplo, “un formalismo para la teoría de la computación *define implícitamente* todos y cada uno de sus estados computacionales por la totalidad de sus relaciones computacionales (de sucesión o de sucesión probabilística, por ejemplo) con todos los demás estados del sistema en cuestión” (*ThC*, 35) Así que la definición implícita simultánea de todo el conjunto de estados es lo que individualiza a cada uno de ellos como tal, distinguiéndolo de todos los demás. Pero no hay teoría psicológica que defina o individualice sus estados de una manera ni remotamente semejante, en absoluto. Y si nos empeñamos en imaginarnos una que sí que funcione como un sistema computacional, *no sabemos muy bien en qué demonios estamos pensando*. Por eso Putnam nos pone en guardia contra el hábito de dejarnos llevar por las situaciones que en realidad sólo tienen su lugar en la ciencia ficción.

La misma idea de reducir la psicología del sentido común, nuestro entramado de prácticas de atribución de estados mentales, a física *cum* ciencia computacional, es hoy por hoy una idea sin ningún contenido. No disponemos ni de la más mínima noción del aspecto que podría corresponder a la teoría a la que se supone vamos a hacer la reducción, y sólo tenemos una muy problemática de la teoría que vamos a reducir. En general, como ya vimos, el reduccionismo se ve enfrentado a obstáculos insalvables, como los que se derivan del externismo semántico o de la múltiple realizabilidad de lo mental. Pero además, observa en este momento Putnam, tenemos que concluir que el reduccionismo fisicalista es *incoherente*, incluyendo en él desde luego a su anterior funcionalismo. No sabemos en el fondo de qué estamos hablando cuando nos preguntamos si *pensar en la belleza de las rosas de A Coruña* es o no es “reducible a” (o “nomológicamente coextensivo con”) una propiedad física. Porque, en definitiva, a la noción de “identidad” no se le ha dado en contextos como este ningún sentido definido (85). No se trata, no se puede tratar, al contrario de lo que tantos aseguran, del modelo de identificación teórica derivado de la famosa reducción de la termodinámica a la mecánica estadística (“la temperatura *es* la energía cinética molecular media”), porque en este caso tanto la teoría reductora como la reducida cuentan con un cuerpo de leyes bien definido. Pero en nuestro

caso nadie tiene la más remota idea de cómo sería un formalismo en el que se pudieran escribir programas que redujeran la psicología del sentido común. Incluso llega a decir Putnam que no tenemos un significado determinado que darle a las propiedades computacionales.

Yendo hasta al final por esta senda destructiva, que compromete en primer lugar al materialismo en cuanto tal al cuestionar directamente a su último y más sofisticado representante, llegamos a ese punto culminante en que el filósofo se dispone a embestir contra la inteligibilidad misma de la noción de “superveniencia”, la que representaba el mínimo de materialismo que todos estaríamos dispuestos a aceptar en filosofía de la mente, en realidad el mínimo que *tendríamos que aceptar*. Se convierte ahora en blanco de la artillería de Putnam, en definitiva, la noción de “estado psicológico interno” o “contenido estrecho”, cuya consistencia científica se le antoja verdaderamente frágil<sup>10</sup>. El internismo intencional se compadecería a la perfección con nuestra intuición central de la causalidad mental, en tanto coherente con la tesis general de la superveniencia. Justamente en este flanco va a irrumpir el descreimiento de nuestro filósofo, con un argumento dirigido en contra de la plausibilidad misma de toda la posición mayoritaria: “¿Tiene realmente sentido suponer que lo que hago cuando adscribo a alguien una creencia, o un deseo, o un recuerdo, es *involucrarme en una pequeño fragmento de especulación sobre las causas internas de su conducta?*” (114).

El funcionalismo computacional daría por bueno, por consiguiente, que la psicología del sentido común es verdaderamente una teoría explicativa. Ahora bien, la noción misma que le permite hacer esto, la de superveniencia, se tambalea cuando consideramos una vez más cómo se individualizan las creencias, por ejemplo, siempre en dependencia del lenguaje que hablamos y en el que pensamos. Lo que el hablante cree en particular en un momento determinado se halla en función de su sistema global de creencias. A la vez, lo que el hablante cree en particular en ese momento determinado estará en función de su entorno físico y social. En cualquier caso, cuando decimos que alguien cree que X, lo que decimos tiene sentido en relación con el contexto en que lo decimos. Por eso la creencia no puede ser en absoluto un estado interno, esto es, un estado que sobrevenga a una propiedad especificable en términos computacionales y neurológicos. De manera que la misma posesión de conceptos no puede ser un estado funcional, y darse cuenta de esto significó el principio del fin para el funcionalismo de Putnam, se nos recuerda ahora en estas páginas (118):

Todo lo que tenemos para apoyar la idea de que la creencia es un ‘estado interno’ es

---

<sup>10</sup> Como es bien sabido, el “contenido estrecho” de los estados mentales supone la respuesta fodoriana inicial al desafío representado por el externismo semántico para la noción de superveniencia, y en general, como vimos, para toda la teoría computacional de la mente.



ciencia ficción—o mejor, afirmaciones que tendrían un papel legítimo como diversión si aparecieran en una obra de ciencia ficción. Pero cuando estas mismas afirmaciones aparecen en los escritos de los filósofos de la ‘ciencia cognitiva’ son profundamente confusas, porque se proponen tener la clase de uso que tiene una hipótesis científica sin que se les haya dado ningún contenido científico en absoluto (119).

Se vuelve a apuntar la sospecha, a la que más arriba dimos la importancia que a nuestro parecer merece en el discurso putnamiano, de que toda la aparentemente decisiva cuestión del estatuto de la intencionalidad, o sea del lugar que puede ocupar el lenguaje intencional en la ciencia, dependería en el fondo de una noción “completamente ideológica” de ciencia, y por lo tanto lo mejor que podemos hacer es dejarla definitivamente de lado. En conexión con esto, y en lo que hace a las pretensiones del materialismo eliminativo, Putnam se limitará a dejar consignado que aprendió de los pragmatistas, y de Wittgenstein, la importante lección de que es preciso tomarse en serio, *metafísicamente* en serio, los modos de hablar que resultan absolutamente indispensables en nuestra vida de todos los días. Valga con este par de escuetas notas por lo que respecta al problema del estatuto científico de las creencias y los significados en general.

Dando un paso adelante, la crítica a la que aquí asistimos pondría de manifiesto la falta de sentido de la posición naturalista en general, la misma idea de *corrección psicofísica* tomada esta vez como su mínimo universalmente aceptable y prácticamente indiscutible. Y no se trata de defender el dualismo, en absoluto, como después para terminar veremos, sino de mostrar que la imagen que hace de las características psicológicas estados internos que tienen que ir correlacionados o no correlacionados con lo que sucede en el interior del cuerpo es, literalmente, una idea que se nos deshace entre las manos. Para ser más claros en lo que toca a la posición definitiva del filósofo:

Los materialistas tienen razón al insistir en nuestra naturaleza incorporada (*embodied*), tienen razón al insistir en que la conexión entre la mente y el cuerpo es demasiado íntima como para que tenga algún sentido hablar de ‘espíritus desencarnados’, ‘novios automáticos’, la ‘tribu de los sin alma’, y otras cosas por el estilo. Pero se equivocan cuando su científicismo les lleva a afirmar que sólo podemos pensar nuestra mente como algo que actúa *en y a través* de nuestros cuerpos *con la condición* de reducir los términos de la psicología natural a los términos de la química, la física, la neurología, la ciencia computacional, etc. (149).

Esta denuncia también se aplicaría a los denominados *mysterians*, los que desesperan de poder responder a la pregunta por la naturaleza de la conciencia, porque tras su escepticismo no se esconderían sino las mismas ansias reduccionistas de siempre: la cuestión “¿cuál es la naturaleza de X?” tiene hoy para muchos sólo un

sentido preciso y digno de discusión: “¿cómo se puede reducir X a la física, la química, la neurofisiología, etc.?”. Es esta pregunta misma la que tendríamos que asegurarnos que tiene sentido en cada caso concreto, como el de la conciencia por ejemplo. Porque al parecer de Putnam no cabe duda de que hemos estado demasiado tiempo en poder de la imagen cartesiana que nos hizo pensar que el dualismo y el materialismo de la identidad eran alternativas *con sentido*, y eran las *únicas* alternativas. Reaccionando por fin contra esta imagen, si lo pensamos bien, nos daremos cuenta de que “decir ‘la ciencia tal vez algún día encuentre el modo de reducir la conciencia (o la referencia, o lo que sea) a la física es, *aquí y ahora*, lo mismo que decir que tal vez la ciencia haga algún día no sabemos qué no sabemos cómo” (173). Lo cual no significa de ningún modo resistirse, u obstinarse en dificultar la brillante investigación de las bases físicas de nuestra vida mental a la que venimos asistiendo en las últimas décadas. Todo lo contrario: el reduccionismo sólo lleva a desenfocar la comprensión de los problemas empíricos.

En segundo término, y a partir de todo lo anterior, ¿qué imagen positiva de lo mental nos ofrecería nuestro pensador como salida del sinsentido reduccionista? Sabemos ya dos cosas como mínimo, que nada tendría que ver con el dualismo (hay simplemente un hecho: que ya no podemos encontrar creíble el dualismo cartesiano), por una parte, y que estará en la línea de lo que James denominaba realismo natural del hombre corriente, por otra. El error de siglos, para decirlo de una vez, habría consistido en pensar la mente como *cosa* (de forma que, “científicamente”, la tenemos que identificar de algún modo con el cerebro). Es decir: habría según esto un ámbito subsistente de experiencias y fenómenos mentales que tienen lugar en la mente/cerebro, en el interior de la cabeza humana. Y será en primer lugar en el terreno de la filosofía de la percepción donde Putnam tomará la decisión de romper con el dilema “teoría de la identidad/dualismo”: haciendo de las experiencias sensoriales experiencias de aspectos del mundo por parte de un ser vivo, y no afectaciones pasivas de un objeto llamado “mente”, llegamos a darnos cuenta de que se trata de un dilema incoherente que carece por completo de sentido (hay que rechazar No-Autómata de la misma manera que rechazamos Autómata): “Hablar de la mente no es hablar de una parte inmaterial de nosotros sino más bien un modo de describir el ejercicio de ciertas aptitudes y capacidades (*abilities*) que poseemos, aptitudes que ‘sobrevienen’ sobre las actividades de nuestros cerebros y sobre todas nuestras diversas transacciones con el medio, pero que no tienen que ser explicadas reductivamente usando el vocabulario de la física y de la biología, o incluso el de la ciencia computacional (37-38). Siguiendo la línea apuntada en 1994 por John McDowell<sup>11</sup>, Putnam llegará a proponer con toda conciencia este “relineamiento metafísico” crucial, que consagrará definitivamente nuestra familiaridad con una pluralidad de recursos conceptuales y de diferentes vocabularios mutuamente irre-

<sup>11</sup> En *Mind and World*.

ducibles, o sea, que privilegiaría el mundo de las prácticas para no ir a parar en los desvaríos monistas. Si Putnam rechaza Autómata, y desde luego esto es lo que aquí más nos ha interesado, no es en absoluto porque defienda No-Autómata.

Realineamiento metafísico que esencialmente implica abandonar la vieja idea de *representación* como *interfaz* entre nosotros y aquello sobre lo que pensamos, entendiendo desde luego que el realismo natural no tiene por qué prescindir por ello de la *actividad de representar* como una de las múltiples actividades en las que estamos comprometidos, lo que sería absurdo. El realismo directo o natural consiste desde este punto de vista en el reconocimiento de que no habría ninguna necesidad de poner representaciones internas en el pensamiento y la percepción (no tendría ningún valor explicativo hacerlo, además). Tal vez por esta razón en estos escritos reunidos y publicados en 1999 Putnam le da una importancia crucial a la filosofía de la percepción en relación con nuestro tema de la filosofía de la mente. Y es que sólo abandonando la imagen de la percepción como mediada por un conjunto de representaciones en un teatro interior, podremos salir, por fin y definitivamente, de ese “indefinido reciclaje” de posiciones y teorías que no funcionan en filosofía de la mente, un reciclaje que ha dudado al menos cuatrocientos años (102).

Antes de terminar conviene hacer presente lo que nuestro autor debe a pensadores como Dewey en esta concepción no reduccionista de la actividad mental, la concepción según la cual “hablar de nuestra mente es hablar de capacidades que tenemos y que involucran al mundo, y de actividades en las que nos ocupamos”. Para decirlo en las propias palabras del pragmatista americano:

‘Mente’ es primariamente un verbo que denota todos los modos en que tratamos consciente y expresamente con las situaciones en las que nosotros mismos nos encontramos. Por desgracia, una influyente manera de pensar ha cambiado las modalidades de la acción en una sustancia subyacente que realiza las actividades en cuestión. Ella ha tratado a la mente como una entidad independiente *que* atiende, tiene propósitos y preocupaciones y recuerdos.<sup>12</sup>

Del hecho de concebir la mente como cosa habría brotado la posibilidad del materialismo reduccionista.

## 5. Conclusión: cambiando de nivel

Como era de esperar, las nuevas posiciones de Putnam en todo lo relativo a la “filosofía de la mente”, tal y como cuajaron en forma al parecer definitiva en las *Dewey Lectures* de 1994 y en las *Royce Lectures* de 1999, las que hemos recogido

---

<sup>12</sup> Fragmento de *Arte y experiencia* citado en *ThC*, 170.

en el apartado anterior, no han tenido que aguardar mucho para encontrar las más enconadas críticas del mismo campo naturalista, el que se identifica simple y llanamente a sí mismo con el sentido común iluminado por el desarrollo científico de vanguardia. El trabajo ya citado de L.M. Pujadas es bien representativo en este aspecto, por cuanto termina su repaso de la evolución putnamiana en filosofía de la mente manifestando que, para el que escribe, el viejo argumento realista de que sólo el realismo “puede explicar el éxito de la ciencia sin recurrir a un milagro”, argumento defendido en primer lugar por el propio pensador americano en los años setenta, “no ha perdido ni un ápice de su fuerza” (Pujadas 2002, 233-234). Y en lo relativo al problema mente-cuerpo, la teoría de la identidad que asimila el propio funcionalismo en forma de *token identity theory* vendría a ser sin duda la respuesta “definitiva” (2002, 275): el análisis conceptual ha ido logrando en los últimos decenios la disminución de la reacción de extrañeza de nuestro sentido común ante los inequívocos avances de las neurociencias. Cuando Pujadas nos revela al final su *credo* nos encontramos en definitiva con el panorama de la neurociencia cognitiva que se ha incorporado al saber filosófico de nuestro tiempo (2002, 283). A decir verdad, lo único que le resulta a este autor problemático es lo que no entra a formar parte del núcleo de la respuesta científica a la cuestión. La verdad es que para llegar a esto más nos hubiera aprovechado estudiar los manuales y los *bestsellers* actuales de las neurociencias.

Respuestas a Putnam que no se limitarían a señalar con dedo firme “el estado presente de nuestro conocimiento científico”, es decir, respuestas más centralmente filosóficas, las podemos encontrar, por tomar un ejemplo también muy significativo, en un trabajo de A. Marras dedicado a refutar, sobre todo, la declaración según la cual ni el materialismo de la identidad ni el de la superveniencia serían capaces de dar verdadero sentido a lo que enuncian.

Para empezar, en lo referente al problema mente-cuerpo, Marras enfatiza la urgencia de cerrar de una vez el capítulo de las insistencias intuitivas del sentido común: *lo indiscutible* es el hecho es que hay una *probabilidad aplastante* (“*it is overwhelmingly likely*”) de que la propiedad de ser consciente “sobrevenga” a ciertas propiedades físico-funcionales del cerebro, lo cual debe ser afirmado rotundamente con independencia de que nosotros no entendamos en absoluto (¿por lo de ahora?) la manera en que estas últimas propiedades “determinan” a las primeras (Marras 2001, 409). (O sea, que no entendemos *cómo* determinados sucesos cerebrales *pueden ser* sucesos conscientes). Pues, en efecto, no hay más remedio que reconocer que (¿todavía?) tenemos con nosotros el “problema duro de la conciencia” (Chalmers) o la “brecha explicativa” (Levine). A pesar de ello, o a lo mejor precisamente por ello, el programa filosófico de la naturalización de la intencionalidad tiene que seguir adelante, renunciando tal vez al reduccionismo estricto de la identidad para limitarse a establecer las condiciones naturalistas sobre las que las ads-

cripciones de propiedades intencionales *sobrevienen* fiablemente (2001, 412). Pero el problema es que Putnam ha cuestionado con mucha seriedad que haya una diferencia real entre un materialismo reduccionista referido a lo mental y otro que no sea reduccionista, es decir, lo que ha cuestionado, como sabemos, es que los términos “identidad” y “superveniencia” adquieran un sentido inteligible y especificable cuando los usamos para fijar la “relación” que se daría entre mente y cerebro. Lo cual me parece que tiene mucho que ver con lo *indiscutible* del hecho ya referido de que no entendemos (¿de momento?) cómo puede ser lo cerebral consciente. Porque si no conseguimos entender esto parece evidente que lo que precisamente no entendemos es qué demonios puede significar que la intencionalidad se *identifique* con o *sobrevenga* en lo físico-funcional. Que la neurociencia nos enseñe una infinidad de nuevos descubrimientos puede derivar, desde luego, en lecciones paralelas, por ejemplo, lo que por lo demás ya sabíamos al reflexionar sobre el hecho de que algunos golpes en la cabeza conllevan la pérdida de la conciencia. La intencionalidad depende, *de algún modo y entre muchas otras cosas*, de mecanismos cerebrales, eso desde luego es otro hecho, pero la filosofía del materialismo referente a lo mental desborda con excesiva amplitud lo que semejante hecho permitiría conjeturar. Más que ciencia es *ideología*.

Pero la estocada mayor de Putnam se ha dirigido en el fondo contra la consistencia conceptual de la noción misma de estado psicológico en cuanto “estado interno”. Que un deseo o una creencia sean estados internos, tal y como lo necesitan desde luego los teóricos de la superveniencia, choca con el modo en que se individualizan el deseo o la creencia (de modo holista y siempre contextual, para empezar). El reto mayúsculo del externismo semántico para toda teoría materialista de lo mental, en segundo lugar, dígame lo que se diga, resulta tan simple de comprender como decir que las creencias se individualizan por su contenido, y el contenido es comúnmente individualizado por condiciones “fuera de la cabeza”; por lo tanto las creencias...no son estados internos. Esta claridad cristalina del dos y dos son cuatro la intenta enturbiar Marras apoyándose en la autoridad de Davidson, con un ejemplo como el siguiente: “el hecho de que identifico que alguien tiene fiebre, una temperatura corporal de más de 37°C, en términos de condiciones externas, la lectura de un termómetro, no demuestra que la fiebre no sea un estado interno de mi cuerpo” (2001, 420). Lo mismo intentaba alcanzar Davidson con el ejemplo de la quemadura de la piel excesivamente expuesta al sol (1987, 451). ¡Así que una creencia es como una quemadura de la mente-cerebro o como un cerebro en estado febril! Ambos atacarían lo obvio para defender la tesis filosófica central del materialismo de lo mental, y en el caso de Davidson, además, lo que quedaría de privilegio epistemológico de los propios contenidos mentales, dado el externismo semántico. Por lo demás, y para decirlo en los mismos términos del citado Marras, tan significativos, habría otras muchas “estrategias en el mercado” para preservar la tesis de la

superveniencia de la amenaza que supone la filosofía última de Putnam. Pero me parece que ninguna de ellas es otra cosa que un producto de marketing barato.

Cuando nos persuadimos de lo escasamente convincente que ha sido la respuesta naturalista a las últimas posiciones de Putnam; de hasta qué punto, para reivindicar lo indiscutible, que desde luego no hace ninguna falta reivindicar, ha pretendido obviar o enturbiar otros aspectos asimismo indiscutibles que sí necesitarían, en cambio, de nuestro especial cuidado, sobre todo entonces volvemos a pensar en la pregunta de por qué a nuestra cultura tecnológica se le ha ido tanto tiempo y esfuerzo en intentar reducir o simplemente negar la intencionalidad que todo lo penetra. Es decir, una visión crítica del materialismo de la identidad en la filosofía de la mente, sobre todo y muy especialmente en su arrolladora versión del funcionalismo computacional, nos transporta muy naturalmente al planteamiento desenmascarador de la lucha contra la ideología. Somos conscientes de que con esto estamos aventurándonos en una dislocación o cambio de nivel del planteamiento que puede parecer no sólo inconveniente sino filosóficamente ilegítimo. Pero casi estamos convencidos de que sólo resulta decididamente *inconveniente*... (para algunos o muchos), porque cuando no se obtiene una respuesta en un terreno tenemos toda la justificación del mundo para buscarla en otro diferente. Eso no lo vamos a hacer aquí, de momento nos preguntaremos simplemente por qué ha venido pareciendo tan convincente e irresistible el modelo computacional a tantos psicólogos, neurocientíficos y filósofos de la mente, por qué el programa de la naturalización de la intencionalidad, en vez de revelarse como la consumación de la *barbarie*, para decirlo con el término elegido por el pensador francés Michel Henry, o sea, como la definitiva destrucción de la cultura, ha venido pasando por significar, para tantos autores brillantes, la verdadera razón de ser de la empresa filosófica, reconocida como tal a la altura de los tiempos.

En esta línea de la sospecha como genuina actitud filosófica mucho se ha reflexionado ya sobre las condiciones de posibilidad y las repercusiones, sociales, culturales y económicas, del *hombre de Turing*: ahí tenemos el brillante trabajo de Bolter, aparecido ya hace más de veinte años. Lo que viene a resumirnos el núcleo de este tipo de reflexiones críticas es ante todo la idea de que hemos terminado por contemplarnos a nosotros mismos con la lente de la *racionalidad tecnocrática* que ha marcado desde hace ya mucho tiempo nuestra relación con el mundo, y esto ha sido principalmente lo que habría determinado el carácter irresistible de la metáfora computacional. Vivimos en la era del ordenador, y su estilo de “pensar” es solidario del estilo de pensar de la elite tecnocrática que ha adquirido una posición absolutamente dominante en nuestra civilización, que es la del nihilismo pasivo del que Nietzsche había llamado “el último hombre”.

Para el punto de vista tecnocrático, que ha sido caracterizado como el del “observador analítico” (Evans 1993), lo que se ganaría sobre todo con la lente com-

putacional que consagra definitivamente el materialismo en la filosofía de la mente sería la absoluta transparencia “desmontable” de lo humano, como conjunto de procesos previsible y mensurables, es decir, perfectamente controlables. En todo ello también juega su papel esa peculiar tendencia occidental a la auto-humillación, que pasa por ser, a los ojos de tantos enigmáticos masoquistas, una de las huellas anímicas más saludables del verdadero progreso científico (ya se sabe: Copérnico, Darwin, Freud, y, al final, el ordenador). El misterio es absolutamente expulsado de nuestra existencia, para universal regocijo de administradores, empresarios y tecnócratas en general. En lo que respecta a los paradigmas psicológicos, del conductismo podíamos salvar, aunque no se lo quisieran creer los conductistas, nuestra propia intimidad: con la psicología cognitiva empezó el definitivo asalto, la colonización de la mente (Evans 1993, 60).

Terminaremos con unas palabras del mencionado Henry, que aciertan a subrayar con total claridad contra quién hemos estado disparando filosóficamente en este trabajo: “No es el saber científico, insisto, lo que pongo en cuestión, es la ideología que se le une a éste hoy en día, según la cual es el único saber posible, que debe eliminar a todos los demás saberes” (Henry 1987, 43).

## Referencias bibliográficas

- BOLTER, J.D.: *Turing's Man: Western Culture in the Computer Age*. Chapel Hill, The University of North Caroline Press, 1984.
- DAVIDSON, D.: “Knowing One's Own Mind”, *Proceedings and Addresses of the American Philosophical Association*, 61, 1987, pp. 441-458.
- EVANS, F.: *Psychology and Nihilism. A Genealogical Critique of the Computational Model of Mind*. State University of New York Press, 1993.
- HENRY, M.: *La Barbarie*. Paris, Bernard Grasset, 1987.
- MARRAS, A.: “On Putnam's Critique of Metaphysical Realism: Mind-Body Identity and Supervenience”, *Synthese*, 126, 2001, pp. 407-426.
- MCDOWELL, J.: *Mind and World*. Cambridge Mass., Harvard University Press, 1996.
- MCGINN, C.: *Problems in Philosophy. The Limits of Inquiry*. Oxford, Blackwell, 1993.
- MOYA, C.J.: *Filosofía de la Mente*. Universitat de València, 2004.
- PUJADAS, L.M.: *La ascensión y caída de la teoría funcionalista de la mente. O de cómo Putnam y Fodor idearon primero, y abandonaron después, el funcionalismo computacional*. Palma, Universitat de les Illes Balears, 2002.
- PUTNAM, H.: “Minds and Machines” (1960), en *Mind, Language and Reality. Philosophical Papers*. Volume 2. Cambridge University Press, 1975, pp. 362-385. *MM*.

- PUTNAM, H.: "Robots: Machines or Artificially Created Life?" (1964), en *Mind, Language and Reality. Philosophical Papers*, pp. 386-407. *R*.
- PUTNAM, H.: "The Mental Life of Some Machines" (1967), en *Mind, Language and Reality. Philosophical Papers*, pp. 408-428. *ML*.
- PUTNAM, H.: "The Nature of Mental States" (1967), en *Mind, Language and Reality. Philosophical Papers*, pp. 429-441. *NMS*.
- PUTNAM, H.: "Logical Positivism and the Philosophy of Mind" (1969), en *Mind, Language and Reality. Philosophical Papers*, pp. 441-452. *LP*.
- PUTNAM, H.: "Reductionism and the Nature of Psychology" (1973), en *Words & Life* (edit. By James Conant), Cambridge Mass., Harvard University Press, 1994, pp. 428-441. *RNP*.
- PUTNAM, H.: *Reason, Truth and History*, Cambridge University Press, 1981. *RTH*.
- PUTNAM, H.: "Three Kinds of Scientific Realism" (1982), en *Words & Life*, pp. 492-498. *SR*.
- PUTNAM, H.: "Reflective Reflections" (1985), en *Words & Life*, pp. 416-427. *RR*.
- PUTNAM, H.: "Artificial Intelligence: Much Ado about Not Very Much" (1988), en *Words & Life*, pp. 391-402. *AI*.
- PUTNAM, H.: *Representation and Reality*. Cambridge Mass, The MIT Press, 1988. *ReRe*.
- PUTNAM, H.: "Why Functionalism Didn't Work" (1992) , en *Words & Life*, pp. 441-459.
- PUTNAM, H.: *Renewing Philosophy* ["Guifford Lectures", 1989]. Harvard University Press. *RPh*.
- PUTNAM, H.: *The Threefold Cord. Mind, Body, and World*. New York, Columbia University Press. *ThC*.

Mariano Rodríguez  
 Departamento de Filosofía IV  
 Universidad Complutense de Madrid  
 marian@psi.ucm.es