

Universidad de Chile
Facultad de Filosofía y Humanidades
Magister en Filosofía con mención en Epistemología

LA REPRESENTACIÓN MENTAL EN LAS TEORÍAS “ENCARNADAS” CONTEMPORÁNEAS DE LA COGNICIÓN

Tesis para optar al grado de Magister en Filosofía; mención Epistemología

Profesor Patrocinante: Guido Vallejos
Alumna: María José Ortúzar

Mayo, 2005

A Eugenio, mi padre

A Diego, mi hermano

AGRADECIMIENTOS

Como siempre, a la familia nuclear. A mi padre, por el apoyo constante y un tanto escéptico. Por la conversación de siempre, por la pregunta inesperada, por devolverme con humor a la realidad. A mi madre, que fue quien me hizo pensar por primera vez que la configuración corporal podía tener algo que ver con cómo nos movíamos por el mundo. A Diego, por la paciencia, el silencio, la música, los almuerzos y las tardes compartidas. A Pabla, por las distancias y las cercanías. Y por ayudarme a hacer el único dibujo de la tesis.

A mis amigas Camila Corvalán, por mandarme tanto texto al que no tenía acceso y Alejandra Maldonado, por las conversaciones prolongadas y por el cable a tierra. A Óscar, por una historia de fragmentos descosidos.

A quienes comparten conmigo el día a día. A Julia Muñoz, por sonreír cada vez que le pido algo. A Graciela Silva, por hacer lo propio. A Jorge Castro, por los cafés de la mañana y de la tarde, y por tanto favor concedido. A Lucía Invernizzi y a Paz Vidaurrázaga, por la generosidad (y la comida). A María Isabel Flisfisch, por la confianza y por el cariño.

Y muy especialmente a Guido Vallejos, por la crítica certera y el amor al detalle, y por la ayuda para pensar este proyecto, formularlo y llevarlo a cabo.

No sólo le costaba entender que el símbolo genérico perro abarcara tantos individuos dispares de diversos tamaños y diversa forma; le molestaba que el perro de las tres y catorce (visto de perfil) tuviera el mismo nombre que el perro de las tres y cuarto (visto de frente).

Jorge Luis Borges

CONTENIDOS

• Agradecimientos	3
• Introducción	7
• Capítulo Primero. La Noción de Representación en la TRM y en el Conexionismo	11
I. La Representación en el Modelo Computacional/ Representacional de la Mente.....	11
II. Críticas a la TRM.....	17
III. La Representación en el Conexionismo.....	19
IV. Críticas al Conexionismo.....	25
V. Precisiones: El lugar de la Representación.....	27
• Capítulo Segundo. Robótica Situada	30
I. Robótica Situada.....	32
II. Robótica Situada vs. GOFAL.....	38
III. Robótica Situada y Representación.....	41
IV. Discusión.....	43
• Capítulo Tercero. Sistemas Dinámicos	47
I. Ejemplos de Sistemas Dinámicos.....	48
II. La Teoría Matemática.....	51
III. Los Fenómenos Cognitivos como Sistemas Dinámicos.....	54
IV. Sistemas Dinámicos vs. el Modelo Computacional / Representacional.....	58
V. Sistemas Dinámicos y Representación.....	61
VI. Discusión.....	63
• Capítulo Cuarto. Biología Y Cognición	67
I. La Mente “Encarnada”.....	68
II. Sistema Senso-motor y el Origen de las Representaciones (Lakoff y Johnson). ..	78
III. Enfoques Encarnados, Modelo Computacional y Representación.....	81
IV. Discusión.....	83
• Capítulo Quinto. Revisión Crítica De Las Teorías “Encarnadas”	86

I.	Aspectos Comunes de los Diferentes Enfoques “Encarnados”.....	86
II.	Enfoques “Encarnados” y Representación.....	90
III.	Vindicación del Rol Explicativo de la Representación.....	96
IV.	Conclusión.....	99
•	Conclusión. Hacia una Idea del Origen de la Representación y su Rol Explicativo.....	101
•	Bibliografía.....	107

INTRODUCCIÓN

Esta investigación busca establecer el rol explicativo que desempeña la noción de representación en las distintas aproximaciones que están actualmente en disputa en el ámbito de la ciencia cognitiva contemporánea.

La noción de “representación” que se aborda aquí es la que se llama a veces “representación interna” o “representación mental”, como opuesta a la “representación externa”. En términos generales, una representación interna es concebida como algo que está en lugar de otra cosa y juega un rol en cómo actuamos. Una representación externa, en cambio, se refiere al conocimiento y las estructuras en el ambiente (por ejemplo, símbolos escritos, las cuentas del ábaco, las dimensiones de un papel, etc.).

La noción de representación mental tuvo un rol central en la explicación de la vida mental por parte de la ciencia cognitiva tradicional, que entendía la inteligencia básicamente en términos del pensamiento deliberativo. La ciencia cognitiva tradicional concibió la mente principalmente como un operador de símbolos, que tenían la capacidad de referirse a otra cosa, y para la definición de la noción de representación mental, no se consideró relevante el substrato físico en el que era implementada (capítulo primero).

Esta visión de la mente, que se formuló filosóficamente como la Teoría Representacional de la Mente, presentó varios problemas para explicar ciertas capacidades cognitivas humanas, como la capacidad de generalizar a casos nuevos y de degradar refinadamente, lo que llevó a pensar que la manera en que las representaciones mentales se daban en la cognición real tenía una relevancia para éstas mismas. El conexionismo presentó una alternativa a la visión más tradicional, en la que la representación siguió siendo un factor importante para la constitución de la vida mental, pero ésta se entendió en términos de redes neurales. En estas redes neurales el almacenamiento y el procesamiento de datos estaban imbricados, los procesamientos se llevaban a cabo transformando las cargas entre conexiones neurales y las representaciones estaban dedicadas a la resolución de tareas específicas, antes que a la resolución de problemas en abstracto de tipo deliberativo (capítulo primero).

La atención a nuevos tipos de problemas, que amplían la definición de la inteligencia humana más allá de los límites del pensamiento deliberativo, ha seguido incorporando rasgos como la “encarnación” –la propiedad de tener un cuerpo– de los agentes cognitivos, la interacción de este cuerpo con el ambiente en el que se anida tanto en la vida cotidiana como en la historia evolutiva y el rol jugado por los recursos externos no biológicos en el desarrollo del éxito cognitivo, tales como la acción corporal, los instrumentos y los artefactos, el ambiente local y las estructuras simbólicas externas. Así, han surgido alternativas teóricas “encarnadas”, que tienen como un factor común el considerar central para el desempeño cognitivo y su explicación la propiedad de tener un cuerpo. El cuerpo específico del agente cognitivo determina cómo se relaciona éste con su ambiente y qué problemas debe resolver para poder conservar su existencia.

La consideración de los agentes cognoscentes como “encarnados” y “situados” ha llevado asimismo a una nueva definición de cuáles son los problemas relevantes para la cognición, lo que ha conducido a replantear la pregunta con respecto a si la cognición humana requiere de representaciones internas. Nociones como el acoplamiento entre sistema nervioso, cuerpo y ambiente, que permean todas las “teorías encarnadas”

de la cognición parecen, en sus visiones extremas, no requerir de una noción de representación para explicar el comportamiento inteligente. En estas visiones, no está claro qué se entiende por “representación”, es decir, qué significa que un sistema (típicamente, un sistema interno, como el sistema nervioso) representa a otro (típicamente, un sistema externo, como el ambiente), si lo que da cuenta del comportamiento es la dinámica del sistema global (que incorpora sistema nervioso, cuerpo y ambiente). Por otra parte, en los ejemplos favorecidos por las visiones encarnadas no parece tener lugar la noción de una representación (que, por ejemplo, ayudaba a resolver un problema como en la Teoría Representacional de la Mente o a reconocer un objeto como en el conexionismo).

Las teorías “encarnadas” aquí expuestas son la robótica situada, la teoría de sistemas dinámicos en ciencia cognitiva y los enfoques que se centran en nuestra constitución biológica para explicar cómo conocemos. La robótica situada tiene como objeto explicar y diseñar agentes situados que sean robustos respecto de un nicho ecológico particular. La construcción de estos agentes se lleva a cabo típicamente por capas que responden más o menos directamente a cambios en el ambiente. En las versiones más radicales, este tipo de construcción parece indicar que, aunque el agente pueda ser entendido como operando con representaciones, en verdad no hay en su actuar sino reacciones al ambiente (capítulo segundo). La teoría de sistemas dinámicos, por su parte, considera el desarrollo en el tiempo del comportamiento de sistemas globales, cuyos componentes se encuentran en una relación de interdependencia y de mutua influencia. Así, se estudia cómo sistema nervioso, cuerpo y mundo están interrelacionados y cómo surgen entre ellos dinámicas particulares, que determinan que el objeto a estudiar no sea ninguno de los componentes, sino la dinámica entre ellos. En este tipo de explicación puede entenderse que no hay lugar para una representación, dado que todo lo que importa para el actuar del agente cognitivo es su *acoplamiento* con los componentes de cuerpo y mundo y no cómo éste se represente su mundo (capítulo tercero). En las teorías biológicas respecto de la cognición se alude a dos posturas: la de Varela, Thompson y Rosch y la de Lakoff y Johnson. En la primera se busca entender cómo opera un agente cognitivo que se ha constituido mediante una historia evolutiva y ontogenética en relación con el mundo. En su formulación más radical, este agente parece no tener representaciones, pues todo lo que hace es intentar mantener su homeostasis organizacional mediante cambios estructurales que responden al interactuar con el mundo (aunque para el sistema nervioso no hay mundo). En la postura de Lakoff y Johnson, en cambio, se descubre una base senso-motora que podría sustentar la noción de representación, al proyectarse sobre el dominio subjetivo (capítulo cuarto).

Estas posturas y su formulación radical respecto de la representación son examinadas en el capítulo quinto. En este capítulo se considera qué hay de propiamente representacional en la cognición humana y si la noción de representación tiene un rol efectivo en diversas clases de acciones cognitivas en las que los agentes se comprometen. Aquí se muestra que de la crítica de los enfoques encarnados no se sigue la vulneración efectiva de la noción de la representación. Ahora bien, esta noción no puede continuar entendiéndose como una entidad separada del mundo, sino que surge precisamente a partir del operar en el mundo con un cuerpo. Esto implica la reformulación de la noción de representación, no sólo en términos de considerar los factores que están más allá de la caja craneana, sino cómo estos factores pueden ser determinantes para cómo se

constituye la representación y para qué sea aquello que se representa. Tal reformulación se basa básicamente como una noción que se construye sobre el operar biológico, entendido éste como la captación de regularidades marginales en los inputs sensoriales.

Esta reformulación de la noción de representación en términos de la captación de regularidades marginales en los inputs sensoriales se explora en la conclusión. Esta captación se lleva a cabo en primera instancia por el cuerpo en cuanto está sensibilizado a su ambiente (distingue del ambiente propiedades básicas –regularidades marginales– funcionales a su supervivencia). A partir de una “arquitectura de incrementación”, cuyos “circuitos” van cobrando nuevas funciones a partir de constreñimientos ecológicos específicos, el sistema nervioso se sensibiliza progresivamente a nuevas regularidades marginales, cada vez más alejadas del ambiente específico en el que tal sistema se anida. Esto permite el surgimiento de representaciones en tanto estructuras orientadas a la acción, pero cuya funcionalidad es el operar off-line.

En esta visión, la propiedad de tener un cuerpo, de estar “encarnado”, juega un rol fundamental para lo que se represente del ambiente, y este cuerpo unido a su historia evolutiva y ontogenética es el que es el substrato para nuestra actividad representadora. El operar en el mundo con un cuerpo es el que determina el contenido de una representación. Las cosas representadas por un agente cognitivo son las cosas que su cuerpo le presenta de su ambiente, en interacciones repetidas con él. Cuerpo y ambiente determinan así qué es útil que el agente represente para sobrevivir y reproducirse, mediante una historia de experiencias ontogenéticas y evolutivas en relación con el objeto que representan.

Este enfoque recupera el rol de la representación en nuestra vida cognitiva, pero no en tanto una entidad objetiva, sino en tanto surge precisamente del operar de un sistema nervioso en el mundo con un cuerpo y se constituye en el operar cognitivo mismo.

CAPÍTULO PRIMERO

LA NOCIÓN DE REPRESENTACIÓN EN LA TRM Y EN EL CONEXIONISMO

En este capítulo se revisan dos teorías de la mente en las cuales la noción de representación juega un rol central. En estas teorías, el problema fundamental es establecer cuál es el estatus de la “representación

interna”. Para la Teoría Representacional de la Mente (en adelante, TRM), dicha noción corresponde a un símbolo en una suerte de lenguaje interno o mentalés. La representación es sustentada por una arquitectura serial, esto es, de procesamiento secuencial a modo de algoritmos o procedimientos efectivos de pasos a seguir linealmente hasta llegar a la solución de un problema. En el conexionismo, la representación es un patrón de activación que emerge en virtud de la fuerza de conectividad de un conjunto de microunidades que simulan ciertas propiedades básicas de las neuronas y que constituyen una red neuronal. La arquitectura en la que se basa esta noción de representación es la de un modelo de “computación neuronal” de procesamiento paralelo y distribuido. Pese a las divergencias que ambas teorías exhiben respecto de su “arquitectura”, la representación juega un rol central en la explicación del funcionamiento de la mente cognitiva, permitiendo entender cómo un agente inteligente resuelve problemas y cómo es capaz de realizar tareas de reconocimiento y recolección.

Ambas posiciones consideran que el procesamiento cognitivo se realiza dentro de la caja craneana. En consecuencia, al modelar los procesos cognitivos no contemplan como explicativamente relevantes aspectos como el tener un cuerpo y estar situado en un mundo.

I. La Representación en el Modelo Computacional/ Representacional de la Mente

La manera más tradicional en ciencia cognitiva de pensar sobre cómo opera un sistema inteligente supone que éste lo hace a través de representaciones, entendiendo por representación algo que está en lugar de otra cosa; así, una representación de un “perro” no es el perro propiamente tal, sino “algo” que está en lugar del perro real (cf. Bechtel et al. 1999). Algo está en lugar de otra cosa, o más bien, es acerca de otra cosa, si porta un contenido, lo que se cumple si algún x es acerca de, simboliza, designa, o de otra forma “está por” alguna y . Cualquier entidad que porta contenido *al interior* de un sistema es una *representación interna*. Los filósofos han hecho uso por largo tiempo de tales entidades en su explicación del procesamiento cognitivo, basándose en la idea de que no puede haber causación interna sin representación explícita. Este compromiso con la noción de representación está aparejado con la noción de computación, dado que los sistemas computacionales requieren representaciones internas como un medio a través del cual llevan a cabo sus computaciones (cf. Stufflebeam 1999: 636).

Este encuentro entre la noción de representación y la de computación permite pensar la mente como un sistema formal. En los sistemas formales, como en la lógica formal, se respetan las relaciones semánticas (tales como la preservación de la verdad: si “A y B” es verdadero, entonces “A” es verdadero) en virtud de procedimientos que no hacen referencia a los significados. Y los detalles físicos específicos de cualquier sistema tal no se consideran importantes, dado que lo que importa es la red de movimientos y transiciones. Tal idea de la mente fue posibilitada por la formalización de la computación misma, realizada por Turing y su “máquina”. La “máquina de Turing” es un dispositivo imaginario que consiste

en una cinta infinita, un procesador simple (un “estado de máquina finito”) y un “lector / escritor”. La cinta actúa como depósito de datos, usando un conjunto fijo de símbolos. El “lector / escritor” puede leer un símbolo de la cinta, moverse un espacio hacia delante o hacia atrás en la cinta, y escribir en la cinta. El estado de máquina finito (una suerte de procesador central) tiene memoria suficiente para recordar qué símbolo fue recientemente leído y en qué estado (el estado finito de máquina) estaba. Ambos determinan la próxima acción, que se realiza por el lector / escritor, y determinan el nuevo estado del estado finito de máquina. Turing mostró que un dispositivo tal, realizando una secuencia de computaciones simples dirigidas por los símbolos en la cinta, podía computar la respuesta a cualquier problema bien definido. Esta noción sumada a los avances en lógica y sistemas formales y la construcción efectiva de una máquina computadora electrónica para propósitos generales y el desarrollo de técnicas flexibles, de programación de alto nivel, permitieron el surgimiento de la ciencia cognitiva clásica, que, como se señaló, tiene en su seno la noción de representación (Clark 2001a: 11).

Es en este marco que se comprende el surgimiento de la Inteligencia Artificial –de ahora en adelante, IA–. La IA tiene como propósito el diseño de agentes inteligentes, entendiendo por un agente una entidad que puede ser vista como percibiendo y actuando en su ambiente. En los primeros años de la IA, ésta se concentró en la habilidad de los agentes para resolver problemas. Para esto, asumió tres compromisos claves. Primero, el uso de un código simbólico como un medio de guardar todo el conocimiento a largo plazo del sistema. Segundo, la descripción de la inteligencia como la habilidad de buscar exitosamente un espacio de problema simbólico. Un sistema físico “ejerce su inteligencia para resolver problemas al buscar –esto es, al generar y modificar progresivamente las estructuras simbólicas hasta que encuentra una estructura de solución” (Newell y Simon, citados Clark, *Ibid.*: 33). Tercero, la inteligencia reside en, o es cercana a, el nivel del pensamiento deliberativo, consistiendo en la recuperación de información guardada simbólicamente para usarla en procesos de búsqueda que permiten la solución de un problema. Tales procesos implican la generación, composición y transformación de las estructuras simbólicas hasta que se encuentran las condiciones especificadas para una solución. Estos tres compromisos se ven reflejados en una arquitectura que se basa en la información en forma de símbolos. La más común de éstas es la de von Neumann, que asume que el procesamiento implica operaciones seriales, gobernadas por reglas sobre representaciones simbólicas. La idea clave, la técnica de “programa almacenado”, permite que un programa y los datos sean almacenados en una misma memoria. Consiste de una unidad de procesamiento central, una unidad de memoria y unidades de inputs y outputs. La información se concibe como un input, que es guardado y transformado algorítmicamente para producir un output.

Si, en términos generales, la cognición se entiende como una actividad de resolución de problemas, una noción importante en esta formulación es la de un “espacio de búsqueda simbólico”. Tal espacio de búsqueda es donde el agente debe encontrar la solución a un problema dado. Esto se hace mediante una búsqueda guiada por la heurística, que considera un objeto (lo que se está buscando) y un ámbito (la región o conjunto de cosas en las cuales se busca el objeto). Dado que

muchos problemas corren el peligro de que una búsqueda exhaustiva de soluciones (la partida perfecta de ajedrez, por ejemplo) tenga como consecuencia una explosión combinatoria, se trata de evitar tal explosión haciendo la búsqueda selectiva. Esto significa estudiar procedimientos computacionales falibles, aunque suficientemente confiables, diseñados para la solución de problemas basados en la exploración por ensayo y error de soluciones posibles. Newell, Shaw y Simon desarrollaron sobre esta base el programa “Solucionador del Problema General” – SPG–, en el cual desarrollaron una estrategia de búsqueda heurística al interior del espacio del problema basada en el análisis de medios y fines. Se entendía en este programa que todo problema consistía básicamente de tres pasos. La resolución de problemas se hace definiendo un estado inicial específico, un estado final específico y un conjunto de operadores, para transformar un estado en otro. La tarea consistía en encontrar una secuencia de operaciones que transformaran el estado inicial en el estado final (Haugeland 1988: 167 y ss.). SPG corresponde asimismo a un tipo de arquitectura de von Neumann, llamado “sistema de producción”. En esta visión, la mente consiste en una memoria que trabaja, un conjunto grande de reglas de producción y un conjunto de reglas de precedencia que determinan el orden de activación de las reglas de producción.

Esta descripción de la inteligencia por parte de la IA se reformuló en una visión filosófica de la mente, entendida como una entidad que opera computando representaciones, por Jerry Fodor en los años '70. A partir de esta noción formal computacional defendió la tesis de que una psicología de actitudes proposicionales (o psicología del sentido común) era compatible con una TRM. Tal como los agentes inteligentes que consideraba la IA resolvían problemas pensando en un estado de problema, su solución y la manera de llegar de uno a otro; la psicología de actitudes proposicionales tiene su origen en la forma estándar de la adscripción de deseos y creencias con tal de explicar acciones. Tales explicaciones consisten en actitudes mentales (creer, desear, temer, etc.) respecto de proposiciones específicas (“que está lloviendo”, “que el café está en la cocina”, etc.) para comprender la acción inteligente. De tal modo, en una oración como “Diego cree que queda pan”, la construcción “que” introduce una proposición (“queda pan”) respecto de la cual se supone que el agente ha de exhibir una actitud (en este caso, creer). Nuestra comprensión cotidiana del comportamiento del otro envuelve sendas dosis de adscripción de actitudes proposicionales: por ejemplo, yo puedo creer que Diego abre el refrigerador para buscar pan (dado que cree que queda pan). Puesto que este modo de hablar (y de pensar) permite un gran éxito predictivo y es la moneda corriente de muchos de nuestros proyectos sociales y prácticos, Fodor intentó vindicarlo entendiéndolo, como ya se señaló, en términos de una TRM.

La TRM fodoriana tiene tres ingredientes fundamentales: primero, las actitudes proposicionales son relaciones computacionales con *representaciones mentales*; segundo, las representaciones mentales forman un *sistema de símbolos*; y tercero, los procesos mentales son procesos *causales* que suponen la manifestación explícita de símbolos desde el sistema de símbolos (Clark 2000: 378).

Estos tres ingredientes hacen que la TRM suscriba una versión computacional / representacional de la cognición, pues supone que la mente es un mecanismo computacional y también un sistema simbólico. El

primer supuesto dice relación con la noción de que los procesos mentales cognitivos se consideran funcionalmente equivalentes con procesos computacionales, siendo un computador entendido como un sistema formal automático, es decir, como un conjunto de procedimientos efectivos de manipulación de símbolos regidos por reglas (ver más arriba). El segundo supuesto apunta al hecho de que las propiedades formales por sí solas no constituyen las representaciones en cuanto tales, sino que su propiedad esencial es ser vehículos de o poseer contenido intencional, es decir, ser acerca de algo que es distinto de ellas mismas.

Así, Fodor (1998) considera que para explicar los procesos mentales – secuencias causales de ejemplificaciones de representaciones mentales– ha de tomarse en cuenta que las instanciaciones de las representaciones mentales son símbolos. Las instanciaciones de símbolos son objetos físicos con propiedades semánticas. Las computaciones son aquellas relaciones causales entre símbolos que representan confiablemente las propiedades semánticas de los relata. Aquí se combina la postulación de representaciones mentales con la “metáfora del computador”. Los computadores muestran cómo las relaciones sintácticas entre símbolos se proyectan sobre el ámbito de las relaciones causales entre instanciaciones físicas de los símbolos y, a su vez, sobre relaciones semánticas entre proposiciones. Si tener una actitud proposicional significa ejemplificar un símbolo que expresa una proposición, las relaciones semánticas entre proposiciones constituyen procesos mentales. Según Fodor, las relaciones causales entre las instanciaciones de los símbolos se proyectan sobre las relaciones semánticas entre las proposiciones expresadas por dichos símbolos mediante la sintaxis. La sintaxis de un símbolo depende de las propiedades físicas de orden superior de dicho símbolo. La estructura sintáctica de un símbolo puede pensarse como un rasgo abstracto de su configuración. Las instanciaciones de símbolos interactúan causalmente en virtud de sus estructuras sintácticas. Así, la sintaxis de un símbolo determina las causas y los efectos de sus instanciaciones, de la misma manera que la geometría de una llave determina qué cerraduras abrirá. La idea es construir una máquina en la que (1) sus operaciones consistan enteramente en transformaciones de símbolos; (2) en el curso de la realización de tales operaciones, la máquina es sensible sólo a las propiedades sintácticas de los símbolos; (3) las operaciones que la máquina realice con los símbolos se limitan enteramente a alterar sus configuraciones sin alterar el contenido que dichos símbolos expresan. Las máquinas, en este caso, computadores, son básicamente entornos en los cuales la sintaxis de un símbolo determina su rol causal respetando su contenido.

Esta capacidad de operar con símbolos que se combinan sistemáticamente por reglas complejas para conformar unidades más complejas, tales como pensamientos, posibilita la concepción de que tenemos un lenguaje mental, el mentalés. Dado que los símbolos son intencionales o representacionales (son acerca de cosas), los estados que ellos componen son intencionales. Dada la estructura constituyente que exhibe el lenguaje del pensamiento, los pensamientos son composicionales en lo que concierne a su semántica. Así, puede decirse que los estados mentales heredan su intencionalidad de sus representaciones mentales constituyentes.

La visión fodoriana de la cognición subraya además la analogía con la implementación software/hardware computacional. Así, se consideró que la mente era a la “carne” neural como el software es al computador. El cerebro puede ser la implementación estándar (local, terrestre, biológica), pero la cognición es una cosa que acontece a nivel de un programa.

Esta versión computacional / representacional de la cognición detenta a su vez el principio del individualismo metodológico. El individualismo metodológico consiste básicamente en la afirmación de que lo que es explicativamente relevante para la psicología son los estados internos que supervienen sobre la estructura neurofisiológica. De este modo, para los efectos de la individuación de las propiedades causalmente relevantes de la conducta inteligente no se tienen presente los factores ambientales. En la TRM los estados mentales se describen funcionalmente, y éstos supervienen sobre estados descritos físicamente, estando estos últimos en el cerebro. Si bien el modelo computacional / representacional establece en última instancia el requerimiento ontológico de la realizabilidad física de lo mental, los detalles de tal realizabilidad no son explicativamente relevantes, pues lo explicativamente relevante es la red de movimientos y transiciones en virtud de procedimientos computacionales efectivos.

II. Críticas a la TRM

Las críticas básicas a la TRM se basan principalmente en su alejamiento de la realidad biológica, en términos del cerebro y del ambiente.

La concepción de la mente como un sistema computacional / representacional conlleva, pese al requerimiento último de una realizabilidad física de lo mental, cierto desprecio por las bases biológicas – cerebrales y ambientales– de la cognición. Así, la realización efectiva de las operaciones cognitivas en el cerebro no determina en medida alguna lo que sean o cómo surjan las representaciones. En este sentido, el “operar” efectivo del sistema neurológico no tiene en la TRM una relevancia explicativa. La consecuencia de esto es que el modo en que nuestro cerebro opera realmente no tiene incidencia alguna en la constitución de nuestras representaciones mentales y en los procesos en los que éstas están involucradas. Ahora bien, como veremos más adelante, ciertos detalles acerca de la implementación física, esto es, del funcionamiento cerebral, pueden tener una implicancia directa en cómo concebimos, almacenamos y recuperamos las representaciones mentales.

Tal alejamiento de la realidad biológica en cuanto al *locus* de lo mental se hace evidente si se considera la resolución de problemas de tipo distinto del deliberativo. Así, la explicación de la velocidad relativa a la cual resolvemos diferentes problemas y las clases de efectos de interferencia que experimentamos cuando tratamos de resolver diversos problemas a la vez (por ejemplo, recordar dos números de teléfono muy similares), pueden bien requerir explicaciones que envuelven detalles muy específicos acerca de cómo las representaciones interiores se depositan y estructuran. Esto último implica que los detalles específicos del cerebro efectivamente son relevantes para la explicación de la cognición, lo que entraña a su vez que la

imagen de la mente como un software no provee ninguna garantía profunda respecto de que la ciencia cognitiva pueda ignorar los hechos del cerebro humano¹ (Clark 2001a: 19).

La imagen computacional / representacional ve a la mente como operando sobre reglas secuenciales, que se aplican una a la vez (o, en términos de Fodor, en virtud de la sintaxis de los símbolos). Dado el foco de la IA clásica en los códigos simbólicos almacenados como el “espacio de búsqueda” donde resolver un problema, ésta encontró asimismo una serie de profundos problemas, centrados mayormente en las dificultades de búsqueda, movimiento y “up-date” por parte de un cuerpo amplio de contenidos almacenados cuasi-lingüísticos. Esto constituye una limitación muy importante al momento de realizar tareas que requieren un gran número de operaciones secuenciales (tales como el análisis de una imagen o la predicción del tiempo). Una búsqueda continua por algoritmos de procesamiento paralelo tiene poco éxito en tales casos. Otra limitación importante se refiere a que el procesamiento simbólico no está localizado: la pérdida o mal funcionamiento de cualquier parte de los símbolos o reglas del sistema resultan en un mal funcionamiento. No se da entonces la capacidad de *degradar refinadamente* cuando los sistemas son parcialmente dañados. Además, este modelo presenta dificultades para *generalizar a casos nuevos* –la capacidad que nos permite aprender– y una ineficiencia creciente a medida que los modelos se hacen más complejos. Una teoría que de cuenta de este funcionar es más cercana a lo que conocemos como inteligencia humana. Además, el dar cuenta de estos detalles de “implementación” supone dar cuenta de cómo se dan efectivamente los procesos mentales y si acaso éstos tienen alguna relevancia para la constitución de las representaciones mentales.

En la visión de la TRM, el medio social en el que se ocupan y se aprenden las representaciones tampoco tiene una relevancia explicativa. Esto es evidente si se considera que en la TRM fodoriana la intencionalidad –la capacidad de representar o ser acerca de algo– de las representaciones es anterior ontológicamente y en el orden de explicación a la intencionalidad de las actitudes proposicionales y a la de los lenguajes naturales. Esto supone que se ha de explicar, primero, cómo la representación mental es acerca de algo y, luego, cómo nos podemos referir a las cosas en el lenguaje natural. Dado que la característica más sobresaliente del entorno social es la de comunicarse a través de un lenguaje, cabe preguntarse si acaso tal entorno social no determina en alguna medida el contenido y/o el uso de tales representaciones (cf. Clark 2001a, 1997a; Clark y Thornton 1997).

La comprensión de la explicación psicológica como una forma de descomposición funcional, donde las capacidades cognitivas sofisticadas eran divididas y explicadas a través de las actividades de los componentes, que a su vez se explicaba por su operación simbólica interna, parece entonces no dar cuenta de varias características del operar cognitivo inteligente. Esto llevó a indagar otras maneras de codificar y acceder a la información, atendiendo a una panoplia de estados más difíciles de interpretar, tal vez, sub-simbólicos, cuyo efecto cumulativo consiste en esculpir el comportamiento en maneras que respetan el

¹ Clark (Ibid.: 23-24) se pregunta en este mismo sentido si acaso las hormonas, los químicos y la materia orgánica juegan algún rol más allá del mero afectar cómo fluye la información y cómo se procesa en el cerebro, es decir, si nuestra naturaleza, tal como la experimentamos, no es sólo una función del flujo de la información; en otras palabras, si acaso la conciencia consciente es un fenómeno *informativo*.

espacio de razones y el sentido semántico (ver más abajo). Aquí el propósito de la IA se distanció de su formulación clásica para entender el comportamiento semánticamente sensible a partir de una enrevesada masa de interacciones en paralelo difíciles de manejar entre elementos externos y recursos internos semánticamente opacos (Clark 1995).

III. La Representación en el Conexionismo

El marco aceptado de representaciones y computaciones que definía el enfoque explicativo dominante en ciencia cognitiva se vio cuestionado en los años '80 por una mayor atención al rol que desempeñaban “los detalles de implementación”, a saber, a los procesos neuronales, por una parte, y a los procesos ambientales, por otra.

Esta atención a los procesos neuronales motivó el modelo conexionista, que guarda una relación (algo distante)² con la arquitectura y el operar del cerebro biológico. Como el cerebro, una red artificial se compone de muchos procesadores vinculados en paralelo por una asombrosa masa de cableado y conectividad. En el cerebro, los “procesadores simples” son neuronas y las conexiones, axones y sinapsis. En las redes conexionistas, los procesadores simples son llamados “unidades” y las conexiones consisten en vínculos numéricamente cargados entre estas unidades. Estos vínculos son las conexiones. En ambos casos, los elementos procesadores simples (neuronas, unidades) son generalmente sensibles sólo a las influencias locales. Cada elemento toma sus inputs (que pueden ser excitatorios o inhibitorios) de un determinado número de “vecinos”, responde a tal input de acuerdo a una función simple de activación y pasa a su vez los outputs a un pequeño (a veces traslapado) grupo de vecinos con los cuales está conectado (Clark 2001a: 62).

En el vocabulario conexionista, los ítems a tener en cuenta no son símbolos o representaciones propiamente tales, sino *patrones complejos de actividad* entre numerosas unidades que conforman una red. La red conexionista típica consiste en tres capas: una de “unidades de input” (cada una capaz de contener valores variables de activación numérica), otra de “unidades ocultas” que media el proceso, y una última de “unidades de output” (capaces de especificar una respuesta en términos de otro conjunto de valores variables de activación numérica). Las capas de unidades se conectan por vínculos de cargas y son éstos los que se adaptan durante el aprendizaje. El procesamiento implica la propagación de la actividad a través de la red en respuesta a la presentación de un conjunto específico de valores de activación para las unidades de input.

Una forma estándar en que las redes aprenden es por medio de la “retro-propagación”. Puesto que el número de conexiones es muy grande, es casi imposible encontrar la conexión apropiada entre unidades calibrando cada una a mano y probando cada una. Para calibrar las cargas entre las conexiones existen procedimientos automáticos como los algoritmos de aprendizaje. El más famoso de éstos es el de retro-propagación. En el aprendizaje de retro-propagación, la red comienza con un conjunto de pesos de conexión

seleccionados al azar (el diseñador fija el diseño, número de unidades, etc.). Esta red es expuesta a una gran cantidad de patrones de input. Por cada patrón de input, se produce un output (inicialmente incorrecto). Un sistema automático de supervisión examina el output, lo compara con el objetivo de output (el correcto) y calcula ajustes pequeños a las cargas de conexión; estos ajustes causarían una mejora menor en la realización del mismo ejercicio si la red fuera sometida al mismo patrón de input. Este procedimiento se repite una y otra vez, exponiendo a la red a un corpus grande y cíclico de casos de entrenamiento. Tras un entrenamiento suficiente, la red a menudo (no siempre) aprende a asignar cargas que efectivamente resuelven el problema, reduciendo el error y produciendo el perfil deseado de input-output (Clark 2001a: 65-66).

En las redes neurales, tal como se revela en el caso de la retro-propagación, el conocimiento almacenado está profundamente imbricado con las características del procesamiento, puesto que ajustar las cargas es al mismo tiempo registrar la información y determinar la manera en que la información ha de usarse en la resolución de problemas siguiente. La información no es un bien pasivo que ha de almacenarse de alguna forma canónica y luego ser utilizado por operaciones de procesamiento distintas e independientes. Por el contrario, la base de conocimiento y el perfil de procesamiento se desarrollan simultáneamente y dependen de un recurso singular y común: el conjunto de unidades y cargas. Dada la profunda interpenetración entre conocimiento y procesamiento, no es conveniente concebir la resolución por parte de las redes neurales conexionistas imaginando algún conjunto de reglas de inferencia definidas que son aplicadas a un cuerpo de ítems pasivos y de una sintaxis similar a textos. En cambio, los conexionistas tratan la inferencia como un caso de transformaciones complejas entre vectores y valores de activación: transformaciones cuya naturaleza no está determinada independientemente de los contenidos informacionales a los que la red se ha adaptado. Esto se debe a que el aprendizaje de estilo conexionista supone encontrar una representación (y una codificación) para un cuerpo de conocimiento *en el contexto concreto* de tratar de realizar una tarea específica. Las representaciones que la red adquiere reflejan así las demandas específicas de una tarea. De este modo, las representaciones se seleccionan como lo dictan las demandas de una tarea concreta y se codifican de una manera que no distingue entre el conocimiento y su uso en el procesamiento. Dado que la búsqueda de información y su uso en el procesamiento ya no son eventos conceptualmente distintos, el razonamiento que utilizan tales representaciones no se concibe como una inferencia tradicional (Clark 1995: 92).

El conocimiento a largo plazo de un sistema conexionista no consiste entonces en un cuerpo de afirmaciones declarativas escritas en una notación formal basada en la estructura del lenguaje natural o de la lógica. El conocimiento, en cambio, es inherente al grupo de cargas de conexión y a la arquitectura de la unidad y muchas de estas conexiones cargadas participan en un gran número de actividades de resolución de problemas del sistema. El conocimiento que tiene lugar –la información activa durante el procesamiento de un input específico– puede igualarse así con los patrones de activación momentáneos que ocurren en el estrato de unidad oculta. Tales patrones a menudo envuelven esquemas codificadores *distribuidos y superposicionales*

² Para las diferencias entre las redes artificiales y las neuronales, ver discusión del conexionismo más abajo.

(Clark 2001a: 66). En su sentido más básico, la *representación distribuida* es una que está de alguna manera “esparcida” sobre una extensión más que mínima de los recursos disponibles para representar. Más concretamente, un ítem de información tiene una representación distribuida si es expresado por la actividad simultánea de cierto número de unidades. Pero lo que hace computacionalmente potente a la representación distribuida no es sólo este hecho, sino el uso sistemático de las distribuciones para codificar mayor información respecto de sutiles similitudes y diferencias. Un patrón distribuido de actividad puede codificar información “microestructural”. Así, por ejemplo, variaciones en el patrón general pueden representar la presencia de un gato negro en el campo visual, mientras que pequeñas variaciones en tal patrón pueden acarrear información sobre la orientación del gato (si acaso está de frente, de lado, etc.). Igualmente, el patrón de activación para una pantera negra puede compartir algo de la sub-estructura del patrón de activación del gato negro, mientras que aquel de un ciervo blanco puede no compartir ninguna. La noción de almacenamiento *superposicional* es precisamente la noción del uso parcialmente traslapado de los recursos distribuidos, en los que el traslape es significativo informacionalmente de las maneras ya mencionadas. Así, ítems semánticamente relacionados se representan por patrones de activación sintácticamente relacionados (parcialmente traslapados). La codificación superposicional distribuida puede ser pensada entonces como un truco para forzar aún más información dentro de un sistema de codificaciones al explotar vehículos sintácticos más altamente estructurados que las palabras. Este truco proporciona varios beneficios adicionales, incluyendo el uso económico de recursos representacionales, la generalización “libre” y la degradación refinada (todos, como se señaló, problemas a los que la visión computacional / representacional no podía responder satisfactoriamente).

Así, la *superposicionalidad* explica que se de la *generalización* en las redes neurales. La generalización sucede debido a que un nuevo patrón de input, si acaso se asemeja a uno más antiguo en algunos aspectos, producirá una respuesta basada en el traslape parcial, de manera que son posibles respuestas “sensibles” a estos nuevos inputs. La *degradación refinada*, como vimos, se refiere a la habilidad de producir repuestas sensibles dado algún daño sistémico. Esto es posible porque el sistema general actúa aquí como una suerte de *completador* de patrones: dado un fragmento lo suficientemente grande de un patrón familiar, recordará la cosa entera. La generalización, la compleción de patrones y la tolerancia al daño son reflejo de cuán poderosa es la estrategia computacional conexionista al usar esquemas de almacenamiento distribuido y superpuesto y al poseer una memoria parcial basada en claves (Clark 2001a: 66-67).

Estos esquemas de almacenamiento distribuido y superposicional han llevado a considerar el conexionismo como un *paradigma sub-simbólico*. La idea esencial es que, mientras los enfoques básicos de símbolos físicos exhiben una suerte de transparencia semántica, tal como que las palabras y las ideas familiares se consideran como simples símbolos internos en forma proposicional, los enfoques conexionistas introducen una distancia mucho mayor entre el habla diaria y los contenidos manipulados por el sistema computacional. Al describir esquemas de representación conexionista como cambiados de dimensión y sub-simbólicos, Smolensky y otros quisieron sugerir que los rasgos que el sistema descubre son de grano más fino y más sutiles que aquellos revelados por las palabras singulares en el lenguaje público. Según Smolensky

(1998), hay un nivel intermedio de estructura entre los niveles neural y simbólico: el sub-simbólico. Cuando la cognición se describe a nivel sub-simbólico corresponde a un sistema conexionista. El nivel sub-simbólico intenta formalizar la clase de procesos que ocurren en el sistema nervioso. En un programa sub-simbólico los elementos con contenido no son los mismos que constituyen objetos regidos por manipulaciones simbólicas según reglas que definen el sistema, como en el modelo computacional / representacional. En el paradigma sub-simbólico, las unidades semánticamente interpretables son, como vimos, *patrones de activación* sobre un gran número de unidades en el sistema. Las reglas tienen la forma de reglas que pasan la activación y son esencialmente distintas en carácter de las reglas de manipulación de símbolos. La manera en que los símbolos se combinan, como vimos, es por *superposición*, no por *concatenación*. A diferencia de la TRM, los elementos con contenido en el nivel sub-simbólico, que explican el operar cognitivo, no corresponden a los mismos que se emplean en el lenguaje natural. La activación de una unidad específica (en un contexto específico) señala así un hecho semántico, pero éste puede ser un hecho que desafíe una descripción fácil al usar las palabras y las frases del lenguaje diario. El significado, por lo tanto, no se sitúa en los símbolos particulares, sino que una función del estado global de un sistema y se vincula a la actuación general en algún dominio específico, tal como el reconocimiento o el aprendizaje (Varela et al., *Ibid.*: 98 y ss.). La estructura semántica representada por un patrón extenso de actividad de unidad puede ser muy rica e incluso sutil, y diferencias menores en tal patrón pueden marcar diferencias igualmente sutiles en cuanto matices contextuales (recuérdese la discusión del gato más arriba). Las diferencias de activación a nivel de unidad pueden, así, reflejar detalles diminutos de las dimensiones visuales, táctiles, funcionales o incluso emotivas de nuestras respuestas a los mismos estímulos en contextos variados de mundo real.

El contraste con el paradigma representacional / computacional no se da sólo a nivel de cómo se conciben las representaciones o símbolos al interior del sistema, sino también en cuanto la “arquitectura” (serial o paralela y distribuida). En la arquitectura conexionista una típica red entrenada no emplea cuerdas gramaticales (semejantes a la que se emplean en el lenguaje natural), ni tampoco operaciones de procesamiento sensibles a la estructura de tales cuerdas como en la TRM. Se encuentran aquí, en cambio, complejos prototípicos de propiedades representadas en un espacio altamente dimensional. El espacio está altamente organizado en el sentido de que los ítems de datos que necesitan ser tratados de manera muy parecida se codifican en regiones vecinas del espacio. Es esta *métrica* semántica la que le permite a la red generalizar y responder bien bajo condiciones en las que los sistemas computacionales / representacionales fallan.

Pese a que un sistema de arquitectura conexionista se desempeña mejor que uno computacional / representacional si lo que se quiere explicar es la degradación refinada y la capacidad de generalizar a casos nuevos, un problema serio que enfrenta tal arquitectura es el de la “sistematicidad”, planteado por Fodor y Pylyshyn. La “sistematicidad” se refiere a un conjunto de capacidades que parecen ser variaciones composicionales sobre un substrato estructurado singular, por ejemplo, la capacidad de un hablante del castellano de decir “Diego quiere a Moka” y “Moka quiere a Diego”; la misma competencia permite la capacidad de generar una variedad de oraciones que envuelven dichas partes. La respuesta conexionista

estándar ha consistido en insistir en que ellos pueden de hecho explicar la sistematicidad sin implementar una arquitectura clásica al usar representaciones distribuidas que contienen estructuras complejas de un modo no concatenado.

IV. Críticas al Conexionismo

Pese a que el conexionismo se motivó desde un inicio por explorar una forma alternativa al estilo TRM de concebir la representación, mucho de la investigación hecha bajo tal paradigma continuó viendo en la noción de representación básicamente un “contenedor” de información, si bien ahora éste estaba distribuido y no se correspondía con las estructuras del lenguaje natural. Y aunque los conexionistas buscaron inspirarse de algún modo en la realidad biológica del funcionamiento cerebral, la abstracción del “mundo real” en la construcción de las redes conexionistas no puede obviarse, siendo debatible cuál es la realidad biológica efectiva del conexionismo³.

Respecto de esto último, una primera pregunta se formula en términos de quien construye la red. Pese a que las redes pueden crear sus propias soluciones a problemas dados, lo que ellas aprenden depende en buena medida de la variedad de opciones hechas por el diseñador humano. Al elegir una variedad determinada de opciones, tal diseñador prescinde del mundo real y del organismo que actúa. Tales simplificaciones pueden oscurecer las clases de soluciones a problemas ecológicamente reales que caracterizan a la inteligencia de agentes encarnados activos, tales como nosotros. Además, tal alejamiento del mundo real supone privar a nuestros sistemas artificiales de la oportunidad de simplificar o de transformar sus tareas de procesamiento de información explotando directamente la estructura del mundo real; explotación que permite usar el mundo como su propio modelo y reestructurar físicamente el ambiente con tal de reducir la complejidad computacional de la resolución de problemas (cf. Clark 1997a, 2001a).

En cuanto al alejamiento de la realidad neurológica, éste se revela en los recursos usados por los conexionistas tempranos para elaborar sus redes: grupos relativamente pequeños de unidades y conexiones comparadas con el cerebro, que abordan problemas relativamente discretos y bien definidos. Estas redes tienden a entrenarse en versiones artificiales de problemas del mundo real. Estas versiones reducen significativamente los vectores de input y output que demandarían los datos sensoriales reales y el control motor. Más aún, habitualmente se enfocan en un solo problema, lo que se revela en la escasa variedad de uso de las representaciones (o patrones de activación) conexionistas, esto es, que las representaciones no sirven para diversos contextos y sólo para tareas específicas⁴. Las redes neurales efectivas, por el contrario, deben lidiar con inputs y outputs de una gran dimensionalidad, y deben de alguna manera arreglárselas con el hecho

³ Recuérdese aquí que esta también era una crítica que se le hacía a la TRM. Ahora bien, en la TRM la realizabilidad física era sólo un desideratum a cumplirse y no tenía injerencia en cómo fuese la representación.

⁴ Esto puede leerse también en la crítica a que los sistemas conexionistas carecerían de composicionalidad de Fodor y Pylyshyn.

de que muchas veces somos asaltados por grupos de datos que corresponden a múltiples problemas y que requieren luego ser escogidos y distribuidos internamente. Las redes biológicas contribuyen así a realizar diversas clases de tareas solucionadoras de problemas. Esta idea de que el cerebro puede operar utilizando una amplia multiplicidad de sub-redes de propósito especial encuentra apoyo creciente en la neurociencia contemporánea, pero la mayoría de las redes artificiales permanecen todavía bastante distantes de los detalles de la investigación neurocientífica real. Los conjuntos neuronales reales exhiben una amplia variedad de propiedades de las que carecen la mayoría de los modelos conexionistas, como por ejemplo los efectos no locales, procesamiento continuo en el tiempo, el uso de una variedad de diferentes tipos de función de activación y de conectividad muy recurrente, pero no simétrica (Clark 2001a: 80-81).

Lo anterior revela que los conexionistas se han enfocado, tal como los proponentes de la TRM, en fragmentos aislados de la competencia cognitiva a nivel humano, construyendo sus redes a partir de inputs bastante alejados de los datos sensoriales y produciendo outputs que no logran generar acciones del tipo encontrado en la cognición biológica (como movimientos de brazos y bloques) o imitan sólo sub-clases de acciones particulares muy restrictivas (como la producción de cuerdas de fonemas o fonemas característicos de una expresión en tiempo pasado) (Clark 1995: 90).

Elman (1999: 496-97) subraya además otras dos deficiencias de los modelos conexionistas, a saber, la pasividad de la red neural y la cognición asocial. La red neural es pasiva en el sentido de que si se abandona a sus propios mecanismos no hace nada muy interesante. Incluso después de aprender a resolver un problema determinado, pocos modelos conexionistas exhiben algún comportamiento que sea generado internamente. A diferencia de la cognición en el “mundo real”, donde incluso organismos biológicos muy primitivos exhiben un comportamiento dirigido a metas, la mayoría de los sistemas conexionistas son reactivos o sus metas están programadas de antemano por una agencia externa (el programador). La cognición asocial se refiere a que la mayoría de los modelos conexionistas consideran la cognición como un fenómeno esencialmente individual: ocurre dentro de la caja craneana. Esto determina que en algún sentido la representación sea vista como “interna” respecto de un agente y, por tanto, los conexionistas también pueden verse como adscribiendo al individualismo metodológico. Así, estos modelos se enfocan en competencias tales como el ajedrez, la resolución de problemas y el reconocimiento de patrones. Pero, sin embargo, en los humanos en particular la cognición es un fenómeno social, y dicha inserción en un medio social puede incluso determinar el uso de las representaciones (cf. Clark 1997a, 2001a).

IV. Precisiones: El Lugar de la Representación

Hemos visto que la ciencia cognitiva tradicional descansaba fuertemente en la noción de representación para comprender la vida mental, no importando para la definición de dicha noción el substrato físico en el que era implementada. Ahora bien, dados los problemas a los que se enfrentaba la TRM, la pregunta por cómo se

daban en la cognición real dichos estados internos se hizo apremiante. Y el conexionismo presentó una alternativa, postulando que sí se daban dichos estados internos, pero que éstos no calzaban tan bien con las estructuras y los contenidos de las actitudes proposicionales. De este modo, el conexionismo demostró que era posible resolver problemas cognitivos sin el aparato asociado con la hipótesis original de un sistema de símbolos físicos, alejándose así de una correspondencia entre la vida mental y la sintaxis del lenguaje natural.

El conexionismo dio entonces una respuesta a varias de las fuentes de insatisfacción que estaban surgiendo respecto de los modelos simbólicos, como la degradación refinada y la generalización a casos nuevos, erigiéndose así en un desafío para la visión clásica del cerebro del tipo de la TRM, esto es, como una máquina sensible a la sintaxis. Este desafío de la noción de representación como una contenedora pasiva e inferencial se basa en tres características, a saber, (1) la fusión conexionista del almacenamiento y del procesamiento de datos, (2) el reemplazo de la inferencia definida sobre ítems cuasi-lingüísticos por transformaciones de vector a vector, y (3) la concepción de las representaciones como representaciones dedicadas a la realización de tareas específicas. Estas tres características hacen de la noción de representación una noción bastante alejada de aquella de la TRM, según la cual procesamiento e información se distinguen con claridad. En el conexionismo, es el uso de un patrón de activación al tratar de ejercer una tarea específica el que determina la existencia misma de una representación y su continuidad, de llevar a cabo la tarea satisfactoriamente. De alguna manera, pese a que las tareas escogidas en el conexionismo se alejan también del ambiente efectivo, la representación como patrón de activación se vincula con tareas “ambientalmente” más plausibles que aquellas con las que opera la TRM, presa de la imagen de la inteligencia deliberativa.

En cuanto al rol del ambiente en la cognición, el trabajo reciente en robótica situada (ver próximo capítulo) desafía la tendencia de la TRM, pero también del conexionismo, a concebir la noción de representación como un “contenedor” de información y como una entidad básicamente interna. Esto se ha logrado prestando mayor énfasis en temas como el control, la acción y la intervención en tiempo real. El trabajo en robótica situada y en sistemas dinámicos toma en serio la idea de que la forma de nuestros recursos internos ha sido modelada y conformada de maneras inesperadas y contra-intuitivas por nuestra encarnación física y por las demandas del mundo real por acciones rápidas y robustas en respuesta a inputs sensoriales directos (Clark 1995: 90).

Para dar una imagen más global de la cognición humana, queda aún por incorporar rasgos que reconocen los roles profundos jugados por los recursos externos y no biológicos en el desarrollo del éxito cognitivo, tales como la acción corporal, los instrumentos y los artefactos, el ambiente local y las estructuras simbólicas externas, tales como el lenguaje.

En los capítulos siguientes, veremos cómo el cuerpo y el ambiente se convierten en factores relevantes en la explicación de la conducta inteligente, cómo varios aspectos cognitivos pueden explicarse atendiendo a sus dinámicas internas y cómo la “encarnación” corporal puede determinar lo que sea la cognición. Estas consideraciones entrañan, en sus formulaciones más extremas, la negación de la noción de representación como una noción necesaria para la

explicación de la vida cognitiva. Se explorará a continuación, entonces, qué hay de propiamente representacional en la cognición humana y si la noción de representación tiene un rol efectivo en diversas clases de acciones cognitivas en las que los agentes se comprometen. Esto implica la reformulación de la noción de representación, no sólo en términos de considerar los factores que están más allá de la caja craneana, sino cómo estos factores pueden ser determinantes para cómo se constituye la representación y para qué sea aquello que se representa. Tal reformulación se hace básicamente en términos de una noción que se construye sobre el operar biológico, entendido éste como la captación de regularidades marginales en los inputs sensoriales.

CAPÍTULO SEGUNDO

ROBÓTICA SITUADA

En el capítulo anterior vimos el modelo representacional / computacional de la mente, en el que la representación tenía un lugar preciso en el operar de los seres inteligentes, operar que se consideraba principalmente como resolver problemas. Esta visión y sus problemas forzaron la aparición de otro marco explicativo, el conexionismo, donde la representación aún tenía importancia como factor explicativo de la

cognición humana, pero era entendida como una entidad emergente y determinada por el carácter superposicional y distribuido de las redes neurales. En este último enfoque, el operar cognitivo se centraba básicamente en tareas de reconocimiento y recolección.

Actualmente, la atención a nuevos tipos de problemas, que amplían la definición de la inteligencia humana más allá de los límites del pensamiento deliberativo, ha incorporado rasgos como la “encarnación” – la propiedad de tener un cuerpo– de los agentes cognitivos, la interacción de este cuerpo con el ambiente en el que se anida tanto en la vida cotidiana como en la historia evolutiva y el rol jugado por los recursos externos no biológicos en el desarrollo del éxito cognitivo, tales como la acción corporal, los instrumentos y los artefactos, el ambiente local y las estructuras simbólicas externas.

Esto ha generado una visión de la cognición y del comportamiento adaptativo como “encarnada”, situada y dinámica, que puede hallarse en ámbitos tan diversos como la robótica, la neurociencia y la psicología del desarrollo cognitivo. La “encarnación” y el “estar situado” subrayan los roles jugados por las circunstancias físicas y contextuales de un agente en la generación de su comportamiento. No sólo importa aquí considerar el ambiente y el cuerpo, sino hasta dónde este tener cuerpo determina qué y cómo percibimos y actuamos adaptativamente sobre el ambiente que nos rodea. Para un agente encarnado, situado, el tomar una acción apropiada se convierte en una preocupación primaria, y la biomecánica del agente, la estructura de su ambiente, la retro-alimentación continua entre el sistema nervioso, el cuerpo y el ambiente, además del contexto social, se convierten a la vez en constreñimientos significativos y en recursos para la acción (Beer 2003).

El comportamiento adaptativo es considerado así como el resultado de una interacción continua entre el sistema nervioso, el cuerpo y el ambiente, cada uno de los cuales tiene una dinámica rica, complicada y altamente estructurada. El rol del sistema nervioso no consiste tanto en dirigir o programar el comportamiento como en darle forma y evocar los patrones apropiados de dinámica a partir de un sistema enteramente acoplado. Como consecuencia, uno no puede dar crédito por el comportamiento adaptativo a ninguna de las piezas de este sistema acoplado. En esta visión, pueden emerger nuevos comportamientos que son propiedades sólo del sistema acoplado.

La consideración de los agentes cognoscentes como “encarnados” y “situados” ha llevado a una nueva definición de cuáles son los problemas relevantes para la cognición, lo que ha conducido a replantear la pregunta con respecto a si la cognición humana requiere de representaciones internas estructuradas. Nociones como el acoplamiento entre sistema nervioso, cuerpo y ambiente, que permean todas las “teorías encarnadas” de la cognición parecen, en sus visiones extremas, no requerir de una noción de representación para explicar el comportamiento inteligente. La razón de esto reside en que no está claro qué se entiende por “representación”, es decir, qué significa que un sistema (típicamente, un sistema interno, como el sistema nervioso) representa a otro (típicamente, un sistema externo, como el ambiente), si lo que da cuenta del comportamiento es la dinámica del sistema global (que incorpora sistema nervioso, cuerpo y ambiente); pero también porque en los ejemplos favorecidos por las visiones encarnadas no parece tener lugar la noción de

una representación (que, por ejemplo, ayudaba a resolver un problema como en la TRM o a reconocer un objeto como en el conexionismo).

Para entender cómo toma forma este escepticismo respecto de la representación interna se revisará en este capítulo y en los dos siguientes tres enfoques, que en sus formas más radicales niegan que la noción de representación tenga algún rol en el operar cognitivo: la robótica situada, los sistemas dinámicos y los enfoques que se centran en nuestra encarnación biológica para explicar la cognición. El trabajo en robótica situada explora la idea de que la forma de nuestros recursos internos ha sido modelada y conformada de maneras inesperadas y contra-intuitivas por nuestra encarnación física y por las demandas del mundo real en cuanto a acciones rápidas y robustas en respuesta a inputs sensoriales directos. La teoría de sistemas dinámicos estudia cómo el desarrollo del comportamiento en el tiempo puede explicarse mediante la interacción entre sistema nervioso, cuerpo y mundo. Las orientaciones “encarnadas” en biología intentan comprender cómo la forma de nuestros recursos internos ha sido modelada además por nuestra historia ontogénica y evolutiva.

I. Robótica Situada

En los años '80 en el laboratorio del MIT se diseñó y construyó un robot llamado Herbert. Su misión consistía en recolectar latas vacías de bebidas que se habían dejado en el laboratorio. Ésta no es una tarea trivial: el robot debe negociar con un ambiente desordenado y cambiante, evitar derribar las cosas, chocar con la gente e identificar y recolectar las latas. Un artefacto producido según la imagen computacional / representacional trataría de resolver este complejo problema del mundo real usando datos visuales minuciosos para generar un mapa interno detallado de los alrededores, aislar las latas y planear una ruta; esto es, básicamente formular y tratar de resolver un problema con recursos simbólicos. Herbert, en cambio, no opera de acuerdo a un plan central, sino que responde a comportamientos simples que incluyen el evitar obstáculos (parar, reorientarse, etc.) y rutinas de locomoción. Así, una rutina de locomoción se interrumpe si se detecta por parte de un sistema visual simple un contorno similar al de una mesa. Cuando Herbert está junto a una mesa, las rutinas de locomoción y de evasión de obstáculos ceden su lugar a otros subsistemas que barren la superficie de la mesa con un láser y una cámara de video. Una vez que el contorno básico de una lata ha sido detectado, el robot rota hasta que el objeto similar a una lata esté en el centro de su campo de visión. En este punto, las ruedas paran y se activa el brazo del robot. El brazo, equipado con sensores de tacto simples, explora gentilmente la superficie de la mesa enfrente de él. Una vez que Herbert encuentra la forma distintiva de una lata, prosigue con un comportamiento de asir: la lata es recolectada y el robot continúa en su búsqueda de latas.

Herbert podría describirse en la jerga robótica actual como un *agente situado*. Un agente situado es un sistema físico real fundado y anidado en un mundo real, aquí y ahora, que actúa y reacciona en tiempo real. Los agentes situados operan no sólo

con acciones situadas, sino también con un razonamiento situado y con una percepción situada. No es un agente simple en un mundo estático, sino que tal agente se enfrenta a múltiples agentes en un mundo dinámico. La inmersión en un mundo dinámico implica que un agente situado debe responder a sus inputs en tiempo real considerando ciertas restricciones ambientales y adaptativas específicas, y éste puede usar su percepción del mundo en vez de un modelo de mundo objetivo. La “encarnación” es crítica aquí por dos razones, a saber, primero, sólo un agente con cuerpo puede validarse completamente en el mundo real, segundo, sólo a través de un fundamento físico un sistema cualquiera puede encontrar su nivel más básico y dar significado al procesamiento que sucede dentro del sistema. Sin una participación en, y una percepción del, mundo, no hay significado para un agente – todo se reduce a símbolos vacíos que se refieren a otros símbolos–. La respuesta temporal, por su parte, se incorpora con la noción de que el agente opera “on-line”, en el sentido de que un circuito computa una huella de output como una transducción de sus huellas de input (Mackworth 1997: 277).

La robótica situada es el estudio de robots anidados en ambientes complejos, que a menudo cambian dinámicamente. Este “estar situado” se refiere precisamente al existir en un ambiente y a que el comportamiento es fuertemente afectado por tal ambiente. La “encarnación” se relaciona con el “estar situado” y se refiere, como ya mencionamos, al tener un cuerpo físico que interactúa con el ambiente. De este modo, la “encarnación” es una forma de “estar situado”: las acciones del agente están directa y fuertemente afectadas por el cuerpo que tiene. Un aspecto importante, entonces, de la robótica situada es que los robots están encarnados: poseen un cuerpo físico para sentir su ambiente y actuar y moverse en él (Mataric 2002). La robótica situada se enmarca dentro del “movimiento situado”, que subraya que la razón no es un fin en sí misma, sino que sirve al propósito de permitir la selección de acciones que afectarían el ambiente del agente que razona de modos favorables. Así, el razonamiento siempre ocurre en un contexto específico y con metas específicas.

En esta interacción entre agente, cuerpo y ambiente un aspecto importante es el control del robot. Mediante tal control se adquiere información sobre el ambiente a través de los sensores del robot y se elabora dicha información para producir un comportamiento. La complejidad del ambiente, esto es, el grado de “estar situado”, se relaciona claramente con la complejidad del control, que a su vez está directamente relacionada con la tarea del robot: si la tarea requiere del robot que éste reaccione rápida, aunque inteligentemente, en un ambiente desafiante y dinámico, el problema del control es muy difícil. Si éste debe reaccionar más lentamente o el ambiente es menos complejo, el problema del control se vuelve más sencillo. Actualmente, hay cuatro diferentes clases fundamentales de metodologías de control de robot, habitualmente encarnadas en cuatro arquitecturas específicas de control de robot, a saber, control reactivo (“no pienses, reacciona”), control deliberativo (“piensa, luego actúa”), control híbrido (“piensa y actúa independientemente y en paralelo”) y control basado en el comportamiento. Estas cuatro clases de control deben confrontarse con tres hechos básicos: el pensar es lento, pero la reacción a menudo debe ser rápida; el pensar permite planear para evitar acciones incorrectas, sin embargo, pensar por mucho tiempo puede ser

peligroso; para pensar, el robot necesita potencialmente una gran cantidad de información precisa que debe mantenerse al día, pero, sin embargo, el mundo continúa cambiando mientras el robot está pensando, de manera que mientras más piensa, más imprecisas son sus soluciones⁵ (Mataric, Ibid.). De estas cuatro clases diferentes de metodología de control de robot, son los robots con *control basado en el comportamiento* los que plantean las dudas más serias respecto de qué sea la representación y cómo ésta se origina, por lo que trataremos este tipo de robots con mayor detalle.

Los robots basados en comportamientos se basan en la idea de Rodney Brooks según la cual se deben considerar animales más simples como el fundamento de un modelo “hacia arriba” para construir la inteligencia. Dejando de lado el razonamiento como primer componente del intelecto del robot, la interacción del robot con su ambiente se estimó como el determinante primario de la estructura de su inteligencia, siendo difícil demarcar qué es propiamente inteligencia y qué, interacción ambiental. La inteligencia misma no tiene en esta visión un lugar determinado, sino que ésta –o la funcionalidad del sistema como un todo– emerge de la interacción entre los componentes del sistema –como en Herbert los módulos de “evitar obstáculos”, “pararse”, “control de mirada”, etc.– (esto es lo que se conoce como emergencia, cf. Elman 1999: 498).

Para estos robots con control basado en el comportamiento, que Brooks denominó “criaturas”, defendió la idea de una *arquitectura de subsumión*, consistente en un diseño que incorpora varios subsistemas cuasi-independientes, cada uno de los cuales es responsable por un aspecto encerrado en sí mismo de la

⁵ En detalle: el control reactivo es una técnica para acoplar ajustadamente los inputs sensoriales y los outputs de efecto, de manera que el robot responde muy rápidamente a ambientes cambiantes y no estructurados. Aquí se privilegia la reacción rápida por sobre la complejidad del razonamiento. En el control deliberativo, en cambio, el robot utiliza toda la información sensorial disponible y todo el conocimiento internamente acumulado para pensar acerca de qué acciones tomar. Habitualmente la toma de decisión resulta del planificar, que requiere la existencia de una representación interna del mundo y permite que el robot mire adelante en el futuro para predecir los resultados de las acciones posibles en diversos estados. El planificar es lento y, si el robot está en un ambiente dinámico y ruidoso, una solución precisa es improbable, por lo que pocos robots situados son puramente deliberativos. El control híbrido combina los mejores aspectos del control reactivo y deliberativo, intentando combinar la respuesta en tiempo real con la racionalidad y la eficiencia de la deliberación. La interacción entre el componente reactivo y el deliberativo no es fácil, por lo que se requiere un componente intermedio que la regule. De ahí que los robots híbridos sean llamados también “sistemas de tres capas”: la capa reactiva, la deliberativa y la intermedia. El control basado en el comportamiento se inspira en la biología y trata de modelar cómo los animales lidian con sus ambientes complejos. Los componentes de estos sistemas se llaman “comportamientos” y consisten en patrones observables de actividad que emergen de las interacciones entre el robot y su ambiente (ver más abajo). Así como los sistemas híbridos, los sistemas basados en comportamiento pueden organizarse en capas, pero éstas no difieren significativamente una de otra en términos de escala temporal y representaciones usadas. Éstos son similares también a los sistemas reactivos en cuanto a su construcción en capas crecientes y módulos distribuidos, pero, a diferencia de éstos, pueden almacenar representaciones de un modo distribuido. De este modo, los componentes de planear y el razonar del sistema basado en el comportamiento usan los mismos mecanismos que los comportamientos de sentir y de actuar, y operan así en una escala de tiempo similar y con una representación similar. En este sentido, el “pensar” está organizado muy similarmente al “actuar” (Mataric, Ibid.).

actividad del agente, lo que puede pensarse también como que cada uno de estos subsistemas tiene su propio propósito. La actividad de un subsistema responde a una reacción por parte del agente a cambios peligrosos o importantes en su ambiente, reacciones que son rápidas precisamente porque no hay representaciones ni razonamiento acerca de ellas. Dado que el propósito de la criatura está implícito en sus propósitos de mayor nivel, no se requiere de una representación explícita de propósitos que sea seleccionada por algún procesador central (o distribuido) para decidir qué es más apropiado que haga a continuación la criatura. De este modo, los subsistemas no están coordinados por ningún sistema central, y son capaces sólo de enviar señales simples que circunvalan, anulan u ocasionalmente modifican la respuesta de otros subsistemas. Al tener múltiples subsistemas en paralelo y al remover la idea de una representación central, hay una posibilidad menor de que cualquier cambio en las propiedades del mundo pueda causar un colapso total del sistema. Un cambio podría incapacitar alguno, pero no todos los niveles de control, de modo que mientras se entra en un mundo más extraño la actuación de la criatura puede degradar refinadamente⁶. Un sistema tal permite un camino creciente (o “arquitectura de incrementación”) que va desde sistemas muy simples a sistemas complejos inteligentes autónomos. En la construcción de capas debe recapitularse la evolución, o una aproximación de ella, como metodología de diseño, en cuanto las mejoras en actuación se producen por añadir un circuito específico respecto de una situación, mientras se deja el antiguo en su lugar, capaz de operar si este nuevo circuito fallara (Brooks 2000, 1997, Clark 1999a)⁷.

Las criaturas de Brooks (2000, 1997; ver también Brooks et al. 1998, Adams et al. 2000) son concebidas como agentes móviles autónomos que coexisten en el mundo con los humanos y que son vistos por esos humanos como seres inteligentes por derecho propio. Tales criaturas deben arreglárselas apropiada y temporalmente con cambios en su ambiente dinámico; deben ser robustas respecto de su ambiente, de manera que cambios menores en las propiedades del mundo no lleven a un colapso total del comportamiento de tales criaturas, sino que ellas debieran cambiar gradualmente en tanto el ambiente se transforma más y más; deben ser capaces de mantener múltiples metas y, dependiendo de las circunstancias que encuentran en sí mismas, cambiar qué metas particulares están persiguiendo activamente, de manera

⁶ Pese a que recuerda a una red conexionista al basarse en redes que subsumen a otras más simples, Brooks (2000) defiende que sus criaturas no son “conexionismo”. Esto, Los nodos de procesamiento conexionista tienden a ser uniformes, y ellos buscan una manera de descubrir a partir del aprendizaje cuál es la mejor manera de interconectarlos. Por el contrario, en las criaturas todos los nodos son estados finitos de máquina únicos, la densidad de conexiones es mucho menor, no es del todo uniforme y es especialmente baja entre capas.

⁷ Esto concuerda con la observación de Philip Agre respecto de que la mayoría de la actividad cotidiana es rutinaria, en el sentido de que requiere poco-si-algo de nuevos razonamientos abstractos. La mayoría de las tareas, una vez aprendidas, puede ser logradas de una manera rutinaria, con pocas variaciones. Agre propuso que una arquitectura de agente eficiente podía basarse en la idea de “recorrer argumentos”. A grandes rasgos, la idea es que, dado que la mayoría de las decisiones es rutinaria, ellas pueden codificarse en una estructura de bajo nivel (como un circuito digital), que sólo necesita una puesta al día periódica, quizás para manejar nuevos tipos de problemas.

que se adapten tanto al ambiente como a las metas específicas, y medrar sobre la base de circunstancias fortuitas⁸.

Otra versión de los robots basados en comportamientos es la de Beer y su equipo (cf. Beer et al. 1997, Chiel et Beer 1997), quienes se han centrado en robots biológicamente inspirados. Ellos han demostrado que al incorporar propiedades biomecánicas similares a aquellas de los insectos a robots hexápodos se puede simplificar su control, permitiéndoles atravesar terrenos irregulares y haciéndolos robustos frente a las lesiones. La inspiración en la biología estriba principalmente en la evolución: los animales han evolucionado para prosperar bajo las condiciones complejas del mundo real y ellos a menudo dependen o incluso explotan activamente dicha complejidad en sus comportamientos⁹.

Este equipo se ha enfocado particularmente en los insectos, puesto que ellos exhiben una rica variedad de comportamientos complejos, incluyendo un comportamiento motor altamente hábil, aprendizaje y memoria, comunicación y comportamiento cooperativo. Pese a esta complejidad, un estudio detallado de los invertebrados y del control neural subyacente es a menudo más realizable técnicamente que el de los vertebrados. A partir de estos estudios, se ha mostrado que la locomoción en estos animales está distribuida. Esto significa que, antes de concentrar la responsabilidad de la locomoción en un sistema centralizado, los animales distribuyen esta responsabilidad a lo largo de las características físicas de sus dispositivos de locomoción, en el sistema de circuitos local reflexivo de cada dispositivo y en el sistema de circuitos que interconecta los controladores de los dispositivos. Antes que tener un módulo singular que dirija los movimientos de los dispositivos de una manera señor-esclavo, en un controlador de locomoción distribuida el caminar estable surge de las interacciones cooperativas entre muchos componentes diferentes. Tal enfoque distribuido es capaz de disminuir significativamente las demandas computacionales puestas en cada componente del sistema de locomoción, prestando flexibilidad y robustez al comportamiento general. En estos robots biológicamente inspirados se han subrayado las aplicaciones de varios principios generales de los insectos, incluyendo el control altamente distribuido, la combinación sinérgica de reflejos locales con el control distribuido del caminar y el acoplamiento de las propiedades mecánicas del cuerpo a las demandas del terreno y del diseño del controlador. La aplicación de tales principios ha resultado en artefactos que pueden moverse efectivamente por sobre terreno complejo con recursos computacionales muy modestos.

⁸ Brooks y su equipo del laboratorio de Inteligencia Artificial del MIT se han abocado además a la construcción efectiva de un robot humanoide, con tal de explorar materias que ellos consideran críticas para la inteligencia humana tal como la estructura de desarrollo, la encarnación y el acoplamiento físico, la integración de sistemas de múltiples sensores y motores y la interacción social. Este proyecto tiene un objetivo ingenieril, a saber, construir un prototipo de robot autónomo flexible y diestro de propósito general, y uno científico, esto es, entender la cognición humana. Hasta el momento, dicho equipo ha construido un robot del torso para arriba llamado “Cog”. Este avance en la concepción de Brooks y su equipo se denomina robótica cognitiva o “cognobótica”.

⁹ Recuérdese aquí que Brooks también suscribe la idea de que la construcción de agentes inteligentes debe seguir el “camino” de la evolución.

II. Robótica Situada vs. GOFAI

La robótica situada muestra un giro importante con respecto a la IA clásica, cuya meta era replicar la inteligencia humana en una máquina. La IA clásica, como vimos en el capítulo anterior, asumió tres compromisos claves. Primero, el uso de un código simbólico como un medio para guardar todo el conocimiento a largo plazo del sistema. Segundo, la descripción de la inteligencia como la habilidad de buscar exitosamente un espacio de problema simbólico. Tercero, la inteligencia reside en, o es cercana a, el nivel del pensamiento deliberativo. La visión de la inteligencia era una que cifraba la inteligencia en la capacidad de resolver problemas, consistiendo en la búsqueda controlada de información almacenada simbólicamente y empleando procesos de búsqueda. Tales procesos suponían la generación, composición y transformación de las estructuras simbólicas hasta encontrar las condiciones especificadas para una solución por parte de un controlador interno. Este controlador interno coordina o controla la cognición y puede entenderse en los términos definidos para el “control deliberativo” descrito en el punto I (ver más arriba).

Esta visión de la inteligencia que dependía profundamente de la representación y del pensamiento deliberativo o resolutor de problemas, es decir, del razonamiento que llevaba a cabo una manipulación basada en reglas de estructuras simbólicas, es lo que algunos autores identifican como GOFAI (siglas en inglés por *good old fashion artificial intelligence*, algo así como la “buena inteligencia artificial tradicional”) (Brooks 1991; Mackworth 1997). La GOFAI deja de lado, o considera a lo más como secundarias, preocupaciones como el modo en que un sistema siente el mundo y actúa en él. Mackworth (Ibid.: 272) caracteriza el modelo típico en IA y en la robótica de la IA tradicional como uno que consiste en tres módulos para la percepción, el razonamiento y la acción. Los supuestos que guían la investigación de esta IA son:

- todo lo que es útil para un agente puede ser descrito en términos de individuos y relaciones entre individuos
- las creencias de un agente acerca del mundo son verdaderas y justificadas
- el conocimiento del mundo por parte de un agente es definido y claro
- el conocimiento del mundo por parte del agente es completo. Esto requiere que todo lo relevante acerca del mundo sea conocido por el agente. Este supuesto de “mundo cerrado” permite al agente suponer con seguridad que un hecho es falso si no puede inferir que es verdadero
- el ambiente es estático a menos que un agente lo cambie
- hay sólo un agente en el mundo
- dada una descripción completa y definida del mundo, el agente puede predecir todos los efectos de sus acciones

- las acciones son discretas y son llevadas a cabo secuencialmente

La inteligencia humana era entonces entendida basándose en modelos internos monolíticos, en control monolítico y en un procesamiento multi-propósito. Esta imagen parece contradecirse con ciertos hechos naturales conocidos actualmente, tales como que las interacciones diarias entre agente y ambiente no parecen depender de la construcción y uso de modelos internos detallados de toda la escena tridimensional, que la percepción de bajo nivel puede “llamar” a rutinas motoras que generan un mejor input perceptual y por ello mejoran la información a seleccionar, que las actividades en el mundo real pueden a veces jugar un rol importante en el proceso computacional mismo, que la representación interna de eventos y estructuras puede ser menos similar a una estructura o descripción de datos pasiva y más similar a una receta directa para la acción (Clark 2001a: 88).

En el mismo sentido, Brooks (1997) afirma que los humanos, al realizar nuestras actividades cotidianas, tendemos a minimizar la representación interna del mundo y no construimos modelos completos y monolíticos del ambiente, sino que representamos sólo aquello que nos es inmediatamente relevante del ambiente, y aquellas representaciones no tienen un acceso completo una por parte de la otra. El modelo de un control monolítico tampoco parece funcionar si se quiere hacer un robot que interactúe dinámicamente con su ambiente (ver más arriba la discusión sobre los diversos tipos de control de robot). Y la afirmación de que la inteligencia es multi-propósito parece contradecirse con la observación de que los humanos parecen ser eficientes sólo en conjuntos particulares de habilidades en desmedro de otras habilidades, a menudo de maneras no obvias. Además, la inteligencia humana parece desempeñarse pobremente en el área de la lógica deductiva y no usar a menudo reglas similares a sub-rutinas para tomar decisiones, siendo éstas más a menudo emocionales que racionales.

La diferencia entre el enfoque de la GOFAI y el de la robótica situada puede verse claramente en los diferentes diseños de inteligencia artificial. Así, en la GOFAI las respuestas de comportamiento no se diseñan en el sistema, sino que más bien ellas emergen de la interacción entre el planificador y las metas dadas y de un modelo particular de mundo, que ha sido construido a partir de datos sensoriales. En los robots basados en comportamiento, en cambio, las respuestas de comportamiento están explícitamente diseñadas en el sistema, pero no hay ninguna meta explícitamente representada, aunque tales metas puedan ser atribuidas por quien observa a un robot tal (por ejemplo, podemos pensar que Herbert tiene como meta recoger latas de bebidas). Así, en la GOFAI no hay nunca una secuencia de acciones explícitamente representada a priori en el programa del robot, sino que tal secuencia es un resultado del razonar acerca del modelo de mundo y de las metas asignadas al sistema. En el enfoque basado en el comportamiento, en cambio, a menudo hay acciones específicas representadas, aparejadas a otras acciones o a condiciones perceptuales. Las metas específicas del robot, sin embargo, nunca están representadas, sino que están implícitas en el acoplamiento de acciones con condiciones perceptuales, y la ejecución aparente de planes se desarrolla en tiempo real mientras un comportamiento cambia la configuración del robot en el mundo, de

tal modo que nuevas condiciones perceptuales gatillan el próximo paso en una secuencia de acciones (Brooks 1997).

La distancia en los modelos artificiales de la GOFAI con respecto a los de la robótica situada se debe, según Brooks (1997), principalmente a dos cambios de perspectiva: el primer cambio y el más radical, evidente en las características de la robótica situada, consistió en abandonar la idea de que los sistemas inteligentes debían tener un modelo central del mundo, típicamente simbólico en cuanto a su naturaleza. El problema con este modelo central es su escasa practicidad: con frecuencia no permite la respuesta rápida, en tiempo real. La información sensorial entrante debe ser convertida en un código simbólico único, de manera que un control central pueda operar con él. Además, los outputs del controlador central deben ser transformados desde este código propio en los diversos formatos requeridos para controlar varios tipos de respuesta motora. Estos pasos de traducción consumen mucho tiempo y son costosos (Clark 1997a: 21). El segundo cambio, más gradual y no del todo aceptado en la comunidad que basa el diseño de sus robots en comportamientos, consistió en abandonar la idea de la inteligencia como un proceso computacional que toma un input y produce un output. En las nuevas visiones, la inteligencia se entiende como qué son capaces de hacer los agentes situados en el mundo, teniendo en cuenta la dinámica de sus interacciones con el mundo junto con sus procesos internos.

El trabajo en agentes autónomos subraya entonces que ha de evitarse un modelamiento excesivo del mundo y que para éste han de tomarse en cuenta las demandas de los sistemas que producen comportamiento en tiempo real. Además, recalca, ha de tenerse en cuenta el encaje muy justo entre las necesidades y los estilos de vida de sistemas específicos (animales, robots o humanos) y las clases de estructuras ambientales portadoras de información a las que ellos responden. Esto apunta a la idea de que nosotros reducimos la carga de procesamiento de información al sensibilizar el sistema a aspectos particulares del mundo, aspectos que tienen una importancia particular dado el nicho ambiental que el sistema habita (Ibid.: 23 y ss.).

III. Robótica Situada y Representación

Brooks, partiendo de estos dos cambios de perspectiva, concluye que:

al examinar inteligencia de nivel muy simple nos encontramos con que las representaciones explícitas y los modelos del mundo simplemente estorban el camino. Resulta mejor dejar que el mundo mismo sirva como su propio modelo (2000: 396).

La hipótesis que sustenta consiste en que “la representación es la unidad de abstracción equivocada al construir la mayor parte de los sistemas inteligentes” (Ibid.). La idea fundamental de Brooks consiste en que, evolutivamente, requerimos en primer lugar saber actuar y reaccionar, y luego se puede llegar en forma más o menos simple al comportamiento que resuelve problemas, al lenguaje, al conocimiento y aplicación expertos y a la razón. Lo esencial para la inteligencia es saber movernos en un ambiente dinámico, sintiendo el entorno en un grado suficiente como para lograr la mantención de la vida y hacer posible la

reproducción. Es en esta parte esencial de la inteligencia donde la evolución ha concentrado su tiempo. En el mundo real, como muestra la robótica situada, no hay una división clara entre percepción (abstracción) y razonamiento. Dado que la “encarnación” y el acoplamiento físico son centrales para la inteligencia humana, Brooks (1997) subraya que el acoplamiento físico entre acción y percepción reduce la necesidad de una representación intermediaria, debido a que, para un sistema “encarnado”, las representaciones pueden estar fundadas en último término en las interacciones senso-motoras con el mundo.

En su versión más radical, Brooks plantea que nuestros sistemas están acoplados físicamente con el mundo y operan directamente en ese mundo, sin ninguna representación explícita de él. Existen representaciones o acumulaciones de estados, pero éstas sólo se refieren a las operaciones internas del sistema, no tienen significado sin la interacción con el mundo externo. Así, mucho de la actividad inteligente, incluso a nivel humano, es una reflexión del mundo a través de mecanismos muy simples y no contempla una representación detallada del mundo (Brooks 2000: 406-7). En el actuar de las criaturas, entonces, se descubre la prescindencia de una actividad que represente al mundo con tal de explicar la acción de ellos: es el “acoplamiento” con el mundo lo que es relevante para la explicación de la actividad de las criaturas. Brooks defiende que no hay representaciones del tipo de la IA tradicional ni a nivel general ni a nivel local, siendo lo más que puede decirse de sus representaciones que “un número ha sido pasado de un proceso a otro. Pero es sólo mirando el estado tanto del primero como del segundo proceso que puede darse a aquel número cualquier representación” (Ibid.: 406). Aquí no hay variables que requieran instancias en procesos de razonamiento, ni reglas que necesiten ser seleccionadas a través de unión de patrones, ni elecciones que hacer. En gran medida, es el estado del mundo el que determina la acción de la criatura, y gran parte de la actividad humana es un reflejo del mundo a través de mecanismos muy simples, que no emplean representaciones detalladas.

De este modo, la “robótica situada” plantea en efecto un desafío a la inteligencia concebida al modo de la TRM o del conexionismo. Ya no hay aquí una suerte de sistema central mediador que define la acción sobre la base de representaciones y computaciones de representaciones, sino que nos enfrentamos ahora a sistemas con múltiples “capas”, donde cada “capa” corresponde a un comportamiento específico. Esto implica que el sistema esté esparcido en múltiples “capas”, que no están conectadas particularmente por vía de los datos o relacionadas según su función, e input y output ya no son fácilmente distinguibles. Dado que cada capa productora de actividad conecta la percepción a la acción directamente, aquí no se requiere la noción de representación para que el robot exhiba un comportamiento inteligente.

El trabajo en vida artificial, entonces, intenta reconfigurar las ciencias de la mente al enfatizar la importancia de factores diferentes de la computación individual y la cogitación. Estos factores incluyen (1) las maneras a menudo inesperadas en los que múltiples factores (neurales, corporales y ambientales) pueden converger en la resolución de problemas naturales, (2) la habilidad de sustentar una respuesta adaptativa robusta sin un control o planeador central y (3) la potencia general de las reglas simples y de rutinas de comportamiento que operan con un rico telón de fondo de otros agentes y de estructura ambiental (Clark 2001a: 109).

IV. Discusión

Las nociones de representación interna, modelos internos de mundo y otros de esta índole fueron introducidos para ayudar a explicar un rango de comportamientos significativamente diferentes de aquellos estudiados por la mayoría de los robotistas situados, a saber, comportamientos asociados a lo que puede ser llamado “razón avanzada” tales como la coordinación de actividad y la elección de alguna acción que supone un estado de cosas distante, imaginario o contrafáctico o la coordinación de actividad y elección de alguna acción con parámetros ambientales cuyas manifestaciones son complejas y no reglamentadas. Ésta es la clase de problemas para los que el enfoque representacionalista está mejor dotado. Este enfoque opera eficazmente cuando se trata de explicar la habilidad de generar una acción y elección apropiadas, pese a la desconexión física. Y esto requiere el uso de algún ítem o proceso cuyo rol es estar en el lugar del estado de cosas ambiental faltante y, luego, apoyar el pensamiento y la acción en ausencia de un input ambiental efectivo (cf. Clark 2001a: 110). Los casos estudiados por la robótica situada, como hemos visto, tienen poco que decir acerca de los casos donde se da una desconexión física efectiva. De ahí que la demostración real de casos que no ocupan representaciones o que ocupan muy pocas no sea un argumento directo a favor del anti-representacionalismo.

Clark (1997a: 22 y ss.) arguye también que el rechazo implícito (o muchas veces explícito, como en el caso de Brooks) de la noción de representación por parte de varios de los contendores de la robótica situada es un error por dos motivos. Primero, no hay ninguna duda de que el cerebro a veces integra múltiples fuentes de información (por ejemplo, a menudo combinamos modalidades, usando el tacto, la vista y el sonido en complejos *loops* interdependientes, donde la información recibida en cada modalidad ayuda a sintonizar y a esclarecer el resto), lo que indicaría que un funcionamiento por capas totalmente independientes no se corresponde con la cognición efectiva. Segundo, la presencia de modelos internos que intervienen entre input y output no siempre constituyen un cuello de botella costoso en cuanto tiempo. Así sucede, por ejemplo, con las señales de propiocepción. La propiocepción es el sentido interno que da cuenta de cómo el propio cuerpo está localizado en el espacio y que ayuda, por ejemplo, a alcanzar un objeto con la mano. Las señales de propiocepción deben viajar de vuelta desde las periferias del cuerpo hacia el cerebro y esto toma tiempo, demasiado tiempo en efecto como para generar movimientos finos de alcance. Se hace necesario entonces una emulación motora, una suerte de conjunto de circuitos que replican ciertos aspectos de las dinámicas temporales del sistema mayor, que toma como input una copia de un comando motor y produce como output una señal idéntica en forma a otra que vuelve de las periferias sensoriales. Lo que hace entonces es predecir cuál debiera ser la retroalimentación propioceptiva. Tal emulador no es un cuello de botella temporal, sino que, por el contrario, facilita el éxito en tiempo real al proveer una “retroalimentación virtual” que se adelanta a la retroalimentación de las periferias sensoriales reales. El rol adaptativo de un emulador depende tanto de su velocidad de operación (su habilidad de adelantarse a la retroalimentación sensorial) como de la información que codifica, y muestra cómo podría empezar a pensarse el surgimiento de representaciones a

partir de la necesidad de emular respuestas senso-motoras que resultan muy lentas para ser adaptativamente viables.

De este modo, modelos internos como el de propiocepción pueden dar una pista acerca de cómo acciones basadas en el comportamiento pueden dar origen a acciones cada vez más alejadas del pensamiento: una suerte de arquitectura de incrementación, donde el incremento no está constituido por nuevos “comportamientos”, sino por nuevos niveles de abstracción respecto del mundo inmediato. Esto se logra tratando el razonamiento off-line como algo similar a un sentir y actuar “simulado”, de manera que se preserve el sabor de la resolución de problemas por parte de una visión “encarnada”, pero se añade un grado más alto de habilidad para alejarse del ambiente. Otro ejemplo de cómo puede implementarse y pensarse una arquitectura tal es el robot METATOTO, construido a partir del robot TOTO. TOTO utiliza sensores para detectar paredes, corredores y otros obstáculos y es capaz de usar un mapa interno de su ambiente, que luego puede usar para visitar nuevamente lugares ya encontrados. El mapa interno de TOTO es bastante singular en el sentido de que guarda información geográfica de una manera “orientada a la acción”, combinando la información acerca del movimiento del robot y su input perceptual correlativo. En este robot es una misma tarea computacional el usar el mapa para encontrar una ruta y el generar un plan para los movimientos reales del robot. Ahora bien, TOTO no es capaz de “pensar” sobre ningún lugar que no haya visitado previamente. METATOTO, en cambio, se construyó sobre la arquitectura original de TOTO –esto es, es capaz de realizar todas las funciones básicas de TOTO– para crear un sistema capaz de encontrar su propio camino, incluso en lugares que nunca antes había visitado. Esto lo hace usando la arquitectura de TOTO off-line, de manera que apoya la exploración en la “imaginación” de un ambiente totalmente virtual. Cuando METATOTO está “imaginando” emplea exactamente la misma maquinaria que normalmente apoya las interacciones físicas con el mundo real. La diferencia estriba en el nivel menor de la interfase: mientras TOTO usa un sonar para actuar y navegar en el mundo real, METATOTO usa un sonar simulado para explorar un mundo virtual (que incluye un cuerpo de robot virtual). METATOTO puede entonces beneficiarse de “experiencias virtuales”, tal como TOTO se beneficia de la experiencia real (cf. Clark 1999b: 347-348).

Ejemplos como el de METATOTO permiten entonces considerar la posibilidad de una arquitectura de incrementación que se constituye de manera que las acciones senso-motoras se vayan aislando cada vez más de su contexto concreto, permitiendo así el surgimiento de representaciones de un orden muy bajo y muy cercanas a las respuestas senso-motoras, que conformarían la base de representaciones de un orden más elevado, es decir, más alejadas del funcionamiento senso-motor, que nos permiten actuar en situaciones contrafácticas. Esto deja pensar que la noción de representación debe replantearse en términos de su origen en la actividad senso-motora, de manera que la experiencia senso-motora determine en alguna medida no sólo el sustrato sobre el cual pueden surgir nuevos “circuitos”, cuya funcionalidad se aleja progresivamente de la interacción directa con el ambiente, sino que también lo que es importante representar por parte del agente (en Herbert, por ejemplo, latas). La noción de representación sería así una que responde por el actuar efectivo del agente, siendo entonces “orientada a la acción”.

Ahora bien, una duda penetrante al respecto es si efectivamente tal movimiento permite explicar o dar cuenta de una inteligencia a nivel humano tal como la conocemos. Tal vez la forma adecuada de considerar la crítica a la representación por parte del enfoque de la robótica situada sea que, efectivamente, la representación no cumple ningún rol explicativo en las actividades adaptativas más básicas de los seres vivos (como el caminar erguido o tomar algo con la mano), pero provee de “circuitos” básicos a partir de los cuales podría emerger eventualmente la noción de representación, tal como en METATOTO. En esta replicación de circuitos básicos, donde se da una funcionalidad distinta al mero operar directamente con el mundo, se basará posteriormente nuestra visión de la noción de representación.

Por otra parte, hay que tener en cuenta que los escenarios privilegiado por los roboticistas no son escenarios “naturales” tal como los conocemos los humanos: hay en este enfoque un privilegio del ambiente “natural ecológico” por sobre el ambiente social. Y es éste ambiente social el que podría, al menos en parte, dar cuenta de cómo la noción de representación adquiere, en combinación con los “circuitos” básicos, su existencia y, por ende, su importancia explicativa (cf. Clark 2001a, 1997a).

CAPÍTULO TERCERO

SISTEMAS DINÁMICOS

No sólo la robótica situada se preocupa por la dimensión temporal de los procesos cognitivos, sino también, aunque en un sentido algo diferente, los enfoques dinámicos de la cognición. Éstos consideran los sistemas cognitivos como sistemas físicos reales que actúan en el mundo real en el tiempo real, centrándose en cómo estudiar tales sistemas interactivos. El enfoque dinámico se centra en la evolución de un sistema en el tiempo y está particularmente bien dotado para tratar con casos en los que un sistema o componente *a* está constantemente afectando y siendo afectado por otro sistema o componente *b* (que puede similarmente ser continuamente sensible al ítem *c*, etc.) (Clark 1999b).

Tal como su nombre lo señala, dichos enfoques tienen como marco general la teoría matemática de sistemas dinámicos y pueden encontrarse en una cantidad creciente de trabajos en psicofísica, percepción, control motor, psicología del desarrollo, psicología cognitiva, robótica situada e investigación de agentes autónomos, inteligencia artificial y psicología social. Ejemplos típicos del uso de sistemas dinámicos para dar cuenta de fenómenos cognitivos son las investigaciones sobre el movimiento de las piernas en niños por parte de Thelen y Smith (Clark 1998, cf. Clark 1997a, 1997b, 2001a; van Gelder 1997; Port 2000), el movimiento de los índices al mismo tiempo de Kelso (van Gelder y Port 1995; Clark 1997b), el problema de la permanencia de objetos planteado en primera instancia por Piaget (Beer 2000: 93-94), la teoría de decisión de campo de Townsend y Busemeyer (van Gelder 1997: 2-4, Port 2000).

Este enfoque, al centrarse en sistemas interactivos, destaca la interrelación entre los diferentes componentes (variables) que conforman tales sistemas y cómo el comportamiento del sistema se desarrolla en el tiempo. El actuar de un agente cognitivo se entiende entonces como un actuar que depende de las otras variables en su ambiente: el cuerpo y el mundo. En estos sistemas interrelacionados la clave para la explicación del agente cognitivo se basa en la interacción entre las diferentes variables. Dicha interacción hace difícil entender qué función cumple la representación, pues todo lo que parece haber es un acoplamiento entre sistema nervioso, cuerpo y mundo. No es tampoco clara la noción de un sistema cognitivo como una unidad explicativa, ya que la imbricación entre los diferentes componentes implica que cualquiera de ellos debe entenderse en su interacción con los otros. Esto supone que no hay una noción clara de que un sistema represente al otro.

I. Ejemplos de Sistemas Dinámicos

Antes de revisar con algún detalle en qué consiste la teoría de sistemas dinámicos y su adopción en ciencia cognitiva, se introducirán dos ejemplos típicamente citados en la literatura. El primero, el estudio de

Thelen y Smith respecto de cómo los niños aprenden a mover sus piernas, permite dar cuenta de qué se habla cuando se adopta una perspectiva “dinámica” para explicar fenómenos cognitivos. El segundo, el del regulador de Watt, no es un ejemplo propio de la ciencia cognitiva, pero ayuda a entender mejor la divergencia entre una perspectiva “dinámica” y una representacional / computacional (o también una conexionista).

Thelen y Smith estudian el movimiento de las piernas en los niños. Un niño recién nacido, cuando está suspendido sobre el suelo, realiza un movimiento reconocible de andar. Después de unos pocos meses esta respuesta desaparece, volviendo a manifestarse entre los 8 y 10 meses, edad en que los niños comienzan a soportar su propio peso. Este patrón de movimiento –concluyen Thelen y Smith– no refleja las etapas predeterminadas de un programa guardado o la influencia de alguna variable o parámetro singular, sino que resulta de la interacción de varias fuerzas que están esparcidas entre cerebro, cuerpo y mundo. Tales fuerzas incluyen rasgos corporales (como la masa de las piernas), efectos mecánicos (extender los músculos y saltar), influencias ecológicas (el agua, un molino) y el impulso cognitivo (la voluntad de moverse). El resultado es una imagen del niño como un sistema donde sus componentes, es decir, el cerebro, el cuerpo y el mundo, conspiran para producir los perfiles de comportamiento que la ciencia cognitiva busca explicar y entender¹⁰ (Clark 1997a: 37). Esta imagen de un “engranaje sutil” en un sistema extendido cerebro/cuerpo/mundo lleva a describir el desarrollo como una creación y disolución sucesiva de atractores en un sistema cuyos componentes orgánicos y ambientales están cambiando en el tiempo.

El regulador de Watt es el ejemplo favorecido por van Gelder, uno de los proponentes del enfoque dinámico en ciencia cognitiva (ver más abajo), para explicar qué es un sistema dinámico (van Gelder 2000: 422 y ss.; cf. Clark 1997a: 95 y ss.; 2001a: 125 y ss.). El trabajo de este regulador consiste en mantener constante la velocidad de un volante que maneja una máquina industrial y que a su vez es manejado por una máquina a vapor¹¹. Dadas las variaciones en la presión del vapor y en la carga actual de trabajo, la velocidad de este volante tiende a fluctuar. Para mantenerla constante, la cantidad de vapor que entra a los pistones es controlada por una válvula reguladora, que se añade a la cañería que lleva vapor de la caldera al pistón. Más vapor resulta en más velocidad, menos vapor, en menos. En algún momento esta válvula debe calibrarse por un ingeniero, pero este proceso puede automatizarse. La solución clásica “computacional” a este problema, no la de Watt, consiste en descomponer el problema en varios sub-problemas, de manera que un mecanismo capaz de resolver el problema del regulador siga un algoritmo (una receta en un orden preciso), por ejemplo:

(1) Comienzo:

¹⁰ Esta investigación y otras, como la percepción y la reacción a pendientes, sugiere además que la combinación entre pensamiento y acción puede ser tan ubicua y fundamental que todo nuestro conocimiento se construye a través de las interacciones temporales de percibir y actuar, esto es, que los niños aprenden del mundo no sólo al realizar acciones, sino también que el conocimiento que adquieren es a menudo específico respecto de la acción.

¹¹ De hecho, el problema real requería que una máquina que hilaba algodón hilara de forma pareja, lo que podía conseguirse sólo si el volante mantenía constante su velocidad.

- (i) Medir la velocidad del volante;
 - (ii) Comparar la velocidad actual con la velocidad deseada.
- (2) Si no hay discrepancia, volver a (1); en caso contrario:
- (i) Medir la presión presente del vapor;
 - (ii) Calcular la alteración deseada en la presión a vapor;
 - (iii) Calcular el necesario ajuste de la válvula reguladora;
 - (iv) Hacer el ajuste de la válvula reguladora
- (3) Volver al paso 1.

Pero esta no fue la forma en que Watt resolvió el problema (tampoco pudo haberlo hecho así, ya que no había nada en el siglo XVIII que pudiera o bien realizar cálculos complejos o bien traducir condiciones físicas en argumentos simbólicos y luego nuevamente en un ajuste físico). Su solución consistió en poner un eje en el volante, de manera que girara a una velocidad que dependía directamente del volante mismo. Unidos al eje por un gozne había dos brazos y al final de cada uno, una bola de metal. Mientras el eje giraba, la fuerza centrífuga conducía las bolas hacia fuera y de ahí, hacia arriba. Mediante un arreglo inteligente, este movimiento de brazos se unía directamente a la válvula reguladora. El resultado: mientras la velocidad del volante aumentaba, los brazos se elevaban, cerrando la válvula y restringiendo el flujo de vapor; mientras la velocidad decrecía, los brazos caían, abriendo la válvula y permitiendo que entrara más vapor. El motor lograba así una velocidad constante, mantenida con una suavidad extraordinaria en presencia de grandes fluctuaciones de presión y carga (ver Fig. 1).

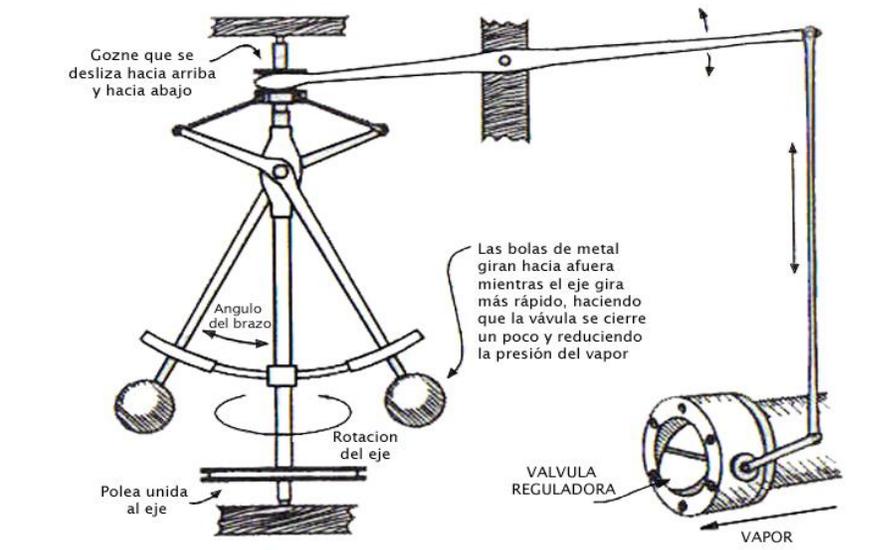


Figura 1

En el regulador de Watt no se presenta ningún rasgo asociado con la solución computacional (como la descrita más arriba), sino que se hace hincapié, en cambio, en las relaciones continuas y simultáneas de influencia causal que hay entre los diversos factores involucrados. Éstas, presente también en la investigación de Thelen y Smith, son lo que se considera como óptimo para un tratamiento en términos de la teoría de sistemas dinámicos. Las relaciones de influencias causales continuas y simultáneas entre diversos factores que describen apropiadamente la acción del regulador no pueden entenderse en un marco computacional / representacional. La noción de representar –de que algo está en lugar de otra cosa– es completamente inadecuada para capturar la interacción real entre el regulador centrífugo y la máquina, o entre el sistema nervioso, el cuerpo y el ambiente del niño. Aquí no parece haber representaciones que han de manipularse, ni manipulaciones que puedan entenderse como operaciones computacionales: no hay pasos discretos ni identificables en los cuales una representación puede convertirse en otra, como en la TRM. Antes bien, toda la operación del sistema es suave y continua y, dado que no hay ni secuencias ni ciclos, no sirve de nada dividir sus cambios en el tiempo en diferentes manipulaciones. También el modelo de input-procesamiento-output es inadecuado: hay un input (el eje es conducido por el motor) y un output (la conexión a la válvula reguladora), pero input, actividad interna y output se imbrican, produciéndose y causándose unos a otros.

Van Gelder y Port (1995) proponen que la manera de Watt de concebir la solución a la regulación de la velocidad del volante¹² es la manera en que también debiera entenderse la cognición, es decir, que la cognición es mejor entendida en términos dinámicos. Antes de revisar en detalle en qué consiste esta afirmación, es menester delinear qué es efectivamente la teoría de sistemas dinámicos.

II. La Teoría Matemática

La teoría de sistemas dinámicos es un marco bien establecido en la física, específicamente en la física de fluidos y también en la meteorología. Se constituyó en un principio para modelar y describir fenómenos que suponían cambio en el tiempo (y cambio en el lapso de tiempo, etc.). Por cierto, la definición más amplia de un sistema dinámico es la de cualquier sistema que cambia en el tiempo. Casi cualquier sistema en el mundo físico puede definirse entonces como un sistema dinámico. Pero tal definición es efectivamente útil sólo cuando los patrones de cambio en el tiempo exhiben cierta medida de complejidad, que es explicable por el aparato técnico de la teoría de sistemas dinámicos (Clark 2001a: 122).

Van Gelder y Port (1995) definen como sistemas dinámicos aquellos sistemas con estados numéricos que se desarrollan en el tiempo de acuerdo a una regla. Un sistema es un conjunto de variables, que pueden estar en diferentes estados en diferentes tiempos. Un sistema no es cualquier conjunto de objetos, sino

¹² Su rasgo clave es el ángulo en el que cuelgan los brazos, pues éste determina cuánto se abre o se cierra la válvula reguladora, y para entender cómo este ángulo fluctúa en el tiempo éste debe entenderse como una función de la velocidad del motor.

que de alguna forma sus componentes están imbricados. Esto significa que los elementos deben interactuar unos con otros, que la manera en que cualquiera de ellos cambia debe depender de la manera en que están los otros y también que, si cualquier elemento del mundo interactúa en el sentido recién descrito con cualquier cosa en el conjunto, entonces claramente es parte del mismo sistema. Así, un conjunto de elementos constituye un sistema si éstos son interactivos y auto-contenidos. El estado general de un sistema en un momento dado corresponde a la manera en que dichas variables están en aquel momento. El comportamiento del sistema es el cambio de su estado general en el tiempo. La totalidad de estados generales en los que el sistema puede estar es el conjunto de estado, denominado frecuentemente espacio de estado (*state space*). Así, el comportamiento del sistema puede pensarse como una secuencia de puntos en su espacio de estado.

Los sistemas dinámicos son sistemas de estado determinado (*state-determined systems*), lo que significa que en dichos sistemas el estado actual siempre determina un comportamiento futuro único. El hecho de que el estado actual determine el comportamiento futuro implica la existencia de alguna regla de evolución temporal que describe el comportamiento del sistema como una función de su estado actual. Y el hecho de que los comportamientos futuros estén determinados únicamente significa que las secuencias del espacio de estado nunca pueden bifurcarse. Los sistemas pueden ser afectados asimismo por factores externos, que no son a su vez afectados por las variables constitutivas del sistema, conocidos comúnmente como parámetros. Los parámetros son relativamente fijos e influyen sólo en la manera en que las variables interactúan, en cuanto inputs ellos son ocasionales y establecen los estados actuales de algunas variables directamente.

Van Gelder y Port (1995) subrayan que un sistema dinámico es aquel en el cual los cambios son una función de las fuerzas que operan dentro de él. Cuandoquiera que las fuerzas se aplican, nos encontramos con aceleraciones y desaceleraciones, esto es, hay un cambio en la tasa de velocidad (*rate*) según la cual los estados están cambiando en un momento dado. Las herramientas matemáticas estándares para describir las velocidades de cambio son las ecuaciones diferenciales, que pueden pensarse como especificando la manera en que un sistema está cambiando en cualquier momento como una función de su estado en dicho momento. De este modo, todos los sistemas dinámicos cambian en el tiempo. El tiempo se considera sólo como un conjunto intrínsecamente ordenado, que ayuda a proveer un orden sobre las otras cosas. El tiempo real de los sistemas concretos es el conjunto de instantes en el que las cosas pueden pasar efectivamente, ordenadas por una prioridad temporal (antes / después). Los eventos concretos son pareados con instantes o períodos de tiempo y por ello están en una relación temporal uno con otro (van Gelder 1997).

Dado que las variables de los sistemas dinámicos son cantidades, se puede hablar de distancias entre los estados totales. De ahí que el conjunto de estados de un sistema dinámico sea, en un sentido interesante, un espacio, dentro del cual cualquier estado es una posición y cualquier comportamiento, una trayectoria. Tal vez el rasgo más distintivo de la teoría de sistemas dinámicos es que provee una forma geométrica de comprensión: los comportamientos se piensan en términos de locaciones, caminos y paisajes en el espacio de estado del sistema (van Gelder 2000: 434-35; van Gelder y Port 2000).

El objetivo de un análisis de sistemas dinámicos es presentar una imagen de un espacio de estado cuya dimensionalidad (las n variables que se desarrollan simultánea y continuamente en el tiempo real) es de un tamaño arbitrario y generar una comprensión de los comportamientos del sistema en términos de la ubicación y el movimiento dentro de ese espacio geométrico abstracto. Para ayudar a tal comprensión se invoca una variedad de otros constructos, que capturan las propiedades distintivas de ciertos puntos o regiones (conjuntos de puntos) en el espacio en tanto están determinados por las matemáticas que los rigen. Éstas especifican típicamente una ley dinámica que determina cómo los valores de un conjunto de variables de estado evolucionan a través del tiempo (tal ley puede consistir en un conjunto de ecuaciones diferenciales¹³). Dado un estado inicial, la secuencia temporal de estados determinada por la ley dinámica constituye una trayectoria a través del espacio. El conjunto de todas las trayectorias que pasan a través de cada punto es denominado el *flujo*, y su forma es el típico objeto de estudio. Para entender la forma de este flujo se ocupan a su vez varios constructos, incluyendo aquel de un *atractor* (un punto o una región) en el espacio. Las leyes que gobiernan el movimiento a través del espacio garantizan que cualquier trayectoria que pase cerca de este atractor será “absorbido” dentro de él. Conceptos relacionados incluyen el de *cuenca de atracción* (*basin of attraction*), el área en el que el atractor ejerce su influencia, y el de *bifurcaciones*, casos donde un pequeño cambio en los valores del parámetro pueden reconfigurar el flujo, produciendo un nuevo “retrato de fase” (*phase portrait*) –esto es, una nueva descripción de la estructura general de cuencas y límites entre cuencas– (Clark 1997a: 100). Otra idea importante es la de *acoplamiento*. Dos variables están acopladas cuando la manera en que cada una cambia en un momento dado depende directamente de la manera en que la otra está en ese momento. En otras palabras, las variables acopladas se desarrollan interdependiente y simultáneamente. Los sistemas dinámicos genuinos exhiben un alto grado de acoplamiento; cada variable está en un cambio continuo, y todos los pares de variables están ya sea directa o indirectamente, determinando mutuamente las formas de los cambios de la otra (van Gelder 2000).

III. Los Fenómenos Cognitivos como Sistemas Dinámicos

De acuerdo a van Gelder y Port (1995) la teoría matemática de los sistemas dinámicos es la que debe utilizarse para explicar los fenómenos cognitivos, dado que los procesos cognitivos:

siempre se desarrollan en el tiempo real; que sus comportamientos están impregnados tanto por continuidades como por discreciones; que están compuestos de múltiples sub-sistemas que están interactuando y están activos simultáneamente; que sus clases distintivas de estructura y complejidad no están

¹³ Una ecuación diferencial es una ecuación que utiliza diferenciales. Una “diferencial” es una función de dos variables que es obtenida de una función dada, $y = f(x)$, y que expresa el incremento aproximado en la función dada como la derivativa de la función por el incremento en la variable independiente, escrita como $dy = f'(x) dx$.

presentes del primer momento, sino que emergen en el tiempo; que los procesos cognitivos operan en muchas escalas temporales y que los eventos a diferentes escalas temporales interactúan; y que ellos están anidados en un cuerpo y en un ambiente real.

Tales características harían del enfoque dinámico el enfoque ideal para comprender el fenómeno de la cognición.

Para los proponentes del enfoque dinámico, la forma tradicional de concebir la actividad en términos representacionales se debería a que varias clases de sistemas que realizan tareas sofisticadas –incluyendo la interacción con un ambiente cambiante–, diferentes en cuanto a su naturaleza y que requieren de diferentes herramientas conceptuales para su comprensión, parecen mostrar que el sistema tiene conocimiento de o razona acerca de su ambiente. Esta manera de concebir la cognición bien puede deberse a preconcepciones engañosamente atrayentes acerca de cómo deben funcionar los sistemas que resuelven tareas complejas. Tal visión tradicional de los fenómenos cognitivos puede estar oscureciendo el hecho de que los sistemas cognitivos pueden efectivamente ser *sistemas dinámicos* y la cognición, el comportamiento de un sistema dinámico no computacional (van Gelder 1999: 432). Esto último es lo que se ha denominado la “hipótesis dinámica en la ciencia cognitiva”, en términos de van Gelder:

Para cada clase de actuación cognitiva exhibida por un agente cognitivo natural, hay algún sistema cuantitativo instanciado por el agente en el nivel más alto relevante de la organización causal, de manera tal que las realizaciones de tal clase son comportamientos de tal sistema; además, la organización causal puede y debiera entenderse generando modelos dinámicos, usando los recursos teóricos de la dinámica, y adoptando una perspectiva dinámica amplia (van Gelder 1997: 13).

Esta hipótesis consta, en realidad, de dos componentes: la hipótesis de naturaleza (los agentes cognitivos son agentes dinámicos) y la hipótesis de conocimiento (los agentes cognitivos deben entenderse dinámicamente) (van Gelder 1997: 1).

La hipótesis de naturaleza nos dice qué es un agente cognitivo al especificar su relación con los sistemas dinámicos. Esta relación no es de identidad, sino de instanciación. Los agentes cognitivos no son ellos mismos sistemas (conjuntos de variables), sino más bien objetos cuyas propiedades pueden conformar sistemas. Los agentes cognitivos instancian varios sistemas en un momento dado. Los sistemas responsables por las actuaciones cognitivas son dinámicos. Esto supone que los agentes cognitivos no son un sistema dinámico en particular, sino que tantos sistemas como se necesita para producir todas las clases diferentes de actuaciones exhibidas por el agente (van Gelder 1997: 9). La hipótesis de conocimiento, que, como vimos, se refiere a entender a los agentes cognitivos como sistemas dinámicos, se refiere simplemente a que la ciencia cognitiva puede y debe tomar una forma dinámica, esto es, los fenómenos debieran

verse desde la óptica de la teoría de sistemas dinámicos y, de este modo, entenderse con las herramientas de dicha teoría¹⁴.

De este modo, un científico cognitivo que quiere explicar algún fenómeno haciendo uso de la teoría de sistemas dinámicos se enfoca en algún aspecto particular de la cognición y propone un sistema dinámico abstracto como un modelo del proceso escogido. El comportamiento del modelo se investiga usando la teoría de sistemas dinámicos, a menudo con la ayuda de una simulación en computadores digitales. Un encaje cercano entre el comportamiento del modelo y los datos empíricos sobre el fenómeno a modelar confirma la hipótesis de si acaso el fenómeno mismo es dinámico en cuanto a su naturaleza y de si acaso puede ser entendido en términos dinámicos (Clark 2001a). Para demostrar, entonces, que algún aspecto del mundo constituye un sistema dinámico se requiere escoger un conjunto relevante de cantidades, y maneras de medirlas, de manera que las trayectorias de su estado de espacio resultante se conforme a alguna regla especificable. La manera distintiva de comprender los comportamientos de un sistema en forma geométrica, lleva naturalmente a los dinamicistas a concebir los procesos cognitivos en términos geométricos. El carácter temporal distintivo de algunos procesos cognitivos, que se refiere a cómo se desarrollan en el tiempo, es un asunto de cómo los estados totales a través de los cuales pasa un sistema están ubicados espacialmente uno con respecto de otro y respecto al paisaje dinámico del sistema. Esto conlleva que lo relevante sea dónde está el estado, antes que de qué está hecho tal estado (van Gelder 1997). De este modo, si estamos interesados en sistemas cognitivos, entonces los comportamientos de interés son sus realizaciones cognitivas (percibir, recordar, conversar, etc.), y son estos comportamientos y sus escalas de tiempo características los que deben desarrollarse de una forma descrita por la regla de evolución. Esta manera de ver los fenómenos ha significado en la práctica que los dinamicistas al interior de la ciencia cognitiva se aboquen a explicar una serie de aseveraciones específicas, para ver en qué sentido ciertos aspectos particulares de la cognición son, en efecto, el comportamiento de sistemas dinámicos, más localizados. Debe notarse que esto es una consecuencia clara de la hipótesis dinámica de naturaleza, que señala que los agentes cognitivos son tantos sistemas como se necesiten para producir todas las diferentes clases de actuaciones exhibidas por el agente.

El enfoque dinámico de la cognición se relaciona cercanamente con ideas acerca de la encarnación de la mente y de la “situacionalidad” ambiental de la cognición humana, dado que subraya lo que hay de común entre, por una parte, el comportamiento en procesos neurales y cognitivos y, por otra, entre eventos fisiológicos y ambientales; considerando siempre que la dimensión del tiempo es común a todos estos ámbitos

¹⁴ Van Gelder (1997: 11) distingue aquí entre el uso de modelamiento dinámico y de la teoría de sistemas dinámicos. El modelamiento dinámico es una rama de las matemáticas aplicadas que concierne a la comprensión de fenómenos naturales mediante la provisión de modelos dinámicos abstractos. La teoría de sistemas dinámicos, en cambio, es una rama de las matemáticas puras y su dominio se extiende a cualquier clase de cambio describable, si bien centra su atención particularmente en sistemas para los cuales no hay una manera conocida de especificar los comportamientos como una función del tiempo, conceptualizando dichos modelos geoméricamente e intentando comprender las propiedades estructurales del flujo, es decir, del rango completo de pasajes posibles.

y que ésta permite a su vez un acoplamiento en tiempo real entre ellos, donde la dinámica de un sistema influencia al otro (Port 2000).

En este sentido, Beer (2000: 97) indica que, desde una perspectiva dinámica, el objetivo principal de un agente situado es tomar una acción apropiada con respecto a sus circunstancias y metas, y la cognición es meramente un recurso entre muchos al servicio de este objetivo. Otros recursos importantes incluyen las propiedades físicas del cuerpo del agente, la estructura de su ambiente inmediato (incluyendo artefactos como listas de compra, calendarios, computadores, etc.) y su contexto social. De ahí que la cognición pueda extenderse más allá del cerebro del agente para estar distribuida en un sistema de gente y objetos dentro de un ambiente.

La concepción de la cognición como estando extendida más allá del cerebro, incorporando efectivamente rasgos corporales y ambientales, parece exigir un lenguaje en el que tal interacción sea comprensible. El lenguaje de los sistemas dinámicos parece en efecto proveer un vocabulario conceptual que permite la unificación de procesos cognitivos en el cerebro con procesos fisiológicos en nuestra periferia corporal y con eventos ambientales externos al organismo. La unificación de procesos a través de estos límites borrosos y en parte artificiales hace posible una comprensión verdaderamente encarnada y situada del comportamiento humano de todos los tipos. De ahí que el enfoque dinámico, al suministrar un lenguaje común para la cognición y los procesos neurofisiológicos que la apoyan, para el comportamiento no humano y para el comportamiento adaptativo de animales más simples, prometa proveer un marco teórico unificado para la ciencia cognitiva, como también una comprensión de la emergencia de la cognición en el desarrollo y la evolución (Beer 2000).

IV. Sistemas Dinámicos vs. el Modelo Computacional / Representacional

La diferencia de esta manera de hacer ciencia cognitiva con el enfoque representacional computacional es clara. Según los dinamicistas, el sistema cognitivo no es un computador, sino un sistema dinámico. No es el cerebro, interno y encapsulado, sino todo el sistema que se compone de variables correspondientes al sistema nervioso, cuerpo y ambiente. El sistema cognitivo no se entiende como un manipulador secuencial discreto de estructuras representacionales estáticas, sino que estas estructuras son vistas como en medio del cambio y teniendo por sí solas poco interés. Sus procesos no tienen lugar en el tiempo arbitrario, discreto de los pasos computacionales, antes bien, ellos se desarrollan en el tiempo real de un cambio en curso del sistema conformado por las variables ambiente, cuerpo y sistema nervioso. Así, el sistema cognitivo no interactúa con otros aspectos del mundo al pasar mensajes u órdenes, sino que evoluciona conjuntamente con él.

La cognición se concibe aquí como una operación que depende de un sistema que se compone a su vez de sistema nervioso, cuerpo y ambiente. Esto se corresponde con el énfasis de la orientación dinámica en el estado total. Los dinamicistas asumen que todos los aspectos del sistema –en este caso, sistema nervioso, cuerpo y ambiente– están cambiando simultáneamente y de esta manera piensan sobre el comportamiento del sistema refiriéndose a cómo el estado total del sistema está cambiando de un momento a otro. Esto los

distingue de los computacionalistas, que ven el cambio como un asunto local, un asunto de reemplazar un símbolo por otro.

Esta mutua influencia entre sistema nervioso, cuerpo y ambiente permite concebir la cognición como una operación que emerge de la interacción dinámica no lineal entre el cerebro o el sistema nervioso, el cuerpo y el ambiente. Para los dinamicistas, la interacción entre estas tres variables puede deberse a parámetros que influyen la forma del cambio, de manera que el input se concibe como una influencia en curso en la dirección del cambio, y el output como una influencia en algo más. Muchas veces la interacción es un asunto de acoplamiento: dos sistemas que simultáneamente conforman los cambios del otro (van Gelder 1997: 13).

Otras razones específicas que apoyan la adopción del enfoque dinámico por sobre una hipótesis computacional / representacional de la mente son que la cognición sucede en el tiempo, la continuidad en el estado, las interacciones simultáneas múltiples, la auto-organización y la emergencia de estructuras.

Que la cognición suceda en el tiempo permite a los modelos dinámicos explicar las funciones cognitivas que dependen del tiempo real (e.g., el procesamiento temporal que da lugar a patrones). Port (2000) señala así que los datos temporales tales como el tiempo de procesamiento (tiempo de reacción, de reconocimiento, etc.), la estructura temporal en los comportamientos motores (patrones motores de alcanzar, producción del lenguaje, locomoción, etc.) y la estructura temporal en la simulación (e.g., para hablar y para la percepción musical, la coordinación interpersonal en jugar un partido de tenis, etc.) pueden, mediante la teoría de sistemas dinámicos, incorporarse conjuntamente en modelos cognitivos si los eventos pueden predecirse en el tiempo. La temporalidad de cualquier operación particular debe respetar la velocidad a la cual se están dando otros procesos cognitivos, corporales y ambientales. En esta incorporación del tiempo, la cognición ya no es vista como si tuviera una estructura secuencial cíclica (sentir-pensar-hacer), sino más bien como un asunto de co-evolución continua. El objetivo en este enfoque es describir y explicar el curso temporal de este comportamiento. El inicio y el final del proceso cognitivo son usualmente sólo de interés secundario, a diferencia de la explicación tipo input-output de una TRM (van Gelder y Port 2000). Esto puede entenderse también dado que el foco de los dinamicistas se centra en cómo suceden en el tiempo los comportamientos, mientras que el de los computacionalistas se centra en qué es el comportamiento, sin considerar los detalles temporales. La sutileza y complejidad de la cognición se encuentra no en estructuras estáticas elaboradas, sino en el tiempo en el flujo mismo del cambio (van Gelder 1999).

La continuidad de estado se refiere a que los sistemas cognitivos naturales cambian de estado a veces de manera continua, a veces de maneras que pueden parecer discretas. El análisis de sistemas dinámicos provee un marco dentro del cual la continuidad y la discreción pueden explicarse –siendo la primera una precondition para la explicación y la aparición de la segunda–, mientras el enfoque computacional sólo puede modelar un sistema como cambiando de un estado discreto a otro. De este modo, el enfoque dinámico es inherentemente más flexible, y más poderoso, que el computacional (van Gelder y Port 1995: 23).

El enfoque dinámico busca entender la cognición humana en tanto una actividad interactiva, múltiple y simultánea. Los sistemas dinámicos consisten justamente en esta actividad simultánea, que se influencia

mutuamente por múltiples partes o aspectos. Esto puede relacionarse también con la visión dinamicista de que los sistemas operan en paralelo, esto es, todos los aspectos cambian inter-dependientemente al mismo tiempo, al contrario de la visión computacionalista que piensa en los sistemas como seriales, esto es, como que la mayoría de las variables permanecen estables en cualquier estado de transición dado.

La multiplicidad de escalas temporales da cuenta de que los procesos cognitivos tienen lugar a escalas temporales diversas: los cambios en el estado de las neuronas pueden tomar sólo milisegundos, el reconocimiento visual o auditivo medio segundo o menos, movimientos coordinados unos pocos segundos, la conversación minutos u horas, la emergencia de capacidades sofisticadas meses u años. Estas escalas temporales están interrelacionadas, los procesos en una escala temporal afectan los procesos en otra. El enfoque dinámico provee maneras de tratar esta variedad e interdependencia de escalas temporales, pensando, por ejemplo, en un sistema de dinámica rápida para las variables de estado en una escala de tiempo corta, una lenta para parámetros –siempre que no se piensen como fijos, sino como cambiando en una escala de tiempo considerablemente más larga que las variables de estado– a escala de tiempo largo, de manera que está dinámica lenta ayude a conformar la rápida (e incluso al revés).

La auto-organización y la emergencia de estructuras se refiere a que los sistemas cognitivos están altamente estructurados tanto en su comportamiento como en la organización interna espacial y temporal. La ciencia cognitiva debe describir dicha estructura y explicar cómo llegó ahí. Una ventaja significativa del enfoque dinámico es que los sistemas dinámicos son capaces de crear estructuras tanto en el espacio como en el tiempo (por ejemplo, una silla o una ola). Dado que la estructura puede llegar a existir sin ningún plan específico o sin un constructor independiente, existe la posibilidad de que muchas estructuras en los cuerpos físicos, como también en la cognición, puedan darse sin fuerzas exteriores impuestas externamente, tal vez ellas simplemente se organizan a sí mismas conformando una estructura, lo que se conoce como *emergencia*¹⁵. Van Gelder y Port (1995) afirman que la cognición es una estructura particular en el espacio y el tiempo, a saber, una que apoya la interacción inteligente con el mundo. El trabajo de la ciencia cognitiva entonces consiste en descubrir cómo tal estructura pudo resultar en un patrón estable del cerebro en el contexto del cuerpo y del ambiente.

V. Sistemas Dinámicos y Representación

Los enfoques dinámicos enfatizan, como hemos visto, la manera en que el comportamiento de un agente surge a partir de la interacción que sucede entre su cerebro, su cuerpo y su ambiente. En este enfoque, el foco cambia de representar un ambiente con precisión a comprometer continuamente aquel ambiente con un agente de manera de estabilizar patrones de comportamiento que son adaptativos para el agente.

En cuanto a la representación, no hay nada que impida que los sistemas dinámicos puedan verse como incorporando alguna forma de representación. Los modelos dinámicos a menudo incorporan

representaciones, pero las conciben como entidades dinámicas (e.g., estados de sistemas o trayectorias formadas por el paisaje de un atractor). Las representaciones tienden a verse como transitorias, estabilidades dependientes del contexto, antes que unidades permanentes estáticas, libres de contexto. Así, van Gelder (1995) afirma que un rasgo estimulante del enfoque dinámico consiste en ofrecer una oportunidad de concebir nuevamente y de una manera radicalmente diferente la naturaleza de la representación en los sistemas cognitivos, incluso dentro de un marco ampliamente no computacional.

Por otra parte, algunos dinamicistas afirman haber creado modelos absolutamente carentes de representación, y hacen conjeturas acerca de que la representación resultará jugar un rol mucho menor en la cognición con respecto al que se le supone actualmente (cf. van Gelder 1997: 13).

En un sentido similar, Beer (2003) plantea la necesaria reconceptualización de la representación interna debido a que el análisis dinámico de agentes desarrollados sugiere un conjunto más amplio de posibilidades. El único requerimiento sobre el sistema nervioso de un agente es que éste esté dotado con un potencial oculto para comprometerse en los patrones de comportamiento necesarios para la supervivencia y la reproducción necesarias para el agente. Sólo cuando tal sistema nervioso está dentro del cuerpo de un agente particular es que su potencial se realiza efectivamente a través de la interacción dinámica resultante. La creencia de que el sistema nervioso del agente debe representar explícitamente cualesquiera regularidades que el comportamiento del sistema completo cerebro / cuerpo / ambiente exhiba, ésta es una hipótesis teórica adicional, sujeta a análisis crítico y a un posible rechazo. Esto supone que la noción de representación ha de ser redefinida en términos que consideren la interacción entre sistema nervioso, cuerpo y ambiente o ha de rechazarse como constructo explicativo en ciencia cognitiva.

Este escepticismo acerca de la representación surge dado que lo que se estudia es la inteligencia encarnada real, que es vista, como señalamos, como un medio de involucrarse con el mundo –de usar estrategias activas que dejan mucho de la información afuera en el mundo y usan astutamente secuencias repetidas y en tiempo real de interacciones cuerpo-mundo para resolver problemas de una manera robusta y flexible–. En este “compromiso” con el mundo se descubre la imagen de dos (o más) sistemas acoplados complejos (el agente y el ambiente), cuya actividad conjunta resuelve el problema, de manera que no hay un sentido claro de qué pueda significar que un sistema *representa* el otro (Clark 1997a: 98). En este sentido, la explicación dinámica implica que, o bien hay que suscribir una noción anti-representacionista de la cognición –en el sistema acoplado no hay nada que esté en el lugar de o sea acerca de otra cosa–, o bien redefinirla en términos de que de alguna manera la interacción entre los diversos sistemas permita entender cómo una de las variables llega a tener una representación de otra. Aquí puede ser decisiva la noción de “cuerpo”, pues puede pensarse en éste como el límite que determina un adentro y afuera, que a su vez permite entender que un sistema “representa” al otro para la realización de ciertas actividades cognitivas.

¹⁵ Para una discusión sobre el concepto de “emergencia” ver Clark 1997a: 103 y ss; 2001: 112 y ss.

VI. Discusión

Una buena explicación dinámica está a medio camino de lo que tradicionalmente se llamaría una mera descripción de patrones de eventos y una explicación real de por qué los eventos se desarrollan como lo hacen. No es una mera descripción porque los parámetros necesitan ser escogidos muy cuidadosamente de manera que el modelo resultante tenga fuerza predictiva: nos dice lo suficiente para saber cómo se comportará en varias circunstancias no presentes. Sin embargo, difiere de las explicaciones cognitivas más tradicionales en tanto varía en una medida considerable con respecto al comportamiento de los componentes sistémicos individuales (Port 2000). Ahora bien, no es claro qué tipo de comprensión provee dicho análisis, pues amenaza con constituir una descripción abstracta antes que una explicación total. Aprendemos qué hace un sistema y cuándo lo hace y qué patrones de evolución temporal exhibe su comportamiento; pero esta comprensión, aunque valiosa, no parece ser exhaustiva. Clark (Ibid.) afirma que, al centrarnos en un sistema global, siempre nos quedamos con una comprensión empobrecida del rol adaptativo de los componentes y de la organización funcional interna del sistema. En otras palabras, la descripción dinámica no es plenamente explicativa, sino más bien descriptiva y no queda claro cuál es la función del agente cognitivo al interior del sistema que conforma. Es más, dada la noción de acoplamiento, no se distingue cuál sea la unidad cognitiva particular a la que hay que atender. Si la cognición efectivamente está dispersa en el acoplamiento entre sistema nervioso, cuerpo y mundo, el organismo que realiza efectivamente una operación cognitiva no se distingue.

Con respecto a la representación, Clark (1997a: 101) plantea que la posición radical de los sistemas dinámicos respecto del rol explicativo de ésta enfrenta, además del problema respecto del tipo de explicación que provee, un desafío importante en cuanto a la escala y a la “tratabilidad”. Sistemas dinámicos simples, por ejemplo, un controlador de un dispositivo de locomoción de un animat –los robots de Beer– compuesto de treinta neuronas, son de tal complejidad que nuestra comprensión geométrica intuitiva no puede aprehenderla. Las matemáticas detalladas de la teoría de sistemas dinámicos se hace continuamente menos tratable a medida que el número de parámetros y el tamaño del espacio de estado aumenta. De ahí que la aplicabilidad real de la teoría de sistemas dinámicos a sistemas acoplados altamente complejos, altamente dimensionales (como el cerebro humano), se ponga seriamente en duda.

Este aspecto es también abordado por Eliasmith (1997: 314) en el sentido de que los dinamicistas han sido incapaces de identificar o de cuantificar suficientemente las variables de un modelo dinámico. Esto, pues es extremadamente difícil, examinar un sistema cognitivo complejo y seleccionar qué comportamientos son apareados apropiadamente con variables para ser usados en un modelo dinámico. A este respecto, Beer (2003) señala que la visualización del estado de espacio completo es rara vez necesaria o deseable y que ninguna de las técnicas locales de los sistemas dinámicos requieren de visualización para su aplicación, pues

pueden encontrarse conjuntos límites¹⁶ y estudiarse sus estabilidades y bifurcaciones locales en muchos miles de dimensiones.

Clark (2001a: 134 y ss.) señala además que el problema más profundo con respecto a la alternativa dinámica se refiere a su tratamiento del cerebro como un factor más en la compleja red general de influencias causales. En cierto sentido es claro que factores internos y externos conspiran para apoyar muchas clases de éxito adaptativo. Pero en otro, es o falso o nuestra visión de mundo habría de cambiar de una manera muy radical, dado que suponemos que es la impresionante complejidad y variabilidad estructural del cerebro la que es la clave para comprender la ruta basada en la inteligencia del éxito evolutivo. Esto puede formularse en términos de si se le hace justicia a la idea de una distinción fundada entre sistemas basados en conocimiento y sistemas meramente físicos tratando a ambos con un enfoque dinámico. Esto es, si la acción humana puede entenderse en los mismos términos que el regulador de Watt. Hay, en efecto, una duda acerca de la propiedad de un vocabulario como el de la teoría de sistemas dinámicos al ser aplicado a fenómenos cognitivos que van más allá de aquellos de “bajo nivel”, es decir, a las ideas de acción y de elección guiada por información. Tales ideas, y con ellas, la idea de representación, parecen amenazadas en el marco de una imagen del cerebro, el cuerpo y el mundo como sistemas singulares, densamente acoplados.

En un sentido positivo, se puede descubrir en la postura más radical respecto de la representación en las teorías de sistemas dinámicos la posibilidad efectiva de prescindir de dicha noción en fenómenos adaptativos básicos, tal como el mover las piernas o el alcanzar un objeto. Esto no prueba, sin embargo, que tal noción no juegue un rol en la explicación cognitiva de hechos como, por ejemplo, el realizar una larga división, en los que los humanos parecen efectivamente confiar en operaciones serialmente ordenadas sobre símbolos digitales (cf. Port 2000).

Además de esto, en la teoría de sistemas dinámicos no queda claro cómo puede surgir un comportamiento coherente por parte de un agente en un nivel un poco alejado de la realidad senso-motora, es decir, de la interacción básica entre sistema nervioso, cuerpo y ambiente. Una teoría como la de sistemas dinámicos, que se centra en explicaciones de fenómenos locales, parece inadecuada para entender cómo, a partir del acoplamiento sistema nervioso, cuerpo y mundo, se diferencien objetos a representar. Una solución puede encontrarse considerando que las representaciones mentales podrían emerger precisamente cuando parte del acoplarse a un medio se refiere no ya a un medio donde la tarea básica es luchar, huir o comer, sino donde el propio medio, en este caso, uno social, determina que sea “adaptativo” poder razonar off-line. La teoría de sistemas dinámicos en ciencia cognitiva podría estar indicando que el camino para entender el surgimiento de cosas tales como representaciones debe buscarse (y puede explicarse) a partir de mecanismos

¹⁶ Cuando se aborda el comportamiento a largo plazo de un sistema dinámico, sucede que el estado de muchos sistemas dinámicos eventualmente termina en un pequeño subconjunto de estado de espacio llamado *conjunto límite*. Un conjunto límite es invariante respecto de la dinámica, de manera que si el sistema cae alguna vez en un conjunto límite, la dinámica actuará para mantenerlo allí indefinidamente. Dos tipos simples de conjuntos límites son los *puntos de equilibrio* y los *ciclos límites*. Un punto de equilibrio es un conjunto límite que consiste en un solo punto, produciendo un comportamiento constante, y un ciclo límite es una trayectoria que se encierra en sí misma, produciendo un comportamiento rítmico sin fin (Beer 2000: 93).

más básicos y enraizados corporalmente. Una analogía interesante en este sentido puede encontrarse en las hipótesis sobre el crecimiento evolutivo del cerebro humano y el uso de herramientas: al parecer, el bipedalismo con su consecuente liberación de las manos para el uso de herramientas fue el que permitió el posterior crecimiento del cerebro humano, y no al revés como se pensaba tradicionalmente. Así, puede pensarse la interacción sistema nervioso, cuerpo y mundo precisamente como lo que proveyó el substrato del cual emergieron estructuras básicas (digamos, patrones de activación neural en el cerebro) que se correlacionaron con aspectos definidos (más o menos estables) del mundo. Estas a su vez permitieron pensar acerca de la relación entre cuerpo y mundo, pensando que la correlación era anterior a la existencia misma de la interacción entre estas variables. Así, entonces, se puede postular que la representación surgió posteriormente a la interacción primera entre sistema nervioso, cuerpo y mundo, haciendo uso precisamente de dicha interacción como el substrato sobre el cual habría de originarse.

CAPÍTULO CUARTO

BIOLOGÍA Y COGNICIÓN

Para terminar esta revisión de las teorías "encarnadas" de la cognición y el lugar de la representación en éstas, se hará referencia en este capítulo a visiones que consideran cómo la cognición se entiende a partir del sustrato biológico de los seres vivos, en general, y del hombre, en particular. A diferencia del pasaje dedicado a la robótica situada, aquí no se considera un modelo biológico para producir un agente cognitivo, en el sentido de que la arquitectura de dicho agente recapitularía el recorrido de la evolución. Se estudia, en cambio, cómo el sustrato físico-biológico mismo de los seres vivos permite la elaboración de facultades cognitivas básicas y desarrolladas. Se revisarán dos posturas: la desarrollada por Varela et al. (1991), fundamentada en parte en su trabajo anterior con Maturana (1980; 1984), y la de Lakoff y Johnson (1999).

La postura de Varela et al. considera nuestras facultades cognitivas como facultades básicamente biológicas, destinadas a la mantención de la homeostasis organizacional de un ser vivo. Aquí, el punto de referencia para entender la percepción no es un mundo dado, sino la estructura senso-motora del perceptor, que además determina cómo el perceptor puede actuar y ser modulado por los eventos ambientales. Esto supone que la cognición se centra en tal estructura y en cómo varía de acuerdo a perturbaciones ambientales. La cognición se entiende como una historia de acoplamientos estructurales que elabora un mundo y opera a través de una red que consiste en múltiples niveles de sub-redes (el sistema nervioso). Ahora bien, desde el punto de vista del agente cognitivo entendido como un sistema nervioso encarnado, no hay sino correlaciones entre perturbaciones y cambios estructurales. Esto supone la prescindencia de la noción de representación para explicar el comportamiento cognitivo.

Lakoff y Johnson se centran también en la estructura senso-motora del perceptor, pero, a diferencia de Varela et al., descubren que tal estructura es la misma que permite, mediante la experiencia, que tengamos conceptos. En esta última alternativa, es posible entender cómo se construyen las representaciones a partir de un sustrato biológico básico. Esto posibilita una comprensión de cómo es posible dar cuenta de las representaciones y de su rol cognitivo a partir de un enfoque encarnado de la cognición.

I. La Mente "Encarnada"

Considérese, primero, el caso de un sapo. El sapo presenta la siguiente conducta alimenticia: el animal se orienta a la presa, tira su lengua larga y pegajosa y al retraerla con la presa adherida a ella, engulle con rapidez. Si al renacuajo se le corta el borde del ojo, respetando su nervio óptico, y se lo gira en 180 grados, cuando adulto el sapo es capaz de hacer un blanco perfecto si utiliza ambos ojos. Ahora bien, si se le tapa el ojo normal, el animal tira su lengua con una desviación exacta de 180 grados. Esto se debe a que, si la presa está abajo y al frente del animal, como su ojo mira un poco hacia el lado, éste gira y tira la lengua a lo que en efecto es atrás y arriba. El error se repite cada vez que se lleva a cabo el experimento: el animal lanza su lengua como si la zona de la retina donde se forma la imagen de la presa estuviera en su posición normal (en Maturana y Varela 1984: 83-84). Este experimento revela, según Maturana y Varela, que para el animal no existe el arriba o el abajo, el adelante o el atrás referidos al mundo exterior. Sólo hay una "correlación interna entre el lugar donde la retina recibe una perturbación determinada, y las contracciones musculares que mueven la lengua, la boca, el cuello y en último término todo el

cuerpo del sapo" (Ibid.). Esto sucede porque el animal, que tiene el ojo rotado, percibe una perturbación visual en la retina arriba y atrás cuando la presa está, efectivamente, abajo y adelante. No tiene, entonces, una representación sobre el mundo, como sería común explicarlo.

De acuerdo a estos autores, este experimento puede ser visto como evidencia de que la conducta del sapo es producto del operar del sistema nervioso, no el resultado de una representación acerca del mundo. Esta conducta surge según el modo cómo se establecen en el animal sus relaciones internas de actividad. La explicación de cómo opera tal animal tiene entonces bases biológicas. Maturana y Varela (1980; 1984) desarrollan una visión global de cómo han de entenderse dichas bases, y del lugar de la representación a partir de ellas.

Para entender esta visión global se ha de partir de la distinción entre *organización* y *estructura*. La "estructura" y la "organización" son dos niveles diferentes de abstracción. Las relaciones abstractas que definen una unidad como un miembro de una clase particular constituyen la organización del sistema. Cuando se distingue científicamente del universo total una unidad para estudio, se hace en virtud de su organización. Las propiedades materiales específicas de los componentes y los procesos que realizan la organización constituyen su estructura. Si un observador científico quiere medir o manipular una unidad, lo hace a través de su estructura. Un rasgo clave de esta distinción es que la misma organización puede ser instanciada por diferentes estructuras (Beer 2004: 6; cf. Maturana y Varela 1984: 25 y ss.).

Los seres vivos se caracterizan porque se producen continuamente a sí mismos, es decir, tienen una *organización autopoietica*. Un sistema autopoietico es una red de procesos que producen componentes y que, además, detenta la propiedad de que las interacciones entre los componentes generan la misma red

de procesos que los produjo y los constituye como una entidad distinta en el espacio en el que existe. El ejemplo paradigmático de autopoiesis es la célula, en la que los componentes son moléculas, las interacciones son reacciones químicas y la membrana de la célula sirve como un límite físico que sitúa espacialmente estas reacciones en una entidad (o "unidad") que se distingue de su ambiente. Se sigue de esto que una "máquina" autopoietica continuamente genera y especifica su propia organización a través de su operación en cuanto sistema de producción de sus propios componentes, y hace esto en una rotación sin fin de sus componentes bajo condiciones de perturbaciones continuas y compensación de tales perturbaciones. De ahí que una máquina autopoietica sea un sistema homeostático que tiene su propia organización, que define la red de relaciones, como la variable fundamental que mantiene constante. La vida misma es vista así como una organización específica de procesos físicos que tiene como su producto principal la mantención de su propia organización: una *homeostasis organizacional*¹⁷ (Beer 2004: 3; Maturana y Varela 1980: 79).

Esta unidad autopoietica interactúa con su ambiente. Este ambiente es el que determina las condiciones de perturbación de la unidad y sus cambios para mantener la homeostasis. Dicha interacción puede ser destructiva si la unidad es incapaz de compensar los cambios estructurales inducidos por una perturbación ambiental, desintegrándose. También puede ser no-destructiva si la organización de la unidad se preserva, aunque su estructura puede ser afectada. Pese a las perturbaciones que recibe, la unidad no es un títere del ambiente. Antes bien, una unidad determina activamente su propio dominio de interacción. En cualquier

¹⁷ Esta noción de vida se condice con aquella de la Vida Artificial (A-Life), según la cual el rasgo básico de las cosas vivientes es la auto-organización. La auto-organización supone la "emergencia" espontánea y la mantención de un origen que está ordenado en un grado menor. No es así un simple cambio superficial, sino un desarrollo estructural fundamental que resulta del carácter intrínseco del sistema (a menudo en interacción con el ambiente), antes que ser impuesto por alguna fuerza o diseñador externos.

momento, la estructura de la unidad especifica tanto el subconjunto de estados ambientales que pueden influenciarla como las interacciones con aquellos estados en los que puede comprometerse sin desintegrarse. Un ambiente sólo puede seleccionar de entre las interacciones disponibles; no puede en general colocar una unidad en un estado arbitrario deseado (Ibid.: 9; cf. Maturana y Varela 1980: xx y ss.; 77).

La marca de lo cognitivo consiste en la habilidad de establecer distinciones a través de una respuesta selectiva a las perturbaciones. El dominio de interacciones en las que una unidad puede comprometerse sin desintegrarse es su *dominio cognitivo*. Esta habilidad está fundada en la autopoiesis, por lo tanto, la cognición es un fenómeno esencialmente biológico, dado que las unidades biológicas pueden comprometerse sólo en interacciones que afectan su estructura sin destruir su organización biológica. En virtud de los cambios que una perturbación introduce en una unidad, el efecto de una segunda perturbación puede ser diferente al que hubiera sido de ser otro el caso. De este modo, cada perturbación que una unidad experimenta, como también los cambios estructurales que sufre incluso en ausencia de perturbaciones, influye en su sensibilidad y respuesta a las perturbaciones subsiguientes.

Al pasar por una sucesión de perturbaciones y de los correspondientes cambios estructurales, que no afectan su organización, cualquier unidad que persista debe necesariamente exhibir una cierta congruencia o "encaje" con su ambiente. De entre todas las secuencias posibles de cambios estructurales por los que una unidad puede pasar, su interacción con un ambiente particular selecciona una vía particular de interacción (y al menos hasta algún punto, dado el estado de espacio mucho mayor del ambiente, viceversa). Esto es lo que se designa como *acoplamiento estructural*: el proceso por el cual los cambios estructurales de una unidad se coordinan con aquellos de su ambiente. La noción de acoplamiento

estructural sirve para recordarnos que una perspectiva centrada en la unidad de la perturbación ambiental no es la única que puede tomarse. Estructuralmente, podemos ver también la unidad y su ambiente como un sistema dinámico singular cuyo estado está meramente desarrollándose de acuerdo a las leyes físicas subyacentes del universo, tal como en los sistemas dinámicos.

Un caso especialmente interesante e importante de acoplamiento estructural acontece cuando múltiples unidades comparten el mismo ambiente. Tales unidades no sólo interactúan con su ambiente, sino que ellas también sirven como fuentes mutuas de perturbación; por cierto, para cualquier unidad particular, las otras unidades son literalmente parte del ambiente. Las interacciones actuales entre múltiples unidades pueden llevar al acoplamiento estructural entre ellas, de manera que las vías que ellas toman a través de sus interacciones se coordinan. Tal comunidad de unidades que interactúan constituyen un *dominio consensual*, en el que las interacciones sirven para orientar a los otros agentes respecto de posibilidades similares de acción futura. Es esta mutua orientación con dominios compartidos lo que forma las bases de las interacciones lingüísticas (Beer 2004: 10 y ss.).

En esta perspectiva, si se quiere dar cuenta de qué es lo que "pasa dentro" o qué "piensa" la unidad se debe recurrir a la red de relaciones que la unidad mantiene para mantener su propia organización. En el caso de los organismos multicelulares ha de recurrirse al sistema neuronal entendido como una red de relaciones que permiten el mantenimiento de la organización propia. El sistema neuronal se halla inserto en el organismo a través de múltiples conexiones con muchos tipos celulares, formando una red tal que entre la superficie sensorial y la motora siempre hay una red de interconexiones neuronales, y constituyendo en conjunto lo que llamamos el sistema nervioso. Estrictamente, los animales multicelulares son unidades de segundo orden, dado que ellos son instanciados por

redes de sistemas autopoieticos de primer orden (células). Tales sistemas de segundo orden son organizacionalmente homeostáticos y mantienen un límite físico. Tienen una estructura y una organización. Pueden ser perturbados por sus ambientes ya destructivamente ya no-destructivamente, y por ello poseen un dominio cognitivo. Exhiben también una sensibilidad diferencial estado-dependiente a las perturbaciones y pueden por eso comprometerse en el acoplamiento estructural con sus ambientes y unos con otros. Así, los sistemas de segundo orden exhiben todas las características claves de un sistema autopoietico (Beer 2004: 12).

En estos sistemas de segundo orden, por fuerza, sistemas multicelulares, hay que recurrir, como recién se dijo, al sistema nervioso si se ha de dar cuenta de lo que pasa dentro del sistema. Recordemos aquí que el sistema nervioso constituye en tales sistemas la red de relaciones mantenida por la unidad para, a su vez, mantener su homeostasis organizacional. El sistema nervioso, una red neuronal, tiene una *clausura operacional*: está constituido de tal manera que, cualquiera sean sus cambios, éstos generan otros cambios dentro de él mismo, y su operar consiste en mantener ciertas relaciones entre sus componentes invariantes frente a las continuas perturbaciones que generan en él tanto la dinámica interna como las interacciones del organismo que integra. En otras palabras el sistema nervioso opera como una red cerrada de cambios de relaciones de actividad entre sus componentes. La noción de clausura operacional es una forma de especificar clases de procesos que, en su misma operación, se vuelven sobre ellos mismos para formar redes autónomas. Tales redes no caen en la clase de sistemas definidos por mecanismos externos de control (heteronomía), sino más bien en la clase de sistemas definidos por mecanismos internos de auto-organización (autonomía). Esto entraña que, en cuanto su operar, el sistema nervioso no discrimina entre adentro y afuera, sino sólo mantiene correlaciones

propias que están en continuo cambio. El estado estructural del sistema nervioso es el que especifica cuáles perturbaciones son posibles y qué cambios gatillan ellas en su dinámica de estados. Esto significa que no tiene inputs o outputs en el sentido tradicional. El sistema nervioso se concibe entonces como una unidad definida por sus relaciones internas, en las que las interacciones con el mundo sólo actúan modulando su dinámica estructural, esto es, como una unidad con clausura operacional.

El sistema nervioso es un sistema en continuo cambio estructural, es decir, tiene plasticidad. Los cambios estructurales ocurren no en las conexiones que unen grupos de neuronas, sino en las características locales de esas conexiones. Los cambios entonces ocurren a nivel de las ramificaciones finales y en las sinapsis. Las neuronas, el organismo que integran, y el medio en el que éste interactúa, operan recíprocamente como selectores de sus correspondientes cambios estructurales, y se acoplan estructuralmente entre sí: el operar del organismo, incluyendo el sistema nervioso, selecciona los cambios estructurales que le permiten seguir operando o la obligan a desintegrarse. Para el operar del sistema nervioso sólo hay una deriva estructural continua que sigue el curso en el que se conserva el acoplamiento estructural (adaptación) del organismo a su medio de interacción (Maturana y Varela 1984: 104 y ss.).

Pese a la importancia del sistema nervioso en cuanto mantenedor de la organización de una unidad, en el caso de un sistema cognitivo éste no es lo único explicativamente relevante para la cognición. En virtud de la sensibilidad diferencial estado-dependiente a la perturbación, cualquier sistema biológico (incluyendo una célula o una planta) es capaz de interactuar selectivamente con su ambiente y por ello posee al menos un dominio cognitivo rudimentario. Sin embargo, no hay duda que los sistemas nerviosos enriquecen significativamente los dominios cognitivos de los animales que los poseen. Al aumentar el estado interno

que puede mantenerse y, así, los cambios estructurales que pueden tolerarse, los sistemas nerviosos expanden enormemente el rango de interacciones en el que un organismo puede comprometerse sin perder su organización.

Conclusiones como ésta llevan a Varela, Thompson y Rosch (1991) a adoptar un enfoque "enactivo", una visión alternativa a la del modelo representacional / computacional. Tal visión supone que el rol del ambiente pasa a ser uno de trasfondo, no una fuente de inputs. De hecho, los sistemas cognitivos ya no se comprenden en base a sus relaciones de inputs y outputs, sino en base a lo que más arriba se ha denominado su *clausura operacional*. El punto clave es que tales sistemas no operan por representaciones. En vez de representar un mundo independiente, ellos *enactúan* un mundo como un dominio de distinciones que es inseparable de la estructura encarnada del sistema cognitivo. Este enfoque tiene como conceptos claves los de conexionismo, emergencia, auto-organización, asociatividad, dinámica de redes. Aquí se pone en cuestión el supuesto de que la cognición consiste en la representación de un mundo que es independiente de nuestras capacidades cognitivas y perceptuales por parte de un sistema cognitivo que existe independientemente del mundo. Este enfoque se refiere, más precisamente, a que la cognición no es la representación de un mundo dado por una mente dada, sino que es más bien la enacción de un mundo y una mente en base a una historia de variedades de acciones que realiza un ser en el mundo.

La impronta conexionista en este enfoque es clara: la cognición se entiende como la emergencia de estados globales en una red de componentes simples, que opera mediante reglas locales para la operación individual y reglas para los cambios en la conectividad entre los elementos. Un sistema tal funciona adecuadamente cuando la propiedades emergentes (y la estructura resultante) pueden verse como correspondiendo a capacidades cognitivas específicas, esto es, una solución exitosa a una tarea específica. Los ítems a tener en cuenta no son

símbolos, sino patrones complejos de actividad entre las unidades numerosas que conforman la red. Estas redes que artificialmente pueden servir para el reconocimiento y el aprendizaje tienen su correlato neurológico: las neuronas han de estudiarse como grandes conjuntos que están constantemente desapareciendo y surgiendo a través de sus interacciones cooperativas y en los que cada neurona tiene respuestas múltiples y cambiantes de una manera que depende del contexto. El cerebro es visto como un sistema altamente cooperativo: las interconexiones entre sus componentes neuronales entrañan que eventualmente todo lo que pasa en el cerebro es una función de lo que están haciendo los componentes. Esta clase de cooperación se sostiene tanto local como globalmente: funciona dentro de subsistemas del cerebro y a nivel de las conexiones entre estos subsistemas (Ibid.: 94).

Los cerebros, entonces, son sistemas altamente cooperativos. Sin embargo, ellos no son redes estructuradas uniformemente, dado que consisten de muchas redes que están ellas mismas conectadas de diversas maneras. El sistema entero se asemeja a un "patchwork"¹⁸ de sub-redes unidas por un proceso complejo de ajustes, antes que a un sistema que resulta de un diseño claro o unificado. La mente no es una entidad unificada u homogénea, ni siquiera en cuanto una colección de entidades, sino más bien una colección de redes de procesos desunificada y heterogénea. Esta clase de arquitectura plantea consecuentemente que, en vez de buscar un gran modelo unificado para todos los comportamientos de redes, se debiera estudiar redes cuyas habilidades están restringidas a actividades cognitivas específicas y luego buscar maneras de conectar tales redes (Ibid.: 105 y ss.).

Este cerebro, compuesto por redes neurales conectadas de diversas maneras, está "encarnado" en un cuerpo. En esta visión, se puede entender que

¹⁸ Un conjunto de retazos unidos para conformar una unidad.

el cuerpo es al hombre como la membrana, a la célula: es lo que proporciona un límite mediante el cual el sistema nervioso interactúa con el ambiente. Esto implica una visión de la cognición como dependiente de las clases de experiencia que provienen de tener un cuerpo con varias capacidades senso-motoras y, segundo, que estas capacidades senso-motoras individuales están anidadas en un contexto más abarcador biológico, psicológico y cultural. Los procesos sensoriales y motores, la percepción y la acción, son fundamentalmente inseparables en la cognición vivida. Tal unión no es contingente, sino que éstos han evolucionado conjuntamente. El enfoque enactivo consiste así en dos puntos: (1) la percepción consiste en acción guiada perceptualmente y (2) las estructuras cognitivas emergen de los patrones senso-motores recurrentes que permiten que la acción este guiada perceptualmente.

Al estudiar cómo el perceptor puede guiar sus acciones en su situación local, se descubre que tales situaciones locales cambian constantemente como resultado de las actividades del perceptor. Esto supone que el punto de referencia para entender la percepción no es ya un mundo dado, independiente del perceptor, sino más bien la estructura senso-motora del perceptor (la manera en que el sistema nervioso vincula las superficies sensoriales y motoras). Esta estructura -la manera en que el perceptor está encarnado- es la que determina cómo el perceptor puede actuar y ser modulado por los eventos ambientales. El enfoque enactivo de la percepción no se ocupa de cómo es recuperada una información acerca de un mundo dado, sino de determinar los principios comunes de uniones legales entre sistemas sensoriales y motores que explican cómo la acción puede ser guiada perceptualmente en un mundo dependiente del perceptor (Ibid.: 173).

Ya se ha mencionado que la unión entre procesos sensoriales, percepción y acción son inseparables en la cognición vivida y que tal unión no era contingente,

sino que se debía a la evolución. La evolución es lo que explica porque los seres vivos -incluidos nosotros- tenemos un operar que es exitoso en el mundo, puesto que la evolución determina las historias anteriores de acoplamiento estructural. La evolución no es considerada aquí como una solución "óptima" a un "problema" establecido evolutivamente. Esta visión tradicional de la evolución la entiende como la explicación principal en el decurso de los seres vivos, en el sentido de que cada estructura, mecanismo, rasgo o disposición de los seres vivos no pueden ser explicados sino teniendo una contribución directa a la supervivencia. En la visión de Varela et al. esta visión prescriptiva de la evolución -donde todo ha de tener valor de supervivencia- se transforma, de manera que la evolución es vista como básicamente proscriptiva: se pasa de la idea de que aquello que no está permitido está prohibido a la idea de que aquello que no está prohibido está permitido. El proceso de selección natural entonces descarta sólo lo que no es compatible con la supervivencia y la reproducción. Los organismos y las poblaciones ofrecen una variedad que la selección natural modula, garantizando sólo aquello que satisface los constreñimientos de la supervivencia y la reproducción. El proceso evolutivo se entiende como satisfaciendo tales requerimientos (tomando una solución satisfactoria, pero no óptima) antes que optimizando: la selección opera como un filtro amplio de supervivencia que admite cualquier estructura que tiene una integridad suficiente como para persistir. El problema evolutivo no consiste en cómo forzar una trayectoria precisa bajo los requerimientos de un encaje óptimo, sino más bien en cómo podar la multiplicidad de trayectorias viables que existen en cualquier punto dado.

Una de las consecuencias de este cambio de selección óptima a viabilidad es que la precisión y la especificidad de rasgos morfológicos o fisiológicos, o de capacidades cognitivas, son totalmente compatibles con su irrelevancia evidente para la supervivencia. Gran parte de cómo un organismo se ve y es está

completamente *sub-determinado* por los constreñimientos de la supervivencia y de la reproducción. Esta visión, denominada "evolución por deriva natural", depende de tres condiciones, a saber, la riqueza de las capacidades auto-organizativas en las redes biológicas, un modo de acoplamiento estructural que permita la satisfacción de trayectorias viables y la modularidad de redes de procesos independientes que interactúan una con otra por ajustamiento.

La *deriva natural* puede articularse en cuatro puntos:

- la unidad de la evolución (en cualquier nivel) es una red capaz de un rico repertorio de configuraciones que se organiza a sí mismas
- existe un bajo acoplamiento estructural con un medio: la selección es un proceso vigente de satisfacción que gatilla (pero no especifica) cambios en la forma de trayectorias viables
- la trayectoria específica (no única) o modo de cambio de la unidad de selección es el resultado entretelado (no óptimo) de múltiples niveles de redes de repertorios seleccionados auto-organizados
- la oposición entre factores causales interiores y exteriores es reemplazada por una relación co-implicativa, dado que el organismo y el medio se especifican uno al otro.

Desde esta perspectiva, la noción misma de lo que el ambiente es no puede separarse de aquella de lo que son los organismos y qué hacen. Las especies presentan y especifican sus propios dominios de problemas que deben resolverse al satisfacerlos. Tal dominio no existe allí afuera en un ambiente que actúa como una pista de aterrizaje para organismos que de alguna forma "caen en el mundo". En cambio, los seres vivos y sus ambientes están en una relación uno con otro mediante la *especificación mutua* o la *co-determinación*. De este modo, las regularidades ambientales no son rasgos externos que han sido internalizados,

sino que son el resultado de una historia conjunta, una congruencia que se desarrolla a partir de una larga historia de co-determinación. Así, el organismo es a la vez el sujeto y el objeto de la evolución (Ibid.: 195 y ss.).

En este sentido, podemos pensar que la evolución determina en cierta medida el acoplamiento estructural entre una unidad y su ambiente, determinando así el *dominio cognitivo* de la unidad. La cognición consiste así en la *enacción* o en la elaboración de un mundo mediante una historia viable de acoplamiento estructural. Tales historias, como ya se vio, no son óptimas, sino simplemente viables. Para que el acoplamiento sea viable, sin embargo, la acción perceptualmente guiada de un sistema debe simplemente facilitar la integridad continua de un sistema (ontogenia) y de su linaje (filogenia). De este modo, cualquier acción emprendida por el sistema está permitida en cuanto no viole el constreñimiento de mantener la integridad del sistema o de su linaje (Ibid.: 205).

La cognición, entonces, es *enacción*: una historia de acoplamientos estructurales que elabora un mundo. Ésta opera a través de una red que consiste en múltiples niveles de sub-redes senso-motoras interconectadas. Un sistema está funcionando adecuadamente cuando se transforma en parte de un mundo existente efectivo (como hacen los jóvenes de cada especie) o conforma uno nuevo (como sucede en la historia evolutiva). En esta visión, la representación ya no juega un rol central y el ambiente como fuente de input retrocede para ser sólo un trasfondo. Entra sólo en las explicaciones en aquellas ocasiones en que los sistemas sufren quiebres o eventos que no pueden ser satisfechos por sus estructuras. Igualmente, la inteligencia no se concibe ya como la capacidad de resolver un problema, sino como la capacidad de entrar en un mundo compartido de significación.

II. Sistema Senso-motor y el Origen de las Representaciones (Lakoff y Johnson)

Hemos visto que, de acuerdo a Varela et al., el sistema nervioso determina que los animales enactúan un mundo como un dominio de distinciones que es inseparable de la estructura encarnada del sistema cognitivo. Así, la cognición se entiende como la *enacción* de un mundo y de una mente en base a una historia de variedades de acciones que realiza un ser en el mundo. Esto supone que la percepción consiste en acción guiada perceptualmente y que las estructuras cognitivas emergen de los patrones senso-motores recurrentes que permiten que la acción esté guiada perceptualmente.

La propuesta de Lakoff y Johnson en términos del surgimiento de las representaciones puede considerarse como una ampliación del concepto de *enacción* como acción guiada perceptualmente. Estos autores toman en serio la idea de que nuestra encarnación y nuestra experiencia en cuanto seres encarnados determinan lo que podemos concebir y que, por ende, la experiencia senso-motora básica determina nuestros conceptos.

Según Lakoff y Johnson, los hombres tienen estructuras cognitivas muy generales denominadas *esquemas kinestésicos de imagen*, como los esquemas del "contenedor", de "parte-todo" y el de "fuente-camino-meta". Dichos esquemas se originan en la experiencia corporal, pueden definirse en términos de ciertos elementos estructurales, tienen una lógica básica y pueden proyectarse metafóricamente para dar una estructura a una gran variedad de dominios cognitivos. Así, los elementos estructurales del esquema del contenedor son "interior, límite, exterior", su lógica básica "dentro o fuera" y sus proyecciones metafóricas estructuran las conceptualizaciones del campo visual (las cosas se "escapan" a la vista), relaciones personales (uno entra o sale de una relación),

teoría de conjuntos (los conjuntos contienen sus miembros), etc. (Varela et al. 1991: 177-78).

A partir de un estudio detallado de estas clases de ejemplos, Johnson afirma que los esquemas de imagen emergen de ciertas formas básicas de actividades e interacciones senso-motoras y de este modo proveen una estructura pre-conceptual a nuestra experiencia. Dado que nuestra comprensión conceptual está modelada por la experiencia, nosotros tenemos conceptos esquemáticos de imagen. Estos conceptos tienen una lógica básica, que imparte una estructura a los dominios cognitivos en los cuales ellos se proyectan imaginativamente. Dichas proyecciones no son arbitrarias, sino que se logran a través de procedimientos de proyección metafórica y metonímica, que están a su vez motivados por las estructuras de la experiencia corporal (Ibid.). A continuación, se detallará un poco más este planteamiento.

Este planteamiento tiene tres supuestos: la mente es inherentemente encarnada, el pensamiento es mayormente inconsciente y los conceptos abstractos son fundamentalmente metafóricos. Esto significa que la razón surge de la naturaleza del cerebro, cuerpos y experiencia corporal nuestros en el sentido de que:

la mera estructura de la razón misma proviene de los detalles de nuestra encarnación. Los mismos mecanismos neuronales y cognitivos que nos permiten percibir y movernos crean también nuestros sistemas conceptuales y modos de razonamiento (Lakoff y Johnson 1999: 4).

De este modo, las inferencias racionales se computan por la misma arquitectura neural que se usa en la percepción o en el movimiento corporal. Esto significa que la naturaleza peculiar de nuestros cuerpos conforma las posibilidades mismas de

conceptualizar, esto es, de tener representaciones. La categorización, así, no es una materia puramente intelectual, que ocurre después del hecho de la experiencia, sino que la formación y el uso de las categorías es la materia de la experiencia.

Los conceptos se conciben como estructuras neurales, que nos permiten hacer alguna clase de tarea inferencial o imaginativa relativa a una categoría. Las categorías son cosas en el mundo que determinan lo que nosotros juzgamos como real (árboles, rocas, etc.) y nuestros conceptos determinan cómo pensamos acerca de dichas categorías. Para funcionar realísticamente en el mundo, nuestras categorías y nuestras formas de razonamiento deben “operar” muy bien en conjunto; nuestros conceptos deben caracterizar la estructura de nuestras categorías suficientemente bien como para que operemos.

La “encarnación” de la razón mediante el sistema senso-motor es de gran importancia. Es, como en la visión de Varela et al., una parte crucial de la explicación de porqué es posible que nuestros conceptos calcen tan bien con la manera en que operamos en el mundo. Ellos calzan tan bien porque han evolucionado a partir de nuestro sistema senso-motor, que, a su vez, ha evolucionado para permitirnos operar bien en nuestro ambiente físico.

La “herramienta” que permite que la imagería mental convencional de los dominios senso-motores sea usada por los dominios de la experiencia subjetiva es, como se señaló más arriba, la metáfora. La metáfora es la que permite que la experiencia senso-motora se pueda “traducir” en un lenguaje subjetivo, es decir, es la experiencia senso-motora la que nos permite representar. Esto se produce por lo que Lakoff y Johnson denominan “la teoría de la metáfora primaria”, que consta de otros cuatro principios teóricos:

- La teoría de la mezcla: los niños no distinguen entre experiencia no senso-motora (subjetiva) y senso-motora.
- La teoría de la metáfora primaria: las metáforas complejas son “moleculares”, hechas de partes metafóricas “atómicas” (primarias)
- La teoría neural de la metáfora: las “asociaciones” hechas durante el período de mezcla se realizan neuralmente en activaciones simultáneas que resultan en conexiones neurales permanentes, hechas a través de redes neurales que definen conceptos de

dominio. Estas conexiones forman la base anatómica de la activación de la fuente al objetivo, que constituye las implicaciones metafóricas.

- La teoría de la mezcla conceptual: dominios conceptuales distintos pueden co-activarse y, bajo ciertas condiciones, pueden formarse las conexiones a través de los diferentes dominios, llevando a nuevas inferencias –este sería el mecanismo por el cual dos o más metáforas primarias pueden ser traídas conjuntamente para formar un complejo de metáforas mayor–.

Las metáforas primarias son inevitables, parte de lo que Lakoff y Johnson llaman el “inconsciente cognitivo”, término que describe todas las operaciones mentales inconscientes que se refieren a sistemas conceptuales, significado, inferencia y lenguaje. El “inconsciente cognitivo” no sólo incluye las operaciones cognitivas automáticas, sino también todo nuestro conocimiento implícito. Todos nuestros conocimientos y creencias se enmarcan en términos de un sistema conceptual que reside principalmente en el inconsciente cognitivo.

De este modo, las metáforas primarias son parte del inconsciente cognitivo. Las adquirimos automática e inconscientemente mediante el proceso normal de aprendizaje neural y podemos ser inconscientes de que las tenemos. Éstas se forman cuando un dominio de experiencia subjetiva es regularmente activado en conjunción con un dominio senso-motor, de modo que se establecen conexiones neuronales permanentes mediante cambios en las cargas sinápticas. Así, desde una perspectiva neural, las metáforas primarias son conexiones neurales aprendidas por co-activación. Ellas se extienden a través de partes del cerebro entre áreas dedicadas a la experiencia senso-motoras y áreas dedicadas a la experiencia subjetiva. La complejidad inferencial mayor de los dominios sensoriales y motores le dan a las metáforas un carácter asimétrico, con inferencias que fluyen sólo en una dirección. Desde un punto de vista conceptual, las metáforas primarias son proyecciones que cruzan dominios, y que van de un dominio-fuente (el dominio senso-motor) a un dominio-objetivo (el dominio de la experiencia subjetiva), preservando la inferencia. Esto último es la propiedad más característica de las metáforas conceptuales (Ibid.: 55 y ss). Tales metáforas se usan activamente en el razonamiento, tal como vimos en el ejemplo del contenedor.

III. Enfoques Encarnados, Modelo Computacional y Representación

La ciencia cognitiva enactiva, entonces, es un estudio de la mente que no describe la cognición como el reflejo interno de un mundo objetivo externo, sino que define las interacciones senso-motoras repetidas con el mundo como el lugar básico de interés científico y explicativo. El enfoque de Varela et al. se opone a la idea de que la cognición es fundamentalmente representación, pues la cognición se entiende como la realización existosa de tareas específicas en el mundo, no como un representarse el mundo para poder actuar en él o resolver problemas. Lakoff y Johnson, en cambio, consideran la existencia de la representación, entendida como una proyección metafórica del ámbito senso-motor.

En el caso de Varela et al., las ideas de acoplamiento estructural y de dominio cognitivo se alejan de la noción de la mente como un aparato de input y output que procesa información, mientras la idea de la mente como una red autónoma y emergente de relaciones cobra un lugar central. Las unidades autopoieticas, a diferencia del agente inteligente resolvidor de problemas de la GOFAI, no tienen inputs o outputs. Como ya se vio, éstas pueden ser perturbadas por eventos independientes y sufrir cambios estructurales internos que compensan aquellas perturbaciones. Ahora bien, cualesquiera sean los cambios internos que tienen lugar, éstos siempre están subordinados a la mantención de la organización de la unidad. De este modo, cualquier relación que se observe entre estos cambios y el curso de las perturbaciones no corresponde a la organización de la unidad, sino al dominio en el que la unidad es observada (Maturana y Varela 1980: 81). No hay aquí una noción de que la unidad se esté representando variaciones en las condiciones ambientales, comparándolas con las nociones necesarias para mantener su organización, y luego haciendo los cambios estructurales necesarios en relación al ideal de homeostasis organizacional. Esto es, en este marco explicativo, no tiene mucho sentido hablar de una unidad resolviendo un problema, y menos, resolviéndolo con representaciones.

Lo importante aquí para explicar la existencia de un dominio cognitivo en el cual puede operar una unidad no es que tal unidad pueda razonar de manera que logre resolver problemas o representarse situaciones contrafácticas. La interacción con el ambiente y el éxito de la unidad en mantener su homeostasis organizacional se da en base a un *acoplamiento estructural*, que se explica por la historia ontogenética y filogenética de la unidad.

Esta visión se opone a una noción de la inteligencia en un sentido computacional / representacional por dos motivos. Primero, el éxito cognitivo de la unidad no requiere para su explicación de una entidad similar a la “representación”. Una entidad tal parece estar demás si queremos explicar qué es el operar cognitivo desde un punto de vista enactivo. Segundo, esta manera de explicar sólo da cuenta de las correlaciones entre sistema nervioso y ambiente, de manera que los límites interior / exterior no son importantes para el operar del sistema nervioso. No se entiende bien aquí cómo podría surgir una noción de representación, en el sentido de que no hay algo para ser representado “fuera” del sistema nervioso.

El enfoque de Lakoff y Johnson permite pensar, en cambio, una salida a estas consideraciones biológicas, entendiendo cómo a partir del sustrato biológico y de la experiencia de tener un cuerpo surgiría la noción de representación. Las representaciones surgirían a partir de proyecciones metafóricas desde un ámbito senso-motor a uno subjetivo. Ahora bien, también aquí la noción de cognición es muy distinta a aquella sustentada por la noción computacional / representacional de la mente, puesto que se entiende que lo que da cuenta de gran parte de la actividad cognitiva humana es inconsciente: no hay aquí una visión de un operador que actúa de acuerdo a metas y fines, buscando soluciones en un espacio simbólico.

IV. Discusión

Los enfoques “encarnados” en biología permiten entender la discusión sobre la existencia de la representación de una manera más “realista” que la robótica situada y la tesis dinámica en ciencia cognitiva. Esta visión se sustenta en las historias evolutivas de los seres vivos. Así, la deriva genética de Maturana y Varela posibilita una visión del operar exitoso de los seres vivos como proscriptiva, lo que implica correlaciones entre una unidad autopoietica y su ambiente, pero dichas correlaciones sólo deben cumplir requerimientos mínimos que le permitan a la unidad subsistir, pese a las perturbaciones que puede sufrir.

De este modo, la evolución parece estar constreñida y liberada en cuanto al sustrato vital sobre el cual opera. Está constreñida para construir sus soluciones incrementalmente sobre la base de formas ancestrales más simples. Y está liberada al ser capaz de descubrir soluciones “confusas”, pero no obvias, que pueden, entre otras cosas, explotar las interacciones ambientales y los *loops* de retro-alimentación de maneras complejas. Las soluciones naturales pueden explotar casi cualquier mezcla de recursos neurales, corporales y ambientales junto con las interacciones complejas entre ellos (Clark 2001a: 86).

Trabajos como los de Varela et al. (1991) y de Lakoff y Johnson (1999) parecen estar apuntando precisamente a un punto de vista más “natural” que “ingenieril” (en el sentido de “resolvidor de problemas”) y recalcar la distancia que existe entre mirar desde el punto de vista de la “naturaleza” y desde el punto de vista humano. En el nivel explorado por Varela et al. se destaca la importancia del sistema nervioso y su interacción, mediante el cuerpo, con el ambiente. En este mismo enfoque, los detalles sobre cómo opera el sistema nervioso permiten entender que, en una primera instancia, no hay una representación del mundo, dado que no hay sino una mantención de un equilibrio homeostático entre unidad –sistema nervioso– y mundo –entendido básicamente como la fuente de perturbaciones–.

El punto de vista enactivo considera la percepción como acción guiada perceptualmente y las estructuras cognitivas como emergiendo de los patrones senso-motores recurrentes (para dirigir la acción guiada perceptualmente). Esto se condice con los postulados de Lakoff y Johnson, según los cuales nuestros conceptos surgen a partir de esquemas experienciales, que a su vez están basados en la experiencia senso-motora. Si bien el planteamiento de Varela et al. parece a veces negar la noción de representación y el de

Lakoff y Johnson no explica cómo se “proyectan” las metáforas de un dominio senso-motor a uno subjetivo, ambas propuestas permiten entender cómo la historia evolutiva y / o la ontogenética operan sobre un material ya habido para desarrollar nuevas soluciones. Este enfoque se relaciona con el postulado de Brooks sobre la construcción por capas de las “criaturas” y METATOTO puede verse asimismo como un ejemplo de la posición de Lakoff y Johnson. METATOTO, recordemos, puede comprometerse en exploraciones off-line de su ambiente, explorando un ambiente virtual con la misma maquinaria que el robot más básico TOTO emplea para moverse en el mundo real. Puede pensarse, entonces, que las proyecciones metafóricas posibilitan precisamente una suerte de transformación funcional de la arquitectura básica on-line de nuestra constitución biológica. Esto se condice también en alguna medida con un enfoque enactivo, dado que las representaciones surgirían aquí de una acción guiada perceptualmente –de una arquitectura diseñada en primera instancia para las interacciones con el mundo–, permitida por patrones senso-motores recurrentes –que permitirían vincular experiencias senso-motoras con conceptos de nuestra experiencia subjetiva–.

En estos postulados, entonces, se puede encontrar una base biológica para nuestro pensamiento representacional: la experiencia senso-motora es transformada crecientemente en experiencia subjetiva, que se desliga del operar reactivo y se puede usar para pensar en cosas que no están inmediatamente presentes para el agente (por ejemplo, en situaciones contrafácticas). Ahora bien, puede que esto sea tal vez sólo una parte de la historia, y que tal “traducción” no sea suficiente para explicar las clases de pensamientos cada vez más lejanos de realidades concretas en los que nos comprometemos (la otra puede ser el ambiente social), como propone Clark (1997a, 2001a; cf. Clark y Chalmers 1998).

CAPÍTULO QUINTO

REVISIÓN CRÍTICA DE LAS TEORÍAS “ENCARNADAS”

En los tres capítulos anteriores se han presentado tres teorías que, en sus visiones más extremas (como en el caso de Brooks en la robótica situada, de algunos dinamicistas y de Varela et al.), ponen en cuestión la noción de representación. En este capítulo se revisa qué tienen en común los tres enfoques presentados, considerando en primera instancia que todos conciben la cognición como la operación de un agente encarnado y ambientalmente situado. Se examina también por qué, en su versión más radical, estos enfoques ponen en cuestión la noción de representación y si tal cuestionamiento se sostiene. Se presentan además casos de operaciones cognitivas en los que parece requerirse la noción de representación para poder explicarlos. Finalmente, se delinea cómo ha de concebirse la noción de representación a la luz de las críticas de los enfoques “encarnados”.

Aspectos Comunes de los Diferentes Enfoques “Encarnados”

La robótica situada, la teoría de sistemas dinámicos y los enfoques biológicos respecto de la cognición tienen varios rasgos que los distinguen de las explicaciones de los fenómenos cognitivos en términos computacionales / representacionales. Un rasgo compartido fundamental se refiere a que el agente cognitivo se considera como un agente con un cuerpo situado en un ambiente específico. De ahí que todos estos enfoques intenten modelar y entender la interacción acoplada entre organismo y ambiente. En esta interacción se descubre crecientemente el rol de trucos múltiples y localmente efectivos, de heurísticas y de atajos (que a menudo suponen acción) en el apoyo del éxito adaptativo y la atención a varias formas de resolución de problemas distribuidas ecológicamente¹⁹, que permiten “descargar” parte importante de la actividad de resolver problemas al ambiente local. Esto ha llevado a reinventar la racionalidad como un logro activo, distribuido y que explota el ambiente.

Así, estas tres visiones se alejan bastante de la imagen tradicional de la mente como un procesador de representaciones, lo que implica que, en ciertas conclusiones más radicales, se considere que la noción de representación no tiene una relevancia explicativa para la vida cognitiva de un agente. Como señalamos, estas

¹⁹ Sobre la consideración del ambiente principalmente en tanto “cultural”, ver Clark (1997a, 2001a).

tres visiones presentan ciertas características comunes como el énfasis en la “mente” situada, es decir, en la mente concebida como anidada tanto en un cuerpo como en un ambiente. La palabra *acoplamiento* subraya precisamente esa característica de la mente, señalando además el hecho de que este anidamiento en el cuerpo y en el mundo se ha dado mediante una historia de interacciones conjuntas, ya sea visto esto en términos de subsistemas que responden al mundo, de variables que se co-determinan o de una historia ontogenética y evolutiva.

Estas tres visiones presentan además una visión marcada por la fragmentación de la cognición: lo que se conoce no es un mundo detallado, sino aspectos parciales del mundo que son útiles en tanto se interactúa con ellos. De ahí que las tareas que lleva a cabo un sistema cognitivo sean también fragmentarias: por las capacidades cognitivas responden o circuitos específicos diseñados para la realización de una tarea en robótica situada, o las ecuaciones dinámicas particulares que dan cuenta de un dominio parcial de la cognición en sistemas dinámicos, o sub-redes senso-motoras interconectadas que se forman en múltiples niveles y que a su vez conforman una red en los enfoques biológicos de la cognición. Esto entraña que la mente ya no se ve como dependiente de un operador central que ordena nuestras sensaciones y nuestras acciones.

Esta negación de la mente como constituida por una unidad ejecutiva central va de la mano con la negación de la representación como una entidad que media entre inputs sensoriales y outputs motores. Dada la imbricación de un sistema nervioso en un cuerpo y en un ambiente, donde cada uno influye en el operar del otro, la separación entre inputs-representación-outputs resulta artificial. Esta separación, además, es engañosa en el sentido de que la percepción y la acción no pueden concebirse más como entidades separadas, dado que las acciones mismas modulan y ayudan a la percepción, y la percepción no es neutral, sino que se orienta a la realización de una acción específica.

Estos rasgos en común son los que ponen en serios aprietos a la noción de la mente como un computador, pues las visiones “encarnadas” de la mente presentan posibilidades diversas de entender los fenómenos cognitivos. La mayoría de los enfoques alternativos revisados en los tres capítulos anteriores implementan modelos que demuestran el uso posible de arquitecturas cognitivas diferentes a la simbólica como, por ejemplo, la construcción de robots con arquitecturas de subsumisión en robótica situada. Un sentido similar tienen las explicaciones dinámicas del movimiento de las piernas de los niños en sistemas dinámicos y el operar de una “unidad cognitiva” mediante un “acoplamiento estructural” como en el caso del sapo de Maturana y Varela o el operar a través de metáforas que son en primera instancia senso-motoras, como en los planteamientos de Lakoff y Johnson.

En estas explicaciones alternativas la representación mental parece jugar un rol menor o ninguno. La explicación de los fenómenos cognitivos se entiende en términos de un acoplamiento entre agente y ambiente, en donde no es necesaria una representación del mundo por parte del agente. Tampoco hay una noción clara de un sistema cerrado, que recibe un input y produce un output, sino una causación mutua continua. En palabras de Clark (1999b: 346), lo que se da es una “dinámica de co-ordinación” entre los mundos externo e interno que reemplaza las nociones de representación interna y de computación por la de estrategias menos costosas,

como reaccionar simplemente a ciertas “pistas” en el ambiente, cuya tarea no es primero representar un mundo y luego razonar sobre la base de la representación, sino mantener una suerte de poderoso equilibrio adaptativo que acopla conjuntamente el agente y el ambiente.

En el mismo sentido, muchos fenómenos pueden entonces entenderse como procesos de causación recíproca, donde lo importante a tener en cuenta no es el agente cognitivo por sí mismo, sino sus relaciones con el ambiente en el que éste se encuentra y viceversa. Estos fenómenos subrayan que, pese a que sea posible encontrar una o más unidades con claridad, la comprensión de dichos fenómenos no debe llevarse a cabo en términos de aislar una unidad de su ambiente local trazando límites importantes de transducción o acción. La envoltura espacial que demarca cada componente es de poca importancia si nuestro objetivo es comprender la evolución de comportamientos del mundo real. La unidad adecuada de análisis, entonces, es el sistema acoplado mayor que ellos constituyen conjuntamente. Este sistema acoplado mayor supone que podemos enfrentarnos a desarrollos de comportamiento que resisten una explicación en términos de inputs y outputs de alguna supuesta maquina cognitiva aislada (cf. Clark 1999a: 514-15).

El acoplamiento entraña que la distinción entre mente, cuerpo y ambiente se vuelva problemática y, por ende, la noción de un agente adquiriendo “información” sobre el mundo –tal como en la visión computacional / representacional– también se vulnera. No hay un sentido claro de cómo pueda darse tal información y de qué signifique que los sistemas operen sobre una base informacional, pues no hay una información que se esté obteniendo del “mundo”, sino reacciones a eventos del mundo (como las criaturas de Brooks), dinámicas sutiles de acoplamiento (como en Thelen y Smith) o procesos de homeostasis interna (al decir de Varela).

Un problema serio en este sentido, es que, dada la noción de acoplamiento, no se distingue cuál sea la unidad cognitiva particular a la que hay que atender. El punto es que es difícil distinguir el organismo que realiza efectivamente una operación cognitiva, si la cognición se entiende como dispersa, dado el acoplamiento entre sistema nervioso, cuerpo y ambiente. Esto implica además que es problemático entender qué sea la cognición, pues no hay una unidad cuyo comportamiento se quiera explicar.

La diferencia entre los enfoques “encarnados” y el representacional, que, a primera vista, exhiben tipos muy diferentes de explicación, podría deberse a que los primeros se centran en clases de capacidades y comportamientos que son menos “racionales” y más “reactivos” o “automáticos” (como seguir paredes, recoger latas de bebidas, los movimientos rítmicos de los dedos, el movimiento de la lengua del sapo) que el segundo (centrado a su vez

en casos como la conversación, la comprensión de cuentos, el razonamiento guiado por metas). Cabría pensar entonces que los enfoques encarnados y el computacional / representacional se enfocan en dimensiones bastante diferentes del mismo sistema. Así, al ver un agente humano, el enfoque encarnado se centraría en la realización de tareas como coger algo y cómo sistema nervioso, cuerpo y mundo proveen conjuntamente la solución a tal tarea; el enfoque computacional, en cambio, intentaría explicar por qué quiere alcanzar algo, adscribiendo creencias y deseos que explican la intención tras el movimiento del brazo.

Sin embargo, esto no es así. Los hechos acerca de la encarnación no impactan sólo nuestras ideas acerca de los procesos senso-motores de bajo nivel. En el caso humano se presenta a todo nivel una mezcla de estrategias altamente encarnadas y anidadas en el ambiente, junto con estrategias aparentemente mucho más abstractas y aparentemente desacopladas y con la creación y manipulación de ítems simbólicos externos que a menudo funcionan como una clase de puente entre ambas. Parece probable que una clave para comprender la naturaleza y la potencia del pensamiento y la razón humana descansa precisamente en la comprensión de las relaciones e interacciones complejas entre estos varios tipos de estrategia y recursos (Clark 1999b: 350; cf. Clark 2001b).

Todos los enfoques “encarnados” ponen así de relieve que el cuerpo y el mundo pueden operar como mucho más que un mecanismo o escenario para acciones prácticas simples (así sucede, entre otros casos, con TOTO y METATOTO, con el caminar de los niños, etc.). Cuerpo y mundo extienden ampliamente el espacio cognitivo y representacional al proveer de un espacio operacional donde las propiedades de guardar información y de transformar información pueden complementar aquellas del cerebro biológico.

Enfoques “Encarnados” y Representación

Las visiones encarnadas pueden verse al menos de dos maneras (Clark 1999b: 348 y ss.). La primera, la *encarnación simple*, trata el hecho de que nuestro sistema nervioso esté anidado en un cuerpo y en un ambiente principalmente como un constreñimiento para una teoría de la organización y del procesamiento

internos. Desde este punto de vista, la racionalidad humana es considerada como un producto indirecto de los modos biológicos de respuesta adaptativa y depende además de los efectos de una clase especial de hibridización, en la cual los cerebros humanos entran en una poderosa cascada de relaciones genuinamente simbióticas con artefactos ricos en conocimiento y con tecnologías. Esta perspectiva no conlleva un rechazo radical de la noción de representación interna, aunque sí exige que dicha noción sea reformulada en términos que consideren el doble anidamiento de los seres humanos en el cuerpo y en el ambiente.

La segunda manera es llamada *encarnación radical* y considera que los hechos de la “encarnación” y del “anidamiento ambiental” han de transformar profundamente la materia principal y el marco teórico de la ciencia cognitiva. Las formulaciones de la tesis encarnada radical implican además una doctrina de “incremento cognitivo biológico”, según la cual la racionalidad humana a larga escala es alcanzada, de un modo bastante directo, por algunas series de alteraciones sobre los modelos biológicos básicos de respuesta adaptativa (Clark 2001b). En la “encarnación radical”, el rol explicativo de la representación se ve seriamente vulnerado, puesto que se plantea como una alternativa explicativa respecto del marco computacional / representacional de la mente. La *encarnación radical* sustenta lo que puede denominarse la *tesis radicalmente encarnada de la cognición*, que Clark formula así:

Las visiones estructuradas, simbólicas, representacionales y computacionales de la cognición están equivocadas. La cognición encarnada es estudiada mejor al usar ideas no computacionales y no representacionales y esquemas explicativos que incluyen, por ejemplo, las herramientas de la teoría de sistemas dinámicos (1997b: 461).

Dicha tesis supone que la comprensión de la interrelación compleja entre cerebro, cuerpo y mundo requiere herramientas analíticas y métodos nuevos, tales como aquellos de la teoría de sistemas dinámicos; que las nociones tradicionales de representación interna y de computación son innecesarias e inadecuadas y que la descomposición típica de un sistema en una variedad de subsistemas neurales o funcionales es a menudo equívoca, y nos ciega a la posibilidad de descomposiciones alternativas y más explicativas que atraviesan las divisiones tradicionales mente-cerebro y mundo (Clark 1999b: 349).

Clark (1997b: 471) descubre tres supuestos básicos en esta tesis, formulada respecto del desafío dinamicista, pero que bien puede extenderse además a las otras dos visiones encarnadas aquí revisadas. El primero consiste en la “hipótesis de continuidad”, que afirma que toda la cognición es continua con sus fundamentos motores y de desarrollo, y de ahí que la conformación de soluciones (emergentes, situadas, de engranaje sutil) a problemas básicos motores senso-motores, como el caminar o el alcanzar un objeto, serán recapituladas en todos los dominios cognitivos supuestamente más altos. El segundo se refiere al supuesto de la naturaleza “objetivista” de cualquier representación interna putativa. Y el tercero, a que hay una cierta clase de “sistema controlador del comportamiento” implicada en las explicaciones computacionales.

La hipótesis de continuidad va de la mano con la idea de que la manera de entender correctamente cómo surge la racionalidad en los seres humanos es llevar a cabo un proceso incrementativo de investigación que comience por explicar los aspectos más básicos de la respuesta adaptativa (como el evitar obstáculos, el reconocimiento del predador, etc.). Tal proceso sustenta una tesis de “continuidad fuerte”, de acuerdo a la cual:

Vida y mente tienen un patrón abstracto común o un conjunto de propiedades organizacionales básicas. Las propiedades funcionales características de la mente son una versión enriquecida de las propiedades funcionales que son fundamentales para la vida en general. La mente es, literalmente, similar a la vida (*life-like*) (Godfrey-Smith citado por Clark 2001b).

En este sentido, Randall Beer (2000: 94) ha planteado que el estudio de la cognición debiera partir definiendo un “comportamiento cognitivo mínimo”, entendiéndose por esto el comportamiento más simple que pueda plantear preguntas de interés cognitivo. El estudiar modelos más simples que exhiben las características básicas de un agente cognitivo situado, pero que están abiertas a un análisis experimental y teórico detallado, es una estrategia poderosa para explorar las implicaciones de una perspectiva encarnada²⁰. En este mismo sentido, Beer (2003) plantea la necesaria re-conceptualización de la representación interna, dado que el único requerimiento sobre el sistema nervioso de un agente es que esté dotado con un potencial oculto para comprometerse en los patrones de comportamiento necesarios para su supervivencia y reproducción. Sólo cuando tal sistema nervioso está dentro del cuerpo de un agente particular, su potencial se realiza efectivamente a través de la interacción dinámica resultante; y de aquí la idea de que el sistema nervioso del agente debe representar explícitamente cualesquiera regularidades que el comportamiento del sistema completo cerebro/ cuerpo/ ambiente exhiba es una hipótesis teórica adicional. Beer, así, establece que, en principio, la interacción dinámica es la que da cuenta de la actividad cognitiva de un agente encarnado y que cualquier enfoque representacional respecto de la cognición debe demostrar cómo se implementa o se da la “representación” en el operar cognitivo; esto es, pone la “carga de la prueba” en los proponentes de una visión representacional de la mente.

A este respecto, Clark (Ibid.; cf. 2001: 130) señala que la noción de “continuidad” cubre una variedad significativamente amplia de posibles uniones, pasajes y cosas en común. La continuidad ontogenética propuesta por los dinamicistas seguramente existe de alguna forma u otra, pero esto no asegura que a partir de una explicación del desarrollo motor temprano se pueda reconstruir toda la cognición humana (como en el planteamiento de Lakoff y Johnson, según el cual el sistema senso-motor “proyecta” su dominio cognitivo sobre uno subjetivo). Es más, puede darse una continuidad arquitectural, a la manera de los robots de

²⁰ Cómparese con la idea de Brooks que el desarrollo de las “criaturas” debe emular el desarrollo evolutivo.

Brooks, sin que se de una continuidad en la solución a problemas, de modo que entren a jugar un papel representaciones internas, como muestra el robot METATOTO (cf. Clark 1999b: 347-348). De este modo, se puede discernir una suerte de continuidad arquitectural junto con una discontinuidad funcional bastante radical. Como resultado, algunas funciones cognitivas pueden depender no sólo de ajustes menores de procesamientos senso-motores básicos, sino también del desarrollo de nuevas clases de procesos funcionales relativamente independientes; esto implica que puede que los mecanismos neurales del pensamiento más elevado sean del todo continuos con los mecanismos de control de la acción on-line, pero también pueden no serlo; o ser una mezcla de estrategias en las cuales nuevas clases de rutinas procesadoras de información coexisten y a veces explotan o cooptan sistemas más primitivos (ver la discusión acerca de las señales de propiocepción y de METATOTO en el capítulo segundo).

En cuanto al segundo supuesto, la naturaleza objetivista de la representación interna, parece que el objetivo de mucho del escepticismo dinámico no es la representación interna *per se*, sino un tipo particular de representación interna, llamada a veces objetivista, que aparece como un modelo detallado, independiente del punto de vista y similar a un mapa de algún aspecto del mundo. Se puede pensar, sin embargo, en otros tipos de representación (y en otro tipo de continuidad con respecto a las raíces motoras y de desarrollo): puede haber representaciones internas cuyos contenidos están fuertemente equipados para apoyar la realización de clases importantes de acciones en el mundo real, en tiempo real. Tales contenidos pueden ser manipulados off-line, pero aún así siguen siendo tipos de contenido que están especialmente ajustados para el control y la coordinación de la acción real en tiempo real. Pese a que la cognición se da a partir de una experiencia en el mundo y su economía interna haya sido esculpida por las necesidades en tiempo real y específicas respecto de tareas, ésta no necesita ser siempre interactiva (Clark 1997b, cf. Clark 2001: 131-32). Si se considera el origen de las representaciones como senso-motor, éstas ya no tendrían que considerarse como representaciones objetivas, en el sentido de que representan un mundo objetivo, exterior al hombre; sino que cobran su significado precisamente a partir de la acción en el mundo, de manera que representan al mundo siendo a la medida “de la experiencia” de nuestro sistema nervioso con un cuerpo y un mundo determinados.

El último supuesto se refiere a que los modelos computacionales entrañan la acumulación y el uso de complejas estructuras de control interno que delinean, en detalle explícito, todos los valores y escenarios de todas las variables físicas involucradas en una acción dada. La concepción opuesta es la que señala que el comportamiento emerge de una secuencia de estados estables momentáneos en un sistema complejo con una dinámica intrínseca ricamente interdependiente. Esta visión puede caracterizarse con el eslogan “patrones sin programas”. Ahora bien, no se necesita ver aquí una dicotomía (programado versus no programado), sino que bien puede verse un *continuum*: mientras menos información se especifique (mientras mayor sea el trabajo llevado a cabo por la dinámica intrínseca más amplia de un sistema), menor valor hay en tratar las contribuciones neurales como una suerte de programa. En medio de este *continuum* hay posiblemente un espacio para “programas parciales”: conjuntos de instrucciones mínimas que explotan al máximo las dinámicas inherentes (del cuerpo y del ambiente) del sistema controlado (Clark 1997b). Desde este punto de

vista menos radical, las investigaciones “encarnadas” pueden estar indicando que problemas como mover las piernas, alcanzar un objeto y similares pueden requerir menos a la manera de conjuntos de instrucciones detalladas interiormente que lo que hasta ahora se había supuesto, gracias a las dinámicas sinérgicas bastante complejas ya implementadas en, por ejemplo, los brazos y los músculos mismos (Clark 1997a: 158). Esto supone que en una amplia variedad de casos no hay necesidad, como en el modelo representacional de la mente, de un concepto de representación que figure en una secuencia de instrucciones. En vez de concebir un control central que indica cómo alzar los brazos o mover las piernas con tal de, por ejemplo, moverse hasta un lugar deseado, lo que se da es una interacción simple entre sistema nervioso, cuerpo y mundo que no requiere de un conjunto de representaciones y de procesamiento de representaciones para operar.

La hipótesis de continuidad, la crítica a la naturaleza “objetiva” de cualquier representación interna putativa y a la existencia de sistemas controladores de comportamiento no son entonces razones válidas para abandonar la noción de representación interna. Indican, sin embargo, que la noción de representación debe ser rearticulada en términos que den cuenta de la condición de la cognición como doblemente situada: en el cuerpo y en el ambiente. Esta rearticulación es necesaria porque hay fenómenos cognitivos donde la representación juega claramente un rol importante, como en el pensar sobre situaciones posibles o explicar la conducta de los otros para tomar las propias decisiones.

Así, la noción misma de representación puede reformularse en un sentido que respete un origen experiencial senso-motor, al modo de “representaciones dirigidas a la acción”, teniendo en cuenta que es posible pensar una continuidad que se dé en términos arquitecturales, pero no funcionales (y esta distinta funcionalidad daría lugar a la representación). Igualmente, los sistemas controladores de comportamientos pueden descubrirse como estructuras que surgen en relación con la cantidad de información especificada para la realización de una tarea. Más precisamente, la noción de representación debe concebirse como surgiendo a partir de la interacción entre sistema nervioso, cuerpo y mundo, pero refiriéndose a una funcionalidad distinta de este operar más reactivo. Si al sistema nervioso se lo fuerza a funcionar en un mundo donde la interacción off-line es importante para la supervivencia, éste debe recurrir a su propia arquitectura de base con este fin (como sucede con las señales de propiocepción estudiadas en el capítulo segundo). Dado que la base de esta representación se halla en un nivel senso-motor más básico, la acción en el mundo es la que le da su significado a la representación. Esto es, el actuar con un cuerpo determinado en un ambiente determinado es lo que hace que el agente tenga una percepción orientada a la acción y que los casos que esta doble determinación establece sean los que son representados. En este sentido, el significado de una representación depende en cierta medida del mundo, pero esta medida está determinada por el cuerpo que posee el agente, pues la interacción del sistema nervioso con el cuerpo en un mundo determina qué es útil que el agente represente para sobrevivir y reproducirse.

III. Vindicación del Rol Explicativo de la Representación

Pese a los numerosos ejemplos de los proponentes de los enfoques encarnados que parecen mostrar que la representación no juega ningún rol en la

explicación cognitiva, la carga de la prueba no necesariamente está del lado de los proponentes de una “encarnación simple”. Y esto precisamente por la noción de representación. De acuerdo a Clark (1997b: 463 y ss.), todas las nociones sustantivas de representación tienen en su núcleo alguna idea de estados internos (o procesos) cuyo rol funcional real es “estar dentro” por otro objeto, evento o estado de cosas (usualmente extra-neural). Los sistemas que usan representaciones son aquellos que logran alguna clase de coordinación con los rasgos ambientales mediante el método especial de tener alguna otra cosa (en vez de una señal directamente recibida del ambiente) que “está dentro” y guía el comportamiento en su lugar. Para que un sistema cognitivo que utilice representaciones sea interesante cognitivamente, éste debe presentar la capacidad de acceder a tales estructuras internas off-line y así sustentar el planeamiento y la resolución de problemas en ausencia de ricos intercambios ambientales efectivos. No importa el mecanismo por el cual dichos sistemas ocupen tales representaciones, sino sólo que cumplan tres rasgos: (1) deben implicar estados internos o procesos cuyo rol funcional es coordinar la actividad del sistema con el mundo (no meras correlaciones), (2) debemos ser capaces de identificar estados o procesos internos específicos con roles representacionales específicos, esto es, aislar parámetros o procesos internos que “están dentro” con estados de cosas extra-neurales particulares y (3) el sistema debe ser capaz de usar estos estados o procesos internos para resolver problemas off-line, para comprometerse en exploraciones vicarias de ciertos dominios, etc.

Una noción tal parece ser central para la realización de tareas cognitivas al menos en dos casos: casos que suponen el razonamiento acerca de estados de cosas ausentes, no-existentes o contrafácticos y casos que suponen una sensibilidad selectiva a estados de cosas, cuyas manifestaciones físicas son complejas y no regladas (Clark 1997a: 167 y ss.).

La primera clase de casos incluye pensamientos acerca de eventos temporal o espacialmente distantes y pensamientos acerca de los resultados potenciales de acciones imaginadas. En tales casos, es difícil evitar la conclusión de que un razonamiento exitoso entraña crear alguna clase de “estar en” previa, que pueda identificarse con el fenómeno ausente –substitutos internos que hacen posible una coordinación de comportamiento apropiada sin la guía provista por el input externo constante–.

La segunda clase de casos es también familiar, aunque más difícil de describir. Ésta se refiere a ocasiones en que el sistema cognitivo debe responder selectivamente a estados de cosas cuyas manifestaciones físicas son muy erráticas, es decir, estados de cosas que se unifican a un nivel bastante abstracto, pero cuyos correlatos físicos tienen poco en común. Un ejemplo de esto sería la habilidad de escoger todos los ítems valiosos en una pieza. Es muy difícil ver cómo se hace razonar a un sistema sobre esas cosas sin que todos los inputs diversos, superficialmente diferentes, sean asimilados primero a un estado o proceso interior común, de tal modo que un procesamiento ulterior pueda definirse sobre el correlato interno: un ítem interno, un patrón o procesos cuyos contenidos corresponden luego a la propiedad abstracta. En tales casos, el éxito del comportamiento parece depender de nuestra habilidad para comprimir o dilatar un espacio sensorial de input. El agente exitoso debe aprender a tratar los inputs cuyas

codificaciones tempranas (en las periferias sensoriales) son muy diferentes como para incluirse en la misma clasificación o, al revés, para tratar inputs cuyas codificaciones tempranas son muy similares como para requerir diferentes clasificaciones. Los estados internos desarrollados para tales fines son representaciones internas, cuyos contenidos conciernen a los estados de cosas requeridos en cuestión.

En ambos casos (el ausente y el no reglado) el rasgo en común es la necesidad de generar un estado interno adicional cuyo rol adaptativo de procesamiento de información es guiar el comportamiento pese a la hostilidad de las señales ambientales (o bien no hay o bien requieren una computación importante para producir guías de acción útiles). En estos casos “hambrientos de representación” el sistema debe, al parecer, crear alguna suerte de ítem interno, patrón o proceso cuyo rol es “estar” por el estado de las cosas elusivo. Éstos, entonces, son los casos en que puede esperarse encontrar representaciones internas. Para poder explicar la cognición sin representación, se requeriría entonces una demostración de que, incluso en los “casos hambrientos de representación”, no se requieren representaciones y pueden explicarse con herramientas tales como las de los sistemas dinámicos. Y hasta el momento tanto los enfoques dinámicos como los de los agentes situados parecen estar muy lejos de explicar tales casos.

Tomando en cuenta que los enfoques encarnados sí parecen explicar fenómenos cognitivos no “hambrientos de representación”, Clark (1997a: 174) desarrolla la noción de “representacionalismo mínimo”, entendiendo por esto las condiciones mínimas bajo las cuales el habla de representaciones internas es útil. Tales condiciones se obtienen cuandoquiera que, una vez descontada la función explicativa de la red causal compleja de influencia, se revela el rol adaptativo procesador de información de algún sistema de estados o procesos –un sistema que puede envolver tanta distribución espacial o complejidad temporal como es compatible con identificar exitosamente las configuraciones físicas que se dan por estados de cosas específicos–. Esto supone, en un mínimo, el uso de un estado interior que permita al agente iniciar y organizar su comportamiento sin un input ambiental inmediato o para anticipar inputs ambientales futuros, etc. Agentes puramente reactivos no satisfacen esta condición de racionalidad mínima.

Los enfoques “encarnados” destacan una noción de inteligencia como un resultado integrado de la percepción, la cognición y la acción, en la que la percepción misma está a menudo imbricada con las posibilidades para la acción y está influenciada continuamente por factores cognitivos, contextuales y motores. En esta visión no se necesita producir un modelo interior rico, detallado y neutral respecto de la acción que espera los servicios de la “cognición central” para deducir las acciones apropiadas. Las antiguas distinciones entre percepción, cognición y acción pueden a veces oscurecer antes que iluminar el flujo verdadero de los eventos. En cierto sentido, el cerebro se revela entonces no como primeramente una máquina de la razón, sino como un órgano del *control ambientalmente situado* (Clark 2001: 95). Aquí, la idea es que inteligencia no depende de la traducción de información que llega a un código unitario que es luego utilizado por la inferencia lógica de propósito general, sino que nos

enfrentamos con un “bolsa de trucos”, donde se mezclan códigos relativos a propósitos especiales y estratagemas cuyo efecto general es el de apoyar las necesidades particulares de ciertas clases de criaturas que ocupan un nicho ambiental específico (Clark 2001: 100).

IV. Conclusión

Pese a que la crítica de la encarnación “radical” no logra vulnerar la noción de representación, puesto que hay casos como los “hambrientos de representación” que requieren de una noción que está en lugar de otra cosa para su explicación, sí indican que esta noción no puede concebirse como un mero símbolo en la cabeza, “neutral” respecto de la acción y que representa al mundo “tal como es”. Las condiciones bajo las cuales la noción de representación es necesaria ya no dicen relación con la noción de una mente aislada tratando de hacerse de una imagen de mundo. De ahí que la propuesta de Clark (1997a: 149 y ss.) sea pensar las representaciones sobre una base corpo-ambiental, lo que determina que, si el cerebro biológico opera con algo que pueda describirse útilmente como una “representación interna”, una gran cantidad de dichas representaciones serán locales y orientadas a la acción, antes que objetivas e independientes de la acción. Las clases de representaciones más orientadas a la acción pueden suponerse razonablemente como las clases evolutivas y de desarrollo más básicas (como las representaciones que podríamos suponer en METATOTO). Y puede incluso ser el caso que la gran mayoría de la resolución de problemas y la acción cotidianas rápida y fluida dependan de ellas²¹. Para Clark (Ibid.) las representaciones orientadas a la acción pueden definirse como aquellas que describen el mundo en términos de acciones posibles. Los productos de la actividad perceptual, entonces, no son descripciones neutrales respecto de la realidad externa, sino que pueden en cambio constituir recetas directas para actuar e intervenir en tal realidad.

Es aquí donde cobran relevancia los postulados de los enfoques alternativos. De un operar cognitivo automático y no representacional (como el del sapo de Maturana y Varela) se debe pasar a un operar cognitivo que requiere recoger información de su ambiente para tomar decisiones o resolver problemas, esto es, operar con representaciones. La materia que opera con representaciones es la misma que opera “automáticamente” y, en un primer paso hacia un comportamiento representacional, parece plausible adscribir una suerte de “hipótesis de continuidad” o criterio de arquitectura de incrementación. Esto supone que en un nivel superior la misma arquitectura neural que sustenta, por ejemplo, el comportamiento del sapo, debe desplazarse en el sentido de sustentar el operar cognitivo en situaciones off-line. Los planteamientos evolutivos de Brooks en la construcción de sus criaturas y la construcción efectiva de un robot que piensa acerca de sus desplazamientos (METATOTO) parecen apoyar la hipótesis de que, al menos en un nivel básico, el uso de circuitos orientados a la acción con el mundo puedan dar cuenta de un nivel más alejado de este actuar. La hipótesis dinámica parece señalar a su vez una condición metodológica para el apareamiento de tales estructuras: si el comportamiento no puede ser explicado en términos de la interacción entre sistema nervioso, cuerpo y mundo, dado que la

²¹ Este punto se condice con la observación de Agre, según la cual la mayoría de la actividad cotidiana es rutinaria, en el sentido de que requiere poco-si-algo de nuevos razonamientos abstractos y de que la mayoría de las tareas, una vez aprendidas, pueden ser logradas de una manera rutinaria, con pocas variaciones.

interacción es menor o no hay, es necesaria otra clase de explicación o, incluso, si se quiere, de un paso evolutivo que de cuenta de los procesos cognitivos cuando la interacción con el mundo está suspendida. El enfoque *enactivo* de Varela et al. puede considerarse en el sentido de cómo dar el paso de una acción totalmente determinada por el acoplamiento entre sistema nervioso, cuerpo, mundo a una acción donde la interacción sea menor. La percepción como acción guiada perceptualmente permite entender que las estructuras cognitivas emergen de los patrones senso-motores recurrentes, que permiten que a su vez la acción sea guiada perceptualmente. Un paso más allá hacia la comprensión de cómo surgen las representaciones estaría dado por los postulados de Lakoff y Johnson, quienes explican que estos patrones senso-motores recurrentes, que se dan en la acción guiada perceptualmente, posibilitan, a su debido tiempo, dar cuenta de la experiencia subjetiva, vía proyección del dominio senso-motor al dominio subjetivo.

Esto puede ayudarnos entonces a entender la base sobre la cual tiene que constituirse la noción de representación. Esta base determina también que la “representación” no puede entenderse sino como un producto de nuestro actuar biológicamente plausible, lo que supone en su estructura un fuerte componente que la vincula con la interacción con el cuerpo y con el mundo. De este modo, la representación tiene un origen en la acción (o en el comportamiento, al decir de Brooks), no es, por consiguiente, “neutral” respecto de la acción y de lo que representa del mundo, pues representa un mundo en tanto nuestro sistema nervioso y nuestro cuerpo puedan actuar en él²².

²² Una posibilidad que quedará inexplorada es la de la función de los ambientes culturales en la conformación de nuestra racionalidad (al respecto ver Clark 1997a, 2001a).

CONCLUSIÓN

HACIA UNA IDEA DEL ORIGEN DE LA REPRESENTACIÓN Y SU ROL EXPLICATIVO

La idea de von Uexkull acerca del *Umwelt* –literalmente, ambiente–, se refiere a un conjunto de características ambientales respecto de las cuales está sensibilizado un animal específico. Esta idea se refiere a que el ambiente efectivo se define por los parámetros que son relevantes para un animal con un estilo de vida específico. La cognición biológica es muy selectiva, y puede sensibilizar a un organismo respecto de cualquier parámetro (a menudo simple) que confiablemente especifica estados de cosas que importan para la forma de vida específica. De este modo, una garrapata que se guía en su actuar por la detección de ácido bórico en la piel de un mamífero actúa de manera similar al robot Herbert, que se guía por un detector de latas (Clark 1997a: 24-25; cf. Brooks 1998: 400). De este modo, cada especie animal y robot tiene su propio *Merkwelt* –algo así como “mundo de percepciones”–. La idea que se sigue de esto, naturalmente, es la de cómo el mundo percibido está similarmente predisposto y constreñido por las propias características del animal.

Una idea algo similar es la de Gibson (en Varela et al. 1991: 203 y ss.), según la cual el que percibe es concebido como un agente activo, que usa su propio movimiento para recolectar información sobre el ambiente. No todos los organismos recogen la misma información del ambiente, sino que más bien *resuenan* con la información que está coordinada con su propio potencial de acción. La acción entonces está perceptualmente guiada, de manera que, dependiendo del animal, en el ambiente se encuentran ciertas propiedades que no se encuentran en el mundo físico *per se*. Las propiedades más importantes consisten en aquellas que se le ofrecen al animal, que Gibson llama *affordances*. En términos precisos, dichas *affordances* consisten en las oportunidades para la interacción que poseen las cosas en el ambiente con relación a las capacidades senso-motoras del animal. Así, dependiendo de ciertos animales, algunas cosas, así como los árboles, son “trepables” o vale la pena treparlos. De este modo, las *affordances* son características ecológicas distintivas del mundo²³, que son percibidas si el animal está sensibilizado a ellos por su accionar.

En otros términos, una concepción similar puede descubrirse en Clark y Thornton (1997), quienes investigan en redes conexionistas cómo ciertos datos de input, que tienen una regularidad marginal, deben ser re-configurados de manera que rutinas primitivas de aprendizaje computacional

²³ Gibson además ofrece una teoría de la percepción para explicar cómo se percibe el ambiente, para más detalles ver Varela et al. (1991: 203).

puedan encontrar un mapeo adecuado, de manera de poder dar sentido a dichos datos de input. Estas rutinas deben tratar las propiedades relacionales de inputs como definiendo características nuevas, de un orden superior. En este caso, es improbable que cualquier aprendizaje no informado (uno que no recibe algún impulso extra o presión para permitirle escoger la re-codificación adecuada a partir de una infinidad de re-codificaciones posibles) distinga la regularidad.

Pese a que el objetivo de Clark y Thornton en este artículo es abogar por el lenguaje como aquel impulso que permite la re-codificación de inputs que presentan regularidades marginales, teniendo en cuenta afirmaciones como las de von Uexkull y de Gibson, podemos entender que los seres vivos tienen ya un potencial decodificador primario. Así, la garrapata decodifica de entre todos los inputs posibles la presencia de ácido bórico y, en robótica situada, Herbert identifica en su mundo las latas. Ambas decodificaciones permiten desencadenar la acción. Tales respuestas simples, pero consistentes, a inputs marginales, son las que estudian las visiones “encarnadas” de la cognición revisadas en esta tesis. De este modo, la robótica situada cifra el actuar de robots inteligentes en comportamientos que se basan en rasgos simples del mundo; la teoría de sistemas dinámicos destaca casos en que los comportamientos pueden entenderse como un producto del desarrollo temporal de la interacción entre sistema nervioso, cuerpo y mundo; la postura *enactiva* entiende que las estructuras cognitivas emergen de patrones senso-motores recurrentes que ayudan a una unidad a mantener su homeostasis organizacional. De alguna manera, la naturaleza misma (esto es, la evolución) se hace cargo de codificar estos inputs marginales para dar origen a una acción adaptativa robusta por parte de un organismo. Hasta este punto, no se necesita una noción como la de representación para dar cuenta del comportamiento inteligente.

Ahora bien, como hemos venido discutiendo (ver especialmente capítulo anterior), no toda la cognición puede explicarse sobre estas bases. Un agente cognitivo humano exhibe ciertos comportamientos que no son explicables por un acoplamiento entre sistema nervioso, cuerpo, mundo, como el resolver problemas abstractos o razonar contrafácticamente, por ejemplo. Retomando la noción de estos inputs cuyas regularidades son marginales y la solución biológicamente económica de responder encontrando ciertas “pistas” a las que reaccionar, se puede considerar que, una vez codificados ciertos inputs como siendo de tal clase, se pueden codificar nuevos inputs cuya regularidad es marginal, pero operando sobre la base decodificadora primera. La necesidad de codificar nuevos inputs debe pensarse como producto de constreñimientos biológicos específicos –en términos de Maturana y Varela, nuevas perturbaciones a las que el sistema nervioso debe reaccionar si no quiere desintegrarse–. Dentro de estos constreñimientos podrían contarse, por ejemplo, la vida en un ambiente hostil, donde no es fácil encontrar comida, la imposibilidad física

de arrancar corriendo de un depredador veloz, el largo período de gestación de los humanos, el largo período que le toma a un niño convertirse en adulto, etc.

Frente a tales constreñimientos, puede pensarse que el organismo, echando mano a los recursos que ya tenía, generó una sensibilidad a nuevas regularidades marginales presentes en los inputs sobre su sistema nervioso. La misma arquitectura (digamos, neural) diseñada para “olfatear” regularidades más básicas se replica, de manera de pensar otras regularidades marginales o regularidades de regularidades (aún más marginales) o regularidades con las que el organismo no está en ese preciso momento involucrado. Esto es, requerimientos biológicos pueden haber dado origen a nuevos conjuntos de “circuitos” que proveyeran, entre otras cosas, una exploración vicaria a partir de ambientes ya conocidos con herramientas similares a las que se emplean cuando el organismo está comprometido con la acción real (al modo de METATOTO). Considerando la vieja idea de que la ontogenia recapitula la filogenia, podemos entender que el desarrollo en el niño de estructuras subjetivas cada vez más alejadas de su origen senso-motor recapitula, precisamente, este paso evolutivo.

De cualquier forma, en la cognición efectiva nos encontramos con que nuestras representaciones se originan a partir de la misma arquitectura que permite nuestros movimientos más básicos: la misma arquitectura neural que nos ayuda a coger un objeto es la que soporta nuestro pensamiento deliberativo (cf. Lakoff y Johnson 1999). La *enacción* de Varela et al. (1991), que considera que las estructuras cognitivas emergen de patrones senso-motores recurrentes que permiten a su vez que la acción esté guiada perceptualmente y, en un nivel superior, las ideas de Lakoff y Johnson respecto de que los procesos cognitivos subjetivos son proyecciones metafóricas de un dominio senso-motor, nos dan una base para entender las representaciones como una elaboración a un nivel más abstracto de las arquitecturas neurales que permiten actuar en el mundo percibiendo “pistas” o “señales”, esto es, regularidades marginales en los inputs, a las que nuestro cuerpo y nuestro sistema nervioso pueden responder.

En esta visión, entonces, hay un procesamiento progresivo de los inputs cuya regularidad es marginal, en el que un sistema nervioso se sensibiliza progresivamente a nuevas regularidades marginales que cada vez se alejan más del ambiente. Así, habría una sensibilidad primera a inputs muy básicos (como de ácido bórico), que generarían respuestas automáticas. Seres más avanzados responderían a diversos inputs marginales (calor, frío más ácido bórico, por ejemplo), seguirían exhibiendo respuestas automáticas y posiblemente una arquitectura de incrementación, cuyos circuitos no poseen una diferenciación funcional. Dados los constreñimientos ecológicos necesarios, estos circuitos se replicarían, pero con una función algo distinta: captar las regularidades marginales

de la información ya codificada por circuitos más básicos, lo que permitiría respuestas menos “automáticas”. Eventualmente, como en METATOTO o como en las señales de propiocepción, se daría origen así a algo funcionalmente similar a una representación interna que respondiera por un objeto externo, que se relaciona probablemente con muchas codificaciones anteriores de sus inputs marginales (así, se identifica una mesa amarilla, pequeña, sólida, etc.). Un alejamiento continuo de la base senso-motora, esto es, “circuitos” que se alejan continuamente de la interacción directa con el mundo y operan en actividades off-line, permitiría, a su debido tiempo, pensar en eventos que no tienen una relación directa con nuestra encarnación biológica o nuestra situación ambiental.

De ahí que la representación sea una noción que se condice con nuestra “encarnación” y con que estemos situado en un mundo específico: es nuestra encarnación la que nos ha sensibilizado a ciertos aspectos del mundo y no a otros. Y de ahí también que no sea “neutral” respecto de la acción, sino más bien se oriente a la realización de tareas específicas: la representación surge de procesamientos senso-motores básicos y responde a la necesidad de percibir ciertas regularidades en los inputs ambientales que nos permiten actuar en el ambiente.

En esta noción de representación vuelve a jugar un rol la noción de información: es a partir de la distinción de regularidades en los inputs, esto es, de recoger cierta información presente en los inputs, que puede irse conformando una noción interna que esté por algo ausente. Ahora bien, este recoger cierta información se hace por parte de un organismo sensible a cierto tipo determinado de información –por la historia evolutiva, por la historia ontogenética o incluso por la educación–. Y como vimos, la representación tiene lugar a partir de una historia informacional de este tipo: a partir de lo que capta nuestro sistema nervioso se va haciendo posible alejarse cada vez más de la acción directa con el mundo, recodificando la información ya presente y desentrañando nueva información a partir de la información ya recodificada.

El contenido de las representaciones responde también a la codificación de inputs y a la posibilidad de captar las regularidades marginales entre ellos. Como se expuso en el capítulo anterior, dado que la base representacional se halla en el nivel senso-motor, la acción en el mundo, que se guía por esta capacidad de captar regularidades marginales en el ambiente, es la que le da su significado a la representación. Las cosas que son representadas en primera instancia por el agente son aquellas que la propia encarnación dentro de un ambiente específico le muestra como útil para la supervivencia, y como útil para razonar acerca de ellas incluso cuando no estén presentes (para pensar en cómo obtenerlas, o en medios más efectivos de cazarlas, por ejemplo). Así, el significado de una representación depende del mundo, pero está sub-determinado por éste, en el sentido que es la encarnación finalmente la que decide qué es o no es útil representarse del mundo y cómo, esto es,

es la propia encarnación la que codifica los inputs ambientales y es capaz de encontrar regularidades en ellos.

Esta noción de representación se aleja de la noción de representación de la ciencia cognitiva tradicional, pues, aunque juega un rol cognitivo, no se concibe en primera instancia como un símbolo en un lenguaje interno que se relaciona con los otros símbolos en virtud de su sintaxis. Tal como en el conexionismo, lo que determina la existencia misma de la representación y su continuidad es su uso al tratar de realizar una tarea específica. A diferencia del conexionismo, el uso y la continuidad no se ven como operaciones que se realizan preferentemente dentro de la caja craneana, sino que esta noción de representación toma en serio las ideas de “encarnación” y de “situacionalidad”. Como vimos, el sistema nervioso se sensibiliza progresivamente a inputs marginales del mundo que le presenta en primera instancia el cuerpo. La experiencia de un ambiente por parte de un cuerpo, tanto evolutiva como ontogenética, es la que determina qué es representado de tal ambiente y cómo es representado.

En este enfoque, entonces, la representación vuelve a jugar un rol cognitivo. Dicho rol cognitivo se constituye a partir de una operación biológica básica, a saber, la captación de regularidades marginales en los inputs sensoriales. Tal visión nos permite dar cuenta de los hechos subrayados por las visiones encarnadas de la cognición, esto es, de la importancia del cuerpo y del ambiente para la resolución de diversas tareas cognitivas. Ahora bien, queda aún por explicar cómo se de precisamente la diferenciación funcional que permita que los mismos “circuitos” neurales se comprometan en operaciones que no tienen que ver con el operar directo de un cuerpo en el mundo y cómo este proceso básicamente biológico se complementa con otro hecho singular de la cognición humana: el medio cultural en el que los seres humanos nos hallamos también anidados.

REFERENCIAS BIBLIOGRÁFICAS

- Adams, Bryan, Cynthia Breazel, Rodney A. Brooks et Brian Scassellatti (2000) “Humanoid Robots: A New Kind of Tool”. En: *IEEE Intelligent Systems*.
- Bechtel, William, Adele Abrahamsen et George Graham (1999) “The Life of Cognitive Science”. En: *A Companion to Cognitive Science*, W. Bechtel et G. Graham, eds., Massachussets- Oxford, Blackwell Publishers, pp. 1-104.
- Beer, Randall D. (2004) “Autopoiesis and Cognition in the Game of Life”. En: *Artificial Life*, Vol. 10, No. 3, pp. 309-326.
— (2003) “Arches and Stones in Cognitive Architecture. A Response to Commentary”. En: *Adaptive Behavior*, Vol. 11, No. 4, pp. 299-305.
— (2000) “Dynamical Approaches to Cognitive Sciences”. En: *Trends in Cognitive Science*, Vol. 4, No. 3, pp. 91-99.
- Beer, Randall D., Roger D. Quinn, Hillel J. Chiel et Roy E. Ritzmann (1997) “Biologically Inspired Approaches to Robotics. What can we learn from Insects?”. En: *ACM*, Vol. 40, No. 3, pp. 31-38
- Brooks, Rodney A. (2000) “Intelligence without representation”. En: (2000) *Mind Readings. Introductory Selections on Cognitive Science*, P. Thagard, ed., Cambridge, Massachusetts, MIT Press, pp. 395-420.
— (1997) “From Earwigs to Humans”. En: *Robotics and Autonomous Systems*, Vol. 20, Nos. 2–4, pp. 291–304.
- Brooks, Rodney A., Cynthia Breazeal, Matthew Marjanovic, Brian Scassellatti, Matthew M. Williamson (1998) “The Cog Project: Building a Humanoid Robot”. En: *Computation for Metaphors, Analogy and Agents*, C. Nehaniv, ed., Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag.
- Chiel, Hillel J. et Randall D. Beer (1997) “The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment”. En: *Trends in Neuroscience*, Vol. 20, No. 12, pp. 553-557.
Clark, Andy (2001a) *Midware: An introduction to the Philosophy of Cognitive Science*, New York-Oxford, Oxford University Press.
— (2001b) “Reasons, Robots and the Extended Mind (Rationality for the New Millenium)”. En: *Mind and Language*, Vol. 16, No.2, pp. 121-145.

- (2000) “The Presence of a Symbol”. En: *Mind Readings. Introductory Selections on Cognitive Science*, Paul Thagard, ed. , Cambridge, Massachusetts, MIT Press, pp. 377-393.
- (1999a) “Embodied, situated, and distributed cognition”. En: *A Companion to Cognitive Science*, W. Bechtel et G. Graham, eds., Massachussets- Oxford, Blackwell Publishers, pp. 506-517.
- (1999b) “An embodied cognitive science?”. En: *Trends in Cognitive Science*, Vol. 3, No. 9, pp. 345-351.
- (1997a) *Being There. Putting Brain, Body, and World Together Again*. Cambridge, Massachussets, MIT Press.
- (1997b) “The Dynamical Challenge”. En: *Cognitive Science*, Vol. 21, No. 4, pp. 461-481.
- (1995) “Moving Minds: Situating Content in the Service of Real-Time Success”. En: *Philosophical Perspectives 9, AI, Connectionism, and Philosophical Psychology*, pp. 89-104.
- (s.f) “Memento’s Revenge: The Extended Mind, Extended”. To appear in R. Menary, ed., *Papers on the Extended Mind*.
- (in press) “Beyond the Flesh: Some Lessons from a Mole Cricket”. En: *Artificial Life*.
- Clark, Andy et David J. Chalmers (1998) “The Extended Mind”. En: *Analysis*, Vol. 58, No. 1, pp. 7-19.
 - Clark, Andy et Chris Eliasmith (2002) “Philosophical Issues in Brain Theory and Connectionism” (Revised Version), (entry for) *Handbook of Brain Theory and Neural Networks*, M. Arbib, ed., MIT Press.
 - Clark, Andy et Chris Thornton (1997) “Trading Spaces: Connectionism and the Limits of Uninformed Learning”. En: *Behavioral and Brain Sciences*, Vol. 20, No.1, pp.57-67
 - Eliasmith, Chris (1997) “The Third Contender: A Critical Examination of the Dynamicist Theory of Cognition”. En: *Mind Design II. Philosophy, Psychology, Artificial Intelligence*, J. Haugeland, ed., Cambridge, Massachusetts, MIT Press, pp. 303-333.
 - Elman, Jeffrey (1999) “Connectionism, artificial life, an dynamical systems”. En: *A Companion to Cognitive Science*, W. Bechtel et G. Graham, eds., Massachussets - Oxford, Blackwell Publishers.
 - Fodor, Jerry A. (1998) *Concepts: Where Cognitive Science went Wrong*, Clarendon Press, Oxford.
 - Haugeland, John (1988) *La inteligencia artificial*, México, Siglo Veintiuno Editores.

- (1997) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*, Cambridge, Massachusetts, MIT Press.
- Lakoff, George et Mark Johnson (1999) *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*, New York, Basic Books.
 - Mackworth, Alan (1997) “On seeing robots”. En: *Mind Design II. Philosophy, Psychology, Artificial Intelligence*, J. Haugeland, ed., Cambridge, Massachusetts, MIT Press, pp. 271-286.
 - Mataric, Maja J. (2002) “Situated Robotics”. En: *Encyclopedia of Cognitive Science*, Nature Publishers Group, Macmillan Reference Ltda.
 - Maturana, Humberto et Francisco Varela (1980) *Autopoiesis and Cognition. The Realization of the Living*, Dordrecht, Holland, D. Reidel Publishing Company.
 - (1984) *El Árbol del Conocimiento. Las Bases Biológicas del Entendimiento Humano*, Santiago de Chile, Editorial Universitaria.
 - Port, Robert F. (2000) “Dynamical Systems Hypothesis in Cognitive Science” (draft entry for) *Encyclopedia of Cognitive Science*, Amy Lockyer, Assoc. Ed, London, Macmillan Reference Ltda.
 - Rumelhart, David E. (1997) “The Architecture of the Mind: A Connectionist Approach”. En: *Mind Design II. Philosophy, Psychology, Artificial Intelligence*, J. Haugeland, ed., Cambridge, Massachusetts, MIT Press, pp. 207-238.
 - Smolensky (1998) “Connectionist Modeling: Neural Computation / Mental Connections”. En: *Mind Readings. Introductory Selections on Cognitive Science*, Paul Thagard, ed., Cambridge, Massachusetts, MIT Press, pp. 233-250.
 - Stufflebeam, Robert S. (1999) “Representation and Computation”. En: *A Companion to Cognitive Science*, W. Bechtel et G. Graham, eds., Massachusetts - Oxford, Blackwell Publishers, pp. 636-648.
 - Thagard, Paul (ed.) (2000) *Mind Readings. Introductory Selections on Cognitive Science*, Cambridge, Massachusetts, MIT Press.
 - van Gelder, Timothy (1997) “The Dynamical Hypothesis in Cognitive Science”. En: *Behavioral and Brain Sciences*, Vol. 21, pp. 1-14.
 - (1999) “Dynamic Approaches to Cognition”. En: *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, pp. 243-45.
 - (2000) “Dynamics and Cognition”. En: *Mind Readings. Introductory Selections on Cognitive Science*, Paul Thagard, ed., Cambridge, Massachusetts, MIT Press, pp. 421-450.

- van Gelder, Timothy et Robert F. Port (1995) “It’s About Time: An Overview of the Dynamical Approach to Cognition”. En: *Mind as Motion: Explorations in the Dynamics of Cognition*, T. van Gelder et R. Port, eds., Cambridge, Massachusetts, MIT Press, pp. 1-44.
- Varela, Francisco, Evan Thompson et Eleanor Rosch (1991) *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge, The MIT Press.