PERSPECTIVAS ACTUALES EN LA FILOSOFIA DE LA MENTE

D. C. Dennett

La filosofía de la mente es uno de los campos más activos de la filosofía actual, y ha cambiado tan drásticamente en los últimos veinte años, que muchos de los tópicos y teorías tradicionalmente centrales se han transformado hasta resultar casi irreconocibles, mientras surgen nuevos intereses que carecen de antecesores claros en la vieja tradición. Una evaluación del trabajo actual requiere la comprensión de las suposiciones recientemente desarrolladas acerca de las responsabilidades y metas del campo, y la mejor manera de proporcionar tal comprensión es una breve historia de los cambios de perspectiva en años recientes.[1]

I. INVESTIGANDO EL LENGUAJE DE LA MENTE

La nueva era en la filosofía de la mente puede fecharse con la publicación en 1949 de *El concepto de lo mental* de Gilbert Ryle (93). En aquel libro Ryle argüía que la filosofía de la mente descansaba sobre una equivocación colosal, un "error categorial" que había tenido por efecto el nacimiento de todo un campo de investigación –la filosofía de la mente– allí donde no debía existir ninguno; las cuestiones que componían la investigación estaban tan radicalmente malconcebidas que condujeron ineluctablemente al sinsentido a todos los intentos honrados de contestarlas. Antes de Ryle habían existido *teorías* de la mente –'ismos' tan formidables como idealismo, materialismo, monismo neutral, epifenomenalismo, interaccionismo– contendiendo sobre una arena de suposiciones compartidas acerca de la naturaleza del pro-

blema que definía el campo: el problema mente-cuerpo. Ryle sugirió que no existía en absoluto semejante problema, sino sólo una confusión engendrada por un uso –o abuso– insensato e insensible del lenguaje ordinario que utilizamos en la vida de cada día para informar sobre los hechos familiares de la mentalidad que constituyen los datos para cualquier investigación o ciencia de la mente.

Un análisis cuidadoso de las formas de nuestra habla común sobre la mente disiparía las confusiones, disolvería los problemas, haciendo así ociosos tanto al dualismo como a su negación, el monismo, y entre las variedades del monismo tanto al idealismo como a su oponente, el materialismo, es decir, en pocas palabras, todos los puntos de vista metafísicos rivales que habían sido el principal producto del campo. El trabajo de Ryle fue el aporte principal a la filosofía de la mente de lo que terminó llamándose *filosofía del lenguaje ordinario*, y su influencia debe medirse no mediante un censo de conversos (las filosofías raramente despliegan su influencia atrayendo partidarios, sino más bien manteniéndose como fuente de controversia durante largos períodos de tiempo), sino por el hecho de que más de una década después de la aparición de *El concepto de lo mental*, las *teorías* de la mente estuvieron fuera de moda y a punto de extinción. Se sostenía que las teorías eran las creaciones de quienes no acertaban a ver que los problemas –al menos los problemas que los filósofos estaban equipados para plantear– surgían de suposiciones equivocadas e ingenuas sobre la manera como las palabras sobre lo mental funcionaban en el lenguaje. La construcción de teoría fue reemplazada por la actividad más cauta y modesta del "análisis conceptual": el examen delicado y persistente, aunque informal y asistemático, de las expresiones del lenguaje ordinario, cuyo producto colectivo es una amplia colección todavía sin sistematizar de distinciones y matices agudamente observados, aducidos mientras se hacían puntualizaciones usualmente pequeñas sobre varios conceptos mentales. Lo mejor que se debe a este método, en las obras de Ryle, Wittgenstein, Austin y Anscombe (5, 8, 10, 93, 117), es el haber dejado al descubierto profundos temas conceptuales que todavía conforman el pensamiento actual y continuarán haciéndolo así.

Y lo peor que hizo, como sucede en cualquier campo, fue producir montañas de trivialidades, pero en la zona intermedia había una buena cantidad de trabajo muy inteligente y útil, cuyo objeto rara vez era el resolver problemas y menos aún el avanzar teorías generales, sino típicamente el alertar al incauto sobre la existencia de más problemas y distinciones de los que uno podía haber esperado (43, 63, 73, 94, 113, 114, 115, 116). En aquellos días parecía haber poca cosa semejante a una generalización sustancial que pudiera proponerse de manera defendible.

Por lo tanto la táctica básica era el "ascenso semántico" (85), si uno cae en perplejidades al hablar sobre cosas (en este caso; mentes, sensaciones, pensamientos y cosas por el estilo) suele servir de ayuda el desplazar el foco y hablar sobre cómo hablar sobre aquellas cosas, sobre "lo que uno diría ordinariamente" bajo varias circunstancias, o (si no se está entusiasmado con el lenguaje *ordinario*) lo que uno *debería* decir en tales circunstancias. No hay que negar el valor de la táctica, como lo demuestra ampliamente el gran trabajo efectuado en el campo en cuestión, pero, en contra del credo de la mayoría de aquel momento, no ha resultado que *todos* los problemas en filosofía de la mente se hayan evaporado con el análisis lingüístico, y es justo decir que una gran cantidad de investigaciones del idioma ordinario fueron incapaces de producir algo más importante e iluminador que una apreciación intensa de las sutilezas de la expresión inglesa. Más aún, si bien el ascenso semántico excusa de la exposición o defensa de un "ismo", no permite sin embargo que uno efectúe su trabajo manteniéndose inocente de suposiciones, suposiciones que en última instancia implican algo similar a una teoría. Típicamente (para Ryle, Malcolm, Anscombe y muchos otros) la teoría tácita era el "conductismo lógico", el punto de vista de que la verdad en la atribución de estados y eventos mentales implica y está implicada por la verdad de diversos enunciados referidos exclusivamente a comportamiento. Había disidentes de este punto de vista: por ejemplo Strawson (108) se comprometió aparentemente con una reposición críptica de la *teoría del doble aspecto* (una *persona* no debe ser analizada en un cuerpo más una mente, sino que es más bien un sujeto propio de atributos tanto mentales como físicos) y en el

análisis de Shaffer (99) se sostenía que el lenguaje ordinario incorporaba el dualismo.

Esta tendencia a manifestar la teoría en el curso del análisis conceptual tuvo el efecto, al menos en sus formas más virulentas, de malgastar completamente el útil caracter indirecto del ascenso semántico, y emergieron cuerpos de doctrina que se parecían de manera sospechosa a las viejas teorías metafísicas sobre las *cosas* (mentes, sensaciones, pensamientos,...), si bien generados a partir de consideraciones escrupulosamente limitadas a *palabras* ("mente", "sensación", "pensamiento",...). Este no fue un desarrollo feliz. Quizá se pueda aprender todo acerca del *concepto* de caballo estudiando la manera como utilizamos ordinariamente la palabra "caballo", y sin duda se pueden aprender una gran cantidad de cosas sobre caballos estudiando el concepto de caballo (ya que mucho de lo que la gente *piensa* que es verdadero sobre los caballos está de alguna manera incorporado en nuestro concepto), pero al final quedan algunos hechos no considerados sobre los caballos e incluso perplejidades de interés no despreciable que pueden descubrirse únicamente examinando un caballo o dos, o al menos leyendo las obras de quienes se han tomado el trabajo de hacerlo.

II. NATURALIZACION DE LA FILOSOFIA DE LA MENTE

La filosofía de la mente del lenguaje ordinario se encuentra hoy extenuada, hasta el punto de que se la puede considerar cómodamente como un fenómeno histórico. Como disciplina esencialmente crítica y reactiva, estaba condenada a morir con sus propios éxitos al agotarse los errores y confusiones importantes que diagnosticar, mientras que sus debilidades se hicieron más evidentes a medida que iba cayendo en trivialidades. Aunque sus doctrinas y métodos más característicos han sido ampliamente rechazados o abandonados, sus contribuciones al pensamiento actual son positivas y omnipresentes. Lo más importante es que el tratamiento nuevo de las palabras destruyó realmente la manera tradicional de componer una teoría de la mente. Los teóricos tradicio-

nales eran culpables, tal como se les acusaba, de hacer generalizaciones apriorísticas que eran el producto (irreflexivo y no autoconsciente) del análisis conceptual, mezclándolas con un manojo de introspecciones y observaciones casuales sobre las experiencias y poderes de la gente normal, y atribuyendo a esta mezcla el estatuto de verdades metafísicas sobre las esencias de las cosas mentales. De todas formas si tuvieran que existir teorías de la mente, éstas no se producirían según los viejos métodos de sillón; por lo tanto al filósofo de la mente le quedaban tres alternativas: abandonar la filosofía y dedicarse a teorías empíricas en el dominio de la psicología o de la ciencia del cerebro, abandonar la teoría y permanecer en las modestas iluminaciones y curas de confusión propias del análisis puramente lingüístico, o convertirse en una especie de meta-teórico, un crítico conceptual de las teorías adelantadas por las ciencias relevantes. Es esta última concepción de la empresa, a la que se considera como una rama de la filosofía de la ciencia, la que domina el mejor trabajo actual en el campo. Su diferencia más notoria tanto respecto a la teorización tradicional como respecto al enfoque del análisis lingüístico es su interés por los datos y teorías de la psicología, las ciencias del cerebro, la inteligencia artificial y la lingüística.

En 1932, la obra clásica de H. H. Price, *Perception* (77) contenía una apología sucinta de la ignorancia científica del filósofo: nuestras bases para creer las consideraciones fisiológicas sobre la percepción "derivan de la observación, y principal si no enteramente de la observación visual". Pero la confiabilidad de la observación es precisamente lo que está bajo consideración para el epistemólogo, y "puesto que las premisas de la fisiología están entre las proposiciones cuya validez estamos investigando es difícilmente probable que sus conclusiones puedan ayudarnos". En la medida en que uno está empeñado en un intento cartesiano de justificar todo conocimiento desde el principio, desde cualquier fundamento mínimo que pueda ser protegido del escepticismo sistemático, puede mantenerse esta justificación familiar (aunque no proporciona ninguna razón buena para no *echar un vistazo* a la fisiología), pero semejante fundamentalismo está decayendo en epistemología y generalmente en filosofía y siendo reemplazado

por una actitud (no una teoría) "naturalista" que asume desde el principio que nuestras creencias cotidianas son en general verdaderas y garantizadas, que la epistemología puede aprender de la psicología, y que la mejor manera de derivar los cánones de justificación es ver cómo se hace buena ciencia. (La actitud está bien expresada por Quine, uno de sus promotores más influyentes, en "Naturalización de la epistemología" (86).)

El peligro de esta corriente parecería estar en que deja a la filosofía sin una plataforma desde la que lanzar críticas realmente radicales a la ciencia actual, pero se trata probablemente de una preocupación equivocada, pues la historia de la ciencia sugiere que las revoluciones en el pensamiento científico deben ser engendradas internamente. Con todo, la pretensión de que el nuevo naturalismo es una capitulación ante el excesivo prestigio de la ciencia moderna tiene algo a su favor, aunque no debe olvidarse que la amistad presente entre la filosofía y la ciencia no es ninguna novedad. La gran filosofía del siglo XVII, por ejemplo, estaba en comunicación íntima con el nacimiento contemporáneo de la ciencia moderna y contribuyó tanto a esa infancia cuanto ganó en recompensa. En todo caso, los filósofos han descubierto una veta que será explotada muy probablemente para la mutua ilustración de la ciencia y la filosofía.

Aunque esta emergencia fuera del aislamiento intelectual del pasado reciente es un desarrollo lógico a partir de lo mejor de la tradición del análisis lingüístico, sin embargo requería del rechazo de una troika de doctrinas centrales en esa tradición: verificacionismo, conductismo lógico, y lo que podría llamarse conservadurismo conceptual. En su forma más exigente, el verificacionismo es la doctrina según la cual el método de verificar la aplicación de un término *es* precisamente su significado: en sus formas más suaves el verificacionismo mantiene que las declaraciones que en principio no pueden ser verificadas carecen de sentido. Este sobreviviente vástago cerebral de la concepción del positivismo lógico es una doctrina bastante plausible hasta que uno, como suele suceder, se pone severo o doctrinario sobre qué es lo que va a contar como verificabilidad. Consideremos, por ejemplo, este enunciado simplificado del famoso "problema de las otras mentes".

¿Cómo sé que otras personas tienen mente? Yo no puedo ver directamente o sentir sus mentes (de la manera como puedo examinar la mía por introspección); todo lo que tengo como datos son hechos observados sobre su comportamiento. Quizá su comportamiento es una buena *evidencia* inductiva para la existencia de sus mentes. Pero podría no serlo, pues en orden a establecer que era buena evidencia, tendríamos que disponer de casos *confirmados* de la co-ocurrencia de tal comportamiento con otras mentes, y esto requiere, *per impossibile*, un método independiente de verificar la existencia de esas mentes en aquellos instantes. Tal como lo decía el slogan, algo puede ser el *síntoma* de x sólo si algo distinto es el *criterio* de x (1, 54) Se pensaba que los criterios eran indicadores no (meramente) *empíricamente confiables*, sino más bien *lógicamente suficientes* o al menos *decisivos* o "proveedores de certeza", de cualquier cosa de la que fueran criterios. Así pues, como un recurso al lenguaje ordinario muestra que la pretensión de que hay otras mentes no carece de sentido (el hombre de la calle sabe muy bien que hay otras mentes), entonces debe ser verificable, y como la única evidencia para verificarlo es comportamental, y como la evidencia sintomática es lógicamente dependiente de la evidencia criterial, *entonces debe haber criterios puramente comportamentales para toda aserción (significativa) sobre otras mentes*. Así el conductismo lógico nace del verificacionismo. Pero ahora supongamos que algún elemento mental, por ejemplo el *dolor*, tiene criterios puramente conductuales. Esto significa que la afirmación de que alguien tiene dolor es realmente (lógicamente equivalente a) un enunciado acerca del comportamiento o las disposiciones conductuales de esa persona. Pero ningún enunciado sobre acontecimientos fisiológicos internos podría ser lógicamente equivalente a un enunciado justamente sobre el comportamiento abierto de la persona; por lo tanto las verdades de la fisiología, al margen de aquello en que puedan consistir, son *irrelevantes* –excepto sintomáticamente– para la verdad de las declaraciones sobre el dolor. Puesto que el concepto de dolor tiene criterios conductuales, no puede tener también criterios fisiológicos. Si los científicos propusieran criterios fisiológicos para el dolor, estarían "proponiendo un nuevo concepto" de dolor, y cualquier cosa que

nos dijeran sobre *su* tipo de dolor no tendría nada que ver con nuestro concepto *ordinario* de dolor. La ciencia no puede revisar o mejorar los conceptos ordinarios, sino que está obligada a respetar los criterios del uso encerrados en el lenguaje ordinario, bajo pena de cambiar el tópico o de hablar sin sentido. Así es como el conservadurismo conceptual nace del conductismo lógico.

El desarrollo de esta línea de pensamiento fue inconmensurablemente más sutil, precavido, y atenuado por condiciones y reconocimientos que el esbozo que hemos dado aquí (ver e. g., (54)) y desde luego había mucho de verdad en ella, pero prestó apoyo a una tesis dudosa: si los psicólogos y neurofisiólogos pensaron que podían estudiar a la mente, se equivocaron; el estudio de la mente era el estudio de los *conceptos* mentales ordinarios, y puesto que éstos tenían criterios conductuales ordinarios para su aplicación, una vez que tales criterios habían sido encontrados por los filósofos, no quedaba ya nada por hacer. Este mensaje era intolerable incluso para muchos de los partidarios del método de análisis del lenguaje ordinario que había conducido a tan embarazoso resultado. Desmantelar el edificio de argumentos y presuposiciones que había llevado a este impasse fue el trabajo de muchas manos, pero todo el proceso se encuentra resumido gráficamente en el ataque clásico de Putnam (79) a Malcolm (59, 60), en la polémica de Fodor con Ryle (38, 39) y en Chihara y Fodor (16).

III. LA TEORIA DE LA IDENTIDAD Y SUS DESCENDIENTES

La primera alternativa presentada al conductismo lógico capaz de llamar seriamente la atención fue la teoría de la identidad de la mente: las mentes son cerebros, y los contenidos de las mentes –dolores, pensamientos, sensaciones y cosas por el estilo– *son* simplemente (idénticos a) diversos acontecimientos, procesos y estados de nuestros cerebros. Los artículos iniciales en favor de la teoría escritos por Place, Feigl, Smart y Armstrong (7, 31, 32, 76, 101, 103), tenían el sabor de manifiestos, y su idea era la de asegurar de la manera más directa posible lo que se estimaba como el fundamento conceptual requerido para una ciencia de la

mente puramente fisicalista o materialista, un baluarte tanto contra las descalificaciones impertinentes del conductismo lógico como contra los excesos metafísicos de las alternativas dualistas.

(Dado que estos días está ampliamente garantizado que el dualismo no es un punto de vista serio contra el que luchar, sino más bien un precipicio hacia el que empujar a nuestros oponentes, quizá convenga una "refutación" en píldora del dualismo para alertar, si no convencer, al no iniciado. Supongamos, con los dualistas, que hay efectos (o acompañamientos) no físicos de eventos cerebrales. Entonces o bien la aparición de estos efectos carece ella misma de *cualquier otro efecto* sobre eventos subsiguientes en el cerebro (y con ello en el comportamiento) de la persona (epifenomenalismo), o bien los tiene (dualismo cartesiano o interaccionista). En el primer caso la postulación de efectos no físicos es claramente ociosa, pues *ex hypothesi* si los efectos cesasen de ocurrir (permaneciendo las demás cosas iguales) la gente seguiría haciendo el mismo tipo de declaraciones introspectivas, reconociendo sus dolores, y tomando tantas aspirinas como de costumbre. Incluso de manera más vívida, si los epifenómenos de una persona se fueran gradualmente retrasando hasta que estuvieran, digamos, diez años detrás de su vida física, ¡ni ella ni nosotros podríamos descubrirlo! En el último caso de dualismo interaccionista, ya que se requería la aparición de eventos *no físicos* (eventos que tienen localización temporal y presumiblemente una dependencia de una persona particular pero que carecen de una localización espacial y de masa-energía) para desencadenar en el cerebro eventos incuestionablemente físicos, se violarían las leyes de conservación de la física. En ambos casos se paga un precio exorbitante por el dualismo).

La teoría de la identidad iba a ser una teoría empírica, diseñada conceptualmente por la filosofía pero cuyos detalles serían completados por la ciencia, y se suponía que su ontología incluiría únicamente entidades bien acreditadas científicamente; ni *élan vital*, ni fuerzas psi, ni ectoplasma, sino solamente células cerebrales con su física y bioquímica. La declaración definitoria de la teoría de la identidad, la tesis de que los eventos mentales no son ni meramente paralelos a, ni coincidentes con, ni causados

por, ni acompañantes de eventos cerebrales, sino que *son* (estrictamente idénticos a) eventos cerebrales, divide a la gente de una manera curiosa. Para unos parece obviamente verdadera (aunque el expresarla apropiadamente pueda suponer pequeñas exigencias con los detalles), y para otros parece falsa de manera igualmente obvia. Los primeros tienden a ver todos los intentos de resistir a la teoría de la identidad como motivados por un miedo irracional al avance de las ciencias físicas, una especie de hylefobia humanista, mientras que los últimos tienden a descalificar a los teóricos de la identidad como cegados por una equivocada adoración a la ciencia para no ver el absurdo manifiesto de la tesis de identidad.

Este antagonismo ha originado una gran literatura en los últimos quince años –mucho de lo mejor de la cual (7, 20, 23, 31, 32, 34, 52, 65, 76, 88, 101, 103, 110), se encuentra en la antología de Borst (15)– y de ella ha surgido una panoplia de sofisticaciones que dejan muy atrás la teoría de identidad original y tosca a la vez que promueve sin modificar su proyecto básico de proporcionar un pedigrí conceptual a las ciencias físicas de la mente. Las dificultades con que se enfrenta la teoría de la identidad pueden dividirse, sin mayor pérdida ni distorsión, en tres áreas básicas: problemas que surgen de la ley de Leibniz, problemas sobre generalización, y rompecabezas lógicos abstractos sobre la relación de identidad.

La ley de Leibniz es la que hace de la identidad una relación lógica más fuerte que la mera semejanza, co-ocurrencia o equivalencia. Ella establece que "x es idéntico a y" implica que *todo* lo que es verdadero de la cosa denotada por "x" es verdadero de la cosa denotada por "y" y viceversa. El principio es inatacable, pues en el caso de cualquier identidad verdadera "x" e "y" denotarán la mismísima cosa, y todo lo que es verdadero acerca de esa cosa es verdadero acerca de ella, cualquiera que sea el nombre que le demos. Pero ahora supongamos que algún pensamiento mío es agudo, o profundo u obsceno; la teoría de la identidad debe entonces sostener que algún proceso o evento cerebral (el proceso o evento cerebral idéntico a ese pensamiento) es agudo, profundo u obsceno, y por lo menos a primera vista los procesos cerebrales no parecen ser el tipo de cosa que pueda ser aguda o pro-

funda u obscena, como tampoco pueden ser la raíz cuadrada de 7 o leales o capitalistas. Una réplica a esta objeción, expresada y defendida de diversa manera, es: eche otro vistazo, y verá que un proceso cerebral *puede* ser agudo o profundo u obsceno en precisamente aquellos casos en los que éste resulta ser un pensamiento. No todos los eventos en el cerebro son pensamientos; algunos son precisamente eventos metabólicos, por ejemplo, y no pueden ser agudos, pero no son los únicos eventos cerebrales. Por supuesto el éxito de esta posición depende enormemente de la existencia de una explicación de qué es lo que podría haber en un proceso o evento cerebral que lo convierta en un pensamiento (y un pensamiento determinado y no otro), y como veremos, hay problemas importantes en este área. (Otra réplica plausible a la objeción distingue el pensamiento como *evento* del pensamiento como *contenido* o *proposición*, y sostiene que propiedades tales como la agudeza se aplican propiamente al contenido, no al evento, pero de nuevo, esta posición no es más fuerte que la teoría que tenga uno sobre la individuación de los eventos por sus contenidos.)

De todas formas, aparentemente hay problemas más difíciles suscitados por la ley de Leibniz (55). Supongamos que soy sometido a un estímulo visual que subsiguientemente produce en mí una post-imagen anaranjada y redonda. No hay por supuesto una imagen anaranjada y redonda sobre mi retina. ¿Tendremos que suponer que hay un estado cerebral o un evento cerebral anaranjado o redondo que sea idéntico a mi post-imagen? Nada que no sea anaranjado y redondo puede ser idéntico a algo que es anaranjado y redondo, por lo tanto o bien mi post-imagen es idéntica a algún estado cerebral redondo y anaranjado (lo que, confío, nadie cree), o bien mi post-imagen no es anaranjada y redonda (¿podríamos preguntar a qué se parecería exactamente?), o bien mi post-imagen es algo (mental) diferente –y adicional a– cualquier evento cerebral, con lo que la teoría de la identidad es falsa, o bien simplemente no existen cosas tales como post-imagenes anaranjadas y redondas. Sin duda es tentador para cualquiera que no esté familiarizado ni impresionado por los horrores conceptuales del dualismo el abandonar en este punto la teoría de la identidad, y admitir a las post-imagenes y su parentela en su ontología

como subproductos no físicos epifenomenales de la actividad del cerebro, pero esto es precisamente lo que un teórico de la identidad rehusa hacer (19). En lugar de esto, el paso favorecido es el último: negar que, estrictamente hablando, haya cosas tales como post-imágenes. Lo que hay, se nos dice, son posesiones-de-post-imágenes, o experiencias-de-imágenes, y *éstas* no son ni anaranjadas ni redondas. La diferencia entre experimentar una post-imagen naranja y experimentar una post-imagen verde no es que la primera experiencia sea naranja mientras que la segunda lo es verde (19, 23, 93, 101, 103, 104). Esta jugada cuestiona lo que podríamos llamar la semántica normal del habla ordinaria sobre la mente. Nosotros hablamos, casual y ordinariamente, usando palabras como "dolor", "imagen", "creencia", "presentimiento", etc., como si estas palabras fueran aproblemáticas al referirse a expresiones que denotan elementos perfectamente reales de nuestras mentes (sean lo que sean). Pero quizá nuestra habla ordinaria incorpora una teoría-mito fosilizada, y una vez que la ciencia nos enseñe qué es lo que realmente sucede en nuestros cerebros, abandonaremos la búsqueda de referentes para nuestros términos mentales ordinarios, de la misma manera que hemos abandonado nuestra búsqueda de sirenas, brujas y demonios. Hablamos como si realmente hubiera cosas tales como post-imágenes, y así parece serlo; también hablamos como si el sol realmente se levantara por el este, y así lo parece, pero la ciencia puede convertir en metafórico al primer modo de hablar, como ha hecho con el segundo. Esta línea de razonamiento puede llegar al extremo de conducir a una variedad de posiciones que a menudo son reunidas bajo la rúbrica de "la forma desaparicional" de la teoría de la identidad: la ciencia no *descubrirá las identidades* de las unidades mentales problemáticas, sino más bien estas unidades desaparecerán como candidatas a la identificación una vez que la vieja visión científica sea superada por otra más sofisticada (23, 34, 88, 109).

El segundo conjunto de problemas de la teoría de la identidad concierne a la generalización. Estos problemas surgen porque el papel principal de las afirmaciones de identidad en las teorías es permitir la generalización. (Si esta nube es idéntica a

una colección de gotas de agua, entonces quizá son así las demás; si este gen es una molécula de DNA, la hipótesis que estamos tentados de contrastar es si todos los genes lo son). Pero no está nada claro que la teoría de la identidad pueda –o deba– proporcionarnos *alguna* generalización que esté más allá de su salvaguardadora pretensión de que todo elemento mental es idéntico a uno u otro elemento cerebral. Supongamos que María piensa en π al mediodía, y el teórico de la identidad sostiene que el pensamiento de María es idéntico a su proceso cerebral π del mediodía (que tiene características físicas definidas F, G, H, ...). No es ni remotamente plausible suponer que todo *pensamiento sobre* π es un proceso cerebral con características F, G, H, ..., aunque sólo sea porque no hay razón para suponer que las criatruras inteligentes en cualquier parte del universo necesitarían compartir nuestra neurofisiología o incluso nuestra bioquímica en orden a pensar en π (81). Ni siquiera es plausible que todo *pensamiento humano sobre* π, o incluso todo *pensamiento de María sobre* π sea idéntico a un proceso cerebral contenido en una clase especificable solamente en términos de las propiedades físicas de los miembros. Parecería un gravamen para cualquier teoría de la mente el que nos dijese qué es lo que hace que los pensamientos sean pensamientos, y qué es lo que hace que los pensamientos sobre π sean pensamientos sobre π, y que no resultase que las características que estamos buscando sean características físicas. La respuesta *débil* que un teórico de la identidad puede dar es que todo lo que él necesita afirmar es que cada evento mental *particular* (cada "instancia") es uno u otro evento cerebral (ningún evento mental es un evento *no* físico, no cerebral). Así pues, distinguimos una teoría de la identidad "instancia" de una teoría de la identidad "tipo", y abandonamos la última en filosofía de la mente. Hoy quizá nadie sostenga que los *tipos* de los elementos mentales puedan ser distinguidos directamente mediante características puramente físicas, pero casi nadie supone ya que ésta fuera una meta razonable del fisicalismo (20). La respuesta *fuerte* va más allá de la confesión de una teoría de la identidad instancia y sostiene que las marcas distintivas de los tipos de elementos mentales buscadas son definibles en términos de los papeles *causales* desempeñados (7, 52), o en términos de los

estados *lógicos* de la máquina abstracta de Turing "realizada" por el sistema nervioso de un ser humano (78, 80, 81, 82, 83), o en términos de los papeles *funcionales* desempeñados (37, 38, 39). A medida que las diferencias entre los tres puntos de vista han sido resueltas, y las diferencias detectadas y corregidas, ha emergido un único punto de vista ampliamente compartido llamado *funcionalismo*: los estados mentales son estados funcionales, esto es, estados individuales por su papel funcional dentro del sistema completo. Decir que un dolor o creencia particular es un estado funcional particular, es decir que cualquier cosa, prescindiendo de su composición química, forma, u otra característica física, que desempeñase el mismo papel funcional en un sistema funcional equivalente, sería el mismo dolor o creencia, y que nada que desempeñase semejante papel funcional puede dejar de ser tal dolor o tal creencia. El funcionalismo se ha convertido en la doctrina dominante en la filosofía de la mente de hoy (esto es, es la única teoría que es ampliamente criticada y defendida en las revistas), y por ello recibirá un tratamiento más detallado posteriormente. Debería ya estar claro que esta clase de funcionalismo tiene poco que ver con la rama del funcionalismo que encontramos en sociología o antropología.

La tercera área de investigación iniciada por la teoría de la identidad concierne a la logica de la relación misma de identidad. Smart y otros defensores tempranos de la teoría de la identidad supusieron inicialmente que el concepto de identidad estaba perfectamente a salvo y bien entendido. Había una táctica común para aclarar la teoría de la identidad dibujando analogías con identidades presumiblemente inocentes y familiares encontradas en dependencias menos enigmáticas, tales como la identidad de los rayos relampagueantes con descargas eléctricas en el aire, la Estrella de la Mañana con la Estrella de la Tarde, genes con moléculas de DNA. Sin embargo éstas no resultaron en absoluto aproblemáticas. Uno puede incluso preguntarse si hay alguna afirmación de identidad que sea aproblemática. Como las desanalogías, las distinciones y perplejidades acerca de la identidad misma comenzaron a multiplicarse, el problema de la identidad adquirió vida propia como un tema lógico y metafísico, y las investigaciones de-

jaron de iluminar de manera específica los problemas de la mente. Esto se debía en parte a la pérdida de confianza por los filósofos de la mente inclinados fisicalísticamente en cualquier noción de identidad. Después de todo, la forma desaparicional de la teoría de la identidad estaba mal bautizada; no se trata realmente de una teoría de la identidad sino solamente de una alternativa fisicalista a la teoría de la identidad, y tal como lo observó Putnam en su primera exposición del funcionalismo (78), la cuestión de si hay que *identificar* un estado lógico de una máquina (realizada) de Turing con una realización concreta en hardware es un tema metafísico relativamente ocioso. Cuando las dos tácticas, el punto de vista de la desaparición y el funcionalismo, se juntan –ya que hay fuertes razones para hacerlo (29, 30)– cualquier identidad que quede todavía por reconocer concierne a entidades más bien curiosas y abstractas. (Por ejemplo uno podría proclamar que el *estado de cosas* de Tom al tener un dolor era idéntico al *estado de cosas* de su cerebro cuando está en cierta situación funcional particular.) En este punto, la motivación inicial de sostener *identidades* que nos salvasen de los temas fantasmales del dualismo ha desaparecido.

Por ello resultó algo anacrónico el que en 1971 Saul Kripke incluyese una asombrosa "refutación" de la teoría de la identidad como un subproducto ilustrativo de su extraordinariamente influyente, incluso revolucionaria, concepción de la necesidad, la designación y la identidad (50, 51). El argumento de Kripke depende de ciertas innovaciones técnicas. Un "designador rígido" es una expresión que designa la misma entidad "en todos los mundos posibles". Así, "Benjamin Franklin" es un designador rígido, mientras que "el inventor de los bifocales" no lo es, aunque en *este* mundo las dos expresiones designan al mismo individuo. Kripke arguye convincentemente que todas las afirmaciones adecuadas de identidad están compuestas de términos que son designadores rígidos y por lo tanto cuando son verdaderas, son verdaderas no contingente, sino necesariamente. La aplicación a la filosofía de la mente viene cuando argumenta que expresiones tales como "mi dolor" y "mi estado cerebral" son ambas designadores rígidos, pero las afirmaciones de identidad compuestas con

ellas no pueden ser *necesariamente* verdaderas, por lo tanto no pueden ser verdaderas en absoluto. (Cfr. 62). El argumento de Kripke es sutil e ingenioso y merece un cuidadoso estudio, pero no ha logrado consenso. Incluso si uno acepta la nueva teoría de Kripke sobre la identidad y la necesidad, sus argumentos para mostrar que los términos relevantes son designadores rígidos no solamente son vulnerables debido a excepciones claras, sino además parecen requerir suposiciones que los modernos materialistas (tales como Armstrong) ya se han tomado el trabajo de negar (33, 58). Retrospectivamente, el argumento de Kripke no parece haber revitalizado el dualismo, sino más bien ha dado a los materialistas más y posiblemente mejores (más profundas) razones para evitar ciertas declaraciones tentadoras de identidad que en su mayor parte ya habían aprendido a evitar.

IV. ¿POR QUE EL FUNCIONALISMO?

Como corresponde a un punto de vista que es el favorito dominante, el funcionalismo tiene mucho que decir en su apoyo. No sólo parece evadir satisfactoriamente las objeciones filosóficas a todas las otras formas de materialismo, sino que está particularmente bien diseñado para servir como puntal para el trabajo actual en psicología, lingüística, y cibernética o inteligencia artificial. Todas estas disciplinas operan un tanto autoconscientemente en un cierto nivel de abstracción, y el funcionalismo proporciona la razón y justificación para semejante estrategia. En cierta forma, no hay nada nuevo en esto. Psicólogos tan diferentes como Freud y Skinner han compartido la táctica funcionalista básica: así como Freud eventualmente cayó en la cuenta que cualesquiera pretensiones que pudiera sostener acerca de la localización física, composición, u operación de sus entidades funcionalmente diferenciadas, el ello, el yo y el superego, eran especulaciones prematuras, también Skinner, aun garantizando que sin duda había *algunos* mecanismos fisiológicos internos que favorecían el refuerzo, renunció a la especulación o el compromiso sobre el tema y se dedicó, con Freud, a organizar las consecuencias predichas de interaccio-

nes funcionalmente caracterizadas bajo una variedad de circunstancias.

(Skinner, por supuesto, ha sido menos claro que Freud respecto al hecho de que él se ha comprometido con un modelo funcionalista de los procesos internos. Piensa que su periferalismo evita aquel "cargo". Que tuviera que preocuparse en absoluto a este respecto resulta embarazoso para los filósofos, porque si él y los demás conductistas no hubieran bebido tan profundamente en el pozo del positivismo lógico y del "operacionalismo" e "instrumentalismo" a la moda en esa época, no se hubieran motivado a constreñir su teorización dentro de una noción de rigor tan paralizante y mal orientada. Veinte años después que el positivismo lógico e incluso sus obituarios han sido olvidados por los filósofos, sus dogmas se encuentran vivos y con buena salud en los libros de texto de los conductistas. Por eso los filósofos de la mente hoy día recomiendan sus doctrinas a sus colegas en otros campos con cierta modestia y turbación. Parece, de paso, que la historia está a punto de repetirse a sí misma. Ahora que los filósofos de la mente han tenido finalmente éxito en desterrar su miedo a entidades internas, "para-mecánicas", un miedo que aprendieron de Ryle (27, 39, 93), Roy Schafer (96) se ha tomado en serio los rigores de Ryle, y nos promete iniciar una era de conductismo lógico ryleano en la teoría psicoanalítica. Los argumentos ahora ya corrientes contra los análisis conductual-disposicionales de Ryle (que son manifiestamente incapaces de ramificarse o generalizarse; pues o bien son obviamente falsos o bien son "salvados" mediante cláusulas *ceteris paribus* que los vuelven vacuos) parecen por lo menos a primera vista pasar como críticas de Schafer).

Más específicamente, el funcionalismo proporciona los puntales conceptuales para el trabajo actual en psicología cognitiva, psicolingüística y modelaje de inteligencia artificial. En estas disciplinas una estrategia de investigación puede caracterizarse en términos de (una versión de) la distinción de Chomsky entre competencia y actuación: dada una especificación de una clase de competencia, digamos una competencia discriminativa, o una competencia lingüística, la tarea es diseñar un modelo de actuación –a menudo un programa de simulación de computador– que

exhiba esa competencia (por lo común artificialmente aislada y afilada de varias formas) y si es posible tenga también una pretensión de "realidad psicológica" (27, 39). Es decir, sólo lograr despellejar el gato puede ser un gran logro, conseguir despellejarlo de la forma en que la gente parece hacerlo es todavía mejor. Esta clase de estrategia de investigación permite la exploración de limitaciones y dificultades altamente abstractas (¿Cómo puede *algo* aprender un lenguaje natural? ¿Cómo puede *algo* lograr una capacidad general para el reconocimiento de patrones en un ambiente no estereotipado?) sin preocuparse de la mecánica y la bioquímica de "realizaciones" concretas en la cabeza, aunque a la vez sin abandonar la limitación fisicalista fundamental de que los sistemas funcionalmente descritos sean *de alguna manera* físicamente realizables (25). En lo más general y abstracto, este tipo de investigación se funde con la investigación en epistemología y filosofía de la mente, y es precisamente en este terreno de encuentro donde se hace hoy el trabajo más prometedor y excitante (e. g., 38, 45, 64, 84, 92, 95).

Por supuesto el funcionalismo tiene sus escépticos y críticos (11, 12, 47, 56, 57, 66, 90, 100), y los problemas más perturbadores visten trajes sorprendentemente tradicionales. Primero están los problemas sobre los *qualia* de la experiencia, la manera como ella siente ser consciente, y segundo, están los problemas, aludidos anteriormente, sobre cómo un estado o evento funcionalmente individuado puede tener *significado o contenido* en algún sentido presumiblemente vigoroso requerido por las entidades mentales. ¿Cómo podría un estado funcional ser un dolor, y cómo podría un estado funcional ser un pensamiento sobre π?

V. QUALIA

No hay una definición satisfactoria de los "qualia", que parezca haberse convertido en el término favorito en las discusiones, pero el sentido general requerido de lo que se supone que son los "qualia" se capta fácilmente mediante un ejemplo que figura de una manera central en la discusión presente. Suponga el lector

que cuando mira un claro cielo "azul" ve lo que yo veo (al menos en lo que toca al color) cuando miro una manzana madura, y así a través de todos los colores del espectro. Su espectro percibido es, digamos una inversión sistemática del mío. Pero como el lector aprendió el uso de las palabras sobre colores de la misma manera que yo (sus padres señalaban el cielo y decían "azul" etc.) nuestro uso de las palabras sobre colores sería indistinguible, y puesto que ambos asociaremos el color del hierro incandescente de la fragua del herrero con el calor, y el color que percibimos que tiene el hielo con el frío, incluso nuestras descripciones secundarias de color (el rojo es un color caliente, el azul es frío, etc.) podrían encajar. De hecho, por lo que respecta a los colores percibidos y otros "qualia" como dolores, picores, sonidos, aromas y gustos, ¿tenemos en absoluto alguna evidencia, o razón para creer, que dos personas cualesquiera experimentan "qualia" similares bajo circunstancias perceptuales similares? Esto, el experimento conceptual del "espectro invertido", no es nada nuevo. Fue un argumento popular entre los verificacionistas que tomaron la "autoevidente" no-verificabilidad-en-principio de la hipótesis como señal de la falta de significado de la suposición inicial de que hay sensaciones internas o —para usar un término que la filosofía tomó de un psicólogo, Tolman— "sensaciones puras" del tipo requerido. Como dijo Wittgenstein, "un 'proceso interno' se encuentra en la necesidad de criterios exteriores" (117), y aunque la exégesis de esta observación es controvertible, es común y fácil interpretarla como una expresión de conductismo lógico, una declaración de la incoherencia de cualquier doctrina que admita sensaciones privadas o experiencias de "qualia".

Se ha señalado a menudo que el funcionalismo de hoy es un descendiente espiritual del conductismo lógico. Donde el conductismo lógico decía que tener un dolor era cuestión de comportarse o tener la disposición de comportarse de una manera particular, los funcionalistas dicen que tener un dolor es cuestión de encontrarse en un estado funcional de determinado tipo, a saber, un estado que *inter alia* predispone a uno a cierto comportamiento bajo ciertas condiciones. Desde una posición panorámica, la única diferencia entre las dos doctrinas parece estar en el deseo

del funcionalista de abandonar el periferalismo (aceptando explícitamente estados funcionales internos), y en el deseo concomitante de definir "función" no sólo en términos de disposiciones para comportarse, sino también en términos de disposiciones a cambiar el estado funcional. Si, como dice Skinner, la piel no es esa frontera tan importante, si el cambio de estado interno cuenta como comportamiento, entonces el funcionalismo es simplemente conductismo lógico con traje nuevo. Por ello no es sorprendente que en la precipitada acometida por salir del verificacionismo propia del pasado reciente, los filósofos intentaran darle la vuelta al argumento del espectro invertido y mostrar que el funcionalismo comete el pecado del verificacionismo al ser incapaz de garantizar el sentido a algo que (como revela la discusión inteligente) realmente tiene sentido: la hipótesis de la inversión del espectro (11, 12, 56, 100). De manera similar, Nagel ha argumentado (66) que hay ciertas hipótesis innegablemente significativas sobre nuestras vidas internas y las vidas internas de los demás, sobre qué es "ser semejante a" una persona o a un perro o a un murcielago, de las cuales el funcionalismo no puede dar cuenta alguna.

La estrategia del debate es transparente. Los amantes de los "qualia" intentan establecer que el funcionalismo inevitablemente *deja algo fuera*: los maravillosos sabores, tonos y colores que hacen que valga la pena vivir. Los funcionalistas intentan mostrar que ellos no han dejado fuera nada *real*, y que la alternativa al funcionalismo sólo puede ser alguna variedad sin fundamento del epifenomenalismo. El asunto no está resuelto todavía, y no será resuelto con la victoria clara de un lado sobre el otro en un debate puramente conceptual. La carga para el funcionalismo es inseparable de la carga correspondiente a la variedad de teorías cognitivistas a las que proporciona apoyo conceptual. Si se desarrolla una teoría psicológica empírica que sea fuertemente confirmada y a la vez predictiva de la rica variedad de fenómenos de conciencia, podemos examinarla para responder la pregunta. Si asigna un papel teórico a algo parecido a los "qualia", "aceptaremos" a éstos en nuestra ontología, pero como entidades teóricas, no como epifenómenos; si no aparece semejante papel para ser desempeñado, entonces la misma capacidad de la teoría minará las

intuiciones que ahora hacen tan antiintuitiva la negación de los "qualia" (26, 29). Si ninguna teoría concebida funcionalísticamente es capaz de manejar los hechos indiscutibles de la conciencia, entonces este fallo de la teoría empírica, y no cualquier argumento puramente filosófico, mostrará que de hecho el funcionalismo deja algo fuera. Las investigaciones filosóficas del tema no son, sin embargo, completamente parasitarias del avance de la investigación empírica; pueden iluminar el terreno, revelando callejones sin salida y escollos, sin intentar dictar la solución.

VI. REPRESENTACION INTERNA Y EL PROBLEMA DEL SIGNIFICADO

Un papel de apoyo similar puede observarse en las contribuciones filosóficas a la otra perplejidad fundamental encarada por las teorías funcionalistas en psicología: el problema del significado o contenido. Ningún problema de la filosofía ha recibido mayor ni más experta atención en los años recientes que el problema del significado, y un resumen del trabajo realizado en este área tendría que ser tan largo como un libro. (Hay varias y buenas antologías del trabajo reciente (21, 22, 36, 46, 106).) Casi todo él es relevante, al menos indirectamente, para los problemas de la mente, y una parte es de importancia central.

Al final del siglo XIX Franz Brentano sostuvo haber descubierto la característica que separaba lo mental de lo físico: la intencionalidad. Los fenómenos mentales, decía, diferían de los fenómenos físicos en que siempre estaban dirigidos a un objeto, (el objeto del pensamiento, o del deseo o de la percepción ...) o relacionados con un contenido (el contenido de una esperanza o pensamiento o creencia ...). Esta era una conexión de un tipo especial, pues los objetos de los fenómenos mentales gozaban de una curiosa clase de "inexistencia". Puedo querer una corbeta sin que exista la corbeta que yo quiero; el objeto de mi deseo es "intencionalmente inexistente" (que *no* significa deliberadamente inexistente: la intencionalidad no tiene directamente nada que ver con lo que uno intenta hacer). En los cincuenta, Chisholm (17)

revivió la noción de intencionalidad de Brentano y (usando la táctica del ascenso semántico con gran efecto) la convirtió en una característica del *lenguaje*: las sentencias que usamos típicamente para hablar de eventos mentales tienen ciertas peculiaridades lógicas. Chisholm intentó caracterizar tales peculiaridades lógicas de manera que la distinción entre *sentencias* intencionales y no-intencionales reflejase la distinción de Brentano entre *fenómenos* intencionales y no-intencionales. En los años siguientes, la intencionalidad, considerada como una característica lógica de cierta clase de proposiciones, ha sido exhaustiva y fructíferamente estudiada, aunque es justo decir que no se ha logrado una unanimidad muy ampliada sobre la definición precisa, estatuto y papel del discurso intencional (18, 24, 48, 53, 85, 97, 98, 112). El interés por la intencionalidad ha sobrevivido a los desacuerdos y dificultades que rodean su definición no sólo porque represente una perplejidad de la teoría lógica todavía no resuelta, sino porque intuitivamente señala una división importante en nuestro esquema conceptual, aunque no exactamente la división que Brentano suponía. Las locuciones intencionales de nuestro lenguaje son aproximadamente lo que Russell llamaba locuciones de actitud proposicional; locuciones que toman típicamente cláusulas –"que" y por lo tanto forman proposiciones complejas a partir de otras, pero sin ser funciones de verdad: p. ej., "Tomás cree que está lloviendo" contiene la proposición "está lloviendo", pero el valor de verdad del todo es independiente del valor de verdad de la proposición encerrada. De nuevo *aproximadamente*, las locuciones intencionales son aquellas locuciones en nuestro lenguaje que relacionan a las personas, a sus partes, sus actos, y sus artefactos con *proposiciones*. Todavía más aproximadamente, usamos las locuciones intencionales para dotar a las cosas –absolutamente a cualquier clase de cosas– de *significado*: si queremos decir lo que Tomás cree, lo que la sentencia significa, lo que el ojo de la rana dice al cerebro de la rana, lo que es innata y tácitamente sabido por el aprendiz infante del lenguaje (107), lo que el ego está tratando de resguardar frente al superego, qué información está almacenada en la cinta, lo que Houston está transmitiendo al Mariner IV, entonces utilizamos el discurso intencional. Una teoría de

la intencionalidad, por lo tanto, debiera ser una teoría que hiciera explícitos los lazos conceptuales entre estos diferentes enfoques, cognitivistas, teóricos de la información, semánticos, y que hiciera constar las limitaciones y suposiciones involucradas en la asignación a las cosas de aspectos intencionalmente caracterizados. Que aquí hay problemas muy profundos puede subrayarse con sólo considerar el caso de la psicología cognitiva.

Lo que une a los psicólogos cognitivos y los distingue mejor de otros teóricos de la psicología es su deseo de ver al individuo no solamente como un sistema de partes o subsistemas funcionalmente individuados, sino de partes individuadas *intencionalmente*, partes cuyas funciones son "decir que *p*", "recordar que *q*", "descifrar que *r*" –codificar, almacenar, transmitir o transformar parcelas de información. De hecho, es el uso de locuciones intencionales en su ciencia lo que mejor les diferencia de los conductistas, en particular Skinner, quien ha concebido equivocadamente esta diferencia como si fuese una diferencia entre dualistas ("mentalistas") y materialistas (24, 25, 28). Los psicólogos cognitivos han pensado generalmente que su uso del habla de la información no sólo es apropiado y está bien fundamentado en el rigor matemático de la teoría de la información y la ciencia de la computación, sino que es positivamente un gran paso adelante en la conceptualización fructífera de la psicología, y creo que al final así será, pero no sin problemas. Decir que la función de cierto sistema es llevar cierta información de *a* a *b* no es lo mismo que decir que la función del tubo es llevar lubricante al cojinete, o que la función del teletipo es llevar cadenas de símbolos de un lugar a otro. Desplazar información no es algo tan fácil de concebir, o si, para algún sentido especial de información, resulta *ser* fácil de concebir, es que se ha comprado la simplicidad al coste de posponer la solución al problema central de la intencionalidad, tal como lo demostrará un pequeño experimento conceptual.

Suponga el lector que se encuentra encerrado en una habitación sin ventanas, con dos paredes cubiertas de luces centelleantes, dos paredes cubiertas de pequeños botones, y una nota diciéndole que se encuentra prisionero en el centro de control de un robot gigante de cuya seguridad depende ahora su propia vida.

Su tarea es simplemente guiar al robot a través de su medio ambiente más bien peligroso, aprendiendo a discriminar y arreglárselas con cualquier cosa que se presente, procurando "alimento" y puerto seguro para el robot en la noche (de manera que el lector pueda dormir) y evitando los peligros. Toda la *información* que uno necesite es transmitida por las luces centelleantes, y la actividad motora del robot se controla pulsando los botones. Sin embargo, el lector advierte, para su consternación, que ninguna de las luces ni ninguno de los botones tiene letreros. No puede decir si la luz de la esquina superior izquierda que centellea insistentemente está avisando de un peligro, indicando una "barriga llena", informándole la posición del sol, o pidiendo grasa para el cojinete del talón. El lector no sabe si cuando oprime un botón y la luz se apaga, ha rascado un picor, cerrado su visión de algo, o destruido un atacante.

Está claro que, si esto es todo lo que le dan a uno para seguir adelante, su tarea es imposible; si tuviera éxito guiando a su robot a lo largo del día sería por pura suerte. Sin embargo en un sentido (y en uno muy familiar a los psicólogos cognitivos) le ha sido proporcionada toda la información que necesita. Ya que nosotros no necesitamos suponer que las luces son meros repetidores de estimulación periferica: su centelleo puede representar los productos de análisis perceptuales de la maquinaria tan sofisticados como uno quiera, y similarmente puede suponer que el output inicia intrincadas acciones guiadas por sistemas jerárquicos de subrutinas informadas por retroalimentación de muchas capas. En pocas palabras, la serie completa de sistemas diseñada por los psicólogos cognitivistas podría estar construida dentro del robot, de manera que transmitiese a su centro de control información altamente mediatizada y refinada, y sin embargo, aunque en un sentido la información estaría allí, en otro sentido más importante, no lo estaría. Aun así, la tarea descrita es en cierto sentido precisamente la tarea del cerebro; carece de ventanas por las que mirar afuera en orden a correlacionar aspectos del mundo con su input.

El problema de la sala de control podría resolverse para el lector, por supuesto, si todas las luces y botones tuvieran letreros correctos (en un lenguaje que él conociese), pero ésta difícilmente

puede ser la solución para el cerebro. El trabajo de lograr que la información input sea *interpretada* correctamente *no* es pues una cuestión de lograr traducir o transcribir la información a un código interno particular, a no ser que obtener la información en ese código sea *ipso facto* colocarla en la posición funcional para gobernar el repertorio conductual de todo el organismo. Este es el problema del significado que eventualmente debe ser enfrentado por cualquier teórico que quiera recurrir a "representaciones internas" como explicativas de los fenómenos psicológicos. Algún trabajo reciente en filosofía se dirige directamente a este punto (23, 27, 29, 39, 45, 64, 84, 92, 95) y esta literatura depende en diverso grado del trabajo fundamental sobre el significado que ha sido desarrollado en respuesta a temas tan centrales como la tesis de Quine sobre la indeterminación de la traducción radical (85), el trabajo de Austin sobre los actos de habla (9), y el tratamiento del significado no natural de Grice (40, 41). Una controversia particularmente activa dentro del área de la representación interna se refiere a la naturaleza de los vehículos de representación supuestos: ¿Son proposicionales (como sentencias) o imaginísticos o analógicos (como cuadros o mapas) o hay otras clases de "estructuras de datos" sin análogos familiares entre los vehículos externos de representación? El trabajo de los filósofos en este área se funde de manera uniforme con el de los psicólogos y cibernéticos (e. g., 64, 84), y puede esperarse que este intercambio interdisciplinar no forzado produzca durante los proximos años algunos avances genuinos en perspectiva.

VII. OTRAS AREAS DE ACTIVIDAD EN EL PRESENTE

Se esperaba que este informe sobre la actividad en la presente situación captase las líneas principales de investigación, pero ha quedado sin mencionar algún trabajo que en modo alguno es periférico. Por ejemplo, el estatuto confuso de la "introspección" ha sido estudiado cuidadosamente en una literatura en aumento acerca del "acceso privilegiado" y de la supuesta "incorregibilidad" de los informes introspectivos (2, 3, 6, 7, 23, 29, 42, 49, 61,

72, 89, 91, 102, 111), y los filósofos han dirigido útilmente su atención a temas específicos que surgen en otros campos, tales como los casos de separación cerebral de Sperry y los diversos puntos de vista que han sido propuestos sobre los papeles diferentes de los hemisferios cerebrales (e. g., 67), y la fisiología del dolor (30, 74, 75). Tambien ha habido un redescubrimiento de Freud (118), y el problema del auto-engaño, en particular, ha provocado algún trabajo excelente (35, 44, 87, 105). La literatura sobre "mentes y máquinas" ha evolucionado desde su preocupación inicial con la cuestión "¿puede pensar una máquina?" (e. g., 4) hacia un examen mucho más informado y detallado de los temas conceptuales que están en el corazón de la investigación actual en inteligencia artificial y teoría de autómatas (e. g., 13, 14, 68, 69, 70, 71). Un pronóstico optimista sería el de que estas líneas de investigación convergirán hacia una comprensión bastante estable y ampliamente aceptada de los puntales conceptuales del enfoque funcionalista y fisicalista de la mente; pero si se tiene en cuenta la gran diversidad de opiniones en el campo, las profundas dificultades que ya ahora puede verse que esperan a este punto de vista, y la lección de la historia, una predicción más realista sería la de que este consenso todavía fragmentario y tenue resultará tan evanescente como sus predecesores y será reemplazado por un conjunto de doctrinas y problemas hoy inimaginable.[1]

Traducción de *José Ruiz Arbeloa*

[1] Esta visión panorámica fue originalmente encargada por el *American Journal of Psychiatry*, que luego la rehusó sobre la base de estar "mucho más allá"de sus lectores, a pesar de mis esfuerzos de explicación. Confío en que el tono algo elemental del trabajo se explique por este hecho relativo a la audiencia a la que originalmente se dirigió. Estoy en deuda con Ned Block, Jeff Titon, Bo Dahlbom, y Sue Stafford por sus valiosas críticas a redacciones anteriores de este trabajo.

BIBLIOGRAFIA

(1) ALBRITTON, Rogers, "On Wittgenstein's Use of the Term 'Criterion'," *Journal of Philosophy*, vol. 56 (1959), pp. 845-57, reimpreso en Pitcher, George (ed.), *Wittgenstein: The Philosophical Investigations* (New York, 1968).

(2) ALSTON, William P., "Varieties of Privileged Access," *American Philosophical Quartely*, vol. 8 (1971), pp. 223-51.

(3) ALSTON, William P., "Self-Warrant: A Neglected Form of Privileged Access," *American Philosophical Quartely*, vol. 13 (1976), pp. 257-72.

(4) ANDERSON, Alan R. (ed.), *Minds and Machines* (Englewood Cliffs, NJ, 1964).

(5) ANSCOMBE, G. E. M., *Intention* (Oxford 1957).

(6) ARBIB, Michael A., "Consciousness: the Secondary Role of Language," *Journal of Philosophy*, vol. 69 (1972), pp. 579-91.

(7) ARMSTRONG, David M., *A materialist Theory of the Mind* (London 1968).

(8) AUSTIN, John L., *Ensayos Filosóficos*. Traducción y presentación de Alfonso García Suárez. Madrid: Revista de Occidente, 1975.

(9) AUSTIN, John L., *How to Do Things With Words* (Oxford, 1962).

(10) AUSTIN, John, L., *Sentido y percepción*. Traducción de Alfonso García Suárez y Luis MI. Valdés Villanueva. Presentación de Alfonso García Suárez. Madrid: Tecnos, Serie de Filosofía y Ensayo, 1981.

(11) BLOCK, Ned J. y FODOR, Jerry, "What Psychological States are Not", *The Philosophical Review*, vol. 81 (1972), pp. 159-81.

(12) BLOCK, Ned J. "Troubles With Functionalism", en C. Wade Savage (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 9

(13) BODEN, Margaret, "Freudian Mechanisms of Defense: A Programming Perspective," in Wollheim, Richard (ed.), Freud: *A Collection of Critical Essays* (Garden City, NJ, 1974).

(14) BODEN, Margaret, *Inteligencia artificial y hombre natural* [1977]. Madrid: Tecnos, Serie de Filosofía y Ensayo (en preparación).

(15) BORST, Clive V. (ed.), *The Mind/Brain Identity Theory* (New York, 1970).

(16) CHIHARA, Charles S. and FODOR, Jerry, "Operationalism and Ordinary Language: A Critique of Wittgenstein," *American Philosophical Quartely*, vol. 2 (1965), pp. 281-95, reimpreso en Pitcher, George (ed.), *Wittgenstein: the Philosophical Investigations* (New York, 1968).

(17) CHISHOLM, Roderick, *Perceiving: A Philosophical Study* (Ithaca, NJ, 1957).

(18) CHISHOLM, Roderick, "On Some Psychological Concepts and the 'Logic' of Intentionality," en Castaneda, Hector-Neri (ed.), *Intentionality, Minds and Perception* (Detroit 1967).

(19) CORNMAN, James W., *Materialism and Sensations* (New Haven, CT, 1971).

(20) DAVIDSON, Donald, "Sucesos mentales", en *Ensayos sobre acciones y sucesos*, Madrid: Tecnos, Serie de Filosofía y Ensayo (en preparación).

(21) DAVIDSON, Donald y HARMAN, Gilbert, (eds.), *Semantics of Natural Language* (Dordrecht, 1972).

(22) DAVIDSON, Donald y HARMAN, Gilbert, *The Logic of Grammar* (Encino, CA, 1975).

(23) DENNETT, Daniel C., *Content and Consciousness* (London, 1969).

(24) DENNETT, Daniel C., "Intentional Systems," *The Journal of Philosophy*, vol. 68 (1971), pp. 87-106.

(25) DENNETT, Daniel C., "Why the Law of Effect Will Not Go Away," *Journal of The Theory of Social Behaviour*, vol. 5 (1975), pp. 169-87.

(26) DENNETT, Daniel C., "Are Dreams Experiences?," *The Philosophical Review*, vol. 85 (1976), pp. 151-71.

(27) DENNETT, Daniel C., "Critical Notice of Jerry Fodor, *The Language of Thought* (New York, 1975)" en *Mind*, vol. 86 (1977), pp. 265-280.

(28) DENNETT, Daniel C., "Skinner Skinned," en Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Montgomery, Vermont, 1978).

(29) DENNETT, Daniel C., "Hacia una teoría cognitiva de la conciencia", en: *Ensayos sobre inteligencia artificial*, Madrid: Cuadernos Teorema.

(30) DENNETT, Daniel C., Por qué no puedes fabricar un computador que sienta dolor, en: *Ensayos sobre inteligencia artificial*, Madrid: Cuadernos Teorema.

(31) FEIGL, Herbert, "The 'Mental' and the 'Physical'," en Feigl, Herbert y Scriven, Michael y Maxwell, Grover (eds.), "Concepts, theories and the Mind-Body Problem," *Minnesota Studies in the Philosophy of Science*, vol. 2 (Minneapolis, 1958).

(32) FEIGL, Herbert, The *"Mental" and the "Physical" The Essay and Postscript* (Minneapolis, 1967).

(33) FELDMAN, Fred, "Kripke and the Identity Theory," *The Journal of Philosophy*, vol. 71 (1974), pp. 665-676.

(34) FEYERABEND, Paul, "Materialism and the Mind-Body Problem," *Review of Metaphysics*, vol. 17 (1963), pp. 49-66.

(35) FINGARETTE, Herbert, *Self-Deception* (London, 1969).

(36) FODOR, Jerry y KATZ, Jerrold J. (eds.), *The Structure of Language* (Englewood Cliffs, NJ, 1964).

(37) FODOR, Jerry, "Explanation in Psychology," en Black, Max (ed.), *Philosophy in America* (Ithaca, NY, 1965).

(38) FODOR, Jerry, *La explicación psicológica*. Trad. de José E. García Albea, Madrid: Cátedra, Colección Teorema, 1980.

(39) FODOR, Jerry, *The Language of Thought* (New York, 1975).

(40) GRICE, H. P., "Meaning," The *Philosophical Review*, vol. 66 (1957), pp. 377-88.

(41) GRICE, H. P., "Utterer's Meaning and Intentions," *The Philosophical Review*, vol. 78 (1969), pp. 147-77.

(42) GUNDERSON, Keith, "Content and Consciousness and the Mind-Body Problem," *The Journal of Philosophy*, vol. 69 (1972), pp. 591-604.

(43) GUSTAFSON, Donald (ed.), *Essays in Philosophical Psychology* (London, 1967).

(44) HAMLYN, D. W. y MOUNCE, H. O., "Self-Deception," *Aristotelian Soc. Supplement* (1971), pp. 45-72.

(45) HARMAN, Gilbert, *Thought* (Princeton, NJ, 1973).

(46) HARMAN, Gilbert, (Compilador), *Sobre Noam Chomsky: Ensayos Críticos*, Madrid: Alianza Universidad.

(47) KALKE, William, "What is Wrong with Fodor and Putnam's Functionalism," *Nous*, vol. 3 (1969), pp. 83-93.

(48) KAPLAN, David, "Quantifying in," *Synthese*, vol. 19 (1968), pp. 178-214.

(49) KLEIN, Barbara V. E., "Some Consequences of Knowing Everything (Essential) There is to Know About One's Mental States," *Review of Metaphysics*, vol. 29 (1975), pp. 3-18.

(50) KRIPKE, Saul, *Identidad y necesidad* Trad. de Margarita M. Valdés, México: Cuadernos Crítica, U.N.A.M., Instituto de Investigaciones Filosóficas 1978.

(51) KRIPKE, Saul, "Naming and Necessity," en Davidson, Donald y Harman, Gilbert (eds.), *Semantics of Natural Language* (Dordrecht, 1972).

(52) LEWIS, David, "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, vol. 50 (1972), pp. 249-58.

(53) LYCAN, William G., "On Intentionality and the Psychological," *American Philosophical Quartely*, vol. 6 (1969), pp. 305-12.

(54) LYCAN, William G., "Noninductive Evidence: Recent Work on Wittgenstein's Criteria," *American Philosophical Quartely*, vol. 8 (1971), pp. 109-125.

(55) LYCAN, William G., "Materialism and Leibniz' Law," *Monist*, vol. 56 (1972), pp. 276-87.

(56) LYCAN, William G., "Inverted Spectrum," *Ratio*, vol. 15 (1973), pp. 315-19.

(57) LYCAN, William G., "Mental States and Putnam's Functionalist Hypothesis," *Australasian Journal of Philosophy*, vol. 52 (1974), pp. 48-62.

(58) LYCAN, William G., "Kripke and the Materialists," *The Journal of Philosophy*, vol. 71 (1974), pp. 677-89.

(59) MALCOLM, Norman, "Dreaming and Skepticism," *The Philosophical Review*, vol. 65 (1956), pp. 14-37.

(60) MALCOLM, Norman, *Dreaming* (London, 1959).

(61) MARGOLIS, Joseph, "Indubitability, Self-Intimating States and Logically Privileged Access," *The Journal of Philosophy*, vol. 67 (1970), pp. 918-31.

(62) MATSON, Wallace, *Sentience* (Bekerley, CA 1976).

(63) MELDEN, A. I., *Free Action* (London, 1961).

(64) MINSKY, Marvin, "Frame Systems: A Framework for Representing Knowledge," MIT Artificial Intelligence Lab Report, 1973.

(65) NAGEL, Thomas, "Physicalism," *The Philosophical Review*, vol. 74 (1965), pp. 339-56.

(66) NAGEL, Thomas, "What Is It Like To Be A Bat?," *The Philosophical Review*, vol. 83 (1974), pp. 435-50.

(67) NAGEL, Thomas, "Brain Bisection and the Unity of Consciousness," *Synthese*, vol. 22 (1971), pp. 396-413.

(68) NELSON, Raymond, "Behaviorism, Finite Automata and Stimulus Response Theory," *Theory and Decision*, vol. 6 (1975), pp. 249-67.

(69) NELSON, Raymond, "On Machine Expectation," *Synthese*, vol. 31 (1975), pp. 129-39.

(70) NELSON, Raymond, "On Mechanical Recognition," *Philosophy of Science*, vol. 43 (1976), pp. 24-52.

(71) NELSON, Raymond, "Mechanism, Functionalism and the Identity Theory," *The Journal of Philosophy*, vol. 73 (1976), pp. 365-85.

(72) PARSONS, Kathryn P., "Mistaking Sensations," *The Philosophical Review*, vol. 79 (1970), pp. 201-13.

(73) PETERS, Richard, *The Concept of Motivation* (London, 1958).

(74) PITCHER, George, "Pain Perception," *The Philosophical Review*, vol. 79 (1970), pp. 368-93.

(75) PITCHER, George, "The Awfulness of Pain," *The Journal of Philosophy*, vol. 67 (1970), pp. 481-492.

(76) PLACE, U. T., "Is Consciousness a Brain Process?," *British Journal of Psychology*, vol. 42 (1956), pp. 44-50.

(77) PRICE, H. H., *Perception* (London, 1932).

(78) PUTNAM, Hilary, "Mentes y Máquinas," en *Mentes y máquinas*, Madrid: Cuadernos Teorema.

(79) PUTNAM, Hilary, "Dreaming and 'Depth Grammer'," en Butler, R. J. (ed.), *Analytical Philosophy* (Oxford, 1962).

(80) PUTNAM, Hilary, "Robots: máquinas o vida creada artificialmente," en *Mentes y máquinas*, Madrid: Cuadernos Teorema.

(81) PUTNAM, Hilary, "Psychological Predicates," en *Art, Mind and Religion* (Pittsburgh, 1967), reimpreso como "The Nature of Mental States" en Putnam, Hilary, "Mind Language and Reality," *Philosophical Papers Vol. II* (Cambridge, 1975).

(82) PUTNAM, Hilary, "The Mental Life of Some Machines," en Castañeda, Hector-Neri (ed.), *Intentionality, Minds and Perception* (Detroit, 1967), reimpreso en la referencia 83.

(83) PUTNAM, Hilary, "Mind, Language and Reality," *Philosophical Papers*, Vol. II (Cambridge, 1975).

(84) PYLYSHYN, Zenon, "What the Mind's Eye Tells the Mind's Brain *Journal of Philosophy*, vol. 61 (1964), pp. 655-66, pp. 1-24.

(85) QUINE, W. V. O., *Palabra y objeto*, traducción de Manuel Sacristán, Barcelona: Labor, 1968.

(86) QUINE, W. V. O., "Naturalización de la epistemología" en *La relatividad ontológica y otros ensayos*, traducción de Manuel Garrido y Josep L. Blasco. Madrid: Tecnos, Serie de Filosofía y Ensayo, 1974.

(87) RORTY, Amelie O., "Belief and Self-Deception," *Inquiry*, vol. 15 (1972), pp. 387-410.

(88) RORTY, Richard, "Mind-Body Identity, Privacy and Categories," *Review of Metaphysics*, vol 19 (1965), pp. 24-54.

(89) RORTY, Richard, "Incorrigibility as the Mark of the Mental," *Journal of Philosophy*, vol. 67 (1970), pp. 399-424.

(90) RORTY, Richard, "Functionalism, Machines and Incorrigibility," *Journal of Philosophy*, vol. 69 (1972), pp. 203-20.

(91) RORTY, Richard, "Dennett on Awareness," *Philosophical Studies*, vol. 23 (1972), pp. 153-62.

(92) ROSENBERG, Jay, *Linguistic Representation* (Dordrecht, 1974).

(93) RYLE, Gilbert, *El concepto de lo mental*. Traducción de Eduardo Rabossi, Buenos Aires: Paidos, 1967.

(94) RYLE, Gilbert, "A Puzzling Element in the Notion of Thinking," *Proceedings of the British Academy*, vol. 44 (1958), pp. 129-44, reimpreso en Strawson, Peter F. (ed.), *Studies in the Philosophy of Thought and Action* (Oxford, 1968).

(95) SAVAGE, C. Wade (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 9 (en preparación).

(96) SCHAFER, Roy, *A New Language for Psychoanalysis* (New Haven, CT, 1975).

(97) SELLARS, Wilfrid, *Ciencia, percepción y realidad*, Madrid: Tecnos, Serie Estructura y Función.

(98) SELLARS, Wilfrid, "Notes on Intentionality," *The Journal of Philosophy*, vol. 61 (1964), pp. 655-66.

(99) SHAFFER, Jerome, "Could Mental States be Brain Processes?," *The Journal of Philosophy*, vol. 58 (1961), pp. 813-822.

(100) SHOEMAKER, Sidney, "Functionalism and Qualia," *Philosophical Studies*, vol. 27 (1975), pp. 291-315.

(101) SMART, J. J. C., "Sensations and Brain Proceses," *The Philosophical Review*, vol 68 (1959), pp. 141-56.

(102) SMART, J. J. C., "Brain Processes and Incorrigibility," *Australasian Journal of Philosophy*, vol. 40 (1962), pp. 68-70.

(103) SMART, J. J. C., *Philosophy and Scientific Realism* (London, 1963).

(104) SMART, J. J. C., "Critical Notice of Content and Consciousness by D. C. Dennett," *Mind*, vol. 79 (1970), pp. 616-623.

(105) DE SOUSA, Ronald, "Self-Deception," *Inquiry*, vol. 13 (1970), pp. 308-34.

(106) STEINBERG, Danny D. y JAKOBOVITZ, Leon A. (eds.), *Semantics* (Cambridge, 1971).

(107) STICH, Steven (ed.), *Innate Ideas* (Berkeley, CA, 1975).

(108) STRAWSON, Peter F., *Individuals* (London, 1959).

(109) TAYLOR, Brandon, "Mental Events: Are There Any?," *Australasian Journal of Philosophy*, vol. 51 (1973), pp. 189-200.

(110) TAYLOR, Charles, "Mind-Body Identity, A Side Issue?," *The Philosophical Review*, vol. 76 (1967), pp. 201-13.

(111) TORMEY, Alan, "Access, Incorrigibility and Identity," *The Journal of Philosophy*, vol. 70 (1973), pp. 115-128.

(112) TROYER, John G. y WHEELER, Samuel (eds.), "Intentionality, Language and Translation," *Synthese*, vol. 23 (1974), pp. 123-456.

(113) URMSON, J. O., "Motives and Causes," *Aristotelian Soc. Suppl.* vol. 26 (1952), pp. 179-94.

(114) WARNOCK, G. J., "Seeing," *Aristotelian Soc. Proc.*, vol. 55 (1954), pp. 201-18.

(115) WHITE, Alan R., "Different Kinds of Need Concepts," *Analysis*, vol. 20 (1956-60), pp. 112-16.

(116) WISDOM, John, "Other Minds," *Aristotelian Soc. Suppl.*, vol. 20 (1946), pp. 122-47.

(117) WITTGENSTEIN, Ludwig, *Philosophical Investigations* (Oxford, 1953).

(118) WOLLHEIM, Richard (ed.), *Freud: A Collection of Critical Essays* (Garden City, N.J., 1974).

La versión original de este artículo apareció en la revista *American Philosophical Quarterly*, vol. 14 (1978), núm.4,pp.249-61, cuyo Director, el Profesor Nicholas Rescher, ha cedido amablemente a *Teorema* los correspondientes derechos.