

## La psicología cognitiva: ensayo sobre un paradigma en transformación

Manuel de Vega  
Universidad de La Laguna

*El presente ensayo explora el estado actual de la psicología cognitiva cuando nos aproximamos a un nuevo siglo. En primer lugar, analiza los fundamentos del paradigma simbólico computacional, que reduce la mente a un sistema de procesamiento de la información análogo a una máquina de Turing o a un ordenador. El paradigma S-C, sin embargo, tiene limitaciones importantes al tratar con la experiencia subjetiva (conciencia) y las características corpóreas del conocimiento (e.g., las emociones). En segundo lugar, se describe el conexionismo como un nuevo paradigma emergente que, superando la metáfora del ordenador, se asienta firmemente en la metáfora del cerebro; además, aporta una nueva noción de computación como un procesamiento distribuido y paralelo. En tercer lugar, la principal contribución de las neurociencias es proporcionar nuevas medidas más precisas de la actividad cerebral durante las tareas cognitivas mediante técnicas de neuroimagen, ERP y el estudio de pacientes con daño cerebral. Las ventajas e inconvenientes del conexionismo y de las neurociencias se revisan a continuación. Finalmente, el artículo propone un reto para la siguiente centuria: el estudio de las representaciones de situaciones y episodios, que no puede reducirse al estudio de la organización de la «memoria semántica». El desarrollo de este tema no sólo mejoraría nuestro entendimiento de cómo las personas codifican las situaciones cotidianas, sino también podría tener consecuencias para el desarrollo de máquinas inteligentes.*

*Palabras clave: paradigma simbólico computacional, conexionismo, neurociencias, conciencia, emociones, modelos de situación, inteligencia artificial.*

*This essay explores the current state of Cognitive Psychology at the verge of a new century. First, it analyzes the foundations of the symbolic-computational paradigm (S-C), which reduces the mind to an information*

*processing system, like a Turing machine or a computer. The S-C paradigm, however, has serious shortcomings in dealing with subjective experience (consciousness) and embodied features of cognition (e.g., emotions). Secondly, connectionism is described as a new emergent paradigm which overcomes the computer metaphor and strongly relies on the brain metaphor; in addition, it involves a new notion of computation as a parallel distributed processes. Thirdly, the main contribution of neuroscience is the new and more accurate measures of brain activity during cognitive tasks by means of neuroimaging, ERP techniques, and brain damage case studies. The advantages and shortcomings of connectionism and neuroscience are briefly considered. Finally, the paper proposes a challenge for the next century: the study of situation and episodic representation, which cannot be reduced to «semantic memory» organization. The development of this subject not only would improve our comprehension of how people encode and update everyday situations, but also would have technological consequences for developing intelligent machines.*

*Key words: Symbolic-computational paradigm, connectionism, neurosciences, consciousness, emotions, situation models, artificial intelligence.*

Hace unos 30 años se estrenó una película titulada *2001: Una odisea en el espacio*, sobre un guión de Arthur Clarke y dirigida por Stanley Kubrick. Se trataba de una fantasía de ciencia ficción, pero era una fantasía razonable, pues el guionista y el director se preocuparon de documentarse entre los expertos de su época acerca de la tecnología que cabría esperar en el año 2001. En la película aparecía un personaje llamado HAL que era ni más ni menos que un ordenador avanzado con capacidades cognitivas prácticamente iguales a las humanas. El maravilloso HAL era capaz de un uso totalmente referencial del lenguaje (v.g., podía comentar las noticias de la televisión con los astronautas), jugaba al ajedrez, tenía metas e intenciones propias (de hecho intenciones malévolas), tenía autoconciencia y emociones y hasta era capaz de «leer» las palabras en los labios. Curiosamente, en la ficción de la película, HAL «había sido construido» en 1997 de modo que es interesante comprobar hasta qué punto se han cumplido las predicciones en tecnología computacional. De hecho, hace unos meses ha sido publicado un libro exactamente con esta intención (Stork, 1997). Un grupo de expertos entre los que se encuentran algunos conocidos científicos cognitivos como Minsky, Dennett, Schank, Wilkins, etc., participaron en este libro para reflexionar sobre el estado actual de la Inteligencia Artificial (IA). El resultado de este análisis es algo decepcionante: disponemos de programas que juegan muy bien al ajedrez, pero las demás capacidades cognitivas de HAL, que lo hacían tan «humano» en la película, nos parecen hoy tan lejanas de nuestras posibilidades tecnológicas como parecían estarlo en los años 60. Incluso más lejanas, ya que en los años 60 los expertos consultados por Clarke y Kubrick eran probablemente mucho más optimistas sobre las posibilidades inmediatas de construir un ordenador con capacidades humanas que los expertos actuales.

En esa misma época, en los años 60, el clima intelectual que reinaba en la naciente psicología cognitiva estaba próximo al optimismo tecnológico de los técnicos de IA. El desarrollo de los ordenadores digitales, la nueva lingüística generativa y la crisis profunda del conductismo, entre otros factores, habían propiciado la irrupción de un prometedor paradigma cognitivo; y el optimismo era patente. Surgió un programa de investigación específico para la psicología cognitiva: los temas clásicos de la memoria, la percepción o la resolución de problemas recibían el impulso de ideas y métodos nuevos; por otra parte, algunos temas hasta entonces prácticamente olvidados se mostraban como territorios nuevos y excitantes al alcance de la indagación científica: la atención, las imágenes mentales, el razonamiento, los procesos del lenguaje, etc. Además, había un programa cognitivo «máximo»: el desarrollo de una ciencia cognitiva general que englobaría tanto los procesos computacionales en los sistemas biológicos (v.g., el cerebro) como en los sistemas artificiales (los ordenadores). Las funciones cognitivas de la mente humana y las funciones computacionales de los ordenadores no serían substancialmente diferentes, según este análisis, sino que ambas podrían explicarse desarrollando una única teoría universal del cómputo. Este programa máximo de la ciencia cognitiva es el que hacía que el HAL de 2001 fuese algo más que una fantasía de escritores y cineastas; era —en todo caso— una fantasía de científicos cognitivos.

Tres décadas después ¿cuál es la situación de la psicología cognitiva? Sin duda aquel empuje inicial mantuvo su continuidad en el tiempo. La segunda mitad del siglo XX ha estado marcada por un dominio en psicología de los planteamientos cognitivistas. La buena salud de la psicología cognitiva se aprecia en multitud de índices externos. Las publicaciones más prestigiosas de la psicología, como la revista *Psychological Review* o las *Journal of Experimental Psychology* incluyen frecuentemente artículos de carácter cognitivo. Existen muchas revistas especializadas en las que las etiquetas «cognitivo» o «cognición» aparecen en la portada. Otras publicaciones especializadas, aun cuando no incluyan estas etiquetas, se refieren claramente a procesos cognitivos tales como la atención, la memoria, el lenguaje, la percepción, el razonamiento, etc. Proliferan los congresos y sociedades internacionales sobre procesos cognitivos que periódicamente editan libros y revistas. Las universidades y centros de investigación más importantes del mundo disponen de departamentos y laboratorios de psicología o de ciencias cognitivas. Finalmente, incluso las investigaciones realizadas en las demás disciplinas psicológicas como la psicología clínica, la psicología social, la psicología evolutiva, la psicología educativa o las neurociencias tienen frecuentemente un planteamiento «cognitivo».

¿Qué hay detrás de este aparente dominio del cognitivismo? ¿Se trata de una simple moda pasajera o, por el contrario, la psicología es y será necesariamente cognitiva? Conviene aclarar previamente qué entendemos por psicología cognitiva. Una posibilidad es identificar la psicología cognitiva (y en general las ciencias cognitivas) con un determinado «paradigma» científico: aquél que surgió a finales de la década de 1950 gracias a algunos pioneros anglosajones como Donald Broadbent, George Miller o Noam Chomsky, y que se consolidó en las décadas siguientes. Esta visión restrictiva implica que la psicología cognitiva

asume un conjunto de prescripciones teóricas y metodológicas particulares que caracterizan a lo que se ha denominado el «paradigma simbólico-computacional». Otra posibilidad, sin embargo, es ampliar la noción de psicología cognitiva a cualquier análisis de los procesos o funciones mentales, cualesquiera que sean las herramientas teóricas o metodológicas utilizadas. En este sentido amplio, la psicología cognitiva abarca movimientos y escuelas diversas, no necesariamente anglosajones. Así, por ejemplo, la psicología genética de Jean Piaget, el enfoque socio-histórico de Lev Vygotski o los estudios factoriales de la inteligencia pueden considerarse con todo rigor como análisis cognitivos. Más aún, el cognitivismo es incluso anterior a la psicología científica, pues algunos filósofos clásicos se habían interesado vivamente por las funciones mentales. Mencionemos, por ejemplo, la temática cognitiva de *Las categorías* de Aristóteles, *La crítica de la razón pura* de Kant, o *La investigación sobre el conocimiento humano* de Hume. Por otra parte, una noción amplia de la psicología cognitiva incorporará también cualquier nuevo enfoque o paradigma que pueda desarrollarse en el futuro con la condición de que estudie los procesos o funciones mentales.

Si asumimos la noción restringida de la psicología cognitiva es fácil prever que el paradigma «simbólico computacional» o del «procesamiento de información» podría tratarse de un producto académico perecedero y tendría un final histórico tarde o temprano. De hecho, hay ya síntomas inequívocos de que el paradigma cognitivo clásico viene arrastrando una crisis en los últimos años, como veremos en estas páginas, y cabe la posibilidad de que esta crisis sea profunda y definitiva. Sin embargo, tal como sugirió Kuhn, un paradigma no se extingue realmente a no ser que surja un paradigma alternativo nuevo o «revolucionario» que lo reemplace. Contemplando la psicología cognitiva de finales de siglo, se observan algunas novedades que quizá se puedan calificar de nuevos paradigmas. En primer lugar, el conexionismo desarrollado en la última década, nació con una cierta vocación revolucionaria, atacando dialécticamente al viejo enfoque simbólico-computacional. En segundo lugar, las neurociencias que intentan reformular la investigación de los procesos cognitivos buscando sus correlatos neuronales, y que han cobrado un gran impulso gracias a los nuevos métodos desarrollados recientemente. Es pronto para saber si el conexionismo o las neurociencias van a substituir a la psicología cognitiva clásica, o son desarrollos complementarios y compatibles con ella. Aunque lo primero fuese cierto, la eventual extinción del paradigma simbólico-computacional no significaría que éste haya sido una «moda pasajera» y que debamos pasar la página sin más. Algunos de los logros del paradigma simbólico-computacional persistirán en los paradigmas emergentes que quizá dominen el siglo XXI. Los principales temas y problemas de investigación, algunos de los métodos empíricos y formales desarrollados, y el optimismo epistemológico acerca de la posibilidad del estudio científico de los procesos mentales son aportaciones extraordinarias, con un valor acumulativo, que debemos a aquellos pioneros del procesamiento de información.

Si asumimos una noción amplia de psicología cognitiva, por otra parte, el futuro es sumamente prometedor para el cognitivismo. El estudio de las funciones de la memoria, la atención, el lenguaje o la conciencia, por poner algunos ejemplos, serán temas centrales de la ciencia, independientemente del para-

digma dominante. Este tipo de cuestiones tienen interés intrínseco y no son simples *puzzles* académicos del momento. Del mismo modo que el origen del universo, la estructura de la materia, o el genoma humano no pueden considerarse modas pasajeras, sino temas básicos de la investigación científica.

En estas páginas revisaré brevemente los problemas y aportaciones de los paradigmas cognitivos, el clásico y los nuevos. También haré una valoración del estado actual de los temas de investigación, destacando tanto aquellos en los que se ha progresado más como los que apenas han avanzado.

## El paradigma simbólico-computacional (S-C)

En 1950 el matemático Alan Turing describió una máquina ideal capaz de realizar operaciones muy simples sobre símbolos discretos, tales como imprimir, borrar o leer un símbolo, adelantar un espacio, etc. Esta máquina formal tendría un poder de cómputo prácticamente ilimitado de modo que, tal como sugirió Turing, podría incluso imitar el comportamiento humano inteligente (v.g., hablar y responder a preguntas) hasta el punto de «engañar» a un observador que sólo tuviese acceso a sus respuestas. La creación de los ordenadores digitales, a la cual contribuyó el propio Turing, hizo que la noción de cómputo universal pasara de ser la simple ilustración de una abstracción matemática a ser una realidad tecnológica. Los ordenadores no son máquinas de Turing pero son funcionalmente equivalentes: pueden resolver cualquier problema computacional a condición de que se sea capaz de reducirlo a un conjunto finito de operaciones o instrucciones. Un paso más lo dieron los psicólogos: no sólo los ordenadores son equivalentes a la máquina universal de Turing sino también el cerebro. Surgió así el paradigma simbólico-computacional como una extensión natural de las tecnologías computacionales. Así, Fodor, Minsky, Simon o Pylyshyn han defendido explícitamente la idea de que lo que llamamos «mente» o «funciones mentales» se reduce a la aplicación de cálculos a representaciones simbólicas.

El paradigma simbólico-computacional (S-C) había nacido y las posibilidades que se abrían a la nueva psicología cognitiva eran fascinantes. Por primera vez los procesos mentales podrían estudiarse con un nivel de explicación mecanicista: el del procesamiento de información. Los símbolos, las reglas sintácticas y los cálculos que subyacen a los procesos mentales son entidades perfectamente mensurables, al igual que las nociones homólogas que manejaban los técnicos en ordenadores. Pero la psicología cognitiva tuvo que pagar un precio: debía asumir una serie de restricciones propias de la noción de cómputo implícita en la tecnología computacional (y por ende en la propia máquina de Turing). Veamos algunas de ellas.

### *El lenguaje de la mente*

La materia prima del procesamiento de información son las representaciones mentales en forma de expresiones simbólicas en el lenguaje de la mente

o *mentales*, tal como lo denominó Jerry Fodor. El *mentalés* es un lenguaje universal (no corresponde a ningún idioma particular) y amodal (válido para representar información verbal, visoespacial o de cualquier otro origen). Las mejores versiones del *mentalés* se desarrollaron empleando el formalismo lógico de las proposiciones. Las proposiciones, en efecto, constituyen un formato ideal para el paradigma S-C, pues son expresiones formales, compuestas de símbolos discretos (un vocabulario), sujetos a unas reglas de composición (una sintaxis). El atractivo psicológico de las proposiciones era evidente: representan el significado abstracto y no las características superficiales de la información, pueden representar cualquier modalidad de información, tienen una estructura interna que permite reflejar la organización de la memoria, son «computables» y por tanto se pueden construir programas de simulación que operan con ellas, etc. El lenguaje proposicional se complementó con formalismos «procedimentales» que reducían los cálculos cognitivos a conjuntos de reglas «si... entonces». Mediante estas reglas o «sistemas de producción» era posible imitar algunas capacidades humanas de resolución de problemas, por ejemplo, se pudo construir programas de ordenador capaces de jugar al ajedrez o de resolver problemas criptoaritméticos.

Sin embargo, casi inmediatamente surgieron problemas: nuestra intuición, pero también los datos experimentales, nos sugieren que algunas representaciones tienen un carácter analógico, es decir que son imágenes mentales. Podemos recordar o imaginarnos los lugares y trayectos de nuestra ciudad, e incluso hacer «paseos mentales» para planificar nuestra ruta; podemos «girar» mentalmente un mueble sin necesidad de moverlo físicamente, con lo cual ahorramos tiempo y energías; o bien podemos resolver de forma creativa problemas que requieren la simulación mental de estructuras visoespaciales dinámicas. Todo ello se nos antoja una forma de representación continua y espacial algo diferente de un cómputo sintáctico sobre símbolos discretos y abstractos. No tiene nada de extraño que los defensores más radicales del núcleo duro del paradigma simbólico-computacional hayan sido muy críticos con la noción de imagen mental, por su carácter amenazador para la ortodoxia.

### *La metáfora del ordenador*

La psicología cognitiva nació, tal como hemos visto, con el desarrollo de los ordenadores y éstos han continuado siendo una guía teórica para los investigadores. La denominada «metáfora del ordenador» ha sido un heurístico de investigación fructífero, inspirando los temas e hipótesis, ofreciendo un vocabulario teórico para interpretar los datos, y sirviendo como herramienta formal en la simulación de modelos. Con todo lo positivo que esto es, conviene reflexionar acerca de una consecuencia de la metáfora del ordenador bastante obvia: *La fundamentación del paradigma S-C es tecnológica, no naturalista.*

La situación es un tanto extraña para una ciencia como la psicología que se supone «natural». Es como si los biólogos investigasen la célula inspirándose en las características del motor de explosión (los dos transforman energía) y ol-

vidándose de indagar sus múltiples funciones fisiológicas en el organismo vivo. Los ordenadores digitales son artefactos culturales de reciente invención, mientras que el cerebro y sus funciones cognitivas son el fruto de la evolución filogenética en nuestros antepasados homínidos. Por tanto, la arquitectura funcional de la cognición y las funciones cognitivas básicas no hay que entenderlas como programas de ordenador arbitrarios, sino como adaptaciones destinadas a resolver tipos de problemas propios de nuestro entorno ecológico e interpersonal.

La desnaturalización de la psicología cognitiva ha sido patente durante años. La metáfora del ordenador y los planteamientos logicistas han inspirado las teorías psicológicas con frecuencia, como si la mente fuese un programa de ordenador arbitrario. Incluso asumiendo esta metáfora, cabe preguntarse ¿para qué sirve ese programa o programas de ordenador?, ¿con qué funciones ha sido «diseñado»? James Gibson, un psicólogo no cognitivo de la percepción, fue el primero en plantearse una cuestión fundamental: ¿Para qué sirve la percepción visual? Su respuesta: la visión constituye un canal para recoger las «invariantes» informativas del medio que subyacen al cambiante flujo sensorial en nuestros receptores. Por ejemplo, vemos «substancias», «superficies» y «sucesos» que tienen un valor adaptativo para nosotros (*affordance*). Compárese esta preocupación por conectar la visión con el mundo, con el análisis clásico de la visión como una suma de sensaciones elementales de colores, formas, tonos, bordes, movimientos, etc.

Afortunadamente, ha habido dentro del propio paradigma cognitivo intentos de restablecer el anclaje naturalista de las funciones cognitivas. David Marr, otro psicólogo de la visión pero esta vez enclavado dentro del paradigma S-C, realizó un planteamiento conciliador: hay varios niveles de análisis en la ciencia cognitiva y todos son legítimos. Uno de ellos es estrictamente formal o *algorítmico*, en el que se puede seguir utilizando el vocabulario y conceptualización S-C (representaciones, símbolos, reglas, etc.); otro es el nivel de *hardware* propio de la neurociencia que trata de determinar el substrato neurológico de los algoritmos; finalmente, hay un nivel *computacional* (aunque el término es confuso y mejor sería decir «ecológico» o, como lo denomina John R. Anderson, «racional») que corresponde a la descripción abstracta de las funciones del algoritmo respecto al entorno; en otras palabras su valor adaptativo.

Intentar desarrollar una teoría «computacional» o «adaptativa» de cada función básica puede ser un ejercicio extremadamente útil para no avanzar a ciegas, evitando líneas e hipótesis de investigación baldías. Así, por ejemplo, seguramente será más fructífero tratar de entender el razonamiento humano preguntándose cómo y para qué razona la gente en situaciones pragmáticas en lugar de la cuestión logicista de si el razonamiento con silogismos formales se ajusta o no a una competencia lógica. También parece más realista preguntarse para qué se usa la memoria en la vida real que basar nuestras teorías en estudios de aprendizaje de listas de palabras como se hizo inicialmente. Lo cierto es que la memoria se utiliza para recordar material verbal mucho más complejo que las listas verbales, por ejemplo, textos o conversaciones, pero también se usa de forma prospectiva para recordar planes e intenciones, y es la base del aprendizaje de destrezas que no son verbales en absoluto. Además, tal como sabemos hoy en día

la memoria no es necesariamente explícita o consciente. Por último, no sólo hay que valorar los módulos de procesamiento del lenguaje que realizan cálculos necesarios, sino que también habrá que explicar el uso del lenguaje en contextos pragmáticos y conversacionales para valorar, por ejemplo, funciones esenciales de la comunicación como la deixis (algunos elementos del discurso como los pronombres sólo adquieren significado en el contexto), las reglas implícitas de la comunicación entre interlocutores (v.g., la valoración del conocimiento compartido y no compartido), etc.

No estamos planteando una mera revisión metodológica, y mucho menos una substitución de los métodos experimentales por otros observacionales o hermenéuticos. Se trata más bien de un cambio de perspectiva teórica: deben seleccionarse tareas relevantes y representativas, parametrizarlas adecuadamente, y generar paradigmas experimentales bien controlados adecuados a los fenómenos seleccionados.

### *El misticismo computacional*

La vertiente filosófica del paradigma S-C pretende haber resuelto de una vez por todas el problema del dualismo cartesiano. La mente humana, al igual que los ordenadores, tiene un componente físico o *hardware* y un componente funcional o *software*. Pero la naturaleza física del *hardware* es irrelevante para comprender los procesos de cómputo. Así, para entender la naturaleza computacional de una tarea que realiza un ordenador podemos desentendernos de la organización material del sistema y estudiar únicamente la descripción del programa en términos de instrucciones, rutinas, etc. Es más, podríamos construir físicamente una máquina de Turing —que es bastante diferente de un ordenador digital— que ejecutase el mismo programa; o bien podríamos manipular un gran número de latas de cerveza cuya posición, de pie o tumbadas, representase los estados binarios 1 y 0 respectivamente; y el ordenador de latas también podría ejecutar el mismo programa. El cerebro sería otro sistema de cómputo cuyas peculiaridades biológicas no tendrían demasiada importancia; sólo el programa o los programas que ejecuta el cerebro serían dignos de interés. Este planteamiento funcionalista permitió un avance considerable de la psicología cognitiva, ya que el análisis de la «arquitectura funcional» del sistema no tuvo que esperar a desentrañar la «arquitectura cerebral». Se podía tratar de desvelar cómo funciona la memoria a corto plazo, por ejemplo, sin preocuparse de la circuitería cerebral responsable de ella.

Una consecuencia inesperada de esta filosofía funcionalista fue la promesa de una «inmortalidad» tecnológica: podría transferirse el contenido de nuestra memoria («bases de datos» y «programas») al disco duro de un ordenador (o su equivalente tecnológico del futuro) con lo cual nuestra identidad continuaría indefinidamente en su nueva residencia artificial. Podríamos crear así «hijos mentales», tal como propone Marvin Minsky, que trascenderían a nuestra existencia biológica. Esta fantasía, por peregrina que pudiera parecer, es una posibilidad teórica que se deriva rigurosamente de las premisas de computación



simbólica y de equivalencia funcional de los sistemas de cómputo. Si la mente no es más que un programa informático entonces podrá transferirse a un ordenador o a un robot.

Sin embargo, me inclino a pensar que este sueño -o pesadilla- emparejado con el mito de Prometeo o el de Frankenstein, sea fruto más bien de nuestra ignorancia (o de la ignorancia de quienes lo han planteado). Bastaría con poner en duda alguno de los dogmas en que se basa para que el misticismo tecnológico fuese inviable. En primer lugar, no está claro que nuestros procesos mentales se reduzcan, después de todo, a meros programas de cómputo aplicados a símbolos discretos. Hay aspectos de nuestros procesos mentales que exceden esta visión restringida. Ya hemos mencionado que algunos datos sugieren la existencia de representaciones analógicas pero, además, habría que mencionar el papel de las emociones en los procesos cognitivos, o el gran enigma de la naturaleza y funciones de la conciencia; nada de esto parece tener cabida en el marco dogmático del paradigma S-C. En segundo lugar, la estructura y organización de la base material del sistema podría tener un impacto enorme sobre la naturaleza de los cómputos. El cerebro es un sistema biológico basado en un conjunto inmenso de procesadores interconectados (las neuronas) que funcionan en paralelo y que se comunican bioquímicamente, mediante una serie de neurotransmisores; el ordenador, por el contrario, es un sistema electrónico incomparablemente más sencillo que el cerebro que realiza cómputos de forma serial utilizando, generalmente, un único microprocesador. Sin entrar en el papel modulador de los neurotransmisores en la cognición, la estructura de procesamiento en paralelo del cerebro marca, sin duda, diferencias funcionales en relación a los ordenadores digitales que ejecutan secuencialmente un programa de *software*. Pretender que las funciones de un procesador paralelo y uno serial son las mismas es como asegurar que los procesos digestivos son lo mismo en una bacteria y en una vaca, por el hecho de que ambas tienen funciones digestivas. Los nuevos avances en el conexionismo y la emergencia de ordenadores con múltiples procesadores en paralelo, nos están mostrando cómo estas nuevas arquitecturas de cómputo son funcionalmente muy diferentes de las de los ordenadores convencionales. Por último, el ordenador es un procesador de información precodificada. Opera a partir de series de símbolos suministrados por un operador humano; pero es una máquina esencialmente «solipsista» que no codifica o sintetiza información directamente a partir del entorno, ni actúa sobre éste. Por el contrario, el cerebro está «corporeizado», puede sintetizar información por sí mismo, a partir de un entorno ecológico e interpersonal no simbólico y, además, gobierna la acción de nuestro cuerpo sobre dicho entorno. Estas diferencias parecen demasiado substanciales como para hacer viable por el momento la promesa de la inmortalidad tecnológica o, si se prefiere, del «enlatado» cognitivo.

### *Las prohibiciones*

Una característica de los paradigmas científicos es que no sólo propician la génesis de hipótesis en ciertos temas de investigación, sino que también prohi-

ben ciertas hipótesis y/o temas de investigación. Así, el conductismo negaba el interés científico e incluso, en ocasiones, la existencia de los procesos mentales. Las restricciones epistemológicas de este tipo también existen en el paradigma S-C. En algunos casos aparecen como una prohibición más o menos explícita, en otros son «olvidos» persistentes o trivializaciones de ciertos fenómenos.

El paradigma S-C tiene una preferencia por los «sistemas modulares» de procesamiento. La situación ideal para el paradigma sería que la naturaleza hubiese estructurado las funciones cognitivas en «módulos» u órganos mentales bien separados, que realizaran funciones especializadas e independientes unos de otros. Esta arquitectura funcional modular es precisamente la que se propone para la percepción visual o el lenguaje. Así, en el caso de la comprensión del lenguaje existiría un módulo «fonológico» encargado del análisis y representación de los fonemas del habla (en el caso de la lectura sería un módulo ortográfico-fonológico), un módulo «léxico» sería responsable del reconocimiento de las palabras, un módulo «semántico» activaría los significados de las palabras y un módulo «sintáctico» construiría una representación formal de las frases. Estos módulos se asegura que operan secuencialmente «de abajo arriba» y están «encapsulados», de modo que cada uno de ellos realiza la tarea asignada sin tener en cuenta lo que hacen los módulos superiores. La psicolingüística ha tenido un gran éxito en su investigación del lenguaje a partir de esta concepción modular, pero hay que reconocer que, en la comprensión del lenguaje, el aparataje modular no lo es todo; tarde o temprano los módulos tienen que comunicarse con lo que Fodor llama los «sistemas centrales» de procesamiento, responsables de la construcción del significado del discurso. Pero estos sistemas centrales ya no son modulares, no están especializados en ninguna operación computacional concreta sino que son enormemente flexibles, globales y multifuncionales. Los sistemas centrales hacen un uso activo de nuestro conocimiento general del mundo, de nuestros recuerdos episódicos y biográficos, de nuestras destrezas físicas y mentales, de nuestros estados emocionales y afectivos, de nuestras metas e intenciones fugaces, etc. Además los sistemas centrales intervienen en la comprensión y producción del lenguaje, en la comprensión de situaciones, en el pensamiento y en la resolución de problemas, en la planificación y en la toma de decisiones, etc. La plácida y ordenada concepción modular y mecanicista no es aplicable a los sistemas centrales. La consecuencia es un pesimismo epistemológico: no es conveniente perder el tiempo estudiando algo que no entendemos, como por ejemplo el razonamiento o la resolución de problemas. Fodor lo expresa así:

Cuanto más global (e.g., más isotrópico) es un proceso cognitivo, menos lo entiende nadie. Los procesos muy globales, como el razonamiento analógico, no se entienden en absoluto (Fodor, 1983, p. 107).

O también la siguiente cita en la que, de paso, se observa el pesimismo de Fodor sobre la inteligencia artificial:

Pienso que, desde luego, el razonamiento y la resolución de problemas son actividades computacionales, ya que la mente es, en cierto sentido, una computadora o,

si se quiere, una máquina de lenguaje del pensamiento. Pero me parece que no tenemos ni una pista acerca de cómo funciona a este respecto. Es un profundo misterio y, hasta que no lo resolvamos, no vamos a tener máquinas inteligentes. Podremos tener máquinas perceptivas, máquinas que analicen oraciones o que puedan recuperar la forma visual de un objeto, pero no creo que vayamos a tener máquinas que piensen. La razón por la que no lo creo es que, para abordar el problema, no van a valer estrategias fragmentarias por las que se enfoque de una manera más local de como aparenta; la razón es que hay algo que no entendemos. (en García-Albea, 1991, p. 39).

Zenon Pylyshyn, por su parte, utiliza de un modo un tanto particular la navaja de Occan en psicología cognitiva: si una función cognitiva es «penetrable» a nuestro conocimiento del mundo, entonces no pertenece a la arquitectura funcional básica. Es decir, que si una determinada función cognitiva se ve influida por nuestras creencias, emociones, metas e intenciones, entonces hay que rechazarla como una entidad primitiva o teóricamente relevante. Cerremos los ojos por un momento e imaginemos una hoja de un árbol y una bola de plomo cayendo al suelo desde la misma altura. Es evidente en nuestra imagen mental que la bola se «desplaza» más deprisa siguiendo una trayectoria vertical y rectilínea, mientras que la hoja «planea» en zigzag. Dicho de otro modo, nuestra imagen mental es penetrable a nuestro conocimiento intuitivo sobre el comportamiento físico de los objetos. Por tanto, nos dirá Pylyshyn, las imágenes mentales no son una buena noción para nuestras teorías cognitivas: no son modulares sino que se ven determinadas por los sistemas centrales de procesamiento.

Fodor y Pylyshyn son coherentes con las restricciones del paradigma S-C y las llevan hasta las últimas consecuencias. Pero no olvidemos que la prohibición o aplazamiento *sine die* de algunos temas de estudio es un indicio de la incapacidad del paradigma, no una demostración de que carezcan de interés científico.

A las dificultades del paradigma S-C para entender los procesos no-modulares hay que añadir sus dificultades para tratar los aspectos «calientes» y cualitativos de la cognición. En particular ¿qué hacer con las emociones y con la conciencia?

Un modo típico de enfrentarse a las emociones dentro del paradigma S-C es considerar que son irrelevantes para la cognición. Una teoría cognitiva simplemente debe tratar de los procesos computacionales y las emociones no son computables (aunque no todos estarían de acuerdo con esta afirmación). En cualquier caso, las emociones constituyen una especie de «ruido» que distorsiona o interfiere con los procesos cognitivos. Esta idea coincide, además, con algunas tradiciones educativas intelectualistas en la sociedad occidental: «para pensar mejor hay que mantener la cabeza fría» y «cuando las emociones irrumpen hay que controlarlas». Sin embargo, de nuevo la realidad no tiene por qué coincidir ni con nuestras preferencias educativas, ni con nuestras restricciones paradigmáticas. Algunos investigadores han mostrado de modo fehaciente que las emociones no sólo no son vecinos indeseables de nuestros procesos cognitivos, sino que son componentes necesarios de la cognición. Por ejemplo, Antonio Damasio (1994) estudió minuciosamente a pacientes con lesiones cerebrales prefrontales que son incapaces de sentir las emociones habituales en

contextos sociales. Estos pacientes tienen, por lo demás, todas sus capacidades cognitivas intactas de modo que son un buen ejemplo de procesadores de «cabeza fría», simples manipuladores racionales de símbolos sin las distorsiones de las emociones. El resultado de esta liberación de las emociones es, sin embargo, catastrófico para estas personas que se muestran incapaces de dirigir sus vidas y tomar decisiones en las situaciones cotidianas más simples. Sus dificultades más notables tienen lugar, precisamente, en el tipo de razonamientos en que una estrategia exclusivamente «racional» o «algorítmica» podría parecer suficiente e incluso deseable. En opinión de Damasio, las emociones establecen vínculos entre los procesos cognitivos y los estados somáticos, es decir, proporcionan una valoración corporal inmediata de los diversos estados cognitivos. De este modo las emociones que acompañan a la cognición proporcionan una impronta positiva o negativa inmediata, previa al desarrollo de cálculos detallados en términos de costes/ beneficios o de una valoración de la coherencia lógica de las acciones. En las situaciones de la vida real casi nunca tenemos datos completos o no tenemos tiempo suficiente para una toma de decisión «algorítmica», de modo que la guía emocional nos ayuda a cortar por lo sano y elegir alternativas o cursos de acción. Cuando elegimos carrera, o decidimos casarnos, o decidimos descomponer una tarea compleja en varias partes y atacar una por una, seguramente las emociones que sentimos (estados corporales agradables o desagradables) en cada punto de decisión son tan importantes como los cálculos algorítmicos. Los pacientes prefrontales a menudo entran en un bucle de consideraciones interminables para la decisión más sencilla (fijar la hora de una cita), precisamente porque se ven abocados a un proceso estrictamente algorítmico de valoración de todas las alternativas posibles y los condicionantes posibles de esas alternativas.

La conciencia es uno de los fenómenos humanos que despierta reacciones encontradas entre los científicos. El debate sobre la conciencia se ha reavivado recientemente incluyendo entre sus filas no sólo psicólogos, sino neurocientíficos, filósofos, biólogos e, incluso físicos y matemáticos. Algunos creen que es un tema científicamente inalcanzable y, por tanto, parecen aplicar la máxima de Wittgenstein: «De lo que no se puede hablar, es mejor callarse». Otros tienden a mitificarla atribuyéndole una entidad física, de momento misteriosa, irreductible a procesos mecanicistas de cómputo. Así, Roger Penrose asegura que la conciencia reside en ciertos fenómenos cuánticos (desconocidos) que ocurren en los microtúbulos de las neuronas, o Christof Koch y Francis Crick aseguran que la conciencia visual se deriva (aunque no sabemos cómo) de los disparos sincrónicos de las neuronas 40 veces por segundo. Otros, en fin, la trivializan considerándola una característica emergente de los sistemas computacionales complejos, pero que no requiere ninguna explicación particular. En principio, un sistema de procesamiento artificial (v.g., un sistema conexionista masivo) podría incorporar conciencia, en opinión de Churchland y Churchland.

Se ha avanzado mucho en la obtención de correlatos funcionales y neuronales de la conciencia. Al menos sabemos que la conciencia está asociada a ciertos procesos cognitivos y no a otros. Está asociada al consumo de recursos cognitivos de la memoria operativa y de la atención, a los procesos controlados que

suelen ir acompañados de «esfuerzo» mental, a los procesos de reconocimiento en la memoria, a las estrategias de resolución de problemas, y a los productos perceptivos de la visión y la audición. La conciencia está asociada también a funciones mucho menos cognitivas como el dolor o las emociones. Por el contrario, no somos conscientes de las rutinas o procesos automáticos, tales como los «módulos» de la visión o del lenguaje, ni lo somos de algunos procesos sensorio-motores como el control vestibular del equilibrio o del movimiento.

Por otra parte también disponemos de correlatos neuronales de la conciencia. Baste recordar cómo determinadas lesiones occipitales determinan el síndrome de los «ciegos videntes», que responden de modo automático ante estímulos visuales y sin embargo no los «ven», o cómo los amnésicos pueden recuperar información previamente aprendida, pero no la «reconocen» como recuerdos. Todo esto pone de manifiesto que la conciencia no podrá considerarse ya más un tema intangible para la especulación filosófica o literario, sino que es un fenómeno con una base fisiológica.

Pero estos avances no resuelven un tema de fondo ¿qué significa la cualidad «subjetiva» o experiencial de la conciencia?, ¿por qué somos conscientes después de todo?, ¿qué necesidad hay de que seamos conscientes? Del mismo modo que algunas de nuestras funciones cognitivas no son conscientes ¿por qué razón no lo son todas? ¿No podría diseñarse un robot «zombi» que ejecutase todas las funciones habituales en la gente, sin la menor conciencia? Algunos programas ejecutados en superordenadores juegan al ajedrez con un nivel de gran maestro pero ¿son por ello conscientes? Una concepción emergentista de la conciencia quizá nos diga que un superordenador suficientemente complejo tendrá algo de conciencia, como mero epifenómeno de la computación pero ¿por qué iba a ser así? ¿Es más consciente un superordenador que mi calculadora de bolsillo? La reducción de la conciencia a un efecto colateral de la complejidad computacional parece, en primer lugar, inexacta y en segundo lugar no es una verdadera explicación.

### *Algunas conclusiones favorables*

La visión un tanto crítica que he desarrollado hasta ahora del paradigma S-C no debe ocultarnos sus muchas virtudes. Desde luego, se ha progresado mucho en los objetivos coincidentes con la ortodoxia paradigmática. Se ha avanzado, por ejemplo, en el conocimiento de los mecanismos modulares del sistema visual o de la comprensión y producción del habla. Pero la cosa no ha quedado ahí. Los propios psicólogos cognitivos, especialmente los de orientación experimental no se han tomado totalmente en serio todas las prescripciones paradigmáticas mencionadas anteriormente. Dichas prescripciones son el fruto de planteamientos filosóficos sobre la fundamentación lógica de la ciencia cognitiva, pero los psicólogos cognitivos no han podido substraerse al estudio de las funciones y tareas que consideraban relevantes, fuesen o no adecuadas a la ortodoxia simbólico-computacional. Así, en los años 80 y 90 muchos psicólogos se han aproximado a una descripción naturalista de los procesos mentales, han estu-

diado la interacción entre emociones y cognición, han investigado la organización de los «sistemas centrales», han intentado desvelar las funciones de la conciencia, han avanzado en el estudio del pensamiento, o han tratado de valorar cómo la arquitectura cerebral impone restricciones a la arquitectura funcional. Por otra parte, como veremos en la próxima sección, la propia noción de cómputo del paradigma S-C está siendo contestada por aquellos que parecen haber dado a luz a un nuevo paradigma cognitivo: los conexionistas.

En conjunto, podemos asegurar que en las tres últimas décadas hemos avanzado en el conocimiento de los procesos mentales más que los tres últimos siglos. Sabemos mucho más que el ciudadano de a pie de modo que nuestros datos y modelos van mucho más allá de las intuiciones de su «psicología ingenua». Pero también sabemos mucho más que los grandes filósofos de antaño, pese a sus capacidades analíticas privilegiadas. Ha sido el método experimental aplicado al estudio de la mente, en mi opinión tanto o más que los avances en la formalización o en *implementación* computacional, el que ha permitido extender nuestro conocimiento sobre la cognición. Una gran parte de los procesos cognitivos son totalmente inconscientes, es decir, inaccesibles al ámbito limitado de nuestra atención. Ello plantea un límite insuperable al análisis introspectivo aunque vaya acompañado de un gran rigor lógico. Simplemente, al ciudadano corriente o al filósofo de la mente le están completamente vedados algunos datos relevantes para entender la naturaleza de los procesos mentales. Sin embargo, las nuevas técnicas de investigación y el ingenio de los investigadores en el diseño de experimentos permite desvelar algunos de estos procesos.

Baste un ejemplo, en nuestro laboratorio investigábamos cómo se produce el reconocimiento de las palabras, y concretamente en qué medida las sílabas son entidades cognitivas necesarias para acceder a las representaciones léxicas (Carreiras, Álvarez y de Vega., 1993). Con esta finalidad, seleccionamos palabras compuestas de sílabas de alta frecuencia y palabras compuestas de sílabas de baja frecuencia. Por ejemplo *copa* está compuesta por 2 sílabas frecuentes (co=9144 /millón y pa= 666/millón), y *guardia* está compuesta por 2 sílabas infrecuentes (guar=201/millón y dia= 282/millón), pero ambas palabras globalmente tienen una frecuencia semejante (62 y 66 por millón, respectivamente). Posteriormente, valoramos el reconocimiento de ambos tipos de palabras en una tarea de decisión léxica. Los resultados fueron inesperados: las palabras compuestas de sílabas de baja frecuencia se reconocían *más deprisa* que las palabras compuestas de sílabas de alta frecuencia. Estos resultados contra-intuitivos tienen, sin embargo, una explicación: cada sílaba funciona como un código de acceso a todas las representaciones léxicas que comparten dicha sílaba. De modo que cuando la primera sílaba es más frecuente activa una cohorte de candidatos léxicos muy numerosa, mientras que si es una sílaba poco frecuente activa pocos candidatos. En cualquier caso, el procesamiento de la segunda sílaba de la palabra envía activación confirmatoria a un solo candidato, mientras que los demás candidatos deben iniciar entonces un proceso de inhibición. El proceso inhibitorio será más lento cuando la cohorte de candidatos es más numerosa, es decir, cuando la primera sílaba fue de alta frecuencia. Las investigaciones posteriores confirmaron la validez de este modelo de activación-inhibición en el procesa-

miento silábico. A partir de la simple reflexión racional, difícilmente se hubiera descubierto este mecanismo, dado que no tenemos ningún acceso consciente a los procesos léxicos. En cambio, los experimentos desvelaron un fenómeno nuevo y forzaron una explicación.

## Los nuevos paradigmas cognitivos

Algo se mueve en la psicología cognitiva en este final de siglo. Ya hemos mencionado que el conexionismo y la neurociencia cognitiva han caldeado de nuevo el ambiente, de modo que el clima intelectual es parecido al de la década de los 60 con el recién estrenado paradigma S-C. No disponemos de la perspectiva histórica de casi 40 años para valorar los nuevos paradigmas<sup>1</sup> y no conviene hacer profecías, de modo que me limitaré a tratar de ellos brevemente y a dese- arles suerte.

### *El conexionismo*

Si el paradigma S-C se desarrolló siguiendo la metáfora del ordenador el conexionismo parece más próximo a la metáfora del cerebro. En lugar de un único procesador central que ejecuta cálculos simbólicos secuencialmente, el conexionismo desarrolla modelos compuestos de un gran número de procesadores (o nodos) interconectados y que funcionan en paralelo. El origen histórico del conexionismo es tan antiguo, al menos, como el del paradigma S-C. Así en la década de 1940, el manifiesto matemático de Warren McCulloch y Walter Pitts sobre redes neurológicas fue seguido de la noción de «asambleas celulares» de Donald Hebb; y en 1962 Rosenblatt causó un gran revuelo al elaborar una prueba de un algoritmo consistente en una única capa de «perceptrones» que podían aprender a responder a varias tareas. Sin embargo, estos esfuerzos tempranos se vieron truncados poco después por el libro de Marvin Minsky y Seymour Papert *Perceptrons*, que mostraba la incapacidad de los perceptrones para resolver cálculos de orden secundario como el famoso problema XOR (exclusivo-or), una función de dos argumentos binarios que produce 1 cuando los *inputs* son diferentes (v.g., cuando uno es 0 y el otro 1) y produce 0 cuando son iguales. Este parón temprano en el desarrollo del conexionismo permitió que el paradigma C-S fuese el dominante durante varias décadas. Pero la situación volvió a cambiar en los 80, cuando David Rumelhart y James McClelland, entre otros, afrontaron el desarrollo de perceptrones multicapa, de modo que entre las unidades de *input* y las de *output*, se introdujeron unidades «ocultas». Ello fue posible gracias al descubrimiento del potente algoritmo de *retro-propagación* capaz de modificar,

1. El término «paradigma» es bastante impreciso y algunos consideran, entre ellos el propio Kuhn, que en psicología no ha habido hasta ahora auténticos paradigmas. Sin embargo, creo que el término es muy expresivo para describir sistemas conceptuales con un vocabulario, una formalización y una metodología bien diferenciadas.

mediante entrenamiento, los pesos de las unidades en estas redes multicapa. Con este equipamiento formal era ya posible resolver el problema XOR y otras cuestiones de interés cognitivo.

A partir de entonces, el conexionismo penetra con fuerza en los circuitos académicos de la psicología así como en los centros de inteligencia artificial. La terminología, las tablas, los gráficos, y las metáforas interpretativas han cambiado substancialmente en relación a la época anterior. Todo hace pensar que el conexionismo no es simplemente un nuevo modelo cognitivo dentro del paradigma S-C, sino un verdadero paradigma alternativo con una concepción muy diferente de las funciones cognitivas. La Tabla 1 recoge algunas de estas diferencias que apenas tenemos espacio para esbozar en este artículo.

TABLA 1. DIFERENCIAS ENTRE EL PARADIGMA S-C Y EL CONEXIONISMO

<i>Paradigma S-C</i>	<i>Conexionismo</i>
Unidades simbólicas	Unidades sub-simbólicas
Macro-procesamiento	Micro-procesamiento
Procesamiento serial	Procesamiento paralelo
Procesamiento discreto	Procesamiento continuo
Reglas explícitas	Reglas emergentes
No aprendizaje	Aprendizaje
Modelos no robustos	Modelos robustos

Las neuronas individuales no representan símbolos (no hay una neurona o conjunto de neuronas que corresponda al símbolo «*mesa*»); más bien es la pauta de actividad de grandes agrupaciones neuronales distribuidas en el cerebro lo que corresponde a una representación. El conexionismo trata de estudiar cómo se adquieren estas pautas de activación de múltiples unidades sub-simbólicas y, en consecuencia, ofrece un análisis de micro-procesamiento en contraste con el carácter «macro» del paradigma simbólico. Por ejemplo, el algoritmo de retro-propagación opera modificando la fuerza de las conexiones asociativas de las unidades subsimbólicas («neuronas» o «perceptrones»). Los sistemas conexionistas, al igual que el cerebro, realizan multitud de cálculos en paralelo realizados por una gran masa de unidades de procesamiento que operan simultáneamente. Además, se trata de un procesamiento continuo a través de la red, muy diferente de la ejecución secuencial en pasos discretos de programación propia del paradigma S-C. Uno de los rasgos más espectaculares del conexionismo es que no existe nada parecido a un conjunto de reglas «*si... entonces*» para explicar la cognición. El investigador conexionista no debe preocuparse por descubrir el «programa» que gobierna una determinada función cognitiva, sino que construye una red con pesos asociativos arbitrarios y luego la «*entrena*» mediante un gran número de ensayos en una determinada tarea, proporcionándole una retro-alimentación a sus respuestas. En cada ensayo, el sistema reajusta los pesos asociativos entre las unidades mediante la regla de retro-propagación u otra similar.



El resultado es que la propia red conexionista se «auto-programa» y las reglas de alto orden «emergen» como una consecuencia de los procesos de aprendizaje en el micro-nivel. Las redes conexionistas al igual que el cerebro son robustas de modo que, una vez entrenadas ante una determinada tarea, son resistentes al ruido. Así, responden correctamente ante estímulos no prototípicos e, incluso, mantienen un buen nivel funcional después de haberse «destruido» algunas unidades de la red. Esto contrasta con la rigidez funcional de los modelos S-C en los que cualquier alejamiento del prototipo de estímulo requiere una re-programación (o re-modelización) del sistema.

El conexionismo está en plena expansión y ha revitalizado la psicología cognitiva. Captura algunas de nuestras intuiciones sobre el funcionamiento del cerebro como computador biológico y pone de manifiesto que algunas de las dificultades del paradigma S-C no son un callejón sin salida. Sin embargo, el conexionismo tiene mucho camino por andar antes de demostrar que es el paradigma cognitivo del próximo siglo. Tanto las defensas como los ataques al conexionismo se nos antojan hasta el momento como planteamientos más bien ideológicos, basados en creencias, expectativas generales y hasta sentimientos románticos. No hay una meta-teoría computacional que nos permita conocer a priori qué es lo que puede hacer el conexionismo y cuáles son sus límites, de modo que el desarrollo del paradigma procede de un modo empírico, por ensayo y error. Tal como aseguran Minsky y Papert:

Ni el profano ni el especialista en computación parece reconocer cuán primitivo y empírico es nuestro estado de conocimiento sobre estos temas. No sabemos en qué medida puede incrementarse la velocidad de cómputo, en general, utilizando máquinas de procesamiento en «paralelo» en lugar de «serial» -o «análogo» en lugar de «digital» (Minsky y Papert, 1969, p. 2).

De momento, el conexionismo explica muy bien los procesos de aprendizaje de patrones, es decir, aquellos procesos en los que nuestro cerebro extrae patrones estadísticos a partir de un gran número de ensayos con ejemplares. Esto ha permitido ofrecer buenos modelos de aprendizaje de conceptos, del procesamiento ortográfico o del análisis gramatical. Sin embargo, no todo lo que hacemos las personas se reduce a un procesamiento «estadístico» del medio. También construimos representaciones singulares de sucesos y episodios específicos, y en estos casos cobra importancia la estructura temporal o secuencial y la configuración única (no iterativa) de conjuntos de parámetros. Por otra parte, los procesos de resolución de problemas tienen asimismo ese carácter secuencial y creativo que difícilmente puede reducirse a la extracción y aplicación de patrones estadísticos.

### *La neurociencia cognitiva*

Las ciencias del cerebro o neurociencias han cobrado un gran impulso y alcanzado enorme prestigio en los últimos años. Las revistas científicas genera-

les más importantes, como *Nature* o *Science*, con frecuencia publican estudios sobre los correlatos neuronales de la cognición y el último congreso internacional de Neurociencia reunió a más de 20.000 participantes. El renovado interés por la neurociencia ha coincidido con el desarrollo de nuevas tecnologías que ofrecen información sobre el cerebro humano en funcionamiento. Por ejemplo, las técnicas de neuro-imagen como PET-scan, y MRI(f) permiten registrar los procesos metabólicos cerebrales mientras una persona está realizando una determinada tarea cognitiva. Ello permite elaborar imágenes de alta resolución espacial sobre las regiones cerebrales que se activan en cada tipo de tarea. Por su parte, los ERP (*Event Related Potentials*) ofrecen un registro de EEG sincronizado con el procesamiento de estímulos, lo cual permite desvelar la «firma» electrofisiológica característica de algunos procesos cognitivos. Otras técnicas más tradicionales se basan en el estudio de los trastornos funcionales adquiridos en los pacientes con lesiones cerebrales, que proporcionan una información útil sobre la organización modular en el cerebro de algunas funciones como el lenguaje o la memoria. Estos datos se complementan con los estudios hechos con animales mediante técnicas de ablación cortical, implantación de microelectrodos, etc.

Estos avances son importantes pero, en mi opinión, la neurociencia cognitiva no constituye un paradigma diferenciado en el sentido en que lo es el conexionismo. La neurociencia constituye el nivel de análisis del *hardware* mientras que la psicología cognitiva supone el nivel *algorítmico*, tal como postulaba Marr. Más que plantear una oposición entre ambos niveles, es mejor considerarlos como formulaciones complementarias que aportan restricciones mutuas. Por una parte, los modelos funcionales sirven necesariamente de guía a las investigaciones de la neurociencia. Las estructuras y funciones postuladas por la psicología cognitiva son necesarias para plantear objetivos específicos a la investigación: las técnicas de neuroimagen son una «ventana» al cerebro, pero debemos disponer de un buen análisis funcional (cognitivo) para saber hacia dónde hay que mirar por esa ventana. En efecto, los sistemas y funciones descritas por los psicólogos, tales como los módulos del procesamiento del lenguaje, la memoria a corto plazo, los procesos atencionales, las imágenes mentales, o la conciencia visual, son algunos de los temas que la neurociencia ha heredado y asumido como propios. Además, los métodos de la neurociencia sólo parcialmente tienen que ver con el cerebro: la medida de los parámetros cerebrales constituye, generalmente, la variable dependiente (v.g.: qué regiones se activan y cuándo se activan), mientras que las variables independientes proceden de diseños de tareas típicamente cognitivos: tareas de decisión léxica, tareas de atención dividida, manipulación de listas de palabras, comprensión de oraciones, rotación de imágenes mentales, etc.

Por su parte la neurociencia contribuye a sancionar los modelos funcionalistas de la psicología cognitiva. Por ejemplo, las lesiones del área de Broca determinan trastornos en la articulación del habla y en el procesamiento sintáctico lo cual favorece determinadas concepciones modulares del lenguaje (v.g., Caramazza y Zurif, 1976); o bien las imágenes de PET-scan muestran una activación de áreas cerebrales similares durante la percepción de un objeto y durante la elaboración de una imagen mental de él, lo que apoya la noción de equivalencia funcional entre imaginación y percepción visual (Kosslyn, 1988). Además, la

neurociencia puede ofrecer sus propios datos que estimulen el análisis funcional; por ejemplo, la especialización hemisférica de algunas funciones, la activación sincrónica de neuronas de 40 ciclos por segundo, la naturaleza de los ERP, o el papel de las zonas de convergencia requerirán explicaciones algorítmicas o funcionalistas para lograr una plena comprensión de ellas.

La psicología cognitiva debe mantener una buena comunicación interdisciplinar con la neurociencia y estar al corriente de los avances de ésta en técnicas de investigación. En realidad, para un psicólogo de formación experimental siempre es informativo el análisis de los fenómenos mediante la «triangulación» que supone utilizar varios métodos convergentes. Siempre se obtendrá una información más rica si a los datos conductuales (v.g., tiempos de reacción o recuerdo) le añadimos los datos de topografía cerebral, los estudios sistemáticos de doble disociación de funciones con pacientes cerebrales, o los registros ERP. Ocasionalmente pueden ser útiles también los estudios de registro de la actividad electrofisiológica de neuronas individuales a partir de micro-electrodos implantados, aunque por razones obvias esto sólo es posible con animales como sujetos de experimentación. Mediante esta técnica se ha avanzado mucho en el estudio de las funciones que compartimos con otros mamíferos, por ejemplo se ha analizado minuciosamente la topografía funcional del sistema visual utilizando gatos y chimpancés como sujetos experimentales.

Es importante recordar, sin embargo, que las técnicas desarrolladas hasta el momento por la neurociencia proporcionan una información parcial sobre los fenómenos investigados. Así, las técnicas de neuroimagen nos aportan información de gran resolución espacial sobre la activación cerebral, pero carecen de suficiente resolución temporal, pues se requieren más de 15 segundos para obtener una imagen funcional. Es evidente que 15 segundos es un grano temporal muy grueso ya que el flujo de información cerebral (y consecuentemente de los procesos cognitivos) es muy rápido. De hecho los estudios cronométricos convencionales de la psicología nos muestran que muchos procesos significativos ocurren en tiempos muy inferiores a un segundo. Por su parte, el registro electrofisiológico de ERP tiene una magnífica resolución temporal del orden de milisegundos, pero en cambio no ofrece suficiente resolución espacial. Cabe esperar que los avances de la electrónica en los próximos años nos permitan un registro de la actividad cerebral durante la ejecución de tareas cognitivas, que tenga al tiempo una buena resolución temporal y espacial. Si esto fuera así, sin duda sería una gran noticia pero en ningún caso los datos neurofisiológicos podrán substituir por sí mismos a la psicología cognitiva (o si se prefiere a la neurociencia cognitiva). Siempre habrá necesidad de disponer de modelos interpretativos de los datos, y cuanto más minuciosos sean éstos más articulados y precisos deberán ser nuestros modelos psicológicos.

### **¿Por qué no existe HAL?**

La tecnología de los ordenadores ha avanzado mucho desde los años 60 y sin embargo éstos no parecen más «inteligentes» hoy en día, salvo en el sentido

limitado en que lo son los programas que juegan al ajedrez o las diversas aplicaciones informáticas. Desde luego no existe HAL ni los robots inteligentes, sólo algunas demostraciones de feria. Por supuesto también existen herramientas robóticas y programas «expertos», pero su funcionamiento rígidamente pre-programado los aleja de lo que entendemos por «inteligencia». Quizá esto no deba preocuparnos a los psicólogos al tratarse de un tema exclusivo de la IA. Pero lo cierto es que la IA es una tecnología cognitiva de modo que la escasez de sus logros refleja, en gran medida, la inmadurez de las ciencias cognitivas entre ellas la psicología.

Creo que parte de las insuficiencias de la psicología cognitiva y subsidiariamente de la IA tienen relación con la marginación del procesamiento episódico en favor del procesamiento semántico. En psicología se ha desarrollado un buen número de modelos de memoria semántica, del léxico mental, o de los esquemas. Todos estos modelos reflejan, sobre todo, el conocimiento que tenemos las personas sobre ciertas regularidades recurrentes del medio, que almacenamos como estructuras permanentes en la memoria. En algunos casos también se intenta explicar la génesis y modificación de estas estructuras de conocimiento, a menudo como un proceso de extracción de regularidades estadísticas a partir de ejemplares o de estímulos singulares. Por ejemplo, aprendemos el concepto *mesa* o el esquema del *restaurante* extrayendo un patrón promedio a partir de multitud de experiencias con mesas y restaurantes individuales, respectivamente. El propio conexionismo supone una continuidad en su preferencia por lo «semántico» de modo que ofrece una explicación computacional detallada del proceso de adquisición de pautas conceptuales genéricas.

Todo ello es importante, pues el cerebro es en gran medida un «procesador estadístico». Pero, además, e igualmente importante, el cerebro es un codificador de situaciones individuales. Así, no sólo tenemos un conocimiento genérico de lo que es un aparcamiento, sino que recordamos (o intentamos recordar) dónde hemos aparcado el coche esta mañana. O bien, tenemos conocimientos prototípicos sobre los roles sociales (camarero, profesor, etc.), sobre tipologías de individuos (un vago, un ambicioso, etc.), o disponemos de teorías implícitas sobre dominios humanos (el papel de la mujer, la familia, las clases sociales, etc.), pero también construimos modelos situacionales únicos sobre un contexto situacional particular. Esto último supone que debemos de codificar, registrar y actualizar un buen número de relaciones cambiantes, tales como «quién dijo qué», «quién hizo qué», «quién sabe qué», «qué intención tiene mi interlocutor», «qué relación hay entre mi interlocutor y otras personas», «en qué lugar está un objeto», «cuándo ocurrió tal suceso», «por qué ocurrió tal suceso», etc. La representación o modelado de situaciones es un problema computacional que las personas resolvemos muy bien y que los modelos psicológicos y de IA apenas tratan.

Naturalmente todos tenemos la intuición de que codificamos y recordamos sucesos o estímulos singulares. Pero en psicología cognitiva lo episódico se ha descrito, generalmente, como un mero proceso de reconocimiento de patrones guiado precisamente por las representaciones semánticas. Así, si yo reconozco una mesa es porque activo la categoría *mesa* ante este estímulo particular; o bien si entiendo un texto en el que se habla de un restaurante es porque activo

o «instancio» el esquema del *restaurante*. Sin embargo, considero este tipo de análisis profundamente erróneo, pues codificar situaciones es algo más que recuperar información de la memoria semántica. Cuando codificamos una situación representamos una configuración *única* de parámetros que es, además, *dinámica* ya que hemos de actualizarla de tiempo en tiempo, a medida que los parámetros situacionales van cambiando. Además la codificación situacional es la que nos garantiza la acción eficaz e inmediata sobre el entorno ecológico o interpersonal. Supongamos por un momento que disponemos de todo el conocimiento general del mundo propio de la memoria semántica (esquemas situacionales, de personas, categorías, destrezas motoras, conocimientos de psicología, etc.) y, sin embargo, no podemos construir representaciones episódicas o situacionales ni retenerlas a largo plazo. No recordaríamos nuestras metas e intenciones, ni nuestras citas, ni llevaríamos un registro de la experiencia que hemos compartido con nuestros amigos, ni sabríamos qué hemos hecho unas horas antes. Nuestras acciones podrían quizá entrar en absurdos bucles iterativos, pues volveríamos a buscar algo que ya hemos encontrado, a reclamar el libro que ya nos han devuelto, a telefonar a alguien con quien acabamos de hablar, a darle a la recién viuda cariñosos recuerdos para su marido, etc. Esta disfunción catastrófica, al tiempo que selectiva, existe de hecho en algunas amnesias retrógradas, producidas por lesiones del hipocampo. En estos pacientes cerebrales se observa una incapacidad en el procesamiento de sucesos cotidianos, pero no en el almacenamiento de conocimientos convencionales repetidos frecuentemente (v.g., Vargha-Khadem, *et al.* 1997).

La tarea del cómputo o modelado de situaciones es ingente porque un sistema que elaborase representaciones estáticas no sería viable. Constantemente se producen alteraciones en uno o varios parámetros situacionales y debemos de actualizar el modelo más o menos en tiempo real. Los sucesos físicos, las acciones de otros agentes animados, o nuestra propia acción sobre el entorno, junto a los cambios en nuestros estados mentales (metas, intenciones o emociones) determinan que la situación se modifique y tengamos que hacer una puesta al día de los modelos situacionales. Pero esto no es una tarea trivial como los teóricos de la IA y los filósofos de la mente han señalado (v.g., Pylyshyn, 1988). Ellos lo denominan el problema del marco (*frame problem*) y podríamos plantearlo así: a cada cambio en un parámetro significativo de la situación, el sistema (v.g., un robot) debe actualizar algunas «creencias» y mantener estables otras, pero no está claro *a priori* cuáles deben modificarse y cuáles deben mantenerse. Por ejemplo, frente a mí hay una taza de café sobre un plato con una cucharilla dentro. Si yo decidiese levantar el plato en alto me consta, y al lector también, que debo actualizar mi modelo situacional modificando la posición del plato, del café, de la taza y de la cucharilla, pues todos ellos se desplazarán juntos. Pero si yo decido levantar únicamente la cucharilla en alto entonces la actualización de mi modelo sólo afectará a la posición espacial de la cucharilla. Un robot debe ser capaz de comprender y anticipar las consecuencias de sus acciones sobre el mundo y de actualizar sus representaciones situacionales. En caso contrario cometería errores tales como coger la cucharilla cuando trata de coger la taza de café. El problema del marco radica

en que los algoritmos deductivos basados en el cálculo de predicados simplemente no podrían determinar qué «creencias» (v.g., enunciados proposicionales) hay que actualizar y cuáles permanecen inalteradas ante los cambios situacionales y las propias acciones.

La representación de episodios singulares no significa que los modelos situacionales no estén sujetos a restricciones. El número de modelos y de sus estados es potencialmente ilimitado (como lo es el número de situaciones posibles) pero los parámetros y pautas legales de modelado son finitos. Endel Tulving propuso en su momento que la memoria episódica se caracteriza por una organización espacial y temporal de la información. Sin duda estas dimensiones «kantianas» son necesarias para organizar la información en nuestros modelos episódicos, pero habría que añadir alguna más. La organización causal constituye un «pegamento» que vincula sucesos y acciones en nuestra representación episódica. Además hay algunas dimensiones interpersonales propias de la «teoría de la mente» que contribuyen a nuestro modelado situacional (Rivière, 1991). Cuando nos representamos situaciones interpersonales (y casi todas lo son en alguna medida) codificamos a los demás en términos de estatus, vínculos interpersonales, intenciones y metas, estados de ánimo, relaciones de pertenencia o propiedad, etc. Muchos de estos parámetros interpersonales son variables, es decir, que su estado cambia durante el desarrollo de la situación y por tanto hemos de actualizarlos.

Los estudios recientes en el marco de la teoría de los modelos situacionales proporcionan evidencias abundantes de que las personas construyen y actualizan este tipo de estructuras episódicas, y que los parámetros privilegiados del modelado son precisamente los que hemos citado (espacio, tiempo, causalidad, creencias, intenciones, etc.). Los estudios de comprensión del discurso son un dominio de investigación particularmente valioso para comprender las propiedades de los modelos situacionales, tal como demuestran los trabajos de Arthur Glenberg, Barbara Tversky o Tony Sanford, entre otros. La ventaja es sobre todo metodológica, ya que el discurso —especialmente si es narrativo— constituye una fuente de información que determina un modelado muy similar al modelado de experiencias con situaciones reales. Pero a diferencia de las situaciones reales, el discurso tiene una estructura relativamente simple (v.g., unimodal, discreta y secuencial) y se puede manipular sistemáticamente, lo cual permite la investigación del curso temporal del modelado.

Hemos de reconocer que no disponemos de nada parecido a una teoría algorítmica de los modelos situacionales y esta es una tarea que no debe demorarse más tiempo. Los procesos de modelado son un tanto paradójicos pues parecen tener una doble naturaleza. Por una parte, los modelos obedecen a un principio de «constricciones múltiples» que nos aproximan a una noción conexionista del cómputo; pero también tienen un carácter direccional, que los aproxima a una formalización en términos de reglas «si... entonces». Veamos estas frases:

- (1) Luis está cerca de Barcelona
- (1') Luis está cerca de la mesa
- (2) Ahora tomaré un café
- (2') Ahora me casaré

Seguramente entendemos que la distancia *cerca de* es de algunos kilómetros en (1) y de metros o centímetros en (1'); mientras que *ahora* corresponde a segundos en (2) y a días o semanas en (2'). Este escalamiento espacial y/o temporal parece posible por que somos capaces de introducir en nuestros modelos parámetros no explícitos que se constriñen mutuamente. Así, el tamaño y el entorno implícito de «Barcelona» y «mesa» permiten establecer las diferentes escalas de distancia, mientras que la persistencia típica de la meta táctica de «tomar café» y la meta estratégica de «casarse» en nuestra memoria, determinan que establezcamos diferentes magnitudes del presente psicológico. Nuestros conocimientos implícitos de las sustancias, los tamaños, las formas, las duraciones de los sucesos, las expresiones faciales, los marcadores gramaticales, etc. seguramente se constriñen mutuamente de modo que establecemos una fina red de microinferencias cada vez que entendemos oraciones y establecemos las escalas dimensionales oportunas. Todo esto tiene un aroma francamente conexionista. Pero al tiempo, los cursos de modelado son anisotrópicos de modo que algunos parámetros no se influyen de modo indiferenciado sino según pautas direccionales y asimétricas. Por ejemplo, las causas preceden a los efectos pero no viceversa; una vez satisfecha una meta ésta debe quedar inactiva; las personas, pero no los objetos, generan acciones; un contenedor debe tener más volumen que un contenido, etc. Los clásicos sistemas de producción, las reglas condicionales, o los postulados de significado parecen más apropiados que los modelos conexionistas para expresar estas relaciones anisotrópicas. Quizá debamos aceptar que la futura teoría de los modelos situacionales puede tener una doble formulación computacional, del mismo modo que la descripción de la naturaleza de la luz se hace en términos de partículas o de ondas, dependiendo del dominio de fenómenos que se pretende explicar.

En cualquier caso, una teoría bien articulada de los modelos situacionales o del procesamiento episódico contribuiría, con toda probabilidad, a resolver varios *puzzles* académicos tales como el ya mencionado problema del marco o la referencialidad del lenguaje y la conceptualización (el problema del *grounding*), sin contar las posibilidades que se abrirían a la inteligencia artificial.

## Consideraciones finales

En estas páginas he ofrecido una visión personal del estado de la psicología cognitiva en este final de siglo. No he pretendido ser exhaustivo sino plantear algunas líneas de reflexión a modo de ensayo, más que de revisión académica convencional. Un análisis más sistemático de la psicología cognitiva actual hubiera requerido muchas más decenas de páginas. Mi reflexión se ha basado, por tanto, en una selección de temas. Más que detenerme a considerar los éxitos de la psicología cognitiva en algunos campos como, por ejemplo, la atención, la memoria, o los procesos del lenguaje, me he planteado algunas de sus limitaciones, y también he hablado de una cierta crisis del paradigma S-C. Pero ni las limitaciones ni las crisis señalan una decadencia del cognitivismo. La psicología cogni-

tiva ha experimentado importantes transformaciones internas pero de ellas ha salido reforzada. La emergencia de la alternativa conexionista y el empujón de las neurociencias son dos muestras de su vitalidad. Quizá no sea hoy tan común como hace unos años utilizar la expresión «psicología cognitiva» o escribir manuales con esta denominación disciplinar genérica. La *Introducción a la Psicología Cognitiva* que publiqué hace unos años (de Vega, 1984) creo que reflejaba con relativa fidelidad el estado de la disciplina a comienzos de la década de los 80. Hoy creo que no sería posible escribir un libro equivalente de sólo 500 páginas. Los nuevos manuales didácticos tratan de temas más especializados como psicolingüística, percepción, memoria, conexionismo, neuropsicología cognitiva, etc. La psicología cognitiva, pues, ha tenido tanto éxito que se ha desmembrado en una serie de disciplinas cognitivas cada una de las cuales tiene sus temas fundamentales, sus debates, sus métodos y sus comunidades de investigadores.

## REFERENCIAS

- Carreiras, M., Álvarez, C. & de Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, 32, 6, 766-780.
- Caramazza, A. & Zurif, E.B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3, 572-582.
- Damasio, A. (1994). *Descartes' error*. New York: Grosset/Putnam Book.
- De Vega, M. (1984). *Introducción a la Psicología Cognitiva*. Madrid: Alianza Editorial.
- Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge, Mass: MIT Press.
- García-Albea, J.E. (1991). Entrevista con Jerry Fodor. *Revista de Occidente*, 119, 17-60.
- Kosslyn, S.M. (1988). Aspects of a cognitive neuroscience of mental imagery. *Science*, 243, 1627-1629.
- Minsky, M.L. & Papert, S.A. (1969). *Perceptrons*. Cambridge, Mass: MIT Press.
- Pylyshyn, Z. (1988) (Ed.). *The robot's dilemma*. Norwood, NJ: Ablex.
- Rivière, A. (1991). *Objetos con mente*. Madrid: Alianza.
- Stork, D.G. (1997, Ed.). *HAL's legacy*. Cambridge, Mass: MIT Press.
- Vargha-Khadem, F., Gadian, D.G., Watkins, K.E., Connelly, A., Van Paesschen, W. & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, Julio, 376-380.