

Sobre *Objetos con mente*: reflexiones para un debate

Ángel Rivière
Universidad Autónoma de Madrid

La mente es nuestro objeto. Los objetos con mente en tanto que la tienen. En los últimos años, la psicología ha renovado su ancestral interés por esos sucesos arcanos, por esos asuntos evanescentes a los que llamamos «funciones mentales». Después de una época en que amplios sectores de la psicología habían entregado su entusiasmo y su paciencia al ideal conductista de depuración conceptual de cualquier clase de residuos mentalistas, el vocabulario de los psicólogos ha vuelto a saturarse de términos que emanan el inequívoco aroma de lo mental: imágenes y esquemas, planes y proposiciones, modelos mentales y redes conceptuales, lázaros resucitados de la mano de la psicología cognitiva. Nociones que, revestidas del prestigio de la ciencia cognitiva, han permitido un asedio a lo mental mucho más riguroso y efectivo que el logrado en muchos siglos de pensamiento sobre la mente. Como ha señalado acertadamente Manuel de Vega (1984, p. 23), en su ya clásica *Introducción a la psicología cognitiva*, «nunca se había dispuesto de tal cantidad de datos relevantes sobre los procesos mentales ni se había progresado tanto en su comprensión teórica» como en los últimos treinta años de desarrollo de la psicología cognitiva.

El cometido obvio y comúnmente aceptado de la psicología cognitiva es, sí, construir una ciencia objetiva de la mente, concebida como sistema de conocimiento. ¿Una ciencia *objetiva* y a la vez de la *mente*? Sin duda, es ésta una tarea difícil, desmesurada, llena de trampas y laberintos conceptuales de difícil solución. En *Objetos con mente* (1991) he querido reflexionar sobre los fundamentos de la psicología cognitiva. A sabiendas, me he aventurado por esos laberintos y, sin quererlo, he tropezado quizás en muchas de sus trampas (eso ya tendrán ocasión de ilustrarlo, más lúcida y distanciadamente que yo, los seis críticos que participan en este número y otros lectores).

Me propongo, en estas páginas, abrir el debate sobre *Objetos con mente*, presentando un resumen de las reflexiones que se contienen en el libro. Creo que hay en él un doble eje, que puede servir para articular esta síntesis: la mente concebida como artefacto natural para la interacción y la mente como sistema de

cómputo. Por una parte, está la mente como *un sistema con el que entendernos a nosotros mismos y entender a los demás* en nuestras relaciones cotidianas. Es decir, tenida (con independencia de todo compromiso ontológico acerca de lo que sea *además* de eso) como algo que *se dice y se piensa* de uno mismo y de los otros: como una estructura conceptual universal de la que nos servimos los humanos para sabernos mutuamente, para entendernos unos con otros. Esa mente de la que hablamos implícitamente cuando, en nuestra inevitable calidad de «psicólogos naturales», decimos que alguien (o nosotros mismos) hizo tal o cual cosa porque *creía* que «x», *deseaba* que «y», *recordaba* que «z» o *pensaba* que «n». Es por así decirlo, «la mente de la calle», la que usamos para andar por casa, para manejarnos en los nichos comunes de los hombres. La que nos permite atribuir razón, emoción y también responsabilidad moral a las acciones propias y ajenas. La que se refleja en el uso ordinario de verbos como creer, recordar, desear o pensar. Todos los lenguajes naturales contienen un amplísimo arsenal de verbos mentales como éstos, aunque incluyan distinciones semánticas diversas en la codificación léxica de lo mental (por ejemplo, en lengua hopi se emplean afijos verbales diferentes para marcar si alguien sabe algo porque lo ha visto, porque lo cree como verdad general, o porque supone que ha sucedido; una distinción que no marca la morfología del castellano).

Sí: cada vez que se explicita, a través de esos verbos, el sistema conceptual de lo que los filósofos llaman «actitudes proposicionales» —es decir, relaciones intencionales con proposiciones— se predica mente, de forma tácita, de los objetos a que se atribuyen tales actitudes. En *Objetos con mente*, se denomina «psicología natural» al sistema conceptual, o si se quiere a la *competencia*, que subyace a ese modo ineludible de ver la acción humana (y buena parte de la acción animal). A esa compulsiva tendencia de nuestro sistema cognitivo a explicar la conducta en términos de creencias y deseos.

Uno de los supuestos básicos del análisis planteado en el libro es el de que esa competencia natural de atribución mentalista, a la que hemos denominado «psicología natural», está ligada irremisiblemente a la fenomenología interna y huidiza de la conciencia humana. Desde la perspectiva del interaccionismo cognitivo que se apunta en *Objetos con mente*, la propia conciencia puede ser entendida, al menos en parte, como una especie de herramienta cuya evolución en nuestra especie debió estar muy ligada a las exigencias de las interacciones psicosociales. Como ha destacado Nicolás Humphrey (1987, p. 14), «fueron las circunstancias de la vida *social* del hombre primitivo —el pertenecer a una comunidad humana con interacciones complejas, su necesidad de ayudarse mientras al mismo tiempo ayuda a los demás— las que, más que nada, hicieron al hombre como especie, la criatura astuta y penetrante que hoy conocemos». Un animal «listo», que precisamente basaría su eficiente estrategia de «homo psychologicus» —por usar el mismo término que Humphrey— en *el empleo de un cuadro privilegiado del propio yo (i.e. en el empleo de la conciencia reflexiva) como modelo de lo que es ser otra persona*.

Si ello fuera así, y en tanto que la psicología natural es heredera (y quizá al mismo tiempo condición) de la evolución de la conciencia, se vería constreñida a establecer la gama limitada de distinciones sobre lo mental que la concien-

cia efectúa, y sería tan miope como ha demostrado históricamente ser la propia conciencia (la introspección), con relación a los mecanismos de producción de lo mental. Después de más de cien años de psicología científica, hay muchas razones para pensar que la conciencia no se desarrolló evolutivamente con la función narcisista de hacerse transparente la maquinaria misma de la mente, sino al cuidado de otras funciones adaptativas más obvias como las de objetivar realidades, seleccionar conocimientos relevantes al contexto, guiar interacciones intraspecíficas, definir relaciones con los objetos y conocimientos (las llamadas «actitudes proposicionales»), planificar la acción y controlar parcialmente la actividad cognitiva. Toda la historia de la psicología introspectiva fue una elocuente demostración, al menos en el primer asalto del combate de la psicología con la mente, de lo peligrosa que puede resultar esa engañosa transparencia con que lo mental simula ofrecerse a sí mismo a través de la conciencia. Sí: la maquinaria de la mente es mucho más oculta de lo que supone la conciencia.

Cuando hablamos de esa maquinaria oculta de la mente, nos situamos, desde una perspectiva cognitiva, en el segundo eje de la reflexión que estamos resumiendo: ahora hablamos de *la mente como sistema de cómputo*. Sin la metáfora o la consideración literal de la mente como una maquinaria que computa representaciones, no es posible entender el desarrollo de la psicología cognitiva. Uno de los intentos de *Objetos con mente* es el de relacionar y cotraponer esos dos sentidos de lo mental: por un lado, la mente como experiencia que se produce esencialmente en primera persona de singular y que se atribuye, en segunda o tercera, en virtud de competencias naturales y universales que poseemos como sistemas cognitivos; por otro, la mente entendida, a la forma propia, estrictamente «cognitiva», como maquinaria de cómputo. Esta perspectiva ha resucitado lo mental, al justificarlo no ya desde la siempre dudosa y escurridiza fenomenología de la conciencia, sino con el marchamo de cientificidad otorgado por una doble operación: la de usar observaciones estrictamente objetivas («extensionales») y externas, en tercera persona de singular, y no en primera) como base empírica de la psicología, y la de construir modelos teóricos con vocación (no siempre lograda) algorítmica y mecanicista. Modelos esencialmente formales, en que ciertas estructuras internas e intencionales (que es como decir «mentales») se formarían y transformarían gracias a los helados trasiegos de los procesos de cómputo.

¿Hasta qué punto ha permitido esta nueva consideración de la mente resolver las viejas aprensiones de la psicología sobre el estatuto científico de lo mental? La respuesta que da a esta pregunta *Objetos* es indecisa, compleja, llena de matices. Creo que no podía ser de otra manera. Es verdad que hemos acumulado más conocimientos específicos que nunca sobre el funcionamiento de la maquinaria de la mente, pero también lo es que el paisaje teórico en cuyo marco se han producido esos nuevos conocimientos puede resultar inaceptablemente limitado, lleno de peligros conceptuales y poco inteligible cuando se considera desde la otra perspectiva a la que hemos apuntado: desde la atalaya privada de la subjetividad y la conciencia, desde esa estructura conceptual, y por así decirlo «de sentido común», que se deriva de la psicología natural, no resultan fáciles de aceptar algunas de las consecuencias menos intuitivas de tomarse en serio la idea de la mente como sistema de cómputo.

Eso no debe extrañarnos. Sabemos que conocer la mente es difícil, y esa es la primera y trivial constatación de la que parte la reflexión hecha en *Objetos*. Es difícil por muchas razones. La primera de ellas es que la mente fenoménica, que se ofrece a una supuesta percepción interna —a la *innere Wahrnehmung* que decía Wundt— de un «ojo interior» (Humphrey, 1993)... esa mente de la que hablaba William James (1890) cuando definía la psicología como la «ciencia de la vida mental», no se identifica con la mente subyacente, que computa representaciones simbólicas, de la que hablamos los psicólogos cognitivos. La que podríamos llamar «mente uno», que nos ofrece su desnuda cohorte de imágenes evanescentes y palabras interiores, como sombras proyectadas en la sutil y huidiza pantalla transparente de la conciencia, no es la misma que la «mente dos»; es decir, que aquélla que se define como una maquinaria actuante en virtud de representaciones mucho más impenetrables y secretas, que el psicólogo cognitivo debería descifrar. Y, siendo así, se le plantea al psicólogo una primera cuestión difícil: ¿cuál es la relación entre la «mente 1» y la «mente 2»?

Ray Jackendoff (1987) ha destacado, de forma brillante, que, del mismo modo que existía un problema tradicional de relaciones entre la mente fenoménica y el cuerpo, también lo hay para relacionar la mente fenoménica con la computacional. Como en aquel caso, caben soluciones alternativas para este nuevo problema de «relación entre mentes»; soluciones interaccionistas (mente 1 y 2 se influyen e interactúan), paralelistas (las mentes se corresponden sin interactuar causalmente), epifenomenistas (sólo tienen virtualidad causal los procesos de cómputo, y la conciencia es su residuo epifenoménico, sin un papel causal) o de identidad (que identificaría a los cómputos con los sucesos conscientes, siendo quizá éstos una expresión molar de aquéllos). ¿Es la conciencia un mero epifenómeno de las computaciones?, ¿son realidades paralelas que no se influyen?, ¿es posible que interactúen? Y, si lo es, ¿cómo? ¿No serán dos formas de hablar de lo mismo?

Lo cierto es que la psicología cognitiva ha conseguido, de entrada, convertir un problema en dos (quizá sea esa una demostración de que es buena ciencia): antes sólo teníamos el viejísimo, y hasta entrañable, asunto de las relaciones entre mente y cuerpo. Ahora tenemos además un problema de relaciones entre mente 1 y mente 2. La situación actual presenta además la paradoja de que se vislumbran, a través del conexionismo, vías de solución al problema de cómo entender la relación entre el sistema nervioso y la mente computacional, pero las relaciones entre ésta y la mente fenoménica siguen siendo oscuras y mal comprendidas (a pesar de Baars, 1988, Johnson-Laird, 1988, y Jackendoff, 1987, y como ha visto lúcidamente Searle, 1992). En definitiva, y como señalaba un poco trágicamente en *Objetos*, «no sabemos bien qué papel asignar a la conciencia en ese guión que dirige el frío drama de las computaciones».

La tragedia es fácil de comprender: el concepto de computación sólo tiene sentido en el marco de la noción de autómatas, y en ese paisaje conceptual, la conciencia no tiene sitio fácil. Eso ya lo veía claro William James (1890), cuando se enfrentaba a las teorías mecanicistas de su época: «¿Cuál podría ser —decía— según este punto de vista la función de la propia conciencia? Si se trata de función *mecánica*, ninguna [...] sería una cadena de sucesos completamente autónoma, y cualquiera que fuera la mente que la acompañara no pasaría de ser un «epi-

fenómeno», un espectador inerte, una especie de «aura, espuma o melodía» (p. 133). La cuestión de fondo con respecto al papel de la conciencia, es ésta: ¿cabe alguna interpretación que no sea la epifenomenista desde una versión fuerte de la mente como sistema de cómputo? La intuición del sentido común sobre la eficacia causal de la conciencia no es fácil de compatibilizar con la noción de autómatas, en tanto que ésta exige necesariamente ser *completa*. Si la mente es un autómata completo (por muy abstracto que sea), guiado por procesos deterministas, no se comprende bien que la intuición de eficacia causal de la conciencia, *en tanto que tal*, no sea más que una ilusión sin fundamento. Pero si los procesos mentales no son derivables de autómatas completos, la noción de cómputo carece literalmente de sentido cuando se aplica a ellos.

Cuando ciertos conceptos científicos, fértiles para la investigación y la comprensión teórica de los fenómenos, no se someten al dictado del sentido común, los científicos suelen tomar una opción clara: ¡peor para el sentido común! (la física cuántica, por ejemplo, es completamente extraña a nuestra «intuición mesocósmica» del mundo, y ahí está). ¿Será también una intuición sesgada (quizás una muestra más de nuestra pobre visión mesocósmica del hombre, de nuestra conmovedora falta de «realismo depresivo», o de nuestras propensiones a sufrir «ilusiones de control») la de la eficacia causal de la conciencia?, ¿cómo hacerla compatible con los nuevos modelos computacionales de la mente?

Estos nuevos modelos presentan además otras dificultades, que se deducen fácilmente de lo que acabamos de decir. Dado que la mente que efectúa los cálculos no se identifica con la que presenta a la mirada o la escucha de la conciencia imágenes mentales o palabras que nos decimos a nosotros mismos, eso significa que no hay ningún testimonio fenoménico *directo* de la mente computacional. Ésta tiene que ser *reconstruida conceptualmente* a partir de la conducta de la gente en situaciones generalmente muy controladas (en experimentos, en que se minimiza el papel de «control ejecutivo y decisión de alto nivel», que acaso pudiera tener la conciencia). Cohen (1977) se ha servido de una parábola muy expresiva para establecer cómo trabaja el psicólogo cognitivo: es semejante a un lector que quisiera descubrir la organización de una biblioteca sin poder entrar en ella. Tendría que emplear una estrategia consistente en sentarse a la puerta de entrada y preguntar a los visitantes por los libros que buscan, anotando cuidadosamente la duración de la búsqueda y el número y tipo de libros buscados. Es dudoso que la información obtenida pudiera darle una idea exacta del sistema y la disposición de la biblioteca, aunque sí podría llegar a desarrollar algunas conjeturas con fundamento.

De modo semejante, el psicólogo cognitivo se ve obligado a sentarse a la puerta del reducto invisible de la mente, y desde allí tratar de inferir trabajosamente su organización y funcionamiento. De la maquinaria causalmente eficaz, de representaciones y cálculos, de la que hablamos los psicólogos cognitivos, no hay por lo general ninguna *innere Wahrnehmung*, ninguna percepción interna, ninguna experiencia predicable en primera persona de singular (por eso dice Dennett que las nociones cognitivas son «subpersonales»). No es extraño que, en esas condiciones, el psicólogo cognitivo se enfrente con un problema de *indeterminación*. ¿Cómo es posible elegir entre modelos alternativos de cómputo so-

bre representaciones en virtud de los datos de conducta? John Anderson (1979) ha defendido rigurosamente la idea de que los modelos cognitivos serían constitutivamente indeterminados (véase debate por Pylyshyn, 1979 y Hayes-Roth, 1979, revisado en Rivière, 1986), expresando así una inquietud frecuente, de los psicólogos cognitivos, acerca de sus inferencias sobre las estructuras de conocimiento y los procesos que se ponen en juego en las actividades que estudian.

Éstas son, por tanto, algunas de las dificultades más obvias para alcanzar un conocimiento objetivo de la mente, desde una perspectiva cognitiva. El difícil problema conceptual de la «relación entre mentes» (y no sólo entre las que hemos llamado «uno» y «dos», sino también entre las que se predicán de individuos diferentes), las dudas sobre el papel causal de la conciencia, las aprensiones sobre el carácter inferencial y quizá especialmente indeterminado de los modelos cognitivos, son sólo una muestra de los problemas, dudas y aprensiones que asaltan, en sus momentos de pesimismo epistemológico, a los psicólogos cognitivos más lúcidos. Algunos, como Johnson-Laird (1983), llegan a plantearse la pregunta de si es posible la psicología. La inquietud de si no será la psicología un intento demasiado pretencioso del hombre por capturar lo inabarcable, como lo era la metafísica para Hume: «una conjetura plausible es que la mente tiene que ser más complicada que cualquier teoría que se proponga para explicarla —dice—: cuanto más compleja sea la teoría, más tendrá que serlo la mente que la pensó por primera vez» (p. 1). ¿Tiene sentido ese intento de asir con las manos la sombra de la mente, de «abarcarse la propia cabeza» o de «subirse a uno mismo sobre los propios hombros»? ¿Acaso no es posible formular en psicología una especie de teorema de incompletitud —a lo Gödel— que establezca que ninguna mente podrá explicarse de forma completa, a sí misma, sirviéndose de su propia complejidad y de su propio lenguaje, es decir, del lenguaje de lo mental?

Los resultados obtenidos en treinta años de investigación cognitiva aconsejan rendirse de entrada a estas melancolías epistemológicas. Es posible que para explicar la mente haya que renunciar finalmente al lenguaje de lo mental. Acaso sólo sea éste un modo metafórico de hablar de propiedades globales y funcionales, que emergen en un nivel molar, del complejo funcionamiento en paralelo del sistema nervioso entendido como sistema de cómputo (a la manera conexionista). Es pronto para decirlo. A mi entender, también es pronto para abandonar del todo el modo tradicional de hablar de lo mental en el enfoque «clásico» de la psicología cognitiva: es decir, en términos de procesos de cómputo realizados sobre objetos intencionales, sobre estructuras de conocimiento. Esa forma de ver ha dado demasiados dividendos, en el conocimiento de lo mental, como para dejarla de lado a las primeras dudas.

Sin embargo, también parece obvio que la psicología actual ha entrado en una fase que presenta todas las características establecidas por Kuhn (1975) para definir las crisis paradigmáticas: hay una conciencia generalizada de los agujeros del paradigma aún dominante en psicología cognitiva, una forma alternativa y emergente de explicar lo mental, y un debate importante sobre las virtualidades respectivas de los enfoques clásico y conexionista (véase, por ejemplo, Fodor y Pylyshyn, 1988; Pinker y Prince, 1988; Smolensy, 1987, 1988; Graubard, 1933). En estas condiciones se impone una reflexión sobre los fundamentos mismos de

la psicología cognitiva, y en ese marco hay que entender la pretensión, quizá excesiva para las fuerzas del autor, de *Objetos con mente*.

Supongo que el primer paso dado en esa reflexión, que consistía en ilustrar la trivialidad de que conocer la mente es difícil, será aceptable para casi todo el mundo. El segundo es, si cabe, aún más obvio. Consiste en detenerse a pensar en la también trivial observación de que *las personas dividimos naturalmente el mundo en dos amplias clases de objetos; de unos pensamos y decimos que tienen mente; a otros no se la atribuimos*. Como usuarios de lo que hemos llamado «psicología natural», predicamos intuitivamente lo mental cada vez que decimos que algo (generalmente «alguien») piensa, cree, percibe, recuerda o siente. Siempre que usamos esos verbos para referirnos a algo (a alguien), estamos atribuyendo mente. No es inexacto decir que lo que hacemos los psicólogos, en esencia, es trabajar con esos verbos: con sus significados y referentes. Tratamos de conocer científicamente en qué consisten esas funciones de ciertos peculiares objetos, tales como desear, pensar, recordar o sentir. Nos interesan los objetos de los que se dicen funciones mentales (Fm), como las personas y los animales, y no aquellos otros de los que no se dicen (-Fm), como los tenedores y los planetas. Nuestro campo de estudio puede definirse, así, de forma muy escueta: Fm; es decir, la investigación de la naturaleza, estructura, génesis y funcionamiento de las funciones mentales.

Tenemos entonces dos categorías de objetos, a las que llamaremos «Fm» (la categoría de los objetos a los que atribuimos mente) y «-Fm» (la formada por objetos de los que no la predicamos). El asunto se complica un poco cuando caemos en la cuenta de que esas categorías no son tan nítidas, en sus bordes, como parecen a primera vista. No son como las clases lógicas bien formadas, que se definen por poseer elementos igualmente representativos —es decir, con el mismo grado de pertenencia—, límites claros e inequívocos, y rasgos que brindan una definición comprensiva neta. La categoría de los objetos con mente tiene, por el contrario, fronteras difusas y una estructura distribucional semejante a la de las llamadas «categorías naturales» (Rosch, 1978). Del mismo modo que los gorriones son más representativos —más prototípicos— de la categoría natural de las aves que los pingüinos (Malt y Smith, 1984), las personas son más representativas de la categoría Fm que las moscas o las bacterias. Una consideración intuitiva de esa categoría lleva a una observación muy significativa: existe en realidad un importante hiato, un corte brusco, entre el grado de tipicidad como objetos con mente que se atribuye a las propias personas, desde la perspectiva de la psicología natural, y la que se otorga a cualquier otro objeto del mundo. ¡Claro... eso es lo que cabía esperar desde la consideración de la mente como instrumento natural al servicio de la interacción con las personas! Tendemos a llamar «sujetos» a los elementos más representativos de la categoría Fm, y en grado eminente a nuestros congéneres. Más allá de los seres considerados propiamente «sujetos», las dudas de todo tipo nos asaltan: ¿tienen mente las moscas?, ¿perciben, piensan, aprenden? Quizá la mejor expresión de estas dudas sea el astuto comentario de una niña pequeña, sobre mentes y perros: «los perros piensan —decía la niña—; lo que no tienen son ideas».

Las fronteras de la categoría Fm no sólo son difusas, sino que han demos-

trado ser movedizas, tanto en el desarrollo de las personas como en la evolución histórica y cultural del hombre. El fenómeno del animismo es una buena demostración de ello. El concepto de «animismo» hace referencia a la tendencia que tienen los niños y los miembros de otras culturas a atribuir funciones mentales (y vida) a objetos que no se incluyen en la categoría Fm de los adultos occidentales. El análisis del animismo ofrece un doble interés: por una parte, tomado como fenómeno cognitivo, demuestra hasta qué punto puede sesgarse el conocimiento de lo real por parte de un sistema cognitivo especialmente diseñado y dotado para la relación intraespecífica, para comprender y atribuir estados mentales (Aston, Harris y Olson, 1988; Whiten, 1991). Por otra, el análisis del animismo como fenómeno cultural ilustra cómo se reduce la extensión de la categoría Fm en aquellos individuos que interiorizan, al menos parcialmente, las consecuencias de la enorme revolución conceptual que tuvo lugar a partir del nacimiento de la ciencia moderna en el renacimiento. Las adherencias animistas del pensamiento cosmológico, presentes en muchos filósofos griegos y en la Edad Media o el renacimiento, fueron barridas por Newton, en su demostración de que las leyes de la cosmología no añadían nada cualitativamente distinto a las de la mecánica, en las que no cabrían atribuciones de funciones mentales.

Sí. En la ciencia renacentista y barroca, la eliminación de los insidiosos residuos animistas para la comprensión del mundo físico —que no se logró del todo hasta Galileo y Newton— permitió establecer una escisión tajante entre los enunciados de la ciencia de la naturaleza, extensionales y objetivos, y los de carácter intencional, que permitirían comprender la conducta humana en términos de creencias y deseos. Esa escisión se cimentó en una visión *mecánica* de la naturaleza. Desde ella, se planteaba un desafío fascinante: ir ganando, para una consideración extensional y mecanicista, dominios cada vez más amplios de conocimiento. Estas coordenadas históricas permiten comprender los dos grandes programas de investigación que dieron origen a la psicología moderna: el de los empiristas, en el mundo privado de unas Fm identificadas completamente con la conciencia y entendidas en términos puramente fenoménicos, trataba de definir pautas de orden y «asociación» semejantes a las del mundo físico (las «leyes de gravitación universal de la mente» de Hume); el de los racionalistas, bajo la sombra del infranqueable muro infinito levantado por Descartes entre lo extenso y lo intencional, sólo tenía dos caminos: o definir de un nuevo modo todo el mundo material, hasta poder incluir la mente como propiedad entera en su fundamento (Leibniz), o incluir a los propios animales (como hizo el propio Descartes) en la categoría -Fm (los objetos sin mente), teniéndolos por meras máquinas sin conocimiento, como máquinas muy perfectas «como hechas por las manos de Dios» (en piadosas palabras de Descartes)... pero tan despojadas de mente como los tenedores y los planetas.

Más acá de estas cuestiones metafísicas, el testamento dejado por mil quinientos años de lucha por comprender la naturaleza, y en especial el legado de la gran revolución paradigmática protagonizada por Galileo y Newton, dejaban al hombre moderno en una situación relativamente cómoda en relación al viejo problema de delimitar los objetos de los que podrían predicarse funciones mentales: serían, en último término, objetos *vivos*, organismos. Fueran lo que fueran

las funciones mentales, serían funciones de un organismo. Es verdad que no estaba nada clara la naturaleza de tales funciones y que, como ha dicho Johnson-Laird, la mente era algo de lo «que se consentía que los adultos pudieran hablar en privado, siempre y cuando comprendieran que, en realidad, no existía» (1988, p. 3). Pero, al menos, estaba clara una cierta distinción entre *cosas* y *no-cosas*. Una de las consecuencias más trascendentes de la influencia de las ciencias de la naturaleza sobre el pensamiento cotidiano fue la de establecer una sima difícilmente salvable entre objetos tales como las montañas, las sillas, los cuchillos de cocina, las nubes y los automóviles, es decir, objetos-cosas de los que no cabría predicar lo mental en absoluto, y seres como las personas, los perros y las moscas, es decir, seres «no-cosas», «objetos más o menos sujetos» a los que podrían atribuirse algunas funciones mentales o todas ellas (en el caso prototípico de las personas). Ya en nuestro siglo, las explicaciones más coherentes y profundas de la función más eminentemente mental, el pensamiento, entendían esa función como prolongación de funciones biológicas de adaptación y organización que poseerían los organismos en tanto que tales (Piaget, 1961).

En esas condiciones históricas, la irrupción de la tecnología del conocimiento implicó una ruptura del *statu quo* alcanzado en relación con las categorías Fm y -Fm. Al mismo tiempo, proporcionó un nuevo impulso al intento de conocer la mente en términos precisos, mecanicistas, estrictamente objetivos y suficientemente complejos como para explicar su delicada maquinaria. Como ha señalado Varela (1988), la tecnología del conocimiento proporciona al hombre una nueva imagen de sí mismo, de su mente: la imagen de una maquinaria que computa representaciones (simbólicas o subsimbólicas), y de forma automática, produce efectos que se parecen mucho, en aspectos relevantes, al pensamiento humano. Se dice en *Objetos* que «el hombre tiene, en los últimos años, la sensación apasionante de que está produciendo funciones mentales (Fm) con sus técnicas. O, como mínimo, de que está produciendo «cogito», pensamiento, que ha sido la marca de fábrica de lo mental por lo menos desde el origen de la epistemología moderna con Descartes» (p. 48).

Como es sabido, la provocativa idea de una máquina que produce pensamiento fue defendida por Alan Turing (1950), en un artículo seminal de toda la ciencia cognitiva: «Computing Machinery and Intelligence». En ese artículo, Turing jugaba a difuminar de nuevo la frontera entre mentes y máquinas, tan trabajosamente lograda por la ciencia de la naturaleza, formulando en esencia la idea de que cualquier pensamiento que pueda ser reducido a un procedimiento efectivo es imitable por una máquina abstracta, universal y muy simple (que no se define por su materialidad, sino por su forma), capaz de resolver toda secuencia de cómputos; una máquina que el propio Turing (1936) había definido en un artículo anterior más técnico, relacionado con el debate sobre la tesis de Gödel (1931) de la incompletitud de cualquier sistema formal. En el artículo de 1950, Turing proponía una especie de juego, consistente en hacer preguntas a dos seres ocultos (una persona y una concreción de su máquina, como por ejemplo un ordenador), y se preguntaba con qué derecho y seriedad sería posible decir que la persona pensaba y la máquina no en caso de que fuera imposible distinguir quién o qué contestaba a las preguntas.

Por debajo de este juego, había tres consideraciones esenciales, que ayudan a comprender la trascendencia que tenía, y que debemos mencionar:

1. La primera es que la máquina de Turing es, se supone, capaz de realizar o imitar la función «mental» prototípica de pensar, y no de forma contingente, sino *por principio*, con tal de que tal función pueda ser reducida a un procedimiento efectivo. Por otra parte, y en el caso de que el pensamiento no pudiera entenderse como un procedimiento efectivo (como sugiere, por ejemplo, Penrose, 1991), se plantearía una seria duda sobre su *inteligibilidad* como proceso psicológico.

2. Esa afirmación de principio establecía un reto profundo para la psicología: ¿cómo sería posible comprender lo mental si no consistiera en procedimientos efectivos?, y, si consistiera en ellos, ¿qué tendría de misterioso?; ¿por qué no estudiarlo, desentrañando los procedimientos efectivos de que se sirve lo mental?

3. Por otra parte, Turing independizaba la actividad cognitiva tanto de sus fundamentos materiales (o biológicos) como de sus concomitantes fenomenológicos, definiendo un nuevo nivel de análisis «independiente del físico pero de estilo mecanicista» (Pylyshyn, 1988, p. 68), en el que iba a situarse la nueva ciencia de la mente, articulada en torno a la noción de cómputo.

¿Cómo era posible que la máquina de Turing, un autómata literalmente amental y definido por cadenas de unos y ceros tuviera unas consecuencias tan profundas para la psicología?, ¿cómo podía borrar un objeto abstracto y simple como ese los contornos de las categorías sobre lo mental, cristalizados desde la genial reducción mecanicista del mundo extensional por Newton? El secreto estaba en la conjunción de la simplicidad, el automatismo y, al mismo tiempo, la ilimitada generalidad del poder de cómputo del autómata. En su propiedad de ser un autómata *universal*, capaz de aplicar cualquier clase de algoritmo. Ya Descartes había señalado, en *El discurso del Método* (1637), que la universalidad constituye el signo de la razón, y se servía de ese argumento para negársela a los animales. Para él, por otra parte, automatismo y razón eran cosas contrapuestas. Los animales, como autómatas muy perfectos, no serían compuestos de dos sustancias separadas por límites infranqueables, sino que serían seres «de una sola pieza» puramente extensa, reductible, en todos sus extremos, a las leyes inexorables de la mecánica del mundo material. Toda la operación dualista de Descartes pivotaba sobre la idea de que el pensamiento es la propiedad específica de lo mental; sobre la idea de una razón universal, concebida como función prototípica de la *res cogitans*. Por su parte, Leibniz, a pesar de ampliar el concepto de mente (esta vez articulada en torno al rasgo conceptual F_m de «percepción») a toda la naturaleza —con lo que salvaba a los animales de ser meros autómatas amentales— seguía entendiendo que lo mental es irreductible a cualquier clase de explicación mecánica.

Lo curioso es que fueran precisamente Descartes y Leibniz, los pensadores racionalistas, los que abrieron un largo camino de pensamiento que iría a parar finalmente en la negación de la tesis de una diferencia insalvable entre mentes y máquinas. Un sendero que fue a dar en objetos tan futuristas como las máquinas de Turing y de Post, el ordenador de von Neumann, y la concepción «mecanicista abstracta» de lo mental, que caracteriza al paradigma clásico de la ciencia

cognitiva. ¿En qué consistió ese camino? Descartes y Leibniz dieron los primeros pasos hacia el ideal de definir un lenguaje lógico *universal*, que pudiera asegurar el rigor deductivo de cualquier razonamiento. Imaginaron así la posibilidad de un autómata abstracto, de un algoritmo general; concibieron implícitamente la hipótesis de un sistema de cómputo, capaz de trasegar, mediante pasos discretos, con cualquier clase de estructuras simbólicas bien definidas. En su sueño racionalista, intuyeron la posibilidad de una máquina abstracta, puramente formal y universal, que realizaría, con inflexible limpieza, las tareas más intrincadas de la deducción humana. Hubieron de pasar trescientos años de pensamiento lógico y matemático para que Turing (1936) hiciera realidad el sueño de Descartes. Una condición para que se cerrara el círculo de las tres ideas que aún permanecían separadas en el discurso racionalista, las de *mente, máquina y lenguaje lógico*, fue despojar a los conceptos de «máquina» y «autómata» de toda connotación material.

Toda la ciencia cognitiva, y la psicología cognitiva en particular, ha girado en torno a ese nuevo plano, el de la noción de cómputo, en que se define un nuevo nivel de análisis que es, al tiempo, mecanicista y no-físico (desde el punto de vista de su descripción completa, aunque no de su ontología). En la tesis del «dualismo funcionalista», que ha predominado en el periodo clásico de nuestra ciencia —y que ha permitido su asentamiento como ciencia independiente—, lo mental podría definirse con independencia del sustrato físico en que se realiza, del mismo modo que es posible hacer, aplicar, conocer, un programa de ordenador sin tener en cuenta los componentes físicos o la estructura de su «sustancia extensa» (el *hardware*). Es cierto que, en un primer momento de desarrollo de la psicología cognitiva, se planteó la posibilidad de desarrollar modelos de redes neurales, que incorporarían principios lógicos en su funcionamiento; modelos de orientación no-dualista, como los de McCulloch y Pitts (1949), Donald Hebb (1949), y Franz Rosenblatt (1962), pero la demostración de ciertas limitaciones lógicas de aquellos autómatas reticulares dejó el camino expedito al predominio de las tesis del funcionalismo dualista, que concebían al sistema cognitivo como una maquinaria físico-simbólica de cómputo, y que defendieron de forma enérgica, lúcida, Newell y Simon (véase, por ejemplo, 1976) y todos los investigadores que aceptaron literalmente el envite de Turing. En *Objetos*, se denomina «Paradigma C-R» (es decir, de computaciones sobre representaciones) a esta versión «fuerte» de la ciencia cognitiva.

¿Qué consecuencias tenía aceptar seriamente el reto de Turing, considerando que, en un sentido literal, la mente es un sistema de cómputo de representaciones simbólicas? En el libro se realiza una reflexión sobre esas consecuencias. La primera ya la hemos comentado; al *des-animar* la consideración de lo mental (es decir, al desligar su análisis del de lo vivo y de cualesquiera propiedades extensionales de un organismo), permitía establecer un dualismo no sustancialista —una especie de «dualismo estratégico»—, basado en la premisa de que existe un plano autónomo y completamente preciso de descripción de funciones mentales, abstraído y olvidado de la descripción del organismo que las realiza como sistema biológico (o de la consideración física de la maquinaria, como el ordenador, que también puede realizar F_m). Esta forma «light» de dualismo parecía

poder avenirse bien con una ontología materialista, tal como la defendida por Newell y Simon (1976) al definir a mentes y ordenadores como sistemas físico-simbólicos (no siendo, al final, los símbolos sino patrones físicos guiados por leyes causales, en una de sus caras, al tiempo que «objetos intencionales», capaces de designación, en la otra). En realidad, los sistemas de cómputo tendrían una especie de «intencionalidad extrínseca» y no intrínseca, ya que no se guiarían en su funcionamiento por el supuesto carácter intencional de los símbolos, ni por su contenido semántico, sino por su forma, por su sintaxis, pues sólo así cabría defender, con sentido, su carácter de máquinas algorítmicas. Pero ello tendría, a su vez, consecuencias importantes para la consideración de lo mental, debido al inflexible determinismo inherente a todo sistema de cómputo de símbolos discretos (al menos, en el concepto clásico de este término), y a la dificultad, a la que ya nos hemos referido, para comprender el papel que pudiera haber a la conciencia en ese tipo de sistemas.

En la psicología cognitiva y la filosofía de la mente, se han dado cuatro respuestas principales al desafío de Turing, consistente en desdibujar —en un plano computacional de descripción— la frontera entre mentes y sistemas artificiales de cómputo. A la pregunta de Turing, «¿pueden pensar las máquinas?» se han dado, en esencia, estas contestaciones:

1. *No*, de ninguna manera. Pensar es una propiedad (privada) de un organismo en tanto que tal. Dejemos las cosas como estaban: ni las máquinas piensan, ni las mentes son máquinas.

2. *Sí*. Hay un sentido literal y serio en que la categoría *Fm* incluye a mentes y máquinas. Las máquinas pueden pensar, si es que «pensar» es algo preciso. *Las mentes son precisamente máquinas simbólicas*.

3. *Metafóricamente sí*, pero sólo metafóricamente. Sin aceptar todas las consecuencias de aceptar que la mente es un sistema de cómputo de representaciones simbólicas, es útil tomarla por tal.

4. *En cierto modo sí*, pero no son precisamente las máquinas como la de Turing las que mejor imitan a la mente, sino otras muy diferentes: las que son semejantes al sistema nervioso, en el plano en que éste se considera como sistema de difusión en paralelo de activaciones distribuidas y no necesariamente conceptualizables como representaciones simbólicas. Pensar, percibir, recordar, etc., pueden ser propiedades holistas y molares de sistemas muy complejos de unidades asociadas, que no se transmiten símbolos sino estados de activación.

(1) La primera posición es la que han defendido algunos filósofos de la mente, y en particular John Searle (1980). La segunda caracteriza al núcleo «duro» del paradigma clásico C-R. La tercera a la mayor parte de la psicología del procesamiento de la información, y la última al conexionismo. En *Objetos*, se analiza detalladamente la crítica de Searle a la psicología cognitiva clásica, en un punto que se titula «La mente de la habitación china y el martirio de la psicología cognitiva». En ese apartado, se diferencian las posiciones críticas de Searle (1980, 1983) con relación a los supuestos clásicos del paradigma C-R, de sus propuestas positivas. Las primeras se expresan poderosamente en la metáfora del propio Searle, que se representa a sí mismo dando «respuestas algorítmicas» ciegas a preguntas en chino (lengua que Searle no conoce), en unos casos, y respuestas inteligentes

e intencionales al inglés, en otros. Mediante esa parábola, Searle critica la confusión entre forma y contenido en los modelos clásicos de la ciencia cognitiva, el carácter falso y residualmente conductista del test de simulación de Turing, y la abstracta indiferencia a lo material que se refleja en el dualismo funcionalista: «en lo que a la IA concierne —dice—, el mismo programa podría ser realizado por una máquina electrónica, una sustancia mental cartesiana, o un espíritu universal hegeliano» (1980, p. 86).

¿Cuál es la parte positiva de la propuesta de Searle? Su núcleo principal consiste en una idea difícil de entender: ese rasgo de intencionalidad, que define a lo mental, no puede darse en seres no-vivos, porque es «un fenómeno biológico, y es tan probable que dependa causalmente de la bioquímica como la lactación, la fotosíntesis y otros fenómenos biológicos» (*ibid.*, pp. 86-87). Para Searle, «no hay ningún vacío que llenar» (p. 50) entre las intuiciones intencionalistas de la psicología natural y las nociones biológicas, extensionales, con que se define, en un nivel bioquímico por ejemplo, el funcionamiento del sistema nervioso. Pero ese aventurado salto en un vacío supuestamente inexistente podría acabar en batacazo, si al vacío tal le diera la mala intención de existir: el salto es semejante al que daría un físico que supusiera que «no hay ningún vacío que llenar» entre la física cuántica, o de partículas elementales, y las intuiciones mesocósmicas y fuertemente sesgadas de la física intuitiva. Searle parece justificar, en un plano psicológico, dos suposiciones que, sin duda, no admitiría en el físico: 1) las teorías cotidianas implícitas son modelos molares de explicaciones científicas más moleculares (es obvio que no lo son), y 2) es posible prescindir de planos de explicación en los que aparecen propiedades, con leyes propias, emergentes del funcionamiento global de los sistemas (en tanto que sistemas atómicos, mecánicos, termodinámicos, etc.). ¿Por qué no está dispuesto Searle a admitir que la intencionalidad es una propiedad derivable, en primer término, de las propiedades cognitivas de los sistemas, y sólo muy indirectamente de sus propiedades físicas o bioquímicas? ¿Predicar intención de las sinapsis es tan anómalo como predicar entropía de los cuantos!

(2) En el extremo opuesto a Searle, se encuentran los tenaces defensores del paradigma C-R, en el más puro estilo MIT. Entre ellos están algunos de los teóricos más consecuentes y lúcidos de las ciencias cognitivas: Noam Chomsky (1980), Jerry Fodor (1976), el Hillary Putnam (1960) neoclásico (antes de sus arrepentimientos epistemológicos recientes), David Marr (1982), Zenon Pylyshyn (1984, 1989), Ray Jackendoff (1987), Fred Dreske (1988) y naturalmente Herbert A. Simon y Allen Newell (1976). Si dejamos de lado posiciones recientes de Putnam y Chomsky, todos estos investigadores mantienen la idea de que puede definirse un plano cognitivo autónomo, en el que la ruptura de la barrera «mentes (Fm)/máquinas (-Fm)» admite una interpretación *literal* y no sólo metafórica. Así, las mentes y las máquinas simbólicas se incluirían en una misma categoría, llámese de «cosas cognoscentes» o «sistemas de cómputo», de la que se derivaría ese plano formal, abstracto y funcional de descripción netamente cognitivo (para ayudarnos a pasar el frío sabor a granizado de limón del trago consistente en admitir que somos primos lejanos de los ordenadores, Pylyshyn, 1988, nos recuerda otros tragos quizá más ácidos: ya tuvimos que asimilar, hace más de un siglo,

la pócima indigesta de que somos parientes de los grandes monos, y, aún antes, la muy pedestre realidad de que damos a la gravedad la misma respuesta monótona que dan las piedras).

Los defensores del paradigma C-R clásico comparten un estilo común de pensamiento, caracterizado por el rigor deductivo y formal, la decisión de aceptar seriamente las consecuencias de tomar a la mente como sistema de cómputo simbólico, y la tendencia a privilegiar los aspectos formales, y por así decirlo «sintácticos» de los modelos, por encima de sus fundamentos semánticos (Johnson-Laird, 1975; De Vega, 1982). Estas características se ponen de manifiesto en su renuncia a admitir la funcionalidad cognitiva de los constructos o fenómenos (como los de conciencia, imagen mental y aprendizaje profundo de conceptos) que no resultan fáciles de encajar en la noción tradicional de cómputo como proceso de transformación discreta de una estructura simbólica, que sería también discreta, compositiva y relevante por su forma, ya que no por su capacidad de designar ni por su expresión fenoménica «en pantalla». ¿Cómo entender, desde esa perspectiva que las imágenes estructuralmente «densas» y continuas, analógicas, puedan ser autónomas y funcionales (Pylyshyn, 1973, 1978, 1980, 1981, 1984)?, ¿qué puede añadir a la cerrada causación de los cálculos el hecho de que éstos dejen algunos residuos en la pantalla de la conciencia (Jackendoff, 1987)?, ¿cómo puede alcanzar la sintaxis subyacente de un sistema de cómputo de símbolos un poder de cómputo mayor que el que ya posee de antemano (Fodor, 1975)?

Como se señala en *Objetos*, «la aportación del paradigma C-R a la psicología y la ciencia cognitiva ha sido, en todo caso, de enorme importancia. Implica una exigencia de rigor conceptual y precisión algorítmica en los modelos de la mente muy beneficiosa para la psicología. Sirve, una y otra vez, de semillero de problemas profundos que debe plantearse cualquier modelo psicológico del conocimiento. Facilita el desarrollo de modelos de la mente que aúnan una amplia generalidad explicativa y una minuciosa precisión computacional, lo cual posibilita el desarrollo de simulaciones... Sin el paradigma C-R, la psicología cognitiva hubiera sido decisivamente diferente y, muy probablemente, mucho menos rigurosa e interesante» (p. 93), pero la psicología es una ciencia experimental y que, frecuentemente, debe hallar soluciones de compromiso entre la coherencia conceptual y la consistencia empírica. Ello explica que la mayoría de los psicólogos cognitivos, dedicados a la investigación empírica, prefieran una versión más débil de la metáfora del ordenador, una interpretación menos literal de la disolución de la frontera tradicional entre mentes y máquinas.

(3) La versión débil de la metáfora de la máquina simbólica define nuestra tercera respuesta al reto de Turing. Es la que adoptan por lo general los psicólogos cognitivos de propensión más empírica que, aún empleando el lenguaje del cómputo y el «procesamiento de la información», no están dispuestos a asumir las consecuencias que se derivan de considerar que la mente es *literalmente* un sistema de cómputo simbólico. Como ha destacado Zenon Pylyshyn (1980, p. 114), «ha habido una resistencia a tomar la computación como una descripción *literal* de la actividad mental, y no sólo como una metáfora con un valor meramente heurístico... La falta de consideración de la tesis de la computación en un sentido literal ha posibilitado una amplia gama de actividades realizadas bajo la rúbrica

de la «teoría del procesamiento de la información, algunas de las cuales representan una desviación significativa de las ideas... básicas en toda teoría computacional de la mente». La aceptación del desafío de Turing como *metáfora*, como juego, ha dado lugar a una psicología cognitiva de estilo más inductivo que deductivo, más basada en criterios de verosimilitud empírica que de coherencia teórica o formal, y en que se emplea el vocabulario de la ciencia cognitiva más clásica (como si las teorías fueran realmente computacionales) para formular modelos que, en muchas ocasiones, no permiten en absoluto definir sistemas completos de cómputo, y en algunos se basan en premisas difíciles de entender o justificar desde la perspectiva del paradigma «duro» C-R. En algunos debates clásicos, como los relacionados con el valor funcional de las imágenes mentales (véase, por ejemplo, Kosslyn, 1980 y Finke, 1989) o la conciencia (Baars, 1988), se expresan con claridad las diferencias sustantivas entre teorías del procesamiento y paradigma C-R, por debajo de la identidad superficial en el vocabulario teórico que emplean. Si en el caso del enfoque literal del reto de Turing valorábamos positivamente el rigor y la coherencia formal, en el de las teorías del procesamiento debemos valorar la ingente tarea empírica realizada, que ha permitido una minuciosa acumulación de detalles sobre el trabajo de la mente, aunque no siempre una visión integrada de su funcionamiento.

(4) ¿Es posible desarrollar una visión integrada alternativa a la del paradigma C-R, y que evite las consecuencias de tomarse literalmente la idea de la mente como maquinaria de cómputo simbólico? Sí, es posible. La alternativa es la cuarta respuesta que se analiza en *Objetos* al reto de Turing: consiste en considerar a la mente, a la manera conexionista, como un conjunto de propiedades molares funcionales que se derivan del funcionamiento masivo, paralelo y distribuido, de enjambres de unidades interconectadas, que se activan (o desactivan) mutuamente. Los sistemas conexionistas, en función de sus propiedades intrínsecas, poseen propiedades deseables para modelar el funcionamiento cognitivo: son relativamente estructurables y flexibles, y soportan con entereza condiciones de degradación a las que frecuentemente se enfrentan los sistemas cognitivos reales. Además, pueden dar cara a demandas múltiples y simultáneas del medio, se modifican adaptativamente en virtud de la experiencia, y pueden completar patrones conceptuales o perceptivos sin necesidad de contar con toda la información lógicamente necesaria para hacerlo.

Los sistemas conexionistas son estructuras esencialmente *inductivas* de procesamiento, frente al carácter deductivo de los sistemas seriales clásicos de cómputo simbólico. A pesar de que aún no hayan alcanzado el grado de «realismo neural» deseable, es obvio que rompen por principio con el dualismo funcionalista que caracterizaba a los enfoques clásicos. Además, y aunque aún sea limitada su capacidad explicativa, han demostrado en varias áreas (especialmente en tareas de tipo inductivo) que pasan aceptablemente la prueba de Turing. Responden más al modelo de máquinas complejas (como las termodinámicas) capaces de «autoorganizarse», en función de sus propiedades intrínsecas, que a las capaces de tratar cualquier clase de algoritmos en virtud de principios extrínsecos y regulaciones externas.

Como es normal en las situaciones de enfrentamiento de paradigmas en

la historia de las ciencias, el enfoque «microcognitivo» del conexionismo no sólo ofrece una imagen global alternativa de lo mental, sino también un modelo diferente de *en qué consiste explicar* en psicología cognitiva, y de los criterios de justificación de las teorías psicológicas. En el nuevo enfoque, explicar la actividad cognitiva no consiste en definir paso a paso una secuencia de cálculos sobre símbolos discretos y supuestamente intencionales (que se corresponden teóricamente con los que realiza el sistema mental), sino en definir de antemano principios de funcionamiento de una estructura en que se efectúan intercambios extensionales (mucho menos intuitivos que los clásicos cálculos sobre representaciones simbólicas, pero también sin el componente relativamente arbitrario o meramente descriptivo que tienen éstos como explicaciones), y que se organiza de forma dinámica, hasta «relajarse» produciendo efectos cognitivos.

Además, al menos idealmente, las teorías conexionistas deberían poseer (aún en un nivel muy abstracto) algún grado de realismo neural, es decir, deberían preservar, en su funcionamiento, propiedades conocidas del funcionamiento del sistema nervioso. Ello establece un criterio adicional para los modelos psicológicos, que restringe la excesiva y peligrosa libertad constructiva que se observa en los modelos clásicos (es decir, el «monismo funcionalista» de los PDP debe tener consecuencias metodológicas importantes en lo que se refiere a los criterios de justificación de los modelos teóricos en psicología). En realidad, los modelos y criterios de explicación conexionista y clásico pueden entenderse, en un sentido metodológico amplio, como complementarios, y no necesariamente como alternativos.

Sin embargo, en lo que se refiere al núcleo mismo de las explicaciones (es decir, a los principios y conceptos básicos, sobre los que se articula la explicación de lo mental), las teorías conexionistas sí pueden entenderse como opciones alternativas a las C-R clásicas. La propia dificultad para entender que puedan manejar reglas en sentido estricto, o estructuras netas de carácter discreto y compositivo es una buena indicación de ello. Sus procesos son de naturaleza esencialmente probabilística, y difíciles de avenir con la noción rigurosa de cálculo. No están guiados por formas, aunque puedan producir estructuras que poseen forma en el sentido laxo —y no en el lógico— del término. Es probable que se manejen especialmente bien para explicar el origen y la funcionalidad de representaciones «densas», tales como las imágenes mentales, por las mismas razones que dificultan la explicación, desde ellos, de las formas más «lingüísticas» y discretas de representación. De este modo, la psicología cognitiva actual se encuentra en una curiosa situación de división, de *split brain*: cuenta con unos recursos en una mano y con otros en otra, pero encuentra serias dificultades para poder emplear las dos manos de forma conjunta e integrada.

En *Objetos* se hace referencia a esa diversidad, aunque en términos algo más optimistas que los planteados aquí. «Nuestros modelos actuales de la mente son así diversos —se dice allí—, y se han beneficiado del rigor conceptual del paradigma C-R, del enorme esfuerzo empírico y de elaboración teórica de la psicología del procesamiento de la información y de las intuiciones y formalizaciones recientes de los conexionistas. La mente es, a la vez, capaz de conocer (y así puede decirse con propiedad que posee «conceptos», «esquemas», etc.) y de autoor-

ganizarse para enfrentarse, de forma cada vez más adaptativa, a un medio cambiante y ruidoso. Actúa de forma «casi reglada», pero no tan reglada como para ser inflexible. Es función de un organismo, pero no puede entenderse como determinada, en sus patrones, por una rígida adherencia a particularidades minuciosas de él. Posee, en cierto sentido metafórico, una sintaxis, pero no la sigue siempre a ciegas... es [...] tan compleja y flexible que sólo podemos entenderla por medio de perspectivas plurales, cuando la miramos desde todas sus esquinas» (p. 107). Esta valoración optimista del pluralismo teórico no debería ocultar, sin embargo, el hecho de que la psicología actual se enfrenta a una escisión entre perspectivas fundamentales sobre lo mental que resultan muy difícilmente compatibles... ¿quizá sirva de consuelo, una vez más, constatar que eso mismo les sucede a los físicos! (la oposición de Einstein a los modelos cuánticos no era meramente estética, sino basada en la intuición de la dificultad de principio para hacerlos compatibles con las teorías clásica y relativista sobre el «mundo mediano» y el macromundo de la cosmología).

Mientras que la segunda sección de *Objetos* ofrece una reflexión sobre el aspecto extensivo del concepto de «mente» (es decir un análisis de los objetos a los que la atribuimos, de forma literal o metafórica), la tercera se dedica a su aspecto intensivo o comprensivo: es decir, a una pregunta sobre la naturaleza misma de las funciones a las que llamamos «mentales». ¿En qué consisten esas funciones? Una manera de enfrentarse a esa cuestión consiste, de nuevo, en relacionar y contraponer las teorías cognitivas con las nociones implícitas sobre las funciones mentales que se apuntan en la psicología natural (y que forman parte de los fundamentos de ella). ¿Cómo se refleja nuestra intuición mentalista, por ejemplo, en el lenguaje ordinario, cuando decimos que alguien piensa, cree, recuerda, desea, etc.? Los enunciados mentalistas incluyen con frecuencia un vocabulario que remite a lo que los filósofos llaman «actitudes proposicionales». Éstas son (1) formas de relación, o acción interna o actitud (en un sentido muy amplio de esta palabra) con respecto a (2) ciertos contenidos (por ejemplo, en «pensar que x», hay un modo de relación, «pensar», y un contenido «x»).

Las expresiones de actitudes proposicionales poseen propiedades que resultan interesantes para el análisis de lo mental: (1) alcanzan un grado máximo de adecuación semántica cuando sus sujetos poseen los rasgos de ser animados y humanos, (2) son muy variables, dependiendo de los verbos que incluyan, en cuanto a su nivel de demanda semántica (los verbos más exigentes son los de pensamiento y lenguaje, que exigen que los sujetos sean personas para formar expresiones semánticamente normales, los menos exigentes los que incluyen el rasgo Fm de «percibir»), (3) los elementos conceptuales básicos de los enunciados de actitudes proposicionales son los verbos mentales, que ciertos autores (Searle, 1983) han tratado de reducir a dos «átomos» semánticos, las creencias y los deseos, en un análisis basado en un lenguaje lógico modal y con cláusulas temporales, (4) todos los lenguajes contienen léxicos mentales ricos y diversos, que sin embargo, (5) es probable que reflejen un sistema conceptual común y universal. La última característica es que (6) los verbos mentales son transitivos, reflejan así lingüísticamente la propiedad de *intencionalidad* (ser «acerca de algo») con la que, desde Brentano (1874) al menos, se define esencialmente lo mental (pero es posi-

ble que haya verbos mentales no intencionales, como pretende Searle, 1983).

Algunos defensores del paradigma C-R (Pylyshyn, 1984; Fodor, 1988) han defendido, en los últimos años, el valor funcional y carácter irreductible del sistema conceptual de las actitudes proposicionales, así como su estrecha conexión con el vocabulario y el sistema explicativo de la psicología cognitiva clásica. En su perspectiva, ese vocabulario mentalista de la psicología natural permite predecir y categorizar la conducta humana (y parte de la animal) de forma diferente a cómo puede discriminarse y concebirse a través de otros vocabularios o conceptos (tales como los físicos, biológicos o puramente conductuales), y expresa la poderosa intuición de que ciertos sistemas complejos regulan su comportamiento en función de representaciones internas, que pueden ser muy independientes de las variaciones proximales de su entorno físico. De este modo, las atribuciones mentalistas de la psicología natural serían, en cierto sentido, también naturalmente «cognitivas», en tanto que presuponen la existencia de seres (a los que se atribuye mente) que, como decía Descartes, «obran por conocimiento y no por la disposición de sus órganos», y de cuya conducta sólo puede darse cuenta recurriendo a mecanismos *top-down*, que implican memoria y predicción, y que permiten un alto grado de autonomía y libertad conductual con relación al medio inmediato.

Sin embargo, la defensa cognitiva del discurso cotidiano de las actitudes proposicionales no debería ocultar que, más allá de las semejanzas, existen diferencias muy profundas entre las intuiciones mentalistas de la psicología natural y las premisas esenciales de las teorías de cómputo. Ya nos referíamos a algunas de esas diferencias al comienzo de este artículo: para los psicólogos naturales, los sucesos de la mente causalmente eficaces transcurren sobre todo en el plano de la conciencia; además su eficacia depende del contenido de las representaciones mentales y no de la forma de ciertas estructuras simbólicas inaccesibles al propio plano de la conciencia. En la intuición cotidiana, la mente es un sistema que actúa de forma finalista, teleológica. Por todo ello, se señala en *Objetos* que «el desarrollo interno de la psicología cognitiva y, más en general, de la psicología científica sólo puede entenderse desde el marco de esa psicología natural que, al tiempo que parece reivindicar la existencia de un plano psicológico autónomo y ofrecer instrumentos conceptuales para un primer acercamiento a él, contiene un vocabulario y unos supuestos implícitos que parecen ofrecer obstáculos importantes para su constitución: un vocabulario ligado a la subjetividad, de naturaleza no-extensional, que permite dar razones más que causas; unos supuestos de teleología en la acción, reducción homuncular de muchas de las explicaciones y libre indeterminación de la conducta... En su ambigua relación con la psicología natural, la psicología cognitiva, al mismo tiempo, la reivindica y la niega, conserva su vocabulario y modifica sutilmente su significado, al introducirle en el engranaje mecanicista de la mente de la que habla» (p. 146).

¿En qué consiste esa psicología natural, cuyo vocabulario y supuestos son explícitamente reivindicados y tácitamente cuestionados por la psicología cognitiva clásica? Uno de los capítulos centrales de *Objetos* se dedica a su análisis: los humanos realizamos con enorme facilidad actividades tales como atribuir intenciones a otros, inferir lo que piensan, creen y desean, comunicarles pensamientos

de forma relevante o comprender las intenciones que subyacen a las ideas que los otros nos comunican, y adaptarnos a lo que saben e ignoran. Por debajo de la aparente facilidad con que hacemos estas tareas delicadas debe existir una maquinaria mental muy compleja (una competencia, si se quiere), que merece por sí misma un serio estudio cognitivo. Al igual que en otros casos, la «suavidad fenoménica» de las funciones puede esconder una gran complicación subyacente. Como se señala en el libro, la psicología cognitiva nos ha enseñado lo útil que puede ser a veces buscar lo computacionalmente complejo más allá de lo fenomenológicamente obvio (la visión es un buen ejemplo de ello; véase Marr, 1982). ¿Cuáles son los mecanismos cognitivos que permiten a los humanos poseer una destreza de atribución e inferencia mentalista —una «psicología natural»— que implica «intencionalidad recursiva» (es decir, poseer estados mentales intencionales, I, sobre otros I, suponiendo, al tiempo, que éstos pueden versar sobre otros I, por ejemplo los nuestros, fomando así estructuras de tipo (I (I (I)))?

En los últimos años, se han hecho esfuerzos considerables para definir tanto la funcionalidad y naturaleza cognitiva de la psicología natural (a la que se ha denominado «teoría de la mente») como su desarrollo en el niño, y posible presencia en otras especies de primates. En síntesis, los resultados de diversas investigaciones hacen viables algunas conclusiones fundamentales:

1. La teoría de la mente, o psicología natural, no debe confundirse con las diversas concepciones y explicaciones que desarrollan las distintas culturas acerca de lo mental. La primera sería una competencia humana universal de atribución, categorización e inferencia de estados mentales propios y ajenos. Las segundas, las «psicologías populares», serían formas histórica y culturalmente variables de dar cuenta y razón de la mente y la conducta.

2. En el hombre, el desarrollo de la teoría de la mente es muy precoz, y tiene un momento esencial entre el segundo año y el quinto (en que se comprende implícitamente el carácter «intensional» de los estados mentales de creencia, es decir, la posibilidad de que sean falsas las creencias que se atribuyen en enunciados mentalistas globalmente verdaderos).

3. Hay muchos argumentos que permiten defender la idea de que la psicología natural, o teoría de la mente (es decir, la competencia que permite realizar actividades de pensamiento «interpersonal») no puede reducirse a otras competencias cognitivas de pensamiento «impersonal», sino que tiene al menos un carácter relativamente independiente y modular (de forma que puede, por ejemplo, alterarse de forma selectiva, como sucede en algunos autistas inteligentes).

4. En antropoides superiores, tales como los chimpancés, hay ya probablemente un cierto desarrollo de un sistema conceptual mentalista, que se expresa en conductas deliberadas, flexibles y espontáneas de engaño interesado.

5. El desarrollo de este sistema —en los humanos y otros primates—, tanto en la ontogénesis como muy probablemente en la filogénesis, se relacionaría mucho con la destreza de engañar de forma espontánea y flexible, y con la capacidad de reconocer el engaño.

6. En lo que se refiere a su naturaleza cognitiva, la teoría de la mente en el hombre depende de una capacidad fundamental de «metarrepresentación». En la formulación de Leslie (1987, 1988), las metarrepresentaciones pueden definir-

se, de forma rápida e intuitiva, como «representaciones que están entre comillas», que no son literales, en las que se suspenden las relaciones ordinarias de referencia y verdad entre las representaciones y las cosas. Así, es una metarrepresentación la del niño que está jugando con una escoba a «esto es un caballo», y lo es también la de la persona que atribuye a otra una creencia, tal como «Carlos cree que “el rey de Francia es calvo”». Los enunciados que implican metarrepresentaciones poseen una marca lógica peculiar: la *intensionalidad* (*con s*). Son enunciados (1) referencialmente opacos (no se puede sustituir lógicamente «Juan cree que “el presidente del Real Madrid veranea en Mallorca”» por «Juan cree que “Ramón Mendoza veranea en Mallorca”», a pesar de que esa sustitución podría hacerse si no estuviera la cláusula de creencia), (2) cuya verdad no compromete con la verdad de las cláusulas sometidas a los verbos intensionales (puede ser verdad que «Juan cree que “Guillermo es honrado”», pero eso no significa que lo sea) y (3) cuya verdad no compromete ni siquiera con la existencia de los objetos definidos en esas cláusulas (puede ser verdad que «Carlos piensa que “el rey de Francia es calvo”», y sin embargo no existir un objeto tal como el rey de Francia). La intensionalidad es una propiedad peculiar de los enunciados de pensamiento y lenguaje. Ello se relaciona con el hecho de que sólo la razón pueda ser portadora de verdad.

7. Así las metarrepresentaciones son las formas de representación subyacentes a los enunciados más prototípicos de atribución mentalista y a los patrones de juego simbólico. Esa intuición permite entrever ciertas relaciones psicológicas profundas entre la capacidad humana de «simular» (en el doble sentido de engañar e inventar objetos imaginarios) y la destreza de atribuir estados mentales.

Las explicaciones cognitivas sobre la psicología natural y las destrezas de metarrepresentación definen bien, como corresponde hacer a este tipo de explicaciones, la *forma* de la estructura cognitiva que permite al hombre ser un animal «listo», propenso a engañar y muy hábil para inventar realidades e inferir estados mentales. Sin embargo, no aclaran la génesis semántica de la teoría de la mente. En *Objetos* se defiende la idea de que la única manera concebible de definir esa génesis implica recurrir a conceptos menos formalistas y más difíciles de acomodar a los modelos de cómputo: nociones tales como las de «conciencia» e «intersubjetividad» (Trevorthen, 1982), sin las cuales no podrían explicarse los procesos que permiten derivar categorías mentalistas de las propias conductas para interpretarlas y predecirlas. Esa constatación conduce, a su vez, a otra que ya señalábamos al principio de este artículo, a saber, que las categorías de que se sirve la «psicología natural» mentalista, que nos sirve de guía en nuestras interacciones, son derivados de aspectos muy básicos de la fenomenología de la conciencia, a los que los humanos accedemos reflexivamente. Esos aspectos se refieren principalmente a la naturaleza «intencional» de los sucesos conscientes, es decir, a su relación con objetos y la naturaleza de dicha relación. Definen cosas tales como si los objetos intencionales poseen una marca temporal (como sucede en el verbo recordar) o no, el grado de probabilidad o certeza que asigna la conciencia a dichos objetos (la diferencia, por ejemplo, entre «suponer» y «creer»), si se da la presuposición de su existencia en el propio proceso mental (diferencia entre «saber» y «creer»), etc.

Situados en la perspectiva de los enunciados mentalistas, podemos hacer una observación interesante, y que nos ayuda a comprender mejor el núcleo del «drama epistemológico» de la psicología: ya hemos visto que, a lo largo de la historia de las ciencias de la naturaleza, las propensiones animistas de un sistema cognitivo esencialmente sesgado a la relación intraespecífica tuvieron que ser contrarrestadas por un serio, largo y sistemático esfuerzo de «desmentalización», gracias al cual fue posible establecer enunciados externos y puramente extensionales al mundo físico. Enunciados de los que pudiera eliminarse cualquier referencia al sujeto conocedor y a sus procesos mentales, y todo verbo mental en el núcleo de los propios enunciados (las proposiciones finales de la ciencia no tienen la forma «Fulano piensa que “p”», sino escuetamente la forma «p»). Tales proposiciones están, por así decirlo, «doblemente desmentalizadas»: ni están sometidas a otras, jerárquicamente superiores, que contengan sujetos o verbos mentales («Fulano *piensa* que...»), ni ellas mismas los tienen (en «p» no hay nada mental). La eliminación de las proposiciones mentalistas de orden superior se logra por los métodos habituales de objetivación científica: asegurando la precisión y replicabilidad de las observaciones, separando a éstas de las inferencias, y garantizando el rigor, la posibilidad de réplica y la universalidad de las propias inferencias, mediante el uso de procedimientos lógicos y matemáticos y de criterios de coherencia. Todo ello permite pasar de proposiciones aún personales (como «veo tal fenómeno») a otras suprapersonales («sucede tal fenómeno»). En cuanto a la «ablación» de sujetos y verbos mentales del interior de las propias proposiciones, ya objetivadas y «universalizadas», se ha realizado a través del largo proceso histórico de superación de las propensiones animistas en el análisis de la naturaleza, proceso al que ya hemos hecho alguna mención en líneas anteriores.

Como se señala en *Objetos*, «los procesos de desanimación de los fenómenos naturales y objetivación de los enunciados científicos (es decir, de eliminación de la atribución de funciones F_m a los fenómenos naturales, por una parte, y de cláusulas F_m en los enunciados observacionales o teóricos de la ciencia, por otra) han sido procesos estrechamente entrelazados, pero que deben distinguirse cuidadosamente. Los primeros permitieron desarrollar una concepción mecanicista de la naturaleza... en que los fenómenos no se comprendían como manifestación de sustancias o mentes, sino, a través de sus relaciones, como especificaciones de leyes de interacción precisas entre propiedades puramente extensionales de la materia y la energía. Los segundos procesos permitieron la justificación de las leyes mediante criterios relacionados con el carácter supra-subjetivo tanto de las observaciones como de las inferencias teóricas» (p. 175). Pero esta observación nos enfrenta, de bruces, a una pregunta inquietante, y que es de una gran importancia para el psicólogo: si la condición de posibilidad de la objetividad científica fue la eliminación de toda clase de predicados de funciones mentales. ¿Cómo es posible una ciencia que trata de ser objetiva, y de serlo precisamente sobre las funciones mentales?

La cuestión está muy lejos de ser meramente retórica. Los enunciados que remiten a funciones mentales (las proposiciones F_m) son especialmente difíciles de manejar por una ciencia objetiva de la naturaleza, por varias razones:

1. Alcanzan su mayor adecuación precisamente cuando se dicen *en prime-*

ra persona de singular. De hecho, no se fundamentan en la misma base observacional cuando se dicen en primera persona que cuando se atribuyen en segunda o tercera persona («tú (él) crees (cree) que "x"») no tiene el mismo fundamento que «yo creo que "x"»). «Cuando un sujeto dice «Yo Fm (p)», «Yo (pienso, creo, intuyo, etc.) que p», remite a un mundo que no es, en sí mismo, exteriorizable más que por sus *síntomas*. Hace un enunciado que, por así decirlo, *nadie, que no sea él mismo, puede corregir*. Un enunciado recluso, incorregible» (p. 177).

2. Sin embargo, los enunciados científicos tienen que ser estrictamente simétricos (no asimétricos) para todas las personas, si quieren ser objetivos. «Así, los enunciados mentalistas en primera persona de singular han presentado siempre una paradójica ambivalencia en psicología; por una parte, está la pregnante y poderosa intuición, que tenemos como psicólogos naturales, de que *sin ellos la psicología no sería posible, ni podría concebirse*; la intuición de que, en cierto sentido, *la primera persona de singular es el «núcleo básico», la «residencia natural» de los enunciados psicológicos*. Por otra parte, y en tanto que tales enunciados no sólo articulan lo que entendemos por subjetividad sino que se encierran en ella, no ofrecen garantías como enunciados observacionales en los que pueda basarse la actividad científica» (p. 178). Ello explica su doble y contradictorio carácter de presuposiciones básicas de la psicología y obstáculos epistemológicos (en el sentido de Bachelard, 1948) para ella.

3. Además, ese carácter obstaculizador se acentúa cuando se toma en consideración qué dicen los enunciados Fm de la psicología natural y qué callan o ignoran: como expresiones que son de la conciencia reflexiva, se ven limitados a reflejar la limitada gama de dimensiones intencionales que la propia conciencia establece, y sólo eso. Pero ofrecen la engañosa imagen de una mente translúcida, «que se ve funcionar a sí misma».

Esta última observación ayuda a entender en qué consistió el drama histórico de la psicología introspectiva, que se examina en el último capítulo del libro, dedicado a un análisis selectivo de algunos aspectos del desarrollo de la psicología científica, en que se emplean los instrumentos conceptuales desarrollados en los capítulos anteriores. El intento de los psicólogos introspectivos era el de pasar de enunciados mentalistas en primera persona, «pienso (percibo, siento, etc.) que x», a otros enunciados nomotéticos, del tipo «mi pensar (percibir, recordar...) consiste en...». *Este salto presupone que en la conciencia de la relación con objetos que la mente elabora, se incluye una conciencia de los procesos o las estructuras que permite esa elaboración*. El problema, para los psicólogos introspectivos, fue el de cómo establecer las condiciones y justificaciones para pasar de los enunciados mentalistas en primera persona a las proposiciones nomotéticas sobre funciones mentales. Mientras que algunos psicólogos introspectivos (como James, los de Würzburgo o incluso Titchener) adoptaron posiciones laxas, abiertas, otros, como Wundt, se mostraron muy reticentes y restrictivos con respecto a las posibilidades de la introspección, limitándolas a las condiciones en que pudiera mantenerse muy cercana de aspectos puramente físicos y extensionales de los estímulos.

En cualquier caso, el intento introspeccionista de formular una psicología Fm, que incluía componentes Fm (es decir, mentalistas) tanto en sus enunciados observacionales de origen introspectivo como en los teóricos, se saldó con un fra-

caso histórico, que se hizo cada vez más evidente, tanto en las dudas crecientes de los grandes creadores (Wundt y James, por ejemplo) sobre las posibilidades de la psicología introspectiva como en episodios tales como el debate del pensamiento sin imágenes. «Podríamos decir que el drama de aquella psicología fue quizás el de confundir un sistema destinado a la previsión de la conducta propia y ajena con un mecanismo de acceso a una maquinaria mental supuestamente transparente» (p. 202). Esa ilusión peligrosa de la psicología había sido ya anticipada por Wundt y James, cuando hablaban por ejemplo del error de objeto, es decir, de la tendencia a confundir los objetos de la conciencia con la conciencia de los objetos.

En los términos analíticos que estamos empleando, el conductismo constituyó un intento de desarrollar una psicología cuyos enunciados, tanto observacionales como teóricos, fueran estricta y escuetamente extensionales. Enunciados completamente simétricos y objetivos, en tercera persona de singular, acerca de un «otro», concebido como un objeto infinitamente distante de la subjetividad propia. En este caso, el drama consistió en que ese formidable intento de «desmentalizar» (desintencionalizar) la psicología nunca pudo lograrse del todo, quizá con la excepción de la obra de Skinner y de sus seguidores. Como ha señalado lúcidamente Yela (1974), en las nociones centrales de estímulo y respuesta de Watson se produce sistemática y subrepticamente un salto entre las definiciones puramente extensionales (los estímulos como «energías», y las respuestas como meros movimientos o patrones musculares) y las implícitamente intencionales (los estímulos como situaciones significativas, las respuestas como conductas significativas). En cualquier caso, la ingente operación conceptual del conductismo implicó un cambio en los criterios de justificación de las entidades admisibles en el vocabulario psicológico (que deberían reducirse a elementos real o potencialmente observables, relacionados de forma serial y asociativa, por medio de procesos bien definidos de aprendizaje y condicionamiento, entendidos en términos laxamente mecanicistas).

¿Qué es lo que define a los modelos cognitivos clásicos (paradigma C-R y procesamiento de la información) en los términos de análisis que estamos empleando? Estos modelos se ven iluminados con una nueva luz cuando se entienden como intentos de construir una psicología rigurosa y mentalista, que aunaría un lenguaje empírico esencialmente extensional con un lenguaje teórico peculiarmente intencional. En efecto, la psicología cognitiva clásica trata de definir una especie de «mecánica de la intencionalidad», y ahí es donde la noción de cómputo juega un papel central (en tanto que los cómputos, por una parte, se insertan por definición en procedimientos efectivos, es decir, en máquinas formales absolutamente precisas y deterministas, y, por otra parte, procesan *símbolos*, que serían estructuras discretas, estructuradas e intencionales). Por otra parte, el enfoque clásico justifica esa mecánica de la intencionalidad en enunciados empíricos *extensionales*, tales como «los sujetos tienen una latencia media de x , en la solución de los problemas y ». Enunciados que no son menos objetivos ni menos extensionales que los clásicos del conductismo, y que están muy lejos de los clásicos enunciados intencionales de la vieja psicología introspectiva (el papel de la introspección en la moderna psicología cognitiva es muy reducido, y

en general más reconocido por los investigadores más cercanos a la IA que acentúan las exigencias sintácticas de los modelos por encima del rigor de las observaciones).

Esa brevísima caracterización de la psicología cognitiva clásica (como una «mecánica objetiva de la intencionalidad») permite dar con el hilo que lleva a una madeja de problemas, que se han ido definiendo en su ya no tan breve historia. Se trataba de una psicología que, al tiempo que se dibujaba a sí misma como rebelión mentalista contra el conductismo, ofrecía aspectos interesantes de continuidad con respecto a su supuesto adversario (continuidad en la extensionalidad de las observaciones, y también, aunque más matizada, en el carácter mecanicista —aunque ahora de «máquinas formales» y no materiales— de los modelos). De una psicología que, a la vez, reivindicaba explícitamente las categorías intencionales del viejo mentalismo natural, y negaba tácitamente muchos de sus supuestos principales (ya hemos hablado de ello). En ella se diferencian, en *Objetos*, dos fases históricas principales: una primera, en que predominaron modelos de corte «sintáctico» y formalista, muy apegados aún a teorías canónicas, lógicas y gramaticales; otra, desde mediados de los setenta, que se caracteriza por el empleo de categorías macroestructurales, la sustitución de muchas explicaciones «sintácticas» por otras semánticas o pragmáticas, el empleo de tareas más ecológicas, y la conciencia creciente de las limitaciones de los modelos computacionales clásicos. «De este modo —se dice en *Objetos*—, la historia relativamente corta de la psicología cognitiva puede entenderse, de forma metafórica pero provechosa, como la historia de unas complejas relaciones entre intencionalidad y forma. O, si se quiere, como un indeciso enfrentamiento entre la ineludible exigencia de definición y claridad sintáctica que requiere cualquier explicación para ser considerada como propiamente computacional, y la inevitable necesidad de referencia intencional que requiere cualquier estado o proceso para ser considerado como propiamente psicológico. Si, la historia de la psicología cognitiva clásica es, en buena medida, la historia de esta agonía de intención y forma» (p. 220).

¿Cómo pueden emerger fenómenos intencionales y formas, tales como las estructuras de lo mental, a partir del funcionamiento estrictamente extensional de un sistema tal como el cerebro? Esa pregunta, que resultaba del todo intratable hasta hace pocos años, comienza a tener una respuesta, aún incipiente e insuficiente pero llena de posibilidades, gracias al enfoque conexionista (Rumelhart, McClelland y PDP, 1986). En este caso (como en el conductismo), nos encontramos ante una psicología cuyos enunciados tanto teóricos como empíricos son netamente extensionales. Pero, a diferencia del conductismo, la versión conexionista de la mente empieza a ser capaz de explicar cómo pueden generarse estructuras intencionales en calidad de propiedades molares del funcionamiento masivo, simultáneo, e inherentemente constreñido por principios de equilibrio, de unidades que computan (?) ... niveles de activación y no símbolos estructurados. Como ha destacado Pagels (1991), las redes neurales y los modelos conexionistas son parte de *Los sueños de la razón* de las dos últimas décadas de nuestro siglo: los sueños de una razón cada vez más consciente de la necesidad de recurrir a conceptos tales como los de orden, caos y complejidad, para explicar aspectos pertinentes del mundo físico y social. Si: desde hace muy pocos años, es un sueño

razonable de la «razón psicológica» el de explicar la propia complejidad de la mente como producto de la organización funcional del cerebro, y mediante nociones que aún suenan extrañas a la mayoría de los psicólogos: nociones tales como resonancia, relajación, bifurcaciones, cuencos de atracción, que pueden modificar profundamente el vocabulario de la psicología cognitiva en los próximos años.

El proyecto de conocer objetivamente la mente es, sin duda, desmesurado. Sin embargo, no sería la primera vez que la propia mente humana realiza, con paciencia y mesura, un proyecto epistémico ingente: recordemos que ha sido esa mente, seguramente sesgada, fácilmente animista y compulsivamente mentalista, esa mente de «animal social», la que ha permitido acumular una gran cantidad de conocimiento «objetivo» y «des-mentalizado» sobre la naturaleza, saltando por encima de sus propios sesgos y limitaciones. ¿Por qué no va a poder realizar un empeño de ese calibre con respecto a sí misma? No es descartable la posibilidad (sugerida alguna vez por Chomsky) de que la mente no esté igualmente preparada para conocer todos los dominios de lo real. Quizás el conocimiento de la propia mente por sí misma sea uno de los dominios más vedados, más clausurados, más rebeldes a un sometimiento riguroso: a veces, tenemos la impresión de que la mente juega a ocultarse a sí misma. Los psicólogos seguimos teniendo la costumbre de lamentar el «atraso» de nuestra disciplina en comparación con las ciencias más viejas de la naturaleza, de forma semejante a como los campesinos tienen la ancestral costumbre de quejarse todos los años de la cosecha (incluso cuando ésta ha sido mejor que nunca). Sin embargo, los rápidos y profundos progresos de la psicología cognitiva en los últimos treinta años, las buenas cosechas obtenidas por los investigadores de los procesos cognitivos, desmienten esos complejos de inferioridad epistémica, y sugieren que quizá la mente no tenga por qué ser necesariamente un misterio para sí misma.

REFERENCIAS

- Anderson, J.R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85, 249-277. (Trad. cast.: Argumentos acerca de las representaciones mediante la capacidad para formar imágenes mentales. En Sebastián, M.V. (Ed.), *Lecturas de Psicología de la Memoria*. Madrid: Alianza Universidad, 1983.)
- Anderson, J.R. (1979). Further arguments concerning representations for mental imagery: A response to Hayes-Roth and Pylyshyn. *Psychological Review*, 86, 249-277.
- Astington, J.V., Harris, P.L. & Olson, D.R. (Eds.) (1988). *Developing Theories of Mind*. Cambridge: Cambridge University Press.
- Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Brentano, F. (1874). *Psychologie von empirischen Standpunkt*. Leipzig (Ed. Hamburgo: Meiner, 1955).
- Cohen, G. (1977). *The Psychology of Cognition*. London: Academic Press. (Trad. cast.: *Psicología cognitiva*. Madrid: Alhambra, 1983.)
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press. (Trad. cast.: *Reglas y representaciones*. México: Fondo de Cultura Económica, 1983.)
- Descartes, R. (1637). *Discours de la Méthode*. (Trad. cast.: *El discurso del método*. Madrid: Akal, 1989.)
- Dretske, F. (1988). *Explaining Behavior. Reasons in a World of Causes*. Cambridge, Mass.: The MIT Press.
- Finke, R.A. (1989). *Principles of Mental Imagery*. Cambridge, Mass.: The MIT Press.
- Fodor, J. (1975). *The Language of Thought*. New York: Harper & Row. (Trad. cast.: *El lenguaje del pensamiento*. Madrid: Alianza, 1984.)

- Fodor, J. (1983). Fijación de creencias y adquisición de conceptos. En Piattelli-Palmerini (Ed.). *Teorías del lenguaje. Teorías del aprendizaje* (pp. 187-193). Barcelona: Crítica.
- Fodor, J.A. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 173-189.
- Graubard, S.R. (Ed.) (1993). *El nuevo debate sobre la inteligencia artificial. Sistemas simbólicos y redes neurales*. Barcelona: Gedisa, 1993.
- Hayes-Roth, F. (1979). Distinguishing theories of representation. A critique of Anderson's «Arguments concerning mental imagery». *Psychological Review*, 86, 376-382.
- Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley. (Trad. cast.: *La organización de la conducta*. Madrid: Debate, 1985.)
- Humphrey, N. (1983). *Consciousness Regained. Chapters in the Development of Mind*. Oxford: Oxford University Press. (Trad. cast.: *La reconquista de la conciencia. Desarrollo de la mente humana*. México: Fondo de Cultura Económica, 1987.)
- Humphrey, N. (1986). *The Inner Eye*. Londres: Faber & Faber. (Trad. cast.: *La mirada interior*. Madrid: Alianza, 1993.)
- Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, Mass.: The MIT Press.
- James, W. (1890). *The Principles of Psychology*. New York: Holt. (Trad. cast.: *Principios de Psicología*. México: Fondo de Cultura Económica, 1990.)
- Johnson-Laird, P.N. (1975). Models of deduction. In Falmagne, R. (Ed.). *Reasoning: Representation and Process in Children and Adults*. Hillsdale, N.J.: Lawrence Erlbaum.
- Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. (1988). *The Computer and The Mind: An introduction to Cognitive Science*. Glasgow: William Collins Sons & Co. Ltd. (Trad. cast.: *El ordenador y la mente. Introducción a la ciencia cognitiva*. Barcelona: Paidós, 1990.)
- Kosslyn, S.M. (1980). *Image and Mind*. Cambridge, Mass.: Harvard University Press.
- Kuhn, Th.S. (1975). *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica.
- Malt, B.C. & Smith, E.E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250-269.
- Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman & Co. (Trad. cast.: *La visión. Una investigación basada en el cálculo acerca de la representación y el procesamiento humano de la información visual*. Madrid: Alianza, 1985.)
- McCulloch, W. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin for Mathematical Biophysics*, 5, 115-155. (Reeditado en Boden, M. (Ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990.)
- Newell, A. & Simon, H.A. (1976). Computer science as empirical enquiry: Symbols and search. Tenth Turing Lecture. In *Communications of the Association for Computing Machinery*, 19. Association for Computing Machinery. (Reeditado en Boden, M. (Ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990.)
- Pagels, H.R. (1991). *Los sueños de la razón*. Barcelona: Gedisa.
- Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press. (Trad. cast.: *La nueva mente del emperador*. Madrid: Mondadori.)
- Piaget, J. (1961). *La naissance de l'intelligence chez l'enfant*. Neuchâtel: Delachaux et Niestlé. (Trad. cast.: *El nacimiento de la inteligencia en el niño*. Madrid: Aguilar, 1969.)
- Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Putnam, H. (1960). Minds and Machines. In Anderson, A.R. (Ed.). *Minds and Machines*. New York: Prentice-Hall. (Trad. cast.: *Mentes y máquinas* (pp. 113-149). Barcelona: Tusquets, 1984.)
- Pylyshyn, Z.W. (1973). What the mind's eye tells to the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80, 1-24.
- Pylyshyn, Z.W. (1978). Imagery and artificial intelligence. In Savage, W. (Ed.). *Perception and Cognition. Issues in the Foundation of Psychology. The Minnesota Studies in the Philosophy of Science*. Vol. IX. Minneapolis: University of Minnesota Press., 15-55.
- Pylyshyn, Z.W. (1979). Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review*, 86, 383-394.
- Pylyshyn, Z.W. (1980). Computation and cognition: Issues in the Foundations of Cognitive Science. *The Behavioral and Brain Sciences*, 3 (1), 111-132.

- Pylyshyn, Z.W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 86, 16-45.
- Pylyshyn, Z.W. (1984). *Computation and Cognition: Toward a Foundation of Cognitive Science*. Cambridge, Mass.: The MIT Press. (Trad. cast.: *Computación y conocimiento. Hacia una fundamentación de la ciencia cognitiva*. Madrid: Debate, 1988.)
- Pylyshyn, Z.W. (1989). Computing in Cognitive Science. In Posner, M.I. (Ed.). *Foundations of Cognitive Science* (pp. 49-92). Cambridge, Mass.: The MIT Press.
- Rivière, Á. (1986). *Razonamiento y representación*. Madrid: Siglo XXI.
- Rivière, Á. (1991). *Objetos con mente*. Madrid: Alianza.
- Rosch, E. (1978). Principles of categorization. In Rosch, E. y Lloyd, B.B. (Eds.). *Cognition and Categorization* (pp. 27-48). Hillsdale, N.J.: Lawrence Erlbaum.
- Rosenblatt, F. (1962). *Principles of Neurodynamics*. New York: Spartan.
- Rumelhart, D.E., McClelland, J.L. & PDP Research Group (Eds.) (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*. Vol. 2: *Psychological and Biological Models*. Cambridge: The MIT Press.
- Searle, J.R. (1980). Minds, brains and programs. *The Behavioral and Brain Sciences*, 3, 413-424. (Reeditado en Boden, M. (Ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990.)
- Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press. (Trad. cast.: *Intencionalidad. Un ensayo en la filosofía de la mente*. Madrid: Tecnos, 1992.)
- Searle, J.R. (1992). *The Rediscovery of the Mind*. Cambridge, Mass.: The MIT Press.
- Somlensky, P. (1987). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy, Supplement*, 26, 137-161.
- Somlensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.
- Turing, A.M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2ª serie, 42, 230-265.
- Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Varela, F.J. (1988). *Cognitive Science. A Cartography of Current Ideas*. (Trad. cast.: *Conocer*. Barcelona: Gedisa, 1990.)
- Vega, M. de (1982). La metáfora del ordenador: implicaciones y límites. En Delclaux, I. y Seoane, J. (Eds.). *Psicología cognitiva y procesamiento de la información* (pp. 63-81). Madrid: Pirámide.
- Vega, M. de (1984). *Introducción a la psicología cognitiva*. Madrid: Alianza.
- Whiten, A. (Ed.) (1991). *Natural Theories of Mind*. Oxford: Basic Blackwell.
- Yela, M. (1974). *La estructura de la conducta. Estímulo, situación y conciencia*. Madrid: Real Academia de Ciencias Morales y Políticas.