

Ordenadores, robots y personas: hacia una teoría general de los sistemas inteligentes*

Mariano Yela**

LA ESTRUCTURA GENERAL DE LA INTELIGENCIA. INTELIGENCIA Y SISTEMA COGNITIVO

El modelo de computación simbólica

La perspectiva general de la inteligencia es considerada principal y predominantemente en nuestros días por la llamada *psicología cognitiva*. Comprende esta psicología muy diversos enfoques, pero todos comparten dos características representativas. La primera es que todos se proponen la *recuperación de la mente* (*The rediscovery of the mind*, Searle, 1991). La segunda es que esta recuperación se intenta conseguir *por la vía del ordenador* (*The computer and the mind*, Johnson-Laird, 1988).

La psicología de las últimas décadas ha replanteado la cuestión de la inteligencia. La ha situado en un plano *mental*. La investigación específica de la inteligencia no consiste en examinar los componentes físicos, por lo demás innegables, de la conducta inteligente. No es lo propio y distintivo de la teoría de la inteligencia la consideración de los estímulos, en tanto que energías físicas, ni de las respuestas, como movimientos mus-



* La primera versión de este artículo fue publicada en el nº 71, año XLVI, de los *Anales* de la Real Academia de Ciencias Morales y Políticas. Reproducimos el texto con su autorización.

** Mariano Yela fue catedrático de Psicología General de la Universidad Complutense de Madrid desde 1957. Falleció en noviembre de 1994.

culares. Ese era, se dice, el objetivo de la *psicología conductista*, predominante hasta mediados de nuestro siglo, al menos en la psicología anglosajona, de influjo imperante en la psicología mundial.

Frente al llamado paradigma conductista E-R (Estímulo-Respuesta), se alza, hacia la mitad del siglo, la revolución cognitivista, que restaura, con nuevas estrategias de investigación, la psicología del conocimiento, la mente y la conciencia, nunca, es verdad, del todo interrumpida en Europa, desde Külpe y la *Gestalttheorie* a Bartlett, Michotte, Piaget o Vigotsky.

Lo decisivo para el psicólogo que estudia la conducta del hombre no son los estímulos iniciales y las respuestas finales, sino lo que acontece entre medias, la elaboración mental de una acción significativa, como respuesta a una situación que significa algo para el sujeto.

El acontecimiento inaugural más saliente del nuevo estilo psicológico fue el Congreso celebrado en el MIT (*Massachusetts Institute of Technology*), en 1956, sobre la *Teoría de la Información*. En ese año aparecen *The logic theory machine*, de Newell y Simon; *Three models of language*, de Chomsky; *The magical number seven, plus or minus two*, de George Miller, y *Study of thinking*, de Bruner, Goodnow y Austin. En ese mismo año John McCarthy acuña la expresión *inteligencia artificial*, que hará rápida fortuna. En la década anterior se había desarrollado la investigación cibernética (Wiener, 1948), la teoría de la comunicación (Shanon y Weaver, 1949) y las nociones sobre las «asambleas neuronales», de Hebb (1949). En los años siguientes se multiplican las contribuciones fundacionales del nuevo paradigma. En 1957 publica Chomsky sus *Syntactic Structures*; en 1958, von Neumann *The computer and the brain*; en 1960, Miller, Galanter y Pribram su *Plans and the structure of behavior*, y el mismo Hebb, *The American revolution*, su discurso presidencial a la *American Psychological Association*, en el que intenta justificar el estudio del conocimiento, la conciencia y la mente desde el enfoque conductista.

El impulso inicial de esta nueva orientación se debe a un matemático genial: Alan Turing. En 1950 propone su célebre *Gedankenversuch* o experimento imaginario: supongamos que en una habitación hay una persona y una máquina, y en otra, un investigador. El investigador formula preguntas y problemas que un mensajero entrega, en la otra habitación, a la persona y a la máquina, de los que recoge las respuestas y se las lleva al investigador. Si éste no puede distinguir entre las respuestas de la máquina y de la persona, ¿con qué derecho podemos decir que la persona piensa y la máquina no? En realidad el problema para Turing no estriba en si una u otra piensa, sino en si las soluciones que ambas producen son indistinguibles. Si lo son, ambas exhiben conductas inteligentes.

Pero ¿es concebible una máquina que pueda imitar el comportamiento inteligente de una persona? La respuesta es sí. La había adelantado el propio Turing en su trabajo de 1936 *On computable numbers*. En él demuestra, fundado en la lógica matemática y en la teoría de la computación, y dentro de las limitaciones del teorema de Gödel, que es matemáticamente plausible una máquina de estados finitos dotada de una cinta de longitud ilimitada sobre la que pueda realizar, según instrucciones precisas, las cinco operaciones de

moverla hacia la derecha o la izquierda, de escribir un «1» o un «0», y de pararse. Una máquina de este tipo puede, en principio, realizar, en código binario, todas las funciones recursivas, es decir, resolver todos los problemas que tengan solución y contestar a todas las cuestiones que puedan ser definidas lógicamente y algorítmicamente.

Pronto otros científicos construyen máquinas computadoras concretas. Johan von Neumann es el primero en idear y construir un ordenador regido por los principios de la máquina universal y abstracta de Turing: cálculo serial de relaciones entre símbolos, representados mediante un código binario.

Está claro que lo que puede hacer un ordenador viene limitado por su *material* o *hardware*. Lo que efectivamente hace, sin embargo, cuando se pone en marcha, depende exclusivamente de su *programa*, *logical* o *software*, es decir, de los datos que se le suministran y las instrucciones que se le dan, sin que para describir y explicar esta actividad haga falta considerar los componentes físicos, los cables, enchufes o chips, que puedan constituir la máquina concreta.

La psicología cognitiva se monta así sobre la *metáfora del ordenador*. La mente es al organismo como el programa al ordenador. El organismo humano pone límites a lo que el hombre puede hacer. Lo que el hombre efectivamente hace es, sin embargo, lo que organiza y dispone su programa mental. Eso es lo que tiene que indagar el psicólogo cognitivo.

Durante las décadas de los años sesenta, setenta y parte de los ochenta, hasta hoy, predomina la psicología cognitiva que acepta, más o menos literalmente, la metáfora del ordenador. Sus máximos artifices, defensores y promotores son Fodor y Pylyshyn (Fodor, 1968, 1975, 1981, 1983, 1988; Fodor y Pylyshyn, 1988; Pylyshyn, 1984). Sus obras representan la más rigurosa expresión del que podemos llamar paradigma C-S, o de *computación de símbolos*. Como el ordenador computa relaciones sintácticas entre representaciones codificadas, la mente humana recibe información, la codifica en forma de símbolos representativos (perceptos, conceptos, proposiciones, esquemas), establece cómputos sintácticos o de relaciones entre estos símbolos, los registra y conserva, los recupera cuando los necesita, los somete a diversas transformaciones mediante múltiples estrategias y, finalmente, por medio de ellos, dirige y evalúa la acción.

La mente actúa como un ordenador serial o secuencial. Sus componentes son símbolos que representan información sobre las entidades a que se refieren, según diversos códigos. En el fondo, los códigos definen los símbolos como clases de equivalencia de señales físicas, que, por eso, por su carácter físico, pueden tener eficacia causal en la actividad del organismo; la cual, a su vez, resulta del tratamiento o «procesamiento» sintáctico de estas clases de equivalencia, mediante la aplicación sucesiva de reglas precisas sobre los símbolos y sus conjuntos.

Con mayor flexibilidad y mesura, el gran representante de este paradigma, creador de múltiples programas concretos e indagador con ellos de muy diversos campos de la psicología, es Herber A. Simon, profesor de psicología y ciencias cognitivas y premio Nobel en Economía por sus trabajos sobre teoría y métodos de toma de decisiones en la conducta administrativa

y económica de las organizaciones. Una sucinta y sustanciosa panorámica de sus publicaciones y una fina interpretación crítica de los aspectos psicológicos de su obra puede encontrarse en Mayor (1993).

Distingue Simon (1980, 1981; Simon y Kaplan, 1989; Ericsson y Simon, 1984) entre *ciencia cognitiva* y *psicología cognitiva*. Aquélla tiene como objeto elaborar una «teoría general de los sistemas inteligentes», naturales y artificiales. Comprende la inteligencia artificial, la robótica, los sistemas de solución general de problemas y los sistemas expertos específicos, los sistemas hombre-máquina, la interpretación de la competencia lingüística, la simulación de escenarios y experimentos en todas las ciencias y la búsqueda de las invariantes comunes a todos los sistemas inteligentes. La psicología cognitiva tiene por objeto simular los procesos mentales de la conducta humana, para explicarla y comprenderla.

Respecto a la ciencia cognitiva, Simon propone las siguientes invariantes, características de todos los sistemas inteligentes.

Primera. *Flexibilidad*. Todos son sistemas flexibles, que tratan de responder a las diferentes y continuamente cambiantes exigencias del medio. Los sistemas naturales, para adaptar el organismo a las demandas del ambiente y éste a las necesidades del organismo. Los sistemas artificiales, para responder a las demandas de la tarea o problema planteado, con los datos e instrucciones de que disponen.

Segunda. *Simbolismo*. Todos los sistemas inteligentes tienen como componentes principales a los símbolos, entendidos como representaciones.

Tercera. *Fisicidad*. Los sistemas inteligentes sustentan los símbolos en fundamentos físicos, por ejemplo, orgánicos o electrónicos.

Cuarta. *Informatividad*. Según la ingeniosa y justa denominación de George Miller (1984), los sistemas inteligentes son *informávoros*: se alimentan de información tanto interna como externa, que codifican y simbolizan. Los naturales, a partir de la percepción sensorial y la recuperación de procesos y datos de la memoria; los artificiales, a partir de la base de datos y reglas del programa, presentes en cada momento o recuperados de sus variadas memorias. Cada vez de forma más amplia, los sistemas artificiales, sobre todo en la robótica, utilizan para captar información la actividad de *sensores* y la retroalimentación de mecanismos *efectores*.

Quinta. *Operatividad*. Todos los sistemas inteligentes operan con símbolos y, mediante la computación de relaciones entre ellos, planean y dirigen la ejecución de las respuestas.

Sexta. *Computabilidad serial*. Todos emplean básicamente los mismos procesos operativos con los símbolos. El proceso de computación es fundamentalmente serial y sucesivo y, en escasa medida, simultáneo y paralelo. Ello se debe a los límites de la memoria de trabajo, que, por ejemplo en el hombre, sólo puede abarcar unas 7 ± 2 unidades significativas o *chunks*.

Séptima. *Búsqueda heurística*. La principal invariante respecto a las demandas de la tarea que hay que emprender o al problema que hay que resolver es la búsqueda heurística (*the heuristic search*). Una vez especificada la tarea y el espacio del problema, con su situación actual y su situación final

o solución, el sistema busca y tantea vías que acerquen la una a la otra, usando información conservada en la memoria y extrayendo nuevas indicaciones que puedan guiar la búsqueda, en el medio interno o externo. Esta búsqueda puede ser estrictamente algorítmica y lógica, cuando un procedimiento computacional determinado conduce con seguridad a la solución de la tarea o problema o, lo que es más frecuente, mixta algorítmico-heurística, en la que se emplea un cálculo lógico-matemático, pero acompañado de una búsqueda tanteante, que evita la explosión combinatoria que se produciría si se exploraran exhaustivamente todas las alternativas posibles, como acontece, por ejemplo, en el juego del ajedrez.

En los últimos años se progresa a ritmo acelerado en la invención de técnicas de inteligencia artificial, tanto en el análisis de estrategias generales para formular cuestiones y plantear problemas, como en el estudio de sistemas de producción y de *elección de medios y fines*, para contestarlas y resolverlos. Se avanza asimismo en la investigación de la inteligencia natural mediante la *simulación*, en programas de ordenador, de la conducta inteligente del hombre y de su capacidad de aprendizaje.

En el futuro, piensa Simon, la indagación habrá de dedicar más esfuerzo e ingenio para simular fielmente los sistemas inteligentes humanos y, sobre todo, las reorganizaciones de los sistemas cognitivos debidas a recursos e innovaciones sociales y culturales. Cada vez está más claro que el funcionamiento de los sistemas inteligentes humanos depende indudablemente de capacidades y limitaciones fisiológicas, pero, sobre todo y más profundamente, de condiciones culturales y sociales.

La ciencia cognitiva debe coordinar los objetivos *descriptivos* y los *normativos*, es decir, la descripción, cada vez más precisa, de los procesos neogenéticos, que conducen a la ampliación del conocimiento, y la especificación rigurosa de las reglas lógicas, que conducen al razonamiento correcto. La meta final, según Simon, es la *optimización normativa*, en la que los sistemas artificiales aprendan el funcionamiento flexible e innovador de los sistemas naturales y los naturales aprendan la precisión lógica de los artificiales.

La investigación de los sistemas inteligentes naturales y, principalmente del hombre, es, como dije, el objeto de la psicología cognitiva. No se interesa ésta directamente por los procesos generales que contribuyen a la solución de tareas y problemas, sino por los que efectivamente acontecen en la mente humana.

Para avanzar hacia su objetivo, la psicología cognitiva debe dar, en sus investigaciones, tres pasos sucesivos: *la hipótesis, la simulación y la verificación*.

La hipótesis. Ante una tarea o problema, el psicólogo cognitivo formula una hipótesis acerca de los componentes, procesos y estrategias que utiliza la mente humana. Concreta finalmente esa hipótesis en forma de programa de ordenador.

La simulación. El investigador pone en funcionamiento el programa en un ordenador y comprueba si el ordenador soluciona o no el problema. Si lo hace, queda empíricamente demostrado que el programa y, en consecuencia, la hipótesis, son válidos.

La verificación. Consiste en comprobar si el programa, además de solucionar el problema, lo hace como el sujeto humano.

Mediante diseños experimentales complejos, en los que suelen figurar como variables dependientes los errores y las latencias, se comparan paso a paso las operaciones del ordenador con los protocolos verbales obtenidos de los sujetos, que hablan mientras piensan en voz alta (*the thinking aloud technique*). Se infiere así, con cierto grado contrastable de probabilidad, si los datos a que el sujeto atiende, las relaciones que percibe, las operaciones que hace, los símbolos que codifica y maneja y las estrategias, submetas y metas que consigue corresponden, o no, a los componentes y pasos del programa del ordenador.

La conclusión final de la psicología cognitiva clásica que se ajusta al paradigma de computación de símbolos es que el organismo humano y, más particularmente, el cerebro, al igual que el ordenador, son casos especiales de la máquina de Turing, funcionalmente equivalentes a ella. Operan sobre unos componentes simbólicos y discretos, mediante ciertos principios sintácticos. El que difieran en su soporte físico es irrelevante. La metáfora del ordenador se ajusta literalmente a la estructura y el funcionamiento de la mente.

Los datos, resultados empíricos y argumentos de los psicólogos cognitivos constituyen un fuerte apoyo a la tesis de este paradigma. Incluso Simon, en 1990, llega a afirmar: «La hipótesis del sistema físico simbólico ha sido comprobada tan ampliamente en los últimos 30 años que puede considerarse ya como completamente aceptada» (Holyoak y Spellman, 1993, p. 206).

Sin embargo, hay que añadir enseguida que la tesis adolece de considerables deficiencias y limitaciones. Creo que las principales son las siguientes (vid. *Anuario de Psicología*, 1993; Johnson-Laird, 1988; Marina, 1993; Rivière, 1991; Rumelhart *et al.*, 1986, con la competente introducción a la versión española de García Madruga, 1992).

Primera. La máquina de Turing dispone de una cinta (una memoria) ilimitada. Los ordenadores digitales, que son sus realizaciones físicas, van utilizando memorias de trabajo cada vez más amplias. La velocidad de procesamiento de los ordenadores va acercándose a los nano y picosegundos. Sin embargo, la atención consciente humana no puede ejercerse más que sobre un pequeño número de unidades o *chunks*, de 5 a 9, como queda dicho. Y la rapidez de los impulsos nerviosos se mide en centésimas o milésimas de segundo. No hay, pues, tiempo material para que el organismo humano pueda realizar secuencialmente, uno después de otro, los cálculos y procesos que la teoría cognitiva clásica reclama. Es preciso admitir procesamientos distribuidos en paralelo, que se realicen sin gastar recursos de atención consciente.

Segunda. La mente humana, a diferencia de la máquina de Turing, es capaz de tratar con eficacia información incompleta y degradada. Si se habla, por ejemplo, de un artista de Hollywood que fue un eminente erudito y llegó a ser Presidente de la mayor potencia mundial, se comprende que se habla de Reagan, aunque se dude o niegue su afirmada erudición.

Tercera. El procesamiento de la información no sólo es, en buena parte, paralelo, sino, además, interactivo. En la comprensión de frases, el significado

de las primeras palabras suele depender de la comprensión de las últimas. Es verdad que en la comprensión y en la producción del lenguaje operan procesos sucesivos y secuenciales, pero los análisis léxicos, sintácticos y semánticos actúan en parte de forma simultánea.

Cuarta. El paradigma de la computación serial de símbolos discretos mediante la aplicación de reglas estrictas supone un sistema mental rígido y poco adaptable, como reconoce Simon, a las cambiantes demandas de las tareas. Dificilmente puede dar razón de procesos tan patentes y ubicuos como los de aprendizaje y desarrollo, realizados, en gran parte, por tanteos, avances y retrocesos. No es extraño que los defensores del paradigma cognitivo clásico, como Fodor o Pylyshyn, propendan a admitir el predominio en la estructura mental básica de componentes y estrategias innatas, con características referencias, desde los primeros trabajos de Chomsky, al innatismo de Descartes.

En realidad este tipo de explicaciones cognitivas se refieren más a la *competencia (competence)* que a la *ejecución (performance)*, es decir, más a la arquitectura supuestamente subyacente al sistema de conocimiento y de lenguaje que a la concreta adquisición y desarrollo del mismo. Lo que provoca problemas a veces insolubles en las ciencias empíricas del lenguaje y el pensamiento, como la psicología.

Sexta. El sistema de computación simbólica, en la medida en que sea aceptable, constituirá, a lo sumo, una nueva y poderosa lógica formal del conocimiento, más bien que una explicación psicológica del mismo.

Séptima. Y tal vez, a mi juicio, la principal. En este sistema, los componentes conscientes corren el riesgo de ser interpretados como epifenómenos subjetivamente importantes, pero causalmente ineficientes. Serían producto de la computación, y no tendrían ningún papel en la cadena causal de la conducta. La efectiva causalidad se debe a la sintaxis de las clases de equivalencia de las señales físicas, que determinan la significación o semántica de los símbolos. Ahora bien, hay razones de peso, como veremos, para admitir que la semántica de los símbolos mentales no se reduce a las relaciones sintácticas entre ellos y que la dimensión consciente desempeña un papel no sólo eficiente sino primordial en el dinamismo mental de la inteligencia humana.

Estas y otras dificultades y objeciones han originado un nuevo enfoque de la psicología cognitiva, que empieza a predominar en la última década. Es el llamado paradigma PDP: *procesamiento distribuido en paralelo o conexionismo*.

Procuraré sintetizar sus características y logros, así como, a mi parecer, sus particulares insuficiencias.

El paradigma PDP

Las limitaciones de la psicología cognitiva clásica provienen de tomar al pie de la letra la metáfora del ordenador, según el modelo universal de la máquina de Turing. Tal orientación puede, en principio, fundamentar una poderosa tecnología de la inteligencia artificial, como en efecto va haciendo. No parece seguro que pueda avanzar en el conocimiento de la psicología de

los procesos inteligentes reales, pues éstos acontecen en seres vivos, en los que la acción inteligente se desarrolla sobre un soporte biológico concreto, que no se puede desatender, so pena de caer en las limitaciones supradichas.

Surgen, por eso, en los últimos años, múltiples intentos de sustituir el modelo de Turing por otro nuevo en el que el procesamiento de la información se realiza por redes complejas de unidades interconectadas, con propiedades similares de activación e inhibición a las conocidas de las neuronas y que operan simultánea y paralelamente, de forma aproximada y redundante.

Estos nuevos modelos constan de unidades de entrada, de salida e intermedias, conectadas entre sí, que actúan cuando reciben impulsos por encima de su umbral y disparan o inhiben la actividad de otras unidades. Cada unidad computa una parte de muchos símbolos y mensajes y cada símbolo y mensaje está procesado por múltiples unidades.

En cada momento, la red tiene un cierto grado de *activación* general y una *fuerza de conexión* entre sus unidades. La distribución de conexiones corresponde al conocimiento del sistema en un momento dado. La configuración de activaciones corresponde al tratamiento de la nueva información. La multiplicidad de procesamiento de los diversos componentes de cada información y, finalmente, de cada símbolo significativo, permite el aprendizaje paulatino o rápido, la comprensión de información deficiente y la transformación del patrón general de conexiones y activaciones en nuevas estructuras de equilibrio complejo, según las leyes termodinámicas de Boltzmann y el dinamismo de las estructuras disipativas de Prigogine.

En términos más sencillos, el modelo conexionista se aproxima a la teoría de la *adaptación inteligente* de Piaget. Mediante el influjo que ejerce el patrón de conexiones existente sobre la información que llega, se explica la *asimilación* de las demandas de la tarea a los esquemas previos del sujeto. Mediante el nuevo patrón de activaciones que provoca la nueva información, se explica el proceso de *acomodación* de los esquemas del sujeto a las demandas de la tarea. El resultado final puede corresponder a la *adaptación* progresiva de la conducta inteligente del sujeto a las exigencias del medio y a la modificación del medio por el sujeto, para responder a sus demandas y proyectos.

Los símbolos conscientes se consideran en este enfoque como propiedades emergentes de los estados globales de equilibrio de la red de unidades. Explican, con algún mayor rigor experimental y matemático (según las leyes y ecuaciones que gobiernan las redes y las dirigen hacia estados globales estacionarios), el fenómeno de la «reestructuración» y la experiencia de la «comprensión súbita», que describió la escuela de la Gestalt, la *Aha! Erlebnis*, de Karl Bühler, o el diálogo interno consigo mismo que han examinado tantas teorías psicológicas, como las clásicas del *verbum mentis* y la más reciente de Vigotsky sobre los titubeos mentales.

Muchos de los resultados funcionales que procuran los mecanismos contruidos y regidos según este modelo, se asemejan a fenómenos capitales verificados por la psicología experimental, como los ya mencionados del aprendizaje lento y con errores y tanteos, la comprensión de información degradada,

la intelección de textos y la solución de problemas que requieren la interacción simultánea de procesos distribuidos, paralelos e interactivos, y la elaboración por el sujeto de categorías borrosas.

Los modelos PDP se apoyan en numerosos antecedentes tanto neurológicos como computacionales. Neurólogos como Jackson, en el siglo XIX y Luria, en el XX, demostraron la localización distribuida y múltiple de las funciones cognitivas, y Lashley, no hace mucho, descubrió la considerable equipotencialidad de grandes zonas de la corteza cerebral. Mi primer trabajo experimental demostró el desplazamiento de la percepción de la dirección del sonido desde las zonas temporales a otros centros vicariantes, cuando se impide o perturba la actividad de los anteriores (Neff, Diamond y Yela, 1956).

En realidad, la revolución cognitiva se inició con investigaciones concretas que se apoyaban en hipótesis de tipo distribuido y paralelo, como los trabajos de McCulloch y Pitts (1943), los ensayos del *perceptron*, de Rosenblatt (1962) y los intentos del *logogen*, de Morton (1969). La investigación inicial de McCulloch y Pitts demostró que las redes neuronales pueden calcular propiedades lógicas y dar cuenta del poder computacional del cerebro. Por ejemplo, una «neurona artificial» conectada con otras dos y con un umbral de excitación que sólo se supera si recibe excitación de cualquiera de ellas o de ambas, computa la «disyunción no excluyente» del cálculo de proposiciones («o p o g o ambas»). Si el umbral requiere la suma de las excitaciones de las dos, computa la «conjunción» (p y q).

El desarrollo directo y sistemático del nuevo enfoque PDP se inicia con los estudios de Hinton y Anderson (1981) y se expone con amplitud en la obra colectiva de Rumelhart, McClelland y el PDP *Research Group* (1986), traducida parcialmente al español en 1992 y, como advertí, con una esclarecedora introducción de García Madruga.

Una exposición crítica de los diversos modelos computacionales y sus limitaciones se expone en el espléndido ensayo de Marina (1993) y en la obra, más atendida a los datos experimentales y a la terminología computacional, de Rivière (1991), así como en los comentarios a la misma que varios psicólogos españoles ofrecemos en el número 56, de 1993, del *Anuario de Psicología* de la Universidad de Barcelona.

En realidad, sin embargo, el paradigma PDP, en sus muy diversas formulaciones, no parece que pueda, de momento, sustituir por entero al de computación de símbolos. La mayor parte de la copiosísima y creciente bibliografía sobre la psicología cognitiva se atiene hoy al enfoque general del *procesamiento de la información*, y recurre, cuando parece pertinente a uno u otro paradigma, o, cuando es preciso, prescinde de los dos.

La razón es que también el modelo PDP adolece de profundas insuficiencias y limitaciones. Las sintetizaré en las siguientes.

Primera. El modelo de redes de unidades abstractas interconectadas y los concretos mecanismos construidos según sus principios, distan todavía mucho del cerebro real. La elaboración de modelos conexionistas es muy reciente y está en continua transformación y crecimiento. Es pronto para decidir. Como señala Manuel de Vega en su *Visita al país de Turing* (1993), hace falta

rellenar de contenido psicológico y fisiológico los modelos conexionistas, que están todavía lejos de corresponder a las propiedades conocidas del funcionamiento cerebral.

Segunda. La actividad mental tiene obvias características seriales manifestadas por la computación de símbolos enteros en muchos procesos de memoria, lenguaje, razonamiento y solución de problemas, que exhiben una clara estructura temporal, no explicada por el procesamiento distribuido y paralelo.

Tercera. El número de unidades intermedias es decisivo para el funcionamiento de los modelos PDP, pero no responde a ninguna ley sistemática derivable de la teoría. Cada micromodelo conexionista lo define *ad hoc*.

Cuarta. No está claro, en estos modelos, cómo acontece la codificación de entrada, es decir, cuáles son los micro-rasgos subsimbólicos que es necesario procesar.

Quinta. El procesamiento mental es *sistemático*. Por ejemplo, si se puede correctamente pensar y decir «Pedro ama a Lola», es necesario que también sea posible y computacional y gramaticalmente hacedero decir y pensar que «Lola ama a Pedro». El paradigma PDP no explica esta sistematicidad.

Sexta. Los modelos PDP no dan cuenta del aprendizaje en un solo ensayo, que es perfectamente explicado por la comprensión consciente de las relaciones entre símbolos discretos. Por ejemplo, basta que se comprenda que un número primo es todo aquel que sólo es divisible por sí mismo y por la unidad para que se haya aprendido de una vez para siempre lo que es un número primo. Basta que me informen de que todos los números de teléfono van a empezar por 5, para que, en adelante, sin tanteos ni procesamientos de microsignos, lo comprenda y retenga.

Séptima. Los modelos PDP no dan cuenta del control metacognitivo (Mayor et al., 1993), es decir, del influjo que en los procesos de inteligencia tiene la comprensión de las actividades inteligentes, de las que el sujeto puede ser consciente y, por eso, ser capaz, al menos en parte, de indagar, modificar y acaso mejorar.

Octava. Estos modelos no explican por qué se retiene lo aprendido y no desaparece en los nuevos patrones de activación y conexión, si bien son más idóneos para dar razón de las múltiples vías y pistas que a veces se siguen para recuperar de la memoria un conocimiento que se resiste a ser recordado.

Novena. La objeción principal es que estos modelos postulan, sin explicarla, la emergencia de símbolos conscientes, efecto resultante de la actividad neural y que podrían interpretarse como simples epifenómenos de la conducta inteligente. Ahora bien, el que los símbolos conscientes emerjan de la actividad neural, no implica necesariamente que carezcan de eficacia causal, como veremos.

El paradigma general del procesamiento de la información

Por todo lo dicho, no parece que el paradigma PDP pueda sustituir sin residuo al modelo C-S o de computación de símbolos. Más bien, si acaso,

podrá complementarlo. Es lo que piensan la mayoría de los psicólogos cognitivos que actualmente realizan investigaciones empíricas y experimentales sobre la estructura y funcionamiento de la inteligencia. Tal vez el psicólogo más sistemático y representativo de este enfoque ecléctico sea Philip Johnson-Laird, del que merece especial atención su obra *The Computer and the Mind*, de 1988.

Parece claro que la inteligencia, como sistema cognitivo, opera en ciertas fases, sobre todo en las iniciales de la acción, mediante procesos paralelos, automáticos e inconscientes. Estos procesos, sin embargo, contienen aspectos interpretativos, que confieren significación al contenido procesado y lo preparan para que, finalmente, se articule en forma de símbolos conscientes. Se ha comprobado, por ejemplo, en los experimentos dicóticos, que la información que se recibe por el oído no atendido influye inconscientemente en el significado de la información captada por el oído atendido. Más simplemente, es lo que acontece en una fiesta. En ella hay un fondo de ruidos y palabras al que no atendemos, mientras fijamos quizás nuestra atención en el diálogo con un amigo. Basta, sin embargo, que en algún lugar se mencione nuestro nombre para que inmediatamente desplacemos hacia ese lugar nuestra atención. Hasta entonces no lo atendíamos, pero de alguna manera era considerado, si bien de forma automática e inconsciente, por nuestra actividad cognitiva. Esta lo estimaba carente de interés y, por eso, no lo atendía, pero, ahora, súbitamente, se percata del cambio ocurrido en esa zona no atendida, lo interpreta como interesante y dirige hacia ella la atención. Todo parece indicar que hay operaciones y componentes mentales significativos, sin que de ellos tengamos conciencia plena o incluso carezcamos completamente de ella. La acción inteligente completa requiere, sin embargo, el tratamiento consciente de símbolos, como cuando atendemos deliberadamente a la significación del diálogo con el amigo o a la conversación en la que, de pronto, ha sonado nuestro nombre.

Una de las propiedades características de la acción inteligente es que en ella, además de otros posibles componentes y procesos automáticos e inconscientes, acontece una relación mental entre un objeto o situación que significa algo para un sujeto y un sujeto para el que ese algo es significado. Esa relación significativa, aunque se incluya en una imprescindible actividad física y abarque aspectos interpretativos automáticos, es de carácter *intencional*; es decir, manifiesta la presencia de un objeto, no en su realidad física, sino en tanto que *ob-jectum*; es decir, en tanto que sale al encuentro del sujeto, como término de un conocimiento en el acto mental de percibirlo, recordarlo, concebirlo o entenderlo. Esa relación intencional es lo que llamamos conciencia, la cual incluye la existencia de símbolos semánticos o significativos en el procesamiento mental.

El uso del término «símbolo» es multívoco en la literatura cognitiva. Su aplicación es a veces incorrecta. Todas las computaciones sintácticas entre los componentes llamados «símbolos» son incapaces de producir un símbolo auténtico. Símbolo es algo que representa otra cosa. Esa referencia representativa es lo que constituye su semántica. Un símbolo sin referencia semántica no es propiamente un símbolo. Los partidarios de la metáfora del ordenador,

o del cerebro, pretenden reducir la semántica a sintaxis. A mi entender, no lo consiguen. Cometten, como dice Johnson-Laird, la *symbolic fallacy*. Las relaciones y transformaciones sintácticas, formales, lógicas o matemáticas, cambian un conjunto de elementos en otro, pero no le confieren significación simbólica. Es como si, para entender un texto escrito en un idioma desconocido, se nos presenta traducido a otro que tampoco entendemos. Sigue sin significar nada para nosotros.

El símbolo tiene una intrínseca referencia a algo distinto de él. Si no, no es un símbolo, sino una mera señal. En el ordenador tipo Turing o en la máquina tipo Boltzmann, el valor referente lo suministra el código impuesto por el inventor del programa. En el sistema inteligente humano la referencia de los símbolos mentales al objeto que representan constituye su intrínseca intencionalidad.

La cual se manifiesta de muy distinta manera. Por ejemplo, como *intención*. Para comprender una frase no basta comprender sus componentes gramaticales y la unidad significativa que en ellos se sustenta; es preciso, además, comprender la intención del que la dice. La expresión «¿puede usted abrir la puerta?», no suele reclamar la contestación «sí puedo», suele más bien significar el ruego de que se abra. Son los componentes de los actos del lenguaje que Austin ha llamado ilocutivos los que modulan el significado final y pragmático de las frases. Lo hacen por muy variados medios perilingüísticos, como la entonación, el gesto o la agregación de sintagmas consabidos, como «por favor». Determinan que la frase signifique ruego, pregunta, orden, súplica, consejo, etc.

Los resultados funcionales de las acciones inteligentes pueden ser simulados por autómatas diversos, como ordenadores o robots. Y los técnicos de sistemas de inteligencia artificial lo van logrando cada vez con mayor rigor y amplitud. La «intención», por ejemplo, en el sentido que queda dicho, puede controlarse en una conversación entre robots, siempre que se programe exacta y explícitamente. Un robot estará programado para decir: «¿Puedo hacerle una pregunta?». El otro responderá: «Sí». El primero dirá entonces: «¿Puede usted abrir la puerta?». A lo que el segundo contestará correctamente: «Sí». O bien, si el programa dicta que el robot diga: «¿Puedo hacerle un ruego?», y el otro contesta que «Sí», al recibir el segundo mensaje «¿Puede usted abrir la puerta?», el otro responderá, no con una palabra, sino abriendo efectivamente la dichosa puerta.

Según los datos que van aportando la psiconeurología y las ciencias cognitivas, no parece imposible que se consigan simular *todos* o *casi todos* los resultados a que llega la conducta inteligente humana. Si, como se va logrando, se programan en un autómata sistemas operativos en una jerarquía en la que los superiores puedan poner en marcha a los inferiores y éstos ejecutar, bajo la vigilancia de aquéllos, los procesos para los que están programados y de los que los sistemas superiores no tienen noticia ni habilidad para realizarlos, aunque sí para ponerlos en marcha, evaluar su rendimiento y pararlos, entonces se conseguirá simular muchos de los resultados que en el hombre logra la operación de la autorregulación y la autoconciencia. Si se

suministra al autómatas un «modelo del mundo» suficientemente rico y preciso, y los sensores y efectores adecuados, se puede simular la formación de proyectos realizables en ese mundo y la elección de rutas y formas de actuación que imiten la elección libre.

Disponemos ya de autómatas que consiguen muchos de esos resultados, como demostrar ciertos teoremas matemáticos con mayor parsimonia y elegancia que Whitehead y Russell, acertar más pronto y mejor que un especialista el diagnóstico de una patología o elegir y emplear las herramientas apropiadas más eficazmente que un obrero. Claro que eso no quiere decir, como señala Johnson-Laird, que el autómatas entienda lo que hace ni más ni mejor que la molécula de ADN, que dirige la formación del organismo, entiende la ciencia genética.

Los diversos enfoques de las ciencias y psicologías cognitivas han supuesto un gigantesco avance en la indagación de la estructura formal del conocimiento, en la construcción de sistemas inteligentes artificiales y en la inauguración de una fase nueva e inédita en la metodología científica, a saber, la posibilidad de simular experimentos y fenómenos que hasta ahora eran irrealizables o inobservables. Abren asimismo nuevas y fecundas vías para el estudio de la mente y la inteligencia humanas.

No dejan, sin embargo, de ofrecer dificultades y limitaciones. Para terminar, yo las compendiaría en las siguientes.

Primera. Las metáforas del ordenador, del cerebro y del procesamiento de la información, son eso: metáforas. Han resultado extraordinariamente fecundas, pero si se toman al pie de la letra conducen, como advirtió Alejandro Luria, a reduccionismos, paralogismos y falacias.

Segunda. Hay dos tipos de intencionalidad: la extrínseca y la intrínseca. La extrínseca es la que impone el programador a los datos que introduce en su programa, según el código convenido. Es una intencionalidad simulada. La intrínseca consiste en la referencia de los símbolos mentales del sistema cognitivo humano a los objetos que representan. Es una intencionalidad auténtica.

Tercera. Existe otra acepción del término intencionalidad, que alude a la intención o propósito. Ambos pueden simularse, hasta límites crecientes cuya amplitud no sabemos precisar, en los sistemas automáticos de computación. Pero en el caso del hombre, a diferencia de lo que sucede en los autómatas, las intenciones y los propósitos hunden sus raíces en las metas que adopta el sujeto, según sus motivaciones y sentimientos, profundamente modulados por su realidad psicosomática y la sociedad y cultura en la que nace y vive.

Cuarta. En los ordenadores y robots hay la posibilidad, todavía apenas actualizada, pero prometedora, de incrementar los procesos de aprendizaje. En los sistemas inteligentes naturales no sólo es obvia la importancia del aprendizaje, sino el hecho más decisivo de la constante modulación del *software* por el *hardware* y viceversa. No sólo los patrones de activación y conexión que se producen en las sinapsis neuronales facilitan nuevas vías de transmisión de impulsos, nuevas actualizaciones de genes y proteínas, una mayor abundancia de botones sinápticos y una transformación de las tasas de neurotrans-

misores, sino que todos estos fenómenos alteran, perfeccionándolo o perturbándolo, el programa de procesamiento mental. Este, a su vez, modifica la estructura de las redes neuronales. Así, a diferencia de los autómatas computadores, se establece en el hombre una intermodulación permanente entre las bases somáticas y las dimensiones mentales de su comportamiento, ambos componentes, el mental y el somático, integrados en la estructura psicoorgánica unitaria de toda la actividad humana.

Quinta. Los sistemas de inteligencia artificial son, en el mejor de los casos, solucionadores de problemas. Lo más característico del sistema inteligente humano es que es también creador e inventor de cuestiones, preguntas y problemas, como muestra con rigor e ingenio José Antonio Marina en su ensayo *Teoría de la inteligencia creadora* (1993): la inteligencia humana es una especie de computador automático en el que ha irrumpido la conciencia intencional, el proyecto y la libertad. El problema capital que es capaz de formular, aunque tal vez no de resolver, consiste en hacerse cuestión de sí mismo. *Quaestio mihi factus sum*, escribió San Agustín.

Sexta. La inteligencia humana no sólo requiere, según argumentan muchos psicólogos cognitivos, como Johnson-Laird, un sistema operativo central, que origine el concepto de un yo propio, sino que descubre en ese sí mismo dos facetas ausentes de los autómatas: la soledad y la intersubjetividad. El yo cognoscente y proyectante se descubre en soledad frente a todo lo conocido y cognoscible, ante todo lo proyectado y proyectable: de un lado, el yo que piensa y proyecta; de otro, toda la realidad efectiva o presunta.

Por otra parte, el yo se percibe habitado por múltiples voces, como sutilmente describe Rivière en su artículo «Las multitudes de la mente» (1993). En nuestra mente, la de cada uno, resuenan, desde su irremediable soledad, las voces incontables de sus prójimos y la de los usos, costumbres e interpretaciones que le brindan o imponen los anónimos forjadores de la sociedad y cultura en las que piensa y decide.

Séptima. El sistema inteligente humano, a diferencia del artificial, no sólo puede formular y resolver cuestiones y problemas; puede también aparentar razones, manifestar respuestas astutas o falaces, proponerse la búsqueda de la ironía, lo insensato y el absurdo, intentar hacer patente, por el lenguaje poético y la obra de arte, lo que hay de latente en la realidad.

Octava. Finalmente, creo que la razón fundamental para rechazar la identificación de los sistemas inteligentes artificiales y humanos es que la conducta inteligente del hombre manifiesta, como su propiedad más distintiva, lo que he llamado *metaconducta*. Toda conducta es respuesta a una situación. La conducta humana es, además, respuesta a y de su respuesta; es decir, puede ser y, en parte, tiene que ser, conducta sobre su propia conducta (Yela, 1992). El hombre se manifiesta, así, en alguna medida, responsable de su conducta. Lo cual confiere a la persona una *dignidad* que la distingue de todas las máquinas que puedan simular o superar los resultados de su actividad inteligente (Yela, 1987 c).

REFERENCIAS

- Aebli, H. (1980-1981). *Denken: Das Ordnen des Tuns*. Dos volúmenes. Stuttgart: Klett-Cotta.
- Anuario de Psicología* (1993). Núm. 56. Universidad de Barcelona.
- Bruner, J.S., Goodnow, J.J. & Austin, G.A. (1956). *A Study of Thinking*. New York: Wiley.
- Carroll, J.B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.
- Chomsky, N. (1956). Three models of the description of language. *IRE Transactions on Information Theory*, 3, 113-126.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton. (*Estructuras sintácticas* 1974. Madrid: Siglo XXI).
- Demetriou, A. & Efklides, A. (1994). *Intelligence, Mind and Reasoning*. Amsterdam: North Holland.
- Ericsson, K.A. & Simon, H.A. (1984). *Protocol Analysis*. Cambridge: Cambridge University Press.
- Fodor, J. (1968). *Psychological explanation: An introduction to the Philosophy of Psychology*. New York: Random House. (*La explicación psicológica*. 1980, Madrid: Cátedra).
- Fodor, J. (1975). *The language of thought*. New York: Harper and Row. (*El lenguaje del pensamiento*. 1984. Madrid: Alianza).
- Fodor, J. (1981). *Representations*. Cambridge: The MIT Press.
- Fodor, J. (1983). *The modularity of mind*. Cambridge: MIT Bradford Books. (*La modularidad de la mente*, 1986, Madrid: Morata).
- Fodor, J. (1988). *Psychosemantics. The Problem of meaning in the philosophy of mind*. Cambridge: The MIT Press.
- Fodor, J. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28, 3-71.
- Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.
- Hebb, D.O. (1960). The american revolution. *American Psychologist*, 15, 735-745.
- Hinton, G.A. & Anderson, J.A. (1981). *Parallel models of associative memory*. Hillsdale, New York: LEA.
- Holyoax, K.J. & Spellman, B.A. (1993). Thinking. *Ann. Review of Psychol.*, 44, 265-315.
- Johnson-Laird, P.N. (1988). *The computer and the mind: an introduction to cognitive science*. Cambridge: Harvard University Press. (*El ordenador y la mente*, 1990, Barcelona. Paidós).
- Journal of Educational Psychology*. Intelligence and its measurement: A symposium, 1921, núm. 12.
- Marina, J.A. (1993). *Teoría de la inteligencia creadora*. Barcelona: Anagrama.
- Martínez Arias, R. y Yela, M. (Eds., 1991). *Pensamiento e Inteligencia*. Vol. 5 del *Tratado de Psicología General de Mayor, J. y Pinillos, J.L.* (Eds.). Madrid: Alhambra.
- Mayor, J. (1993). El solucionador general de problemas de H.A. Simon. En Quiñones, E., Tortosa, F. y Carpintero, H. *Historia de la Psicología. Textos y Comentarios*. 574-582. Madrid: Tecnos.
- Mayor, J., Suengas, A. y González Marqués, J. (1993). *Estructuras metacognitivas*. Madrid: Síntesis.
- McCulloch, W. & Pitts, W. (1943). *A logical calculus of ideas immanent in nervous activity*. *Bulletin Mathemat. Biophysics*, 5, 115-155.
- Miller, G.A., (1956). *The magical number seven, plus or minus two; some limits on our capacity for processing information*. *Psychological Review*, 63, 81-97.
- Miller, G.A. (1984). Informávoros. En Machlup, F. y Mansfield, V. (Eds.) *The study of information: interdisciplinary messages*. N.Y.: Wiley.
- Miller, G.A., Galanter, E. & Pribram, K.H. (1960). *Plans and the structure of behavior*. New York: Holt. (*Planes y la estructura de la conducta*, 1983. Madrid: Debate).
- Montpellier, G. de (1977). *Qu'est-ce que l'intelligence?* Bruxelles: Academie Royale de Belgique.
- Morton, J. (1969). *Interaction of information in word recognition*. *Psychological Review*, 76, 165-178.
- Neff, W.D., Diamond, J.F. & Yela, M. (1956). *Role of the auditory cortex in discrimination requiring localization of sound in space*. *Journal of Neurophysiology*, 19, 500-512.
- Neumann, J. von (1958). *The computer and the brain*. New Haven: Yale University Press.
- Newell, A. & Simon, H.A. (1956). *The logic theory machine*. *IRE Transactions on Information Theory*, 3, 61-79.
- Pylyshyn, Z.W. (1984). *Computation and cognition. Toward a foundation of cognitive science*. Cambridge: MIT Press. (*Computación y conocimiento*. 1988. Madrid: Debate).
- Resnick, L.B. (Ed. 1976). *The nature of intelligence*. Hillsdale, New York: LEA.
- Rivière, A. (1991). *Objetos con mente*. Madrid: Alianza.
- Rivière, A. (1993). Las multitudes de la mente. *Anuario de Psicología*, 56, 112-144.
- Rosenblatt, F. (1962). *Principles of Neurodynamics*. New York: Spartan.

- Rumelhart, D.E., McClelland, J.L. & PDP Research Group. (1986). *Parallel Distributed Processing. Explorations in the microstructure of cognition*. Cambridge: The MIT Press. (Introducción al Procesamiento Distribuido en Paralelo. 1992. Madrid: Alianza).
- Searle, J.R. (1992). *The rediscovery of the mind*. Cambridge: MIT Press.
- Shannon, C.E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: Univ. of Illinois Press.
- Simon, H.A. (1980). Cognitive science: The newest science of the artificial. *Cognitive Science*, 4, 33-46.
- Simon, H.A. (1981). Studying human intelligence by creating artificial intelligence. *American Scientist*, 69, 3, 300-309.
- Simon, H.A. & Kaplan, C.A. (1989). *Foundations of cognitive science*. In Posner, M.L. (Ed.) *Foundations of cognitive science*. Cambridge: The MIT Press.
- Sternberg, R.J. & Detterman, D.K. (Eds., 1986). *What is intelligence, Contemporary view-points on its nature and definition*. Norwood: Ablex. (¿Qué es la inteligencia?, 1988, Madrid: Pirámide).
- Sternberg, R.J. (Ed., 1982, 1988). *Handbook of human intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (1982-1988). *Advances in the psychology of human Intelligence*. Hillsdale, New York: LEA. Vol. 1, 1982; 2, 1985; 3, 1986; 4, 1988.
- Turing, A.M. (1936). On computable numbers. *Proceeding of the London Mathematical Society*, 42, 230-265.
- Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460. Reeditado en Anderson, A.R. (Ed.), 1984. *Controversia sobre mentes y máquinas*. Barcelona: Tusquets.
- Vega, M. de (1993). Visita al país de Turing. ¿Un viaje de ida y vuelta? *Anuario de Psicología*, 56, 102-106.
- Wiener, N. (1948). *Cybernetics*. New York: Wiley.
- Yela, M. (1956). *Psicología de las aptitudes*. Madrid: Gredos.
- Yela, M. (1981a). Ambiente, herencia y conducta. En Jiménez Burillo, F. (Ed.). *Psicología y Medio*. Madrid: MOPU. 69-104.
- Yela, M. (1981b). El progreso de la inteligencia: evolución biológica y desarrollo cultural. *Anales de la Real Academia de Ciencias Morales y Políticas*, 33, 58, 29-60.
- Yela, M. (1982). Inteligencia, estructuras ontogenéticas y dimensiones factoriales. *Revista de Psicología. General y Aplicada*, 37, 215-227.
- Yela, M. (Ed. 1987a). *Estudios sobre inteligencia y lenguaje*. Madrid: Pirámide.
- Yela, M. (1987b). Sobre las raíces de la conducta humana. *Temas de Psicología IV*. Salamanca: Universidad Pontificia de Salamanca.
- Yela, M. (1987c). *Reflexiones de un psicólogo sobre el sentido de la conducta humana*. Salamanca: Universidad Pontificia de Salamanca.
- Yela, M. (1992). Conducta animal y metaconducta humana. *I Congreso Internacional: El hombre y los animales de compañía*. Barcelona: Fundación Purina. 17-35.