

II. CIENCIAS COGNITIVAS

En defensa de la teoría de la representación en ciencias cognitivas

PASCUAL F. MARTÍNEZ-FREIRE
Universidad de Málaga

I. EL CAMPO INTERDISCIPLINAR DE LAS CIENCIAS COGNITIVAS

PODEMOS DECIR, EN TÉRMINOS generales, que las ciencias cognitivas constituyen un campo de investigación interdisciplinar cuyo tema central y común es el estudio del fenómeno de la cognición tanto en seres humanos, como en máquinas o en animales. Y esto nos obliga, claro está, a caracterizar la cognición.¹

Conviene distinguir dos sentidos del término “cognición”, uno inmediato y aproximado, y otro más elaborado y adecuado. En primera instancia, “cognición” puede entenderse como sinónimo de “conocimiento”, entendiéndose en ambos casos recepción de algún tipo de información. Pero, en segunda instancia, “cognición” quiere decir uso y manejo de conocimiento, esto es, uso y manipulación de información. Tanto en el primero como en el segundo sentido de “cognición” nos encontramos con procesos portadores de información, de tal manera que la distinción entre simple conocimiento y cognición propia no es tajante, pero resulta claro que no pertenecen al mismo nivel cognitivo una sensación, por ejemplo de calor asfixiante, y una creencia, por ejemplo que una ducha fría aliviará tal sensación de calor. En el primer caso recogemos datos siguiendo una línea de operación básicamente pasiva, mientras que en el segundo caso manipulamos conocimientos diversos siguiendo una línea de operación básicamente activa.

Dicho de forma más precisa, la cognición propia se caracteriza por ser un fenómeno interno fundamentalmente activo que procesa información. Es un fenómeno interno (al igual que el simple conocimiento) en cuanto es causa interna de una conducta, externa y observable, como cuando mi creencia de

¹ En los dos primeros capítulos de mi libro *La importancia del conocimiento. Filosofía y ciencias cognitivas* (2007) se analiza la naturaleza de las ciencias cognitivas. Algunas ideas de este apartado y del siguiente se toman de estos capítulos.

que lloverá causa que tome un paraguas, aunque puede no manifestarse en conducta, como cuando mi creencia de que debo estudiar no me lleva a estudiar. Pero además, y éste es un punto de vista nuevo aportado por las ciencias cognitivas, la cognición es algo compartido por seres humanos, algunos animales (probablemente desde los vertebrados) y algunas máquinas adecuadamente programadas (computadores y robots); por ejemplo, no sólo hay cognición en un humano que hace planes para sus vacaciones de verano, sino que también hay cognición en un mono que recuerda en qué recipiente se oculta un plátano o en un computador que demuestra un teorema matemático. En realidad, el hecho de haber dado un nuevo sentido al término “cognición”, de uso raro anteriormente, para referirnos a procesos que hubieran podido ser denominados como “conocimiento”, está justificado, en mi opinión, por la doble novedad acompañante, proporcionada por las ciencias cognitivas, de entender que los diversos procesos de conocimiento se explican como procesamiento de información y de entender que tal procesamiento de información se produce tanto en humanos como en algunos animales y en algunas máquinas.

En cuanto a las ciencias integrantes del campo de investigación cognitiva, es preciso distinguir entre aquellas disciplinas científicas que constituyen el núcleo del campo de investigación y aquellas otras disciplinas implicadas de manera instrumental o aplicada, de tal modo que éstas no están incluidas en su totalidad en el campo de investigación cognitiva. Las ciencias cognitivas básicas son la ciencia de la inteligencia artificial y la psicología de orientación cognitiva, no conductista.

La ciencia de la inteligencia artificial o, más brevemente, la inteligencia artificial (IA) puede caracterizarse, en general, como aquella rama de la informática (o ciencia de la computación) que estudia la teoría, diseño y construcción de máquinas inteligentes, entendiéndose por tales los mecanismos artificiales capaces de ejecutar tareas que en los humanos atribuimos a su inteligencia. Por ejemplo, los computadores dotados de programas para jugar al ajedrez o dotados con programas para diagnosticar e indicar el tratamiento para curar una enfermedad, pueden considerarse máquinas inteligentes. En el verano de 1956, en el Dartmouth College, en Hanover (New Hampshire), un grupo de diez matemáticos y lógicos celebraron varias reuniones, con la idea (ingenua y publicitaria) de que los rasgos característicos de la inteligencia humana podían ser descritos de forma tan precisa que podían ser simulados por un computador. Estas reuniones, conocidas como “conferencia Dartmouth”, fueron el inicio definitivo de la inteligencia artificial.

A su vez, la psicología cognitiva puede considerarse constituida en 1960, cuando los psicólogos George Miller y Jerome Bruner fundan el Harvard Center for Cognitive Studies en la Universidad de Harvard. Además en ese mismo año se publica el libro *Plans and the Structure of Behavior*, escrito por George

Miller, Eugene Galanter y Karl Pribram, que puede considerarse como un auténtico manifiesto del cognitivismo en psicología. Unos pocos años más tarde, en 1967, Ulric Neisser publica *Cognitive Psychology*, que es el primer libro de texto importante de la nueva psicología. Para la psicología de orientación cognitiva, los procesos mentales (tanto en humanos como en animales) son procesos cognitivos en el sentido de procesadores de información.

Pero no son la inteligencia artificial y la psicología las únicas ciencias implicadas en el campo de investigación cognitiva. También debe contarse con otras disciplinas que, aunque no agotan sus intereses en el fenómeno de la cognición, constituyen importantes recursos o instrumentos o bien encuentran aplicaciones del estudio de la cognición. Tales ciencias son la neurociencia, la lógica y la lingüística, pero también la pedagogía y la antropología.

Creo que fueron Daniel Bobrow y Allan Collins los primeros en utilizar y divulgar la denominación de “ciencia cognitiva” en el libro que editaron titulado *Representation and Understanding. Studies in Cognitive Science* (1975). En su prefacio dicen: “Este libro contiene estudios en un nuevo campo que llamamos ciencia cognitiva. La ciencia cognitiva incluye elementos de psicología, ciencia de la computación, lingüística, filosofía y educación, pero es más que la intersección de estas disciplinas”. Asimismo en 1975, Donald Norman, David Rumelhart y el Grupo de Investigación LNR publicaron *Explorations in Cognition*, donde dicen al final del libro: “Los esfuerzos concertados de un número de personas procedentes de las disciplinas relacionadas de la lingüística, inteligencia artificial y psicología pueden estar creando un nuevo campo: ciencia cognitiva”.

Finalmente, las ciencias cognitivas se consolidan definitivamente cuando a finales de los años setenta del pasado siglo adquieren un carácter institucional al crearse una revista y una sociedad científica para la investigación cognitiva. En efecto, en 1977 se crea la revista *Cognitive Science*² y dos años más tarde se reunió por primera vez la Cognitive Science Society en La Jolla (California).

II. EL SUPUESTO BÁSICO COMÚN DE LAS CIENCIAS COGNITIVAS

Tanto la psicología como la inteligencia artificial, así como el resto de las ciencias cognitivas en cuanto tales, comparten una misma tesis general o supuesto, a saber, la idea de que cualquier agente cognitivo es un sistema de procesamiento de información (SPI).

2 Posteriormente han sido varias las revistas que se han incorporado a la difusión de la investigación en ciencias cognitivas, como *Behavioral and Brain Sciences*, *Cognition*, *Mind and Language*, *Mind*, *Minds and Machines*, *Cognition and Emotions*, *Cognitive Psychology*, *Artificial Intelligence* o *Cognitive Neuroscience*.

Allen Newell (1927-1992) y Herbert Simon (1916-2001), dos de los fundadores del campo de investigación cognitiva, publicaron en 1972 su obra conjunta *Human Problem Solving*, donde ya aparece una caracterización de un SPI. En el epílogo de este libro se declara que el solucionador humano de problemas y el computador programado son ambas especies del género sistema de procesamiento de información.

Por otro lado, en el trabajo de Newell y Simon titulado “Computer Science as Empirical Inquiry: Symbols and Search” (1976) aparece formulada la denominada “hipótesis del sistema de símbolos físico (SSF)”, que es equivalente a la tesis del sistema de procesamiento de información, de tal manera que un SPI coincide con un SSF. Tal hipótesis sostiene que un sistema de símbolos físico tiene los medios necesarios y suficientes para la acción inteligente general. Con ello se defiende que cualquier sistema que exhiba inteligencia general resultará ser un sistema de símbolos físico y que, conversamente, cualquier sistema de símbolos físico de suficiente tamaño puede organizarse para exhibir³ inteligencia general. A su vez, se entiende por acción inteligente general el mismo alcance de inteligencia que vemos en la acción humana, es decir, una conducta apropiada a los fines del sistema y adaptativa a las exigencias del entorno.

Ahora bien, para una detallada y clara descripción de un SSF debemos dirigirnos al trabajo de Allen Newell titulado precisamente “Physical Symbol Systems” (1980). Aquí se establece que un sistema de símbolos físico consta de una memoria, de un conjunto de operadores (describiéndose diez, que luego indicaremos), de un control, de una entrada y de una salida. Asimismo un SSF es una máquina (en sentido muy general) que existe en un ambiente que consta de objetos distribuidos en un espacio de ubicaciones. Podemos precisar que las entradas al SSF son informaciones sobre los objetos, mientras que las salidas son informaciones sobre la modificación o creación de objetos. La conducta externa del sistema de símbolos consiste en las salidas que produce como función de las entradas; pero el estado interno del sistema consiste en el estado de su memoria y el estado de su control, mientras que su conducta interna consiste en la variación de ese estado interno a lo largo del tiempo.

Es conveniente introducir algunos comentarios. En primer lugar, esta hipótesis del sistema de símbolos físico (SSF) es una formulación más detallada de la tesis del sistema de procesamiento de información (SPI), de tal modo que

3 Me parece importante señalar que la utilización reiterada del término “exhibir” por Newell y Simon puede entenderse como definiendo una tesis filosófica neutra o intermedia en la cuestión de si los computadores son inteligentes. Entre la tesis fuerte de que los computadores son realmente inteligentes y la tesis débil de que los computadores simplemente simulan ser inteligentes, cabe la tesis neutra (o previa) de que los computadores exhiben de alguna manera inteligencia.

un agente cognitivo puede entenderse indiferentemente como un sistema de símbolos o como un sistema de procesamiento de información. Para advertir que ésta es la idea de Newell y Simon, basta con comparar la figura 2.1 de *Human Problem Solving* (1972), donde se presenta la estructura general de un SPI, con la figura 2 de "Physical Symbol Systems" (1980), donde se presenta la estructura de un SSF paradigmático, ya que las coincidencias son obvias. En segundo lugar, se trata de una hipótesis, es decir, se postula que cualquier sistema de símbolos, tal como queda descrito, está dotado de inteligencia, y que cualquier agente inteligente es un sistema de símbolos, quedando abierta a la discusión y a la investigación empírica esta identificación entre sistemas de símbolos y agentes inteligentes.

También conviene analizar con mayor detalle la propia noción de SSF. Para empezar, y frente a ciertas confusiones y acusaciones infundadas, un sistema de símbolos físico no es simplemente un mecanismo sintáctico sino que incluye explícitamente aspectos semánticos y de interpretación. En efecto, el control del sistema interpreta de modo continuo cualquier expresión, interpreta tanto los datos como los programas, y dentro de éstos tanto los símbolos como sus roles. Los operadores descritos por Allen Newell, en su trabajo de 1980, son diez, de los cuales cinco están asociados a la unidad de entrada (receptores) y otros cinco están asociados a la unidad de salida (motores). Los cinco primeros son entrar, leer, escribir, copiar y asignar, mientras que los otros cinco son hacer (internamente), salir-si, continuar-si, citar y actuar (externamente). Entre unos y otros operadores se sitúa la unidad de control, que está en relación permanente con la unidad de memoria. Ahora bien, el operador de asignar desempeña un papel muy importante, ya que establece una relación básica entre un símbolo y la entidad que designa; más aún, el concepto de designación o representación es, para Newell, el concepto más fundamental para un sistema de símbolos. En suma, un SSF se entiende como un mecanismo típicamente representacional.

En segundo lugar, y también frente a ciertas confusiones y acusaciones infundadas, un sistema de símbolos físico no es un agente encerrado en sí mismo, sino un mecanismo dentro de un ambiente y en medio de otros objetos. Justamente por ello dispone de una unidad de entrada y de otra de salida.

Finalmente, resulta claro que la hipótesis del sistema de símbolos físico se aplica por igual a la psicología cognitiva y a la inteligencia artificial. En efecto, un SSF (o SPI) es una descripción abstracta tanto de un sujeto humano o animal como de un computador. En todos estos casos, podemos hablar de unidad de entrada (receptores o sensores), de unidad de control o procesador con sus diversos operadores, de memoria, y de unidad de salida (efectores o motores). Y también en todos estos casos podemos hablar por igual de agentes procesadores de información.

III. ¿ALTERNATIVA CONEXIONISTA?

La inteligencia artificial fundada en la conferencia Dartmouth, en 1956, por John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon, entre otros, es considerada actualmente como la inteligencia artificial clásica. Para ella, los procesos inteligentes son básicamente procesos de razonamiento y, a su vez, razonar es manipular símbolos de manera automática e interna, con lo que para la IA clásica el conocimiento es codificado en estructuras simbólicas.

Ahora bien, desde 1986 se ha consolidado otra poderosa corriente dentro de la inteligencia artificial, que se denomina IA conexionista (o también redes neuronales artificiales), para la cual el conocimiento se codifica en los pesos (o fuerzas) variables de las conexiones entre unidades simples de procesamiento; es decir, el conocimiento no se codifica en estructuras determinadas de símbolos sino en valores numéricos variables.

Otra característica de la IA conexionista es su intención expresa de crear programas en los que el procesamiento de la información imite al cerebro humano, con lo que el procesamiento no será serial o sucesivo (como en la IA clásica) sino paralelo, y además tal procesamiento será realizado por una red de elementos simples (neuronas artificiales o neurodos) masivamente interconectados.

Dejando de lado los trabajos pioneros, las redes neuronales artificiales iniciaron su camino seguro a partir de 1986, cuando David Rumelhart y James McClelland, encabezando un grupo de investigación de psicólogos e informáticos, publicaron dos volúmenes de trabajos titulados *Parallel Distributed Processing*.

Desde el afianzamiento de la IA conexionista cabe distinguir entre programas de computación clásicos y neurocomputaciones.⁴ Maureen Caudill y Charles Butler, en su obra conjunta *Naturally Intelligent Systems* (1990), señalan las nueve características siguientes de una neurocomputación o red neuronal artificial (RNA). En primer lugar, una RNA se compone de un número de elementos de procesamiento muy simples (neurodos o neuronas), que se comunican mediante interconexiones con fuerzas o pesos variables (comparables a las fuerzas sinápticas de las neuronas naturales). En segundo lugar, en una RNA los datos o memorias se almacenan como patrones de pesos de interconexión variables entre los neurodos, de tal manera que la información se procesa mediante patrones de actividad, distribuida entre muchos neurodos, que se propagan y cambian constantemente. En tercer lugar, una RNA es enseñada o entrenada, más que

4 En el capítulo 9, "Historia y Filosofía de la Inteligencia Artificial", de mi libro ya citado *La importancia del conocimiento* (2007) se consideran en detalle el desarrollo y características de la inteligencia artificial. Algunas ideas allí expuestas se reiteran aquí.

propiamente programada, siendo posible que una RNA aprenda por ensayo y error.⁵ En cuarto lugar, las operaciones en una RNA están controladas por tres elementos, a saber, una función de transferencia de los neurodos (que relaciona entradas y salidas), los detalles de las conexiones entre los neurodos, y la ley de aprendizaje que sigue el sistema. En quinto lugar, una RNA actúa de modo natural como una memoria asociativa agrupando ítems similares. En sexto lugar, una RNA es capaz de generalizar a partir de ejemplos concretos. En séptimo lugar, una RNA es capaz de seguir funcionando aunque parte de sus neurodos y conexiones se hayan convertido en defectuosos. En octavo lugar, una RNA procesa patrones espaciales que dependen del tiempo. Y en noveno lugar, una RNA puede ser autoorganizadora, pudiendo generalizar por sí misma.⁶

Comparando computación clásica y computación conexionista, un programa clásico es *digital*, en cuanto opera sobre trozos discretos (determinados) de datos, es *serial*, en cuanto que sigue una secuencia de instrucciones específicas sobre cómo procesar los datos, y tiene cierto carácter *inflexible*, en cuanto que sigue un ciclo fijo de búsqueda de instrucción (junto con los datos necesarios), ejecución de instrucción y almacenamiento del resultado. Justamente estas tres características definen lo que se denomina una “arquitectura Von Neumann”, en honor a John Von Neumann (1903-1957), célebre científico húngaro-americano de la computación. A su vez, las RNA no poseen ninguna de estas tres características, por lo que su estructura no es de tipo Von Neumann. En vez de digitales son *análogas* (con valores continuos o variables), en vez de seriales son *paralelas* (con varias rutas de procesamiento en paralelo) y en vez de inflexibles son *flexibles* (con una gran flexibilidad de ejecución).

Desde la consolidación de la IA conexionista, dentro de la psicología cognitiva cabe distinguir entre los modelos computacionales clásicos y los modelos computacionales conexionistas, correspondiendo a la distinción entre IA clásica e IA conexionista.

Algunas observaciones del filósofo Andy Clark y del psicólogo e informático Paul Smolensky pueden aclararnos el diferente tipo de representación propio del paradigma clásico y del paradigma conexionista.

Para Andy Clark, dada una oración como “la pelota rompió la ventana”, un sistema de IA clásica buscará una representación para “pelota” y otra para “ventana”, mientras que un sistema de IA conexionista buscará como representación de esa oración un patrón de unidades activas, donde cabe la posibilidad

5 Es decir, no hay un programa predeterminado, sino que la RNA construye su propio programa mediante un aprendizaje supervisado o aprendiendo por sí misma.

6 El investigador más destacado en las RNA autoorganizadoras es Teuvo Kohonen, de la Universidad de Tecnología de Helsinki, autor del libro clásico *Self-Organization and Associative Memory* (1989).

de aislar un subconjunto de tal patrón que representaría “pelota”. Ahora bien, tal subconjunto variará de contexto a contexto, de tal manera que “pelota” tal como aparece en “la pelota rompió la ventana” tendrá una representación diferente (aunque parcialmente coincidente) con “pelota” tal como aparece en “el niño sujetó la pelota”.⁷

A su vez, Paul Smolensky, en su trabajo “On the Proper Treatment of Connectionism” (1988), denomina “paradigma simbólico” a la aproximación clásica en IA, ya que en ella las descripciones están constituidas por entidades que son símbolos tanto en el sentido semántico de referir a objetos externos como en el sentido sintáctico de ser operados mediante manipulación simbólica. Por otro lado, Smolensky denomina “paradigma subsimbólico” a la aproximación conexionista en IA, ya que en ella las descripciones están constituidas por entidades que corresponden a los elementos constituyentes de los símbolos usados en el paradigma simbólico, de modo que estas entidades de nivel más profundo, llamados subsímbolos, corresponden a las actividades de las unidades de procesamiento de las redes conexionistas. En una de sus tesis fundamentales, Smolensky sostiene que mientras que en el paradigma simbólico (IA clásica) el contexto de un símbolo se manifiesta en torno a él y consta de otros símbolos, en el paradigma subsimbólico (IA conexionista) el contexto de un símbolo se manifiesta dentro de él y consta de subsímbolos.

Entre 1988 y 1995 se desencadenó un intenso debate entre detractores y defensores de la IA conexionista en psicología. Por una parte, Jerry Fodor, Zenon Pylyshyn y Brian McLaughlin criticaron la psicología conexionista acusándola, en especial, de no explicar adecuadamente todas las propiedades del lenguaje. Por otra parte, Paul Smolensky ha defendido la posición conexionista en diversos trabajos.

Mi punto de vista en la discusión entre psicólogos cognitivos clásicos y conexionistas, tal como sostuve en mi libro *La nueva filosofía de la mente* (1995), es que unos y otros son representantes de recursos y metodologías diferentes dentro de un mismo paradigma general, a saber, el paradigma centrado en la idea de que los procesos mentales son estados internos que manipulan información. Asimismo compete directamente a la IA evaluar ambas metodologías en cuanto técnicas y, a su vez, compete a la psicología cognitiva decidir la aplicación de una u otra según criterios de oportunidad. Por lo tanto, parece claro que el conexionismo no es una alternativa excluyente de la tesis de que cualquier agente cognitivo es un sistema de procesamiento de información.

Además creo que el trabajo de Paul Smolensky titulado “Reply: Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Cognitive Architecture” (1995) contiene elementos suficientes para firmar la paz entre

7 Cf. Andy Clark, *Microcognition* (1989), p. 113.

psicólogos clásicos y psicólogos conexionistas.⁸ En efecto, Smolensky propone una arquitectura integrada conexionista/simbólica que defiende los cinco principios siguientes: 1) en todos los dominios cognitivos, los procesos se describen mediante algoritmos para propagar activación entre unidades conexionistas, 2) en todos los dominios cognitivos, las representaciones son patrones de actividad distribuida, 3) en partes centrales de los dominios cognitivos superiores, las representaciones son estructuras de símbolos, 4) las interpretaciones semánticas de estas representaciones simbólicas están derivadas de modo composicional de su estructura sintáctica, y 5) en partes centrales de los dominios cognitivos superiores, las funciones de input/output computadas por procesos cognitivos son descritas mediante funciones simbólicas (recursivas). Claramente 1) y 2) son principios conexionistas mientras que 3), 4) y 5) son principios simbólicos o clásicos.

Paul Smolensky recibió en 2005 el Premio David Rumelhart, convocado por la Cognitive Science Society, justamente por sus logros en una arquitectura integrada conexionista/simbólica, en la que las representaciones y operaciones simbólicas se manifiestan como abstracciones de las redes conexionistas subyacentes.⁹ John R. Anderson, que recibió el Premio David Rumelhart en 2004, sostenía en su prestigioso manual *Cognitive Psychology and its Implications* (1995) que en la primera mitad de los años noventa se había producido un reconocimiento creciente de que la situación en psicología no debía ser considerada como un enfrentamiento entre explicaciones conexionistas y simbólicas sino que los dos tipos de explicaciones son realmente complementarios.¹⁰

En todo caso la idea de representación, aneja a la tesis del sistema de procesamiento de información, está claramente presente en el conexionismo, tal como puede advertirse, por ejemplo, con suficiente detalle en el trabajo de Paul Smolensky “Representation in Connectionist Networks” (1990). Para este psicólogo, una representación conexionista es una correspondencia o mapeo entre vectores de actividad numérica y elementos del dominio del problema. Además, al igual que en cualquier formalismo, la estructura conexionista tiene sintaxis y semántica; la sintaxis describe el tipo de objetos que existen en el formalismo y cómo interactúan entre sí, mientras que la semántica describe cómo los objetos del formalismo se relacionan con los objetos fuera del formalismo a los que (de algún modo) se refieren. De nuevo parece claro que el conexionismo no es

8 En todo caso, Jerry Fodor no quiso firmar la paz, ya que en 1996 publicó “Connectionism and the Problem of Systematicity (Continued): Why Smolensky’s Solution *Still* Doesn’t Work”.

9 Otro proponente destacado de la integración entre paradigma conexionista y paradigma simbólico es Gary Marcus, autor de *The Algebraic Mind* (2001).

10 John R. Anderson, *Cognitive Psychology and its Implications* (1995), p. 17.

una alternativa excluyente de la tesis de que cualquier agente cognitivo es un sistema de procesamiento de información.

IV. EL ENFOQUE ENACTIVO EN CIENCIAS COGNITIVAS

Francisco J. Varela (1946-2001) fue un notable biólogo chileno, muerto a los 54 años en París, donde desempeñaba el cargo de director de investigación en el Centre National de Recherche Scientifique. Inicialmente Varela había trabajado con su profesor Humberto Maturana, en la Universidad de Chile (Santiago), en su teoría conjunta de la autopoiesis, pero posteriormente se interesó por las ciencias cognitivas, intentando aplicar en ellas algunas de sus tesis biológicas.

En su libro *Connaître: Les sciences cognitives* (1988) y más tarde en su libro conjunto con el filósofo Evan Thompson y la psicóloga Eleanor Rosch, titulado *The Embodied Mind. Cognitive Science and Human Experience* (1991), distingue tres enfoques posibles en las ciencias cognitivas. El primer enfoque es denominado por Varela “cognitivismo”, y con esta etiqueta se refiere a la teoría informacional clásica de la cognición, tal como fue defendida por Herbert Simon o es respaldada por Marvin Minsky (y que hemos considerado en el apartado II). A su vez, el segundo enfoque en ciencias cognitivas es el conexionismo (que hemos analizado en el apartado III). Y finalmente, el tercer enfoque de las ciencias cognitivas es el defendido por Varela, tanto en solitario en su obra de 1988 como junto con Thompson y Rosch en el libro citado de 1991.¹¹

El enfoque enactivo (*enactive approach*) se presenta como una alternativa tanto al cognitivismo como al conexionismo. Para éstos, el criterio de cognición sigue siendo una representación atinada de un mundo externo que está dado de antemano, pero para Francisco Varela las cuestiones relevantes que surgen en nuestras vidas no son predefinidas, sino cuestiones enactuadas,

11 En realidad, salvo algunas diferencias en la presentación de su propio enfoque, las tesis básicas se repiten en ambos libros, aunque el objetivo de uno y de otro es distinto. En el libro de 1988, Varela se propone criticar la tesis clásica del conocimiento como procesamiento de información introduciendo su tesis de la cognición como enacción. En cambio, en el libro de 1991, los autores (aunque asumiendo este enfoque enactivo) proponen construir un puente entre la mente en la ciencia y la mente en la experiencia, mediante la articulación de un diálogo entre la tradición de las ciencias cognitivas occidentales y la psicología meditativa budista. De hecho, el interés por el budismo es común a los tres autores. Por mi parte, tal como ya hice en mi trabajo “El enfoque enactivo en las ciencias cognitivas” (2006), algunas de cuyas ideas reitero en este estudio, me limitaré a presentar y criticar el enfoque enactivo, prescindiendo de las tesis que relacionan ciencias cognitivas y budismo.

emergidas o alumbradas desde un trasfondo.¹² La cognición deja de ser ahora un mecanismo que resuelve los problemas mediante representaciones y, en su lugar, alumbrando un mundo donde el único requisito es que la acción sea efectiva. Entonces el programa de investigación en ciencias cognitivas puede sintetizarse en tres proposiciones: 1) la cognición es acción efectiva (enacción), historia del acoplamiento estructural que alumbrando un mundo, 2) la cognición funciona a través de una red de elementos interconectados capaces de cambios estructurales durante una historia ininterrumpida, y 3) sabemos que un sistema cognitivo funciona adecuadamente cuando se transforma en parte de un mundo o configura un mundo nuevo.

En sus dos libros (aunque con más detalle en el de 1991) Varela presenta el autómata celular Bittorio como un ejemplo de sistema autónomo y a fin de aclarar la noción de acoplamiento estructural. Según nuestro autor, las ciencias cognitivas se están alejando de la idea de la mente como un mecanismo de entrada-salida que procesa información, para llegar a la idea de la mente como una red emergente y autónoma. Y su ejemplo de sistema autónomo es Bittorio, un anillo de autómatas celulares.¹³ Bittorio tiene, por ejemplo, la siguiente forma de acoplamiento estructural con el medio. El anillo se deja caer en un medio de ceros y unos aleatorios; cuando una de las células del autómata se encuentra con un cero o un uno, su estado es reemplazado por la perturbación que encuentra; cualquier secuencia impar de perturbaciones en el mismo lugar conducirá a un cambio en la configuración de estado (total) de Bittorio, mientras que cualquier secuencia par de perturbaciones no producirá cambio. Dada su regla y dada su forma de acoplamiento estructural (con el medio), Bittorio es un “reconocedor de secuencia impar”.

Francisco Varela señala que se trata de un paradigma bastante simple, pero es un ejemplo de cómo un sistema autónomo alumbrando significación a partir de un trasfondo. Este paradigma enactivo se presenta como antirepresentacionista, ya que el mundo de Bittorio, como dice Varela, no está pre-dado y luego recuperado por una representación, sino que es enactuado (alumbrado) a través de una historia de acoplamiento estructural.

12 El término “enaction” ha sido creado por Varela sobre el verbo “enact”, que significa tanto promulgar una ley como representar un papel escénico. En *The Embodied Mind* la enacción es asociada con el verbo “bring forth”, que significa dar a luz o alumbrar. Por ello creo que “enaction” puede traducirse por alumbramiento.

13 John Von Neumann, podemos añadir, introdujo la noción de autómata celular, que definió como un “espacio” computacional constituido por muchas células discretas, cada célula puede estar en uno o en varios estados y cambia o permanece en su estado según reglas determinadas (típicamente locales). En suma, un autómata celular es un diseño matemático simulable por ordenador.

Creo que el enfoque enactivo propuesto por Francisco Varela posee varias deficiencias, por lo cual no puede sustituir a los enfoques clásico y conexionista habituales (ambos representacionales) en las ciencias cognitivas. Resumiré algunas de tales deficiencias.¹⁴

En primer lugar, aunque ya en su libro de 1988 declara que las representaciones abandonan la escena, mantiene lo que denomina “representaciones débiles”. En efecto, nuestro autor distingue dos sentidos de la representación: 1) la representación como interpretación, que indica que la cognición consiste en interpretar o representar el mundo como si fuera de cierta manera, y 2) la hipótesis de que un sistema cognitivo actúa sobre la base de representaciones internas. Varela acepta el primer sentido, que califica de “débil”, mientras que rechaza el segundo sentido, que califica de “fuerte”, ya que tal segundo sentido supone, por una parte, que el mundo está pre-definido antes de cualquier actividad cognitiva y, asimismo, defiende la existencia de representaciones mentales dentro del sistema cognitivo. Desde mi punto de vista, esta defensa de las representaciones mentales internas al sistema cognitivo es algo enteramente razonable, y por otra parte no creo que sea posible separar los conceptos de representación e interpretación, puesto que no existe representación sin interpretación¹⁵

En segundo lugar, en su libro conjunto con Thompson y Rosch, pone a Rodney Brooks (del laboratorio de inteligencia artificial del MIT) como ejemplo de científico cognitivo enactivo. Ahora bien, es cierto que Brooks y Varela coinciden en su declarada posición antirepresentacionista, pero mientras Brooks debería suscribir un realismo ingenuo en cambio Varela sostiene expresamente una postura antirrealista. En efecto, y tal como dice Varela, para Brooks la representación es la unidad de abstracción errónea al construir la mayoría de las partes de los sistemas inteligentes. Pero también Brooks, en su libro *Flesh and Machines* (2002), defiende que sus robots móviles autónomos usan el mundo como su propio mejor modelo; esto supone la aceptación de un acceso directo al mundo, de acuerdo con el realismo ingenuo.¹⁶ Sin embargo, Varela no suscribe el realismo ingenuo, sino que reiteradamente se muestra antirrealista.

En tercer lugar, esta posición antirrealista es tan fuerte que lleva a Varela a no aceptar la psicología ecológica de James Jerome Gibson (1904-1979), aunque ambos rechazan el punto de vista representacionista de la percepción visual. En

14 Pueden verse observaciones críticas más detalladas en mi trabajo ya citado “El enfoque enactivo en las ciencias Cognitivas” (2006).

15 Véase el capítulo 4, titulado “Representación e interpretación”, de mi libro *La importancia del conocimiento* (2007).

16 Para una crítica más detallada de Brooks véase mi trabajo “Being inside: Putting representation, body and world together again” (2005).

efecto, en el libro de 1991 Francisco Varela expone la idea de Gibson, que éste desarrolla en su obra *The Ecological Approach to Visual Perception* (1979), de que en la luz ambiental hay suficiente información para especificar el medio ambiente directamente, sin la mediación de cualquier tipo de representación (simbólica o subsimbólica), es decir, que en la topología de la luz ambiental hay invariaciones que especifican directamente las propiedades del medio ambiente. Ahora bien, Varela declara que esta idea, que define el programa de investigación de Gibson, no es compatible con su propio enfoque enactivo; Varela añade que, mientras Gibson sostiene que el medio ambiente es independiente de la mente, él y sus coautores sostienen que el medio ambiente es enactuado (alumbrado o hecho surgir, podemos decir) por historias de acoplamiento.

En suma, el antirepresentacionismo del enfoque enactivo es solidario de una postura antirealista que no me parece aceptable.

V. LA HIPÓTESIS DINAMICISTA.

En tiempos más recientes, la noción de sistema de procesamiento de información (SPI), y en particular la noción de representación, han sido desafiadas por la noción de sistema dinámico sin representaciones. Según ello, los agentes humanos parecen ser básicamente agentes acoplados en el mundo en vez de representantes del mundo. John Haugeland ha acuñado la nueva situación al hablar de “Mind embodied and embedded”, en el capítulo noveno de su libro *Having Thought* (1998), que podríamos traducir como “mente incorporizada e incrustada”, es decir, una mente en un cuerpo y en el mundo. Realmente esta descripción no sólo es cierta sino también obvia, pero los defensores de que los agentes cognitivos son sistemas dinámicos van mucho más lejos de tal obviedad.

En efecto, Timothy van Gelder intentó derribar las posiciones oficiales al publicar en 1995 su artículo titulado irónicamente “What might cognition be, if not computation?”. Para él, si buscamos un punto de referencia como modelo de cognición, es preferible el regulador de Watt (un mecanismo no-representacional) a la máquina de Turing (la caracterización general de un computador presentada por Alan Turing [1912-1954] ya en 1936, y que es equivalente a un SPI). Van Gelder busca convencernos de que la imagen de la cognición como computación ya no es la única posibilidad, sino que podemos tomar muy en serio una noción alternativa, a saber, que la cognición es evolución en un espacio de estados que se produce en ciertos tipos de sistemas dinámicos no-computacionales. La cognición, según van Gelder, podría ser la conducta de sistemas dinámicos muy semejantes al regulador centrífugo de Watt. A finales del siglo XVIII el ingeniero escocés James Watt (1736-1819) diseñó tal

regulador como solución al problema de mantener constante la velocidad de un volante conectado a una máquina de vapor. La velocidad del volante varía según las fluctuaciones de vapor que tienen lugar en la caldera; para controlar la velocidad del volante debemos controlar la cantidad de vapor que entra en los pistones desde la caldera a través de una válvula (la válvula de cierre); lo que hace el regulador es cerrar esta válvula conforme aumenta la velocidad del volante (de modo que el flujo de vapor se restringe) y abrirla conforme la velocidad del volante disminuye (de manera que deja fluir más vapor); así se consigue que la velocidad del volante sea constante. Para van Gelder, la importancia del modo en que Watt resolvió el problema de regulación consiste en que la tarea se realiza sin representación alguna de la velocidad del volante o de los ajustes de la válvula de cierre.

En un trabajo posterior, “The dynamical hypothesis in cognitive science” (1998), van Gelder compara la hipótesis computacional (HC) y su hipótesis dinámica (HD). La hipótesis computacional dice que los agentes cognitivos son básicamente computadores digitales, mientras que la hipótesis dinámica sostiene que los agentes cognitivos son sistemas dinámicos. A su vez, sigue nuestro autor, HD tiene dos componentes: 1) la hipótesis de *naturaleza*, que especifica que los agentes cognitivos *son* sistemas dinámicos, y 2) la hipótesis de *conocimiento*, que establece que podemos y debemos *comprender* la cognición dinámicamente. Asimismo van Gelder declara que el dominio propio de la HD es el de los agentes cognitivos naturales, es decir, agentes biológicos y evolucionados, tales como las personas y otros animales. Ahora bien, a mi entender, esta delimitación hace que su hipótesis pierda interés al no ser válida en general para las ciencias cognitivas.

Por otra parte, creo que hay serios errores en el antirepresentacionismo de Timothy van Gelder¹⁷ En primer lugar, hay una notable confusión entre *conocimiento* y *habilidad motora* o, lo que es lo mismo, entre sistemas propiamente cognitivos y sistemas que son meramente motores. Si nos preguntamos si los agentes humanos son como el regulador de Watt, creo que la respuesta correcta es negativa. En su “Dynamics and cognition” (1997), trabajo basado en su artículo citado de 1995, van Gelder especifica que un sistema es un conjunto de variables (cosas, aspectos, rasgos, etc.) que cambian en el tiempo, tales que la manera en que cambia cualquier variable en un tiempo dado depende de los estados de las otras variables del sistema en ese tiempo; tomados conjuntamente los estados de todas las variables constituyen el estado del sistema como un todo. A su vez, un sistema dinámico es un conjunto de magnitudes acopladas, donde las variables son cuantitativas, sus cambios son interdependientes en

17 Reitero a continuación algunas ideas críticas ya expuestas en mi trabajo “Being inside: Putting representation, body and world together again” (2005).

tiempo real y los instrumentos para su descripción son ecuaciones diferenciales (es decir, cabe precisar, ecuaciones que trabajan con infinitésimos). Pero la faceta más genuina de un sistema dinámico, sigue van Gelder, es que exhibe un alto grado de acoplamiento (*coupling*): cualquier variable está cambiando todo el tiempo, y todos los pares de variables están, directa o indirectamente, determinando mutuamente las formas de los cambios de los otros pares. Por ejemplo, en el sistema solar (que es, como el regulador de Watt, un sistema dinámico) la posición y la velocidad de cualquier cuerpo está constantemente cambiando, y cualquier variable influye sobre cualquier otra. Según van Gelder, el modo fundamental de interacción de un agente humano con el ambiente no es representarlo, ni intercambiar con él entradas y salidas, sino que la relación se entiende mejor mediante la noción de acoplamiento. En mi opinión, podemos aceptar que el cerebro humano es un sistema dinámico, pero su funcionamiento cognitivo necesita representaciones, y no es suficiente (ni siquiera tiene sentido) un simple acoplamiento con el ambiente. Necesitamos representaciones para tener conocimiento propiamente dicho en vez de simple habilidad motora.

En segundo lugar, el antirepresentacionismo significa un retorno, que es indeseable, al conductismo, ya que la desaparición de representaciones explícitas y de los modelos del cuerpo y del mundo plantea una relación directa entre estímulos y respuestas (sin mente). Sin duda, en muchas ocasiones las relaciones de los agentes humanos con su ambiente pertenecen al tipo de reacciones simples; los cambios en los músculos y en las glándulas se producen frecuentemente sin representaciones, pero en tales casos no deberíamos hablar de conocimiento.

En tercer lugar, la postura antirepresentacionista de van Gelder es tan ingenua como su base metafísica, a saber, el realismo ingenuo. Según tal postura antirepresentacionista nuestra percepción corriente de las cosas del mundo es directa, sin cualquier construcción por la mente, de tal modo que las cosas del mundo son conocidas tal como son en sí mismas (realismo ingenuo). Por ello el rechazo de las representaciones compromete a van Gelder con el realismo ingenuo. No digo que van Gelder defienda expresamente este tipo de realismo, pero sostengo que coherentemente debería ser un defensor suyo.

Timothy van Gelder ha encontrado algunos compañeros y seguidores, pero también numerosos críticos. En el mismo número de la revista *Behavioral and Brain Sciences* donde publicó su artículo "The dynamical hypothesis in cognitive science" (1998) se publicaron también treinta artículos breves (escritos por científicos de la computación, psicólogos y filósofos) bajo el rótulo de "Open Peer Commentary", de los que dieciséis matizan las tesis de van Gelder, once se manifiestan en contra y sólo tres pueden considerarse a favor de la hipótesis dinamicista. El propio van Gelder en su "Author's Response", en el mismo número, resume las objeciones presentadas en tres grupos: 1) objeciones a

sus definiciones de sistema dinámico y de computador digital, 2) objeciones a su formulación de la hipótesis dinámica, y 3) objeciones a la defensa de la hipótesis considerada en sí misma. En un trabajo posterior, “Dynamic Approaches to Cognition” (1999), van Gelder considera la aproximación dinámica a la cognición como una confederación de esfuerzos de investigación, donde incluye la cibernética,¹⁸ la psicología ecológica de Gibson¹⁹ o las redes neuronales (aunque, como vimos, éstas comparten con la computación clásica el enfoque representacional); además declara ahora que las diferencias entre las aproximaciones dinámica y clásica no deben ser exageradas, y reconoce que actualmente muchos aspectos de la cognición están bastante más allá del alcance del tratamiento dinámico. En todo caso, este científico cognitivo australiano, desde el año 2000 y hasta la actualidad, se dedica a proyectos de desarrollo del razonamiento y del pensamiento crítico.

VI. CONCLUSIONES: CONOCER COMO REPRESENTAR

Creo que puede decirse rotundamente que conocer el mundo no es tener el mundo mismo en la mente, sino tener en la mente representaciones suyas. En consecuencia si un agente es conocedor debe ser representador.

Por otra parte, ciertamente el mundo percibido por cada mente humana es diferente del mundo percibido por otras mentes humanas, ya que cada uno de nosotros percibe desde una circunstancia o perspectiva propia, que no sólo es espacio-temporal sino también social y cultural. Sin embargo debemos dejar de lado cualquier relativismo epistemológico puesto que los seres humanos compartimos unos iguales mecanismos de percepción y una igual competencia lingüística. Es decir, aunque los mundos percibidos son diversos para las distintas mentes humanas, tiene sentido hablar de una realidad compartida, ya que compartimos un igual equipo representacional.

Por tanto²⁰ tenemos que combinar realismo ontológico e idealismo epistemológico. Según ello *hay* la realidad independiente de nuestras mentes, pero tal realidad *existe* dividida en tantas construcciones como mentes la representan. Dicho de otro modo, *lo que hay* es el mundo independiente de nuestras repre-

18 La cibernética, cuyo principal representante fue el matemático americano Norbert Wiener (1894-1964), intentó estudiar los mecanismos de control y comunicación en máquinas y animales. Su gran aportación fue el análisis de la noción de retroalimentación, pero los cibernéticos no profundizaron en el estudio de la computación. Por ello, a mi entender, las ciencias cognitivas acabaron desplazando a la cibernética.

19 Acerca de James Jerome Gibson he hablado en el apartado 4.

20 Véase sobre este punto el capítulo titulado “La realidad desde la mente” en mi libro ya citado *La importancia del conocimiento* (2007).

sentaciones (realismo ontológico), pero nuestras representaciones son acerca de *lo que existe*, esto es, acerca de una realidad construida, en variable medida, por nuestra mente (idealismo epistemológico), y obviamente estas representaciones están en nuestra mente.

En cuanto a la noción de representación, podemos decir que R es una representación de un objeto O para un agente A en la medida en que A toma cuenta de O en virtud de la manifestación de R.²¹ Por ejemplo, mi recuerdo visual de la catedral de Málaga es una representación de la catedral de Málaga para mí como agente porque tomo cuenta de tal catedral en virtud de la presencia, en mi mente, de ese recuerdo. Asimismo mi concepto de catedral es una representación de ciertos edificios (en sus rasgos característicos y generales) para mí como agente ya que considero tales edificios en virtud de la presencia, también en mi mente, de ese concepto.

Es verdadero, e incluso un truismo, que los agentes humanos están incorporizados e incrustados, es decir, cada agente humano tiene un cuerpo determinado y está situado en un mundo peculiar. Pero también es verdadero que nuestra mente representa nuestro cuerpo y representa nuestro mundo, lo cual queda patente al advertir que nuestra mente no es idéntica con nuestro cuerpo ni mucho menos es idéntica con nuestro mundo y, sin embargo, es capaz de construir sustitutos internos (representaciones mentales) de uno y de otro.

Desde el punto de vista general de las ciencias cognitivas, la tesis de que conocer es representar y la propia noción de representación también se aplican a animales y a computadores. En el caso de los animales el asunto no es sino una natural extensión de la concepción evolutiva de los seres humanos. En el caso de los computadores el asunto es objeto de una rama, especial y muy importante, de la inteligencia artificial denominada “representación del conocimiento” (*knowledge representation*). El trabajo se desarrolla en dos campos relacionados, a saber, la creación de formalismos representacionales y la invención de técnicas de codificación de información. A su vez, los formalismos representacionales tienen una sintaxis precisa, una semántica adecuada y un procedimiento de inferencia que funcione computacionalmente. Asimismo los formalismos representacionales deben mantener un buen equilibrio entre potencia expresiva (capacidad de representar) y facilidad inferencial, puesto que se trata de establecer representaciones en inteligencia artificial. Los recursos empleados para crear los formalismos representacionales son variados, incluyendo formalismos propios de la lógica, reglas de producción, redes semánticas, marcos y, más recientemente, lógicas de descripciones.²²

21 Pueden verse más aclaraciones en el capítulo titulado “Representación e interpretación” de mi libro citado en la nota anterior.

22 Un texto clásico de representación del conocimiento en computadores es el de Han Reichgelt titulado justamente *Knowledge Representation: An AI Perspective* (1991), mientras que

La disciplina de representación del conocimiento descansa sobre la “hipótesis de representación del conocimiento” que formuló Brian Cantwell Smith en el prólogo a su tesis doctoral, en 1982. Según esta hipótesis, cualquier proceso inteligente realizado mediante un mecanismo estará compuesto de ingredientes estructurales que, en primer lugar, nosotros consideramos de manera natural (como observadores externos) que representan una explicación proposicional del conocimiento que el proceso global exhibe, y que, en segundo lugar, independientemente de tal atribución semántica externa, tales ingredientes juegan un papel formal, pero causal y esencial, en la producción de la conducta que manifiesta ese conocimiento. Es decir, según esta hipótesis de representación del conocimiento, los computadores manifiestan un conocimiento y, al mismo tiempo, poseen ingredientes a los que atribuimos de manera natural la capacidad de representar ese conocimiento, además de producirlo. En todo caso, es justo decir que, aunque el desarrollo de la representación del conocimiento en inteligencia artificial ha sido muy grande, todavía mucho conocimiento humano, de tipo abductivo y de tipo no-monótono (revisable) se resiste a una formalización adecuada.

REFERENCIAS

- ANDERSON, John R., *Cognitive Psychology and its Implications*, W. H. Freeman, Nueva York, 1995 (4ª ed.).
- BAADER, Franz, CALVANESE, Diego, MCGUINNESS, Deborah, NARDI, Daniele y PATEL-SCHNEIDER, Peter, (eds.), *The Description Logic Handbook. Theory, Implementation and Applications*, Cambridge University Press, West Nyack (NY), 2003.
- BOBROW, Daniel, COLLINS, Allan (eds.), *Representation and Understanding. Studies in Cognitive Science*, Academic Press, Orlando, 1975.
- BROOKS, Rodney, *Flesh and Machines: How Robots Will Change Us*, Pantheon Books, Nueva York, 2002.
- CAUDILL, Maureen, BUTLER, Charles, *Naturally Intelligent Systems*, The MIT Press, Cambridge (MA), 1990.
- CLARK, Andy, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, The MIT Press, Cambridge (MA), 1989.
- FODOR, Jerry, “Connectionism and the Problem of Systematicity (Continued): Why Smolensky’s Solution *Still* Doesn’t Work”, *Cognition*, 62, 1996, 109-119.

el texto más completo sobre lógicas descriptivas es el editado por Franz Baader y otros titulado *The Description Logic Handbook* (2003).

- GIBSON, James Jerome, *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates, Hillsdale (NJ), 1986. (Publicado originalmente en 1979).
- HAUGELAND, John, *Having Thought. Essays in the metaphysics of mind*, Harvard University Press, Cambridge (MA), 1998.
- KOHONEN, Teuvo, *Self-Organization and Associative Memory*, Springer, Heidelberg, 1989 (3ª ed.).
- MARCUS, Gary F., *The Algebraic Mind. Integrating Connectionism and Cognitive Science*, The MIT Press, Cambridge (MA), 2001.
- MARTINEZ-FREIRE, Pascual F., *La nueva filosofía de la mente*, Gedisa, Barcelona, 1995 (2002, reimpresión).
- MARTINEZ-FREIRE, Pascual F., "Being inside: Putting representation, body and world together again", P. F. Martínez-Freire (ed.), *Cognición y representación*, Suplemento nº 10 de *Contrastes*, 2005, 39-50.
- MARTINEZ-FREIRE, Pascual F., "El enfoque enactivo en las ciencias cognitivas", *Ludus Vitalis*, XIV, 26, 2006, 129-140.
- MARTÍNEZ-FREIRE, Pascual F., *La importancia del conocimiento. Filosofía y ciencias cognitivas*, 2ª ed., Netbiblo, A Coruña, 2007.
- MCCLELLAND, James L., RUMELHART, David E. y el Grupo de Investigación PDP, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume II: Psychological and Biological Models*, The MIT Press, Cambridge (MA), 1986.
- MILLER, George, GALANTER, Eugene, Pribram, Karl, *Plans and the Structure of Behavior*, Holt, Rinehart and Winston, Nueva York, 1960.
- NEISSER, Ulric, *Cognitive Psychology*, Meredith, Nueva York, 1967.
- NEWELL, Allen, "Physical Symbol Systems", *Cognitive Science*, 2, 1980, 135-183.
- NEWELL, Allen, SIMON, Herbert, *Human Problem Solving*, Prentice-Hall, Englewood Cliffs (NJ), 1972.
- NEWELL, Allen, SIMON, Herbert, "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the Association for Computing Machinery*, 19, 1976, 113-126.
- NORMAN, Donald, RUMELHART, David y el Grupo de Investigación LNR, *Explorations in Cognition*, Freeman, San Francisco, 1975.
- REICHGELT, Han, *Knowledge Representation: An AI Perspective*, Ablex Publishing Corporation, Norwood (NJ), 1991.
- RUMELHART, David E., MCCLELLAND, James L. y el Grupo de Investigación PDP, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume I: Foundations*, The MIT Press, Cambridge (MA), 1986.

- SMITH, Brian Cantwell, "Prologue to *Reflection and Semantics in a Procedural Language*", Ronald Brachman & Hector Levesque (eds.), *Readings in Knowledge Representation*, Morgan Kaufmann, San Mateo (CA), 1985, 31-40.
- SMOLENSKY, Paul, "On the Proper Treatment of Connectionism", *Behavioral and Brain Sciences*, 11, 1988, 1-74.
- SMOLENSKY, Paul, "Representation in Connectionist Networks", *Intellectica*, 9-10, 1990, 127-165.
- SMOLENSKY, Paul, "Reply: Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Cognitive Architecture", Cynthia Macdonald y Graham Macdonald (eds.), *Connectionism: Debates on Psychological Explanation, Volume Two*, Blackwell, Oxford, 1995.
- TURING, Alan Mathison, "On Computable Numbers, with an Application to the *Entscheidungsproblem*", Martin Davis (ed.), *The Undecidable*, Raven Press, Hewlett (Nueva York), 1965. (Publicado originalmente en 1936).
- VAN GELDER, Tim, "What might cognition be, if not computation?", *The Journal of Philosophy*, 92, 7, 1995, 345-381.
- VAN GELDER, Tim, "Dynamics and cognition", John Haugeland (ed.), *Mind Design II*, The MIT Press, Cambridge (MA), 1997, 421-450.
- VAN GELDER, Tim, "The dynamical hypothesis in cognitive science", *Behavioral and Brain Sciences*, 21, 1998, 615-628.
- VAN GELDER, Tim, "Author's Response. Disentangling dynamics, computation, and cognition", *Behavioral and Brain Sciences*, 21, 1998, 654-665.
- VAN GELDER, Tim, "Dynamic Approaches to Cognition", Robert Wilson y Frank Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, Cambridge (MA), 1999, 244-246.
- VARELA, Francisco J., *Connaître: Les sciences cognitives*, Du Seuil, París, 1988.
- VARELA, Francisco J., THOMPSON, Evan, ROSCH, Eleanor, *The Embodied Mind. Cognitive Science and Human Experience*, The MIT Press, Cambridge (MA), 1991.
- VARIOS, "Open Peer Commentary", *Behavioral and Brain Sciences*, 21, 1998, 629-653.