

**PABLO QUINTANILLA, editor**

# **ENSAYOS DE METAFILOSOFÍA**

## **Capítulo 12**



**FONDO  
EDITORIAL**

*Ensayos de Metafilosofía*

© Pablo Quintanilla, editor

Primera edición, marzo de 2009

De esta edición:

© Fondo Editorial de la Pontificia Universidad Católica del Perú, 2009

Av. Universitaria 1801, Lima 32 - Perú

Teléfono: (511) 626-2650

Fax: (511) 626-2913

[feditor@pucp.edu.pe](mailto:feditor@pucp.edu.pe)

[www.pucp.edu.pe/publicaciones](http://www.pucp.edu.pe/publicaciones)

*Prohibida la reproducción de este libro por cualquier medio,  
total o parcialmente, sin permiso expreso de los editores.*

ISBN: 978-9972-42-884-5

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2009-03068

Impreso en el Perú — Printed in Peru

## Comentario a Ricardo Braun

Óscar Augusto García Zárate

Universidad Nacional Mayor de San Marcos, Perú

El problema de la mente tal y como empieza a ser tratado por la filosofía en el siglo XX, esto es, con la exhaustividad propia de lo que ya a estas alturas constituye un ámbito disciplinar bien definido, convoca además de los esfuerzos desplegados en su contexto investigativo, aquellos otros esfuerzos provenientes de esferas afines en cuanto a los objetivos perseguidos. Este es el caso del conocimiento de los procesos mentales. Tenemos así, cubriendo un amplio rango de investigación, la inteligencia artificial, la psicología, la neurociencia, y en general, el despliegue de un esfuerzo multidisciplinar —en alguna medida difuso, pues incluso agrupa también algunas investigaciones aportadas desde la antropología y la lingüística—, todo lo cual conduce a plantear cada vez con mayor insistencia y plausibilidad la concreción de un quehacer novedoso: la llamada ciencia cognitiva.

Aunque la historia de la filosofía ofrece desde antaño claros testimonios de la profunda preocupación manifestada en su ámbito por los problemas relacionados con la naturaleza del pensamiento y con el enigmático mundo interior del ser humano —ya presente de manera acentuada, por mencionar algo, en un texto como el *De ánima* de Aristóteles— lo cierto es que a partir de la irrupción de Descartes es cuando podemos hablar de un problema más acotado, a saber, el problema de la mente humana, o, para ser más precisos, el problema de la relación mente-cuerpo. En efecto, podríamos decir que durante el siglo XVII se «inventa» la mente, en el sentido de que es a partir de entonces que se sientan las bases de la concepción de la mente como aquel espacio interior en que se representan las ideas, espacio imaginado como poseedor de un estatus radicalmente distinto del cuerpo<sup>1</sup>.

---

<sup>1</sup> En el primer capítulo de la primera parte de su clásica y controvertida obra *La filosofía y el espejo de la naturaleza*, Rorty aborda la radical transformación que había de sufrir la concepción de la mente con el

Sin embargo, paulatinamente esta visión hubo de ser radicalmente cuestionada. Muestra de ello son los debates generados a partir, por ejemplo, del planteamiento del problema acerca del lenguaje privado por parte de Wittgenstein, incluso antes de la publicación de las *Investigaciones filosóficas* en 1953<sup>2</sup>. Este nuevo rumbo señalado por sus reflexiones durante la segunda etapa de su labor filosófica, la correspondiente a la desplegada por el así llamado segundo Wittgenstein, constituye un importantísimo punto de inflexión y un referente de primer orden por lo que hace al sólido posicionamiento de la temática filosófica referida a la naturaleza de los procesos mentales. Del mismo modo, ha de ser también considerado como hito importante en la trayectoria seguida por esta temática la publicación, poco antes, en 1949, por parte Gilbert Ryle, de otra obra capital: *El concepto de lo mental*<sup>3</sup>. Pero además, este despliegue auroral del horizonte reflexivo que la pregunta por los asuntos relativos a la mente empezaba a configurar, marchaba de la mano con los desconcertantes avances que se iban logrando, en un principio, en el campo de las investigaciones referidas a la informática y a la Inteligencia Artificial, y un poco más tarde, los no menos cruciales descubrimientos hechos en el terreno de la neurociencia. Todo ello, hace apenas seis o siete décadas. Así pues, es de reciente data el notable impulso interdisciplinario que ha sido dado a la reflexión en torno a la relación entre la mente y el cerebro, tal y como se desenvuelve en las lindes de la filosofía, esto es, adoptando diversos matices y énfasis y a través de acuerdos y desacuerdos, marchas y contramarchas, sin que ello signifique dejar de seguir muy de cerca los aportes de las ciencias.

«Un cálculo lógico de las ideas inmanentes en la actividad nerviosa», artículo escrito por Warren McCulloch y Walter Pitts, y publicado en 1943 en el *Bulletin of*

---

advenimiento de la Modernidad, y que, entre otras cosas, se manifiesta de manera sensible en el alcance que adquiere el término «idea» luego de que se le adjudicara una significación substancialmente distinta de aquella otra que poseía en la Antigüedad e incluso durante la Edad Media. Descartes y luego Locke, por ejemplo, habían de emplear aquel término para referirse a los contenidos de la mente humana. Acerca de este radical giro dado al problema de la relación entre mente y cuerpo, dice Rorty: «La novedad estuvo en la idea de un solo espacio interior en el que eran objeto de cuasi-observación las sensaciones corporales y perceptivas ('ideas confusas del sentido y la imaginación' como diría Descartes), las verdades matemáticas, las reglas morales, la idea de Dios, los talentos depresivos, y todo el resto de lo que llamamos 'mental'. Este escenario interior, con su observador interno, había aparecido insinuado en varios momentos del pensamiento antiguo y medieval pero nunca se había tomado en serio el tiempo suficiente como para servir de base a una problemática. Pero el siglo XVII se tomó con la seriedad necesaria para permitir que se planteara el problema del velo de las ideas... Demostrar que la mente podía imaginarse aparte del cuerpo era... un proyecto totalmente diferente del que se encuentra en la tradición procedente de Aristóteles» (Rorty, R., *La filosofía y el espejo de la naturaleza*, Madrid: Cátedra, 1983, pp. 54-56).

<sup>2</sup> Cf. Wittgenstein, Ludwig, *Investigaciones filosóficas*, Barcelona: Crítica, 2002.

<sup>3</sup> Cf. Ryle, G., *El concepto de lo mental*, Buenos Aires: Paidós, 1967.

*Mathematical Biophysics*<sup>4</sup>, muy probablemente constituye una de las más claras muestras del modo cómo empezó a asumirse la investigación de la denominada Inteligencia Artificial. Sin duda, este artículo posee el carácter de un documento fundacional en este tema y dio origen al surgimiento de dos tendencias en el marco de las investigaciones acerca de la Inteligencia Artificial y la mente humana; por un lado, la llamada variante tradicional, y por otro, el conexionismo<sup>5</sup>.

Al abrigo de este inicial impulso es que se realiza en septiembre de 1948 un simposio en el Instituto de Tecnología de California, el cual tuvo por tema «Los mecanismos cerebrales de la conducta». El evento fue auspiciado por la Fundación Hixson y contó con oradores como John von Neumann y Karl Lashley, al lado de los cuales se encontraba el mismo McCulloch. Asimismo, en el año 1956 tuvo lugar en el *campus* del Dartmouth College un encuentro académico al que asistió un grupo selecto de científicos que realizaban investigaciones en el terreno de la Inteligencia Artificial. Cuatro de ellos —John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Elwood Shannon— fueron los encargados de redactar la propuesta hecha ante la fundación Rockefeller para la realización de dicho evento. Es célebre ya el tenor de su pasaje inicial. En él se plantea el principal propósito perseguido por los investigadores:

Proponemos que se lleve a cabo una revisión de la inteligencia artificial de dos meses y diez hombres en el verano de 1956 en el Dartmouth College de Hanover, New Hampshire. La revisión es para seguir adelante sobre la conjetura de que cualquier aspecto del aprendizaje o cualquier otro rasgo de inteligencia puede en principio ser descrita de forma tan precisa que puede hacerse que una máquina lo simule<sup>6</sup>.

Curiosamente, ese mismo año, se celebraba en el Instituto Tecnológico de Massachusetts (MIT) una reunión que, bajo la denominación de «Simposio sobre Teoría de la Información», se proponía elaborar los lineamientos que había de seguir una ciencia de la mente, siguiendo de cerca e incorporando los conocimientos forjados en los terrenos de la informática. Para muchos, este es el momento en el que oficialmente entra en la escena académica la denominada ciencia cognitiva. Entre los más significativos aportes que se dieron en este evento pueden mencionarse dos. En principio, los correspondientes a la ponencia de Allen Newell y Herbert Simon, que describía el curso seguido por «la máquina lógica» en la primera demostración

---

<sup>4</sup> Cf. McCulloch, W. y W. Pitts, «Un cálculo lógico de las ideas inmanentes en la actividad nerviosa», en: Boden, M. (ed.), *Filosofía de la Inteligencia Artificial*, México D.F: FCE, 1994.

<sup>5</sup> Cf. Boden, M., «Introducción», en: *Filosofía de la Inteligencia Artificial*, pp. 10-11.

<sup>6</sup> McCorduck, P., *Máquinas que piensan. Una incursión personal en la historia y las perspectivas de la inteligencia artificial*, Madrid: Tecnos, 1991, p. 105.

completa de un teorema. Y, además, también las ideas acerca del enfoque transformacional de la gramática formuladas por el destacado lingüista Noam Chomsky<sup>7</sup>.

Éstas son, a vuela pluma, las coordenadas históricas del curso heterogéneo que han seguido las investigaciones que, en lo que podemos llamar un esfuerzo prome-téico, buscan echar luz sobre aquellos aspectos de la mente que parecen obstinarse en permanecer ocultos a la curiosidad humana. El trazado de este panorama a través del cual apreciamos el abigarrado encuentro de diversas perspectivas de investigación, que no obstante, articulan un mismo interés —a saber, el afán de desentrañar el funcionamiento y, si cabe, la naturaleza de la mente— solo tiene un objetivo: situar adecuadamente el tema abordado por nuestro amigo Ricardo Braun a través de su enjundiosa ponencia, que en esta ocasión tenemos el encargo de comentar. Pues el tema en cuestión, precisamente, se encuadra en la intersección de la temática abordada por la psicología, la neurociencia, la ciencia cognitiva, y, claro, la filosofía de la mente.

Un problema que, según nos parece, merece ser considerado en primer lugar, es el relativo a la psicología popular y los embates que ha soportado por parte del eliminativismo. Ricardo Braun aborda este asunto pertinentemente. Aunque en un momento inicial, al referirse al panorama confuso que divisan los estudiantes de psicología, parece mostrarse a favor del programa encabezado por Paul y Patricia Churchland<sup>8</sup>, termina admitiendo que sería muy difícil llevar a efecto lo propugnado por estos dos teóricos y aquellos otros que patrocinan esta corriente, vale decir, dejar de lado absolutamente las actitudes proposicionales en la comprensión y explicación de la conducta humana. En esto último coincidimos con él. Pero, ¿qué podemos decir acerca de la psicología popular? Veamos.

---

<sup>7</sup> Cf. Gardner, H., *La nueva ciencia de la mente. Historia de la revolución cognitiva*, Buenos Aires: Paidós, 1987, p. 44.

<sup>8</sup> Habría, sin embargo, que hacer una necesaria precisión. Aun cuando Patricia y Paul M. Churchland asumen una actitud de frontal cuestionamiento frente a la psicología popular, existe con todo, una marcada diferencia entre el eliminativismo intencional propuesto por Paul Churchland y aquel otro formulado por Patricia Smith Churchland, que, a diferencia de aquél, asume un carácter sentencial. El modelo sentencial, que es cuestionado por Patricia Churchland, postula básicamente la existencia de una estructura interna computacional de naturaleza lingüística como la única posibilidad de justificar la conducta de los sistemas intencionales. Por su parte, Paul Churchland presenta en los siguientes términos los alcances anti-intencionalistas de su materialismo eliminativista: «Hace veinte años, las emociones, los *qualia*, y los 'sentimientos puros' se postulaban como los principales escollos del programa materialista. Con estas barreras derribadas, el lugar de oposición se ha desplazado. Ahora éste es el reino de lo intencional, el reino de la actitud proposicional, que se presenta la mayor parte de las veces como si fuera irreductible e imposible de eliminar en favor de cualquier cosa que se sitúe dentro de un marco materialista» (Churchland, P.M., «Eliminative materialism and the propositional attitudes», en: *The Journal of Philosophy*, LXXVIII, 2 (1981), pp. 67-68).

A diario desenvolvemos una gran cantidad y variedad de acciones que son desencadenadas por ciertos estados mentales. Miedos, deseos y creencias conforman, entre otros estados intencionales, el caleidoscopio interior que constituye el mundo mental del ser humano. A partir de aquellos estados mentales articulamos una serie de explicaciones y predicciones que, sin estar necesariamente basadas en conceptos forjados en los términos propios de una rigurosa investigación científica, sirven para dar cuenta de las acciones de las demás personas y, claro, también de las nuestras. Así, pues, la psicología popular es la base explicativa que, articulada en torno a instancias intencionales, subyace a la visión cotidiana del sentido común. Precisamente por esto último este tipo de psicología —molesto obstáculo tanto para los partidarios de la neurofilosofía<sup>9</sup> como para el eliminativismo radical— también es denominada psicología del sentido común.

Frente a esto, el eliminativismo sostiene que el recurrir a estados intencionales para explicar la conducta de las personas constituye una herramienta conceptual que no posee contenido alguno.

El eliminativismo postula, entre otras cosas, que las propiedades a las que se denominan «psicológicas» no son de ningún modo coextensivas con las propiedades

---

<sup>9</sup> Fue Patricia S. Churchland quien en 1986 publicó *Neurophilosophy*, una obra en que además de exponer un detallado cuadro del avance impresionante que se ha logrado en materia neurocientífica, expresa su pleno y entusiasta convencimiento de que se cuenta ya, por ello mismo, con herramientas científicas que han de hacer posible consolidar lo que la autora llama en algún momento una «ciencia unificada de la mente». Haciendo alusión al nuevo panorama que ella avista en relación con el estudio de la mente, dice: «La convicción que sostiene este libro es que las estrategias *top-down* (que caracterizan a la filosofía, a la psicología cognitiva y a la investigación en inteligencia artificial) y las estrategias *bottom-up* (que caracterizan a las neurociencias) para resolver los misterios de la función mente-cerebro no han de ser desenvueltas en un gélido aislamiento. Lo que se entrevé, por el contrario es una rica interacción entre ambas, de la que se pueda esperar que provoque una fecunda y paralela evolución de teorías, modelos y métodos que se informen, corrijan e inspiren mutuamente» (Churchland, P.S., *Neurophilosophy*, Cambridge (Mass.): MIT Press. 1986, p. 3). Y en otro lugar, refiriéndose también a la neurofilosofía, expresa lo siguiente: «En evolución paralela con la neurociencia, la ciencia cognitiva ha investigado el alcance de funciones de gran escala tales como la atención, la memoria, la percepción y el razonamiento tanto en el adulto cuanto en el desarrollo del niño. Adicionalmente, las ideas computacionales, al vincular fenómenos *cognitivos* de gran escala con fenómenos neurales de pequeña escala, han abierto la puerta a una integración de la neurociencia, la ciencia cognitiva y la filosofía en un marco teórico comprensivo... ¿Qué es la neurofilosofía y cuál es su rol?... La neurofilosofía surge del reconocimiento de que después de mucho, finalmente, las ciencias del cerebro y su tecnología están suficientemente desarrolladas de modo que se puede llegar a alcanzar un real progreso en la comprensión de la relación mente-cerebro. Más aun, ella predice que la filosofía de la mente conducida sin ninguna comprensión de las neuronas y el cerebro sea probablemente estéril. Como resultado de ello, la neurofilosofía centra su atención en problemas que se encuentran en la intersección de una neurociencia que va echando raíces y una filosofía en declive» (Churchland, P.S., *Brain-Wise: Studies in Neurophilosophy*, Cambridge (Mass.): MIT Press, 2002, pp. 2-3).

de naturaleza física. Por otra parte, sostiene que en la medida que los estados neurofisiológicos son estados internos del individuo, los paulatinos descubrimientos que se vienen llevando a cabo en el terreno de la neurociencia proveerán el nivel de conocimientos necesario para dar cuenta, en claros y estrictos términos explicativos, la dinámica profunda de los procesos cognitivos. Sobre la base de esta concepción, los teóricos del eliminativismo aseguran que es posible dejar de postular la existencia de una entidad que permitía explicar un fenómeno —en este caso se trataría del espurio armazón explicativo de tipo intencional con que cuenta la psicología popular— en el preciso momento en que se halle una manera novedosa de explicarlo en substitución de aquélla —también en este caso se trataría de los descubrimientos en materia de neurofisiología—, de modo tal que, en el caso que nos ocupa, la descripción y explicación de un fenómeno mental ha de excluir el uso de un lenguaje con contenido psicológico.

Un aspecto puesto especialmente de relieve por la crítica eliminativista es aquél referido a la presunta condición de teoría empírica que ostentaría la psicología popular. Esto no conduce a otra cosa sino a considerar que una explicación dada en términos de deseos, creencias o expectativas (esto es, una explicación que asume la forma de algunas de estas modalidades del sentido común, o, en otros términos, cada una de las preferencias que dan cuenta de una actitud proposicional) funciona como una instancia particular de una generalización de carácter nomológico. Así, cuando nos referimos al hecho de que Juan estudia con ahínco *diciendo* que la razón de ello es que Juan *desea* ingresar a la universidad, nuestra explicación vendría a ser una instancia particular de una generalización que adoptaría la estructura siguiente:

Para todo individuo  $x$ , todo deseo  $p$  y toda creencia  $q$ ,  
 si  $x$  (Juan) desea  $p$  (ingresar a la universidad) y  $x$  cree que si  $q$  entonces  $p$  (Juan cree que si estudia con ahínco entonces ingresará a la universidad),  
 entonces, a menos que otras creencias o deseos impidan su realización,  
 $x$  (Juan) realiza  $q$  (estudia con ahínco)<sup>10</sup>.

Esta presunta estructura teórica nomológica es puesta en cuestión por el eliminativismo. En la medida en que la psicología apela a instancias «mitológicas» —tales como las actitudes proposicionales— con el objeto de explicar los procesos mentales, y teniendo en cuenta que se trata de una teoría, toda vez que recurre a generaliza-

---

<sup>10</sup> La estructura de esta generalización la he tomado de Toribio Mateas, J., «Eliminativismo y el futuro de la psicología popular», en: Broncano, M. (ed.), *La mente humana*, Madrid: Trotta, 1995, pp. 245-272. Sin embargo, a fin de incorporar alguna dosis de originalidad al ejemplo, he considerado oportuno variar su contenido.



ciones del tipo a que se ha hecho referencia, pues entonces se concluye que la psicología popular no es sino una teoría empírica del todo falsa, y, por ello mismo, debe ser abandonada. No se trata pues de buscar reconducir las nociones intencionales de este ámbito a un terreno científico a través, digamos, de leyes puente —en este caso estaríamos hablando de la aplicación de una estrategia reduccionista<sup>11</sup>—, sino más bien, y como lo pretende Paul Churchland, de reemplazar completamente las explicaciones intencionales por un vocabulario articulado sobre la base de conceptos provenientes de la neurociencia.

Haciendo frente a este intento por acabar de una buena vez con la psicología popular y su entramado intencional de explicaciones se encuentra Jerry Fodor, uno de sus más entusiastas defensores. En el contexto mayor de lo que él denomina «teoría representacional de la mente», Fodor asume la reivindicación de la psicología popular, al punto de aspirar a convertirla en una ciencia respetable. Haciendo gala no solo de solidez argumentativa, sino también de una elegancia expresiva a menudo cautivante, Fodor lleva a término una cerrada defensa de la psicología popular, a la que se refiere como «la vieja psicología de sentido común de las creencias y los deseos»<sup>12</sup>. Esta urdimbre de supuestos, entretejida con deseos y creencias y conectada con acciones, es la que se despliega habitualmente en medio de predicciones que se hacen casi sin notarlos, y de acuerdo a las cuales las más diversas actividades adquieren un orden y un sentido. La defenestrada psicología popular está en todas partes y a todo momento, mostrando con esto que, aunque casi no se note, es notablemente importante.

Luego de esta breve exposición de lo que podría considerarse el estado de la cuestión en lo que respecta al debate entre eliminativistas y aquellos que, como en el caso de Jerry Fodor, otorgan un lugar especial a la intencionalidad en la explicación de los procesos mentales, sería oportuno referirnos a lo que creemos se presenta como una consecuencia nefasta que se desprende del eliminativismo: la concomitante, aunque

---

<sup>11</sup> En este punto, justamente, es donde se separan eliminativismo y materialismo reduccionista. En efecto, mientras aquél rechaza categóricamente que pueda efectuarse una reducción de los principios y de la ontología que implica el empleo del lenguaje intencional a términos provenientes de las investigaciones en neurofisiología, el materialismo reduccionista —que patrocina la denominada tesis de la identidad— estima viable una reducción interteórica de las nociones intencionales de la psicología popular, de modo tal que una vez hecho esto, todo proceso mental que era explicado apelando a dichas nociones es reconducido al interior de un marco teórico vertebrado tomando como eje el lenguaje y las conquistas científicas de la neurociencia (cf. Churchland, P.M., *Materia y conciencia. Introducción contemporánea a la filosofía de la mente*, Barcelona: Gedisa, pp. 50-63).

<sup>12</sup> Cf. Fodor, J., *Psicosemántica, El problema del significado en la filosofía de la mente*, Madrid: Tecnos, 1994.

no explicitada, eliminación de la noción de persona. Pasemos a considerar esta cuestión.

Algún autor<sup>13</sup> ha trazado con acierto y agudeza la tortuosa ruta que ha seguido la noción de persona en el contexto de la filosofía de la mente. De ser convertida en foco de atención y puesta en un lugar de expectativa en autores como Wittgenstein y Strawson, luego de la eclosión de la llamada revolución cognitiva, fue perdiendo protagonismo paulatinamente hasta verse expuesta a la total exclusión en un marco teórico como el formulado por el eliminativismo.

Así pues, los enfoques cognitivos desde la época de aquella revolución han abordado el problema de la relación mente-cuerpo dentro de un marco que puede ser llamado naturalista, y a partir de ello dieron origen básicamente a dos líneas de investigación, ambas siempre dejando de lado la noción de persona. Tales orientaciones fueron el funcionalismo computacional y el eliminativismo, y es éste, como se hace evidente, el que concentra interés en este punto.

El eliminativismo, como tuvimos ocasión de ver, convierte en el centro de su atención más que los avances en materia de Inteligencia Artificial, el papel —que estiman crucial aquellos que están entre sus filas— cumplido por los descubrimientos efectuados por la neurociencia en lo que atañe a las respuestas que, sobre esta base, se podrían alcanzar en referencia al problema de la mente humana. Como también se ha indicado, el propósito perseguido por esta corriente consiste en erradicar las actitudes proposicionales encuadradas en el contexto explicativo de la psicología popular, pues, aducen los eliminativistas, un lenguaje articulado en torno a aquéllas, y por esto mismo típicamente intencional, no es susceptible de ser reducido a los principios y proposiciones que conforman el marco teórico de la neurociencia. Y ello por una sencilla razón: la psicología popular constituye una teoría empírica radicalmente falsa.

Pero he aquí que surge una situación por lo demás alarmante: si el eliminativismo obra en consecuencia, el desenlace de su proceder será simple y llanamente —lo cual es solo un decir— la concomitante eliminación de la noción de persona, con todo lo que ello implica. El riesgo que lleva aparejado una estrategia de este tipo se hace patente al momento de reparar en que estamos frente a la posibilidad de una eventual cosificación de las relaciones humanas.

Pero dejemos las cosas en este punto. El simple propósito en relación con esta temática era dejar bosquejado lo principal de esta incisiva objeción a las implicancias del eliminativismo.

---

<sup>13</sup> Cf. San Félix, V, «Persona y cosificación (A propósito del sentido práctico de la revolución cognitiva)», en Chacón, P. y M. Rodríguez (eds.), *Pensando la mente. Perspectivas en filosofía y psicología*, Madrid: Biblioteca Nueva, 2000, pp. 325-356.

Pasando a tratar ahora otra cuestión, a propósito de la discusión por parte de Ricardo Braun de las implicancias que involucra el individualismo (que es otra manera de aludir al contenido mental restringido, núcleo nocional del internalismo), y a partir de lo cual reseña un experimento mental, a saber, el ideado por Hillary Putnam, aquél conocido como el experimento de la Tierra Gemela, resultaría apropiado aproximarnos al debate particularmente interesante suscitado por el problema concerniente a los estados mentales y el contenido a ellos correspondientes.

Putnam, a través de este ejemplo de ciencia ficción, como él lo denomina, se propone mostrar los problemas que entraña el uso del término «significado», pues estima que éste descansa en una falsa teoría que, a su vez, se encuentra articulada en torno a dos asunciones que no están convenientemente justificadas. Así, señala Putnam, sobre la base de estos dos supuestos, por una parte podemos afirmar que conocer el significado de un término es estar en un determinado estado mental, y, por otra, se sostiene que el significado, en términos de intensión, es el que fija la extensión. Al momento de discutir estas presunciones, Putnam aborda lo que él denomina solipsismo metodológico, es decir, la concepción acerca de que ningún estado mental presupone la existencia de otro individuo a no ser solo la de aquél a quien se adscribe un estado mental. A este tipo de estados los llama estados mentales en sentido restringido. Frente a éstos, Putnam propondrá los estados mentales en sentido amplio, los cuales, a diferencia de aquellos otros, suponen lo que en otro texto llamará «contribución del entorno»<sup>14</sup>, vale decir, la determinación del significado supone la participación de condiciones externas. Dicho de otro modo, y en palabras de Putnam: «¡Los significados no están en la cabeza!»<sup>15</sup>.

En aras de la brevedad, ya que el experimento aludido ha sido ya convenientemente recreado por Braun, solo diremos que dicho experimento, y según lo que venimos diciendo, apunta a poner de relieve el rol decisivo que desempeña el entorno físico, o en otros términos, las condiciones externas, con vistas a determinar el significado asociado a una determinada representación mental. Es importante a este respecto aludir a una cuestión que Braun no menciona, pero que quizá sea necesario recordar. Putnam añade que incluso en el caso de que los personajes de la Tierra y de la Tierra Gemela, Óscar1 y Óscar2, respectivamente, se encontraran en una época anterior al descubrimiento de la composición molecular del agua, digamos en 1750 —que en el caso de la Tierra había de ser determinada como H<sub>2</sub>O; y en el de

---

<sup>14</sup> Putnam, H., *Representación y realidad. Un balance crítico del funcionalismo*, Barcelona: Gedisa, 1995, p. 63.

<sup>15</sup> Putnam, H., *Mind, Language, and Reality: Philosophical Papers II*, Cambridge (Mass.): Cambridge University Press, 1975, p. 227.

la Tierra Gemela como XYZ—, la situación, en términos de significado, no variaría. Óscar1 y su doble seguirían refiriéndose a dos cosas distintas, pues aun cuando se encontraran en estados mentales idénticos, y a la ciencia de sus respectivos mundos les tomara cincuenta años más descubrir la composición del agua, la extensión del término empleada para designarla siempre sería distinto: «agua» para el terrícola Óscar significaría H<sub>2</sub>O, en tanto que para el *Doppelgänger* habitante de la Tierra Gemela sería XYZ. Insistamos en esto: el significado de un término, según Putnam, no queda fijado por los estados mentales de las personas —si fuera así, Óscar1 y Óscar2 estarían en posesión del mismo significado—, sino por la referencia. En el caso de Óscar1 y su sosias idéntico, no es el caso que el significado haya variado en el transcurso de alrededor de cincuenta años (el período aproximado que, desde el año 1750, le tomaría a la ciencia descubrir la composición química del agua), sino son las *condiciones* externas —o como también lo dice Putnam, la relación teórica entre aquello a que llamamos agua y la relación de mismidad (*sameness*) que este término de clase natural establece con aquel tipo de cosas con que se identifica— las que en última instancia han sufrido una modificación; condiciones que por lo demás siempre estarán expuestas a ser modificadas por eventuales descubrimientos posteriores<sup>16</sup>.

El planteamiento de esta temática por parte de Putnam se considera la fuente que genera la explicitación del debate a que nos referimos. Putnam, como se recordará, fue un conspicuo representante temprano del funcionalismo<sup>17</sup>. Propuso él una explicación del significado articulada, como hemos visto, tomando como eje el llamado contenido mental amplio. Se sitúa Putnam de esta manera en las filas del externalismo, adversario teórico del internalismo asumido, entre otros, por autores como Jerry Fodor, y que, en consecuencia, se muestran a favor de la explicación de la problemática relativa a la intencionalidad en términos de lo que se denomina contenido mental restringido.

En este punto no estará de más recordar otra historia de ciencia ficción elaborada por Putnam para explorar la noción de referencia. En este caso se trata de la historia a través de la cual Putnam invita al lector a imaginar que, en realidad, los seres humanos y, en ge-

<sup>16</sup> Cf. *ibid.*, pp. 224-225.

<sup>17</sup> Vendría bien señalar, sin embargo, que Putnam tomó distancia de su original postura, y sometió a un balance crítico los principales supuestos del funcionalismo por él propuesto en la década de los sesenta. Refiriéndose al propósito que persigue en la obra en que manifiesta esta actitud, dice Putnam en la introducción de aquella: «...me propongo demostrar que los argumentos a favor de la concepción computacional, esto es, los mismos argumentos que yo solía esgrimir para demostrar que es incorrecto identificar los estados mentales con estados físico-químicos, pueden generalizarse y extenderse de modo que resulte también incorrecta la identificación directa de los estados mentales con estados *funcionales*, es decir, con estados caracterizados computacionalmente» (Putnam, H., *Representación y realidad. Un balance crítico del funcionalismo*, p. 15).

neral, todo ser que siente, son cerebros en una cubeta conectados a una súper maquinaria que induce en ellos, a través de una sofisticadísima red de *inputs*, la ilusión de un mundo externo en la que son personas de carne y hueso. Estimo pertinente incorporarlo en este contexto porque nos permitirá entrar a abordar luego el otro *Gedankenexperiment* que Braun reseña en su texto: el debido a John Searle. Pues hay, a nuestro juicio, una relación entre estos dos puntos de vista. Tanto Searle como Putnam niegan a toda costa que el nivel de manipulación simbólica en una máquina implique necesariamente un nivel de comprensión de significados y, con esto, de intencionalidad.

Si bien el experimento en cuestión tiene como trasfondo un objetivo distinto —esto es, demostrar, arribando a un desenlace en que el argumento se autorrefuta, que si pensamos que somos cerebros en una cubeta, en realidad no somos cerebros en una cubeta—, no obstante, uno de los aspectos principales en el contexto de la argumentación es la exploración de la noción de referencia. Para los propósitos de la presente exposición bastará tomar en cuenta solo esto último. Veamos.

Putnam, en principio, presenta el fantástico, si bien lógicamente posible caso de una hormiga que, desplazándose parsimoniosamente sobre la arena, trazara una línea de contornos caprichosos que a la postre configurara una caricatura de Winston Churchill. Frente a esto, la pregunta que se podría formular sería la siguiente: ¿dicha caricatura se refiere a Churchill? La respuesta es un rotundo no. Luego se verá cómo se incorpora esta respuesta en el marco argumentativo global que formula Putnam. A partir de esto, y pasando a otro plano de la argumentación, Putnam apela al Test de Turing. Dicho test es una prueba ideada para determinar los márgenes dentro de los cuales resultaría lícito adjudicarle a una computadora procesos de pensamiento semejantes a los que despliega el ser humano y, sobre esta base, señalar en la máquina la presencia de conciencia. Parte del test implica mantener una conversación con la computadora sin estar al tanto, claro está, de que efectivamente se trata de una máquina de este tipo, de modo que no podamos distinguir sus respuestas a nuestras preguntas de aquellas otras que daría una persona. Putnam efectúa una ligera variante al test, pues en su caso él lo tomará en cuenta para determinar si es posible que una computadora logre referirse a algo, es decir, desenvolver procesos de tipo intencional sobre la base de los cuales sea capaz de elaborar representaciones que hablen del mundo tal y como lo haría un ser humano, de modo tal que la posibilidad de hacer una distinción tajante entre el modo de referirse a objetos por parte de la máquina y el correspondiente a una persona —aquello a lo Putnam llama la referencia compartida— quede de plano cancelada. Putnam denomina esta variante de la prueba Test de Turing para la Referencia.

Si imaginamos, dice Putnam, que nos enfrascamos en una conversación con un interlocutor sin saber que se trata de una máquina, al punto de que en el curso de

este juego de la imitación, dicha máquina logra engañarnos, y como producto de ello llegáramos a la errónea conclusión de que se trata de una persona, de lo cual también se desprendería que esa máquina efectivamente ha logrado referirse a los hechos del mundo en los exactos términos en que lo hacen las personas, lo cierto con respecto a todo esto es que, en verdad, aquella máquina no habría llegado a instalarse en un plano representativo, referencial. Y esto por una sencilla razón: la manipulación de símbolos por sí mismo no es garantía de que se produzca un uso referencial de esos símbolos. Y es en esta medida que resulta legítimo decir que la capacidad referencial de una máquina con esas características es la misma que la que, por cuestiones de azar, ha puesto en juego la fantástica hormiguita dibujante que se mencionaba líneas arriba; es decir, dicha capacidad es inexistente.

Parece ser legítimo a este respecto indicar, como ya lo habíamos dejado apuntado antes, que se puede detectar una coincidencia de pareceres dada entre la perspectiva de Putnam y las ideas que animan la concepción de Searle respecto de lo que cabría denominar «estupidez semántica» en relación con los procesos de manipulación simbólica realizados por computadoras. Pues, en efecto, creemos que es posible encontrar un punto de intersección entre la tesis que Searle defiende a través del experimento mental de la habitación china, a saber, que las máquinas no poseen una dimensión semántica que les permita comprender significados —y a través de la cual establece una polémica también con los principales supuestos asumidos en el marco de lo que él llama Inteligencia Artificial fuerte—, y aquella otra defendida por Putnam, que acabamos de exponer, en el sentido de que no es posible atribuirle capacidad referencial a una computadora aun cuando ésta desenvuelva impecables procesos de manipulación sintáctica.

En lo que atañe al célebre experimento mental —conocido como «cuarto chino» y cuyos pormenores ya fueron oportunamente presentados por Braun— a través del cual Searle<sup>18</sup> también recusa la idea procedente de Turing acerca de que una máquina, esto es, una computadora, podría ser capaz de desplegar un tipo de proceso que podría ser considerado como pensamiento, la conclusión a la que se arriba resulta ser categórica: las máquinas de ese tipo solo procesan información en términos sintácticos, del mismo modo como el hombre de su experimento manipula de manera eficaz ideogramas chinos en la habitación en la que se encuentra encerrado, lo que le permite generar un proceso que es considerado eficaz por las personas que se encuentran afuera de la habitación, y esto aun cuando no comprenda absolutamente nada de ese idioma. A diferencia de la ejecución de una tarea como ésta, la comprensión

---

<sup>18</sup> Cf. Searle, J., «Mentes cerebros y programas», en: Boden, M. (ed.), *Filosofía de la inteligencia artificial*, pp. 82-104.

implica la puesta en marcha de un proceso semántico, proceso que las máquinas no están capacitadas para cumplir.

En relación con lo dicho, Searle adopta una postura que sin dejar de reconocer el origen causal de la bioquímica cerebral, otorga del mismo modo su lugar a la intencionalidad, y de acuerdo a ello, defiende la tesis de que los procesos mentales dependen de procesos cerebrales, si bien son irreducibles a éstos<sup>19</sup>. A esta postura algunos autores la han denominado emergentismo. Si bien, dice Searle, la conciencia es un fenómeno que puede ser remitido en su origen a procesos neurológicos, no obstante, aquélla no se agota en éstos. Searle estima que el estudio de la conciencia puede llevarse a término empleando una metodología científica<sup>20</sup> y, según esto, considerando la base neurofisiológica de aquélla, pero sin que ello suponga negar, tal y como lo hace el eliminativismo, la intencionalidad como fenómeno relevante al momento de explicar la naturaleza de los procesos mentales. Asimismo, su desacuerdo es frontal con respecto a la tesis esgrimida por los teóricos de lo que Searle identifica como expresión, algo que ya mencionamos, de la Inteligencia Artificial fuerte, y merced a la cual se busca reproducir y explicar el fenómeno mental a través del diseño de programas de cómputo, pues, sostienen sus teóricos, éstos poseen la particularidad de poder ser ejecutados por cualquier máquina adecuadamente construida. En buena cuenta, tal y como lo hace ver Searle, esta perspectiva supone un dualismo que, aunque no es típicamente cartesiano, de todos modos postula la independencia del programa (la instancia que guarda semejanza con la mente, según los partidarios de la Inteligencia Artificial fuerte) en relación con su concreta ejemplificación por parte de cualquier «cerebro» electrónico. Este dualismo es rechazado por Searle, pues él sostiene más bien que si bien el cerebro humano puede ser considerado una máquina, se trata de una que se relaciona intrínsecamente con los fenómenos mentales. En el campo de la Inteligencia Artificial, por el contrario, no existe ninguna relación intrínseca entre los programas y los «cerebros» que los ponen en marcha. Las únicas máquinas que al igual que el cerebro humano podrían ser asiento de procesos de pensamiento y conciencia serían aquellas que contarán con los mismos poderes causales con que cuenta el cerebro, de modo tal que mal podría decirse, si se asume el marco emergentista, que los procesos mentales

---

<sup>19</sup> Su postura al respecto queda claramente expresada en los siguientes términos: «Yo entiendo inglés y poseo otras formas de intencionalidad, no porque soy el ejemplo concreto de un programa de cómputo... sino porque, hasta donde sabemos, soy un cierto tipo de organismo, con una cierta estructura biológica (es decir, cierta estructura química y física), la cual, en ciertas condiciones, es causalmente capaz de producir percepción, acción, comprensión, aprendizaje y otros fenómenos intencionales» (*ibid.*, p. 97).

<sup>20</sup> Cf. Searle, J., «How to study consciousness scientifically», en: *Philosophical Transactions: Biological Sciences*, CCCLIII, 1377 (1935-1942).

(que pretenden ser reproducidos en alguna escala por los teóricos de la Inteligencia Artificial a través de programas) son independientes de su base física<sup>21</sup>.

Y así pues, y volviendo al punto relacionado con la intencionalidad, esto es, la capacidad referencial propia de la mente humana, Searle niega rotundamente que una máquina de cómputo pueda duplicar los procesos mentales humanos en términos de intencionalidad. Y así, ante la pregunta acerca de si la ejemplificación concreta de un programa a través de una de estas máquinas no humanas podría ser considerada, por sí misma, condición suficiente para el surgimiento de la comprensión, Searle contesta con un categórico no<sup>22</sup>.

De este modo, podemos concluir nuestro comentario que ha pretendido ser, más que una crítica del texto de Braun, una ampliación de sus contenidos en el mismo tono expositivo presente en su ponencia y artículo.

---

<sup>21</sup> Cf. Searle, J., «Mentes cerebros y programas», p. 102

<sup>22</sup> Cf. *ibid.*, p. 98.