

Versión final

TRABAJO DE INVESTIGACIÓN

para la suficiencia investigadora en el programa de doctorado
"Filosofía, Ciencia, Tecnología y Sociedad"

Autonomía e Información en Sistemas Cognitivos

*Una propuesta de naturalización
del fenómeno cognitivo*



.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.

.



Universidad Euskal Herriko
del País Vasco Unibertsitatea

<http://www.ehu.es/ias-research>

Departamento de Lógica y Filosofía de la Ciencia

Xabier E. Barandiaran Fernandez

bajo la dirección de
Alvaro Moreno Bergareche

2004

Autonomía e Información en Sistemas Cognitivos v.1.0

Copyright © 2004 Xabier E. Barandiaran, COPYLEFT:

Siempre y cuando se mantenga esta nota se otorga permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Free Software Foundation; sin secciones invariantes, sin cubierta frontal, sin cubierta posterior. Una copia de esta licencia puede encontrarse en:

<http://www.gnu.org/licenses/fdl.html>.

Una versión electrónica de este texto puede encontrarse en:

<http://barandiaran.net/textos/tesina/>

Agradecimientos

Mi más sincero agradecimiento a Álvaro Moreno por las intensas y continuas discusiones que han dado lugar a casi el conjunto del trabajo que aquí presento. Igualmente agradezco el apoyo de quienes han tenido que soportar el estrés y la tensión que han acompañado el parto y la expresión de muchas de las ideas contenidas en el presente trabajo: a Nuria, Aita, Ama, Xabier e Ibai (sin vosotros nada de lo que hay aquí hubiera podido coger forma).

Este trabajo hubiera sido imposible sin la comunidad del software libre que ha desarrollado las herramientas informáticas necesarias para su realización. Igualmente imprescindibles han sido los recursos a los que la cooperación comunitaria en el ciberespacio me ha permitido acceder. Mi agradecimiento pues a aquellos que han desarrollado la infraestructura de búsqueda de material científico, así como a la parte de la comunidad científica que ha decidido publicar libremente su producción inmaterial y, en general, a todos aquellos que cotidianamente construyen y mantienen el ciberespacio funcionando.

Agradezco también el apoyo económico de la beca de formación de jóvenes investigadores número BFI03371-AE del Gobierno Vasco así como la financiación de los proyectos de investigación 9/UPV 00003.230-13707/2001 de la Universidad del País Vasco – Euskal Herriko Unibertsitatea y el proyecto BFF2002-03294 financiado por del Ministerio de Ciencia y Tecnología y los fondos FEDER.



Índice de contenido

1 Introducción: de la epistemología naturalizada a la naturalización de la cognición.....	6
2 Los Conceptos de Autonomía e Información en la Fundamentación de las Ciencias Cognitivas.....	9
2.1 Información y representación en el funcionalismo cognitivista y evolucionista.....	9
2.2 Situaciones, Cuerpos y Ecuaciones Diferenciales.....	11
2.2.1 De mentes y programas a cuerpos y robots.....	11
2.2.2 Teoría de Sistemas Dinámicos: un pegamento conceptual para cerebro, cuerpo y entorno.....	15
2.3 Autonomía: donde biología y cognición se dan la mano.....	17
2.3.1 Autonomía: el origen del término.....	18
2.3.2 La autonomía básica como fuente de normatividad funcional.....	20
2.3.3 Características comunes de las perspectivas autónomas en ciencias cognitivas... ..	24
3 Autonomía e Información en el Sistema Nervioso: Una Perspectiva Naturalizada.....	27
3.1 El origen evolutivo del SN.....	28
3.1.1 El origen de la motilidad y la agencialidad.....	28
3.1.2 Tensión entre motilidad e incremento de tamaño.....	29
3.1.3 La evolución explora nuevas formas de la agencialidad.....	31
3.2 La organización autónoma del sistema nervioso	33
3.2.1 Desacoplamiento jerárquico y enraizamiento autónomo.....	33
3.2.2 Observables constitutivos del SN.....	38
Observables primarios (impulsos o potenciales de acción).....	39
La matriz de conectividad.....	41
Observables secundarios (neuromoduladores).....	41
3.2.3 Autoorganización y autonomía en el sistema nervioso.....	44
Procesos autoorganizados en el SN.....	46
La autonomía del SN como clausura operacional.....	50
La autonomía interactiva del SN.....	52
3.3 Causalidad informacional en procesos neuro-interactivos.....	55
3.3.1 Los observables como señales.....	56
3.3.2 ¿Cuándo las señales se convierten en información?.....	57
Hacia una noción naturalizada de información.....	58
Información en el SN.....	59
Estudio de Caso (aprendizaje instrumental en la Aplysia).....	63
4 Naturalizar la cognición.....	68
4.1 La emergencia de la cognición desde lo biológico.....	68
4.2 Hipótesis de la Corporización Biológica de la Cognición.....	69
4.3 La especificidad de la cognición.....	70
4.4 Sujeto, entorno y mundo desde la HCB.....	71
Referencias y bibliografía básica.....	75

1 INTRODUCCIÓN: DE LA EPISTEMOLOGÍA NATURALIZADA A LA NATURALIZACIÓN DE LA COGNICIÓN

El marco general en el que se inscribe este trabajo de investigación es el de la epistemología naturalizada (Quine 1969, Campbell 1974) que busca dar cuenta del conocimiento como fenómeno natural y por tanto analizable desde las categorías científicas que hoy operan bajo el nombre de ciencias cognitivas¹. Por naturalismo entendemos que la realidad es un todo unitario en el que no existen sustancias específicas a las que debemos recurrir *ad-hoc* para explicar una serie de fenómenos (como el cognitivo). Entendemos por naturalismo además la contención de la práctica filosófica en la ciencia natural, a saber que la reflexión filosófica ha de tomar su materia de reflexión del cuerpo científico y revertir sobre él su trabajo productivo. Por epistemología naturalizada entendemos, pues, la relación que se establece en un sujeto cognitivo natural entre su conocimiento y las interacciones que establece con su entorno estudiadas a través de las ciencias naturales que se encargan del estudio de los procesos cognitivos. En palabras del propio Quine:

“Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded experimentally controlled input —certain patterns of irradiation in assorted frequencies, for instance— and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and torrential output is a relation that we are propted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence” (Quine, 1969:260)

La naturaleza de esa relación entre inputs y outputs del sujeto cognitivo es el objeto de este trabajo. Pero creemos que la epistemología naturalista no puede dar por sentado el hecho cognitivo sino que ha de dar cuenta del surgimiento de *sujetos* cognitivos *en* la naturaleza y *desde* ella, a saber dar cuenta de las categorías cognitivas en relación a la organización biológica subyacente, a los mecanismos que la sustentan. Por ello el objetivo de este

¹ Como veremos más adelante no toda ciencia cognitiva es susceptible de ofrecer un fundamento naturalista para la epistemología. De hecho el objeto de este trabajo es precisamente el de redefinir los fundamentos de unas ciencias cognitivas capaces de realizar ésta tarea.

trabajo, como prefacio para una epistemología naturalizada propiamente dicha, es el de naturalizar el fenómeno mismo de la cognición, dar cuenta de su especificidad como fenómeno natural pero constitutivo de un dominio propio: el del conocimiento. Esta perspectiva contrasta con el paradigma heredado de la filosofía del lenguaje de corte lógico-positivista que, en forma de funcionalismo representacionista y *folk psychology*, ha construido una fundamentación de las ciencias cognitivas de carácter abstracto y desligada de los mecanismos biológicos subyacentes. Hoy en día, en cambio, está emergiendo una nueva línea de fundamentación de las ciencias cognitivas basada en la teoría de la complejidad, en neurociencias cognitivas y teoría de sistemas dinámicos. Situamos este trabajo de investigación en el marco de las nuevas oportunidades de fundamentación que abren estas nuevas disciplinas para las ciencias cognitivas (ahora sí, propiamente naturalizables).

Dos conceptos (y sus relaciones mutuas) guiarán este proyecto: los conceptos de información y autonomía. A pesar de que las tradiciones científicas y filosóficas de la autonomía y la información han caminado enfrentadas durante los últimos años creemos que una fusión conceptual entre ambas constituye un paso necesario para la naturalización de la cognición como un fenómeno específico del ámbito natural. En concreto el concepto de autonomía ofrece un fundamento naturalizado para la normatividad que, como bien señaló Kim supone uno de los principales obstáculos para la consecución de una epistemología naturalizada (Kim, 1988). El concepto de autonomía permite igualmente dar razón de la emergencia de un dominio fenomenológico nuevo, el de la cognición. La información, en cambio, como una forma causal diferente a los mecanismo de los que emerge se convierte en primitivo teórico específico y constitutivo del dominio cognitivo permitiendo demarcar su especificidad como sistema autónomo.

El trabajo que presentamos a continuación se divide en dos bloques fundamentales. En el primero analizamos los usos de los conceptos de información y autonomía en las ciencias cognitivas contemporáneas en relación a la especificidad de la cognición y su naturalización, con especial atención a las críticas de la nuevas ciencias cognitivas (situada, distribuida, corporizada y dinámica) al concepto de información y representación. En la segunda parte reconstruimos los conceptos de autonomía e información de forma naturalizada. Nuestra propuesta naturalizadora es la de proceder de

abajo arriba (*from the bottom-up*), de lo más simple a lo más complejo, dando cuenta de la emergencia de la cognición en la naturaleza. El resultado no será una imagen cerrada o acabada de la cognición (menos aún de la epistemología) sino la explicitación (de acuerdo al conocimiento científico disponible) del marco mínimo del que emergen los procesos cognitivos; es decir, las bases biológicas necesarias y suficientes que actúan como constricciones posibilitantes del fenómeno cognitivo. Nuestro objetivo es, por tanto, reformular los conceptos fundacionales de la cognición sobre bases naturalistas. La pregunta a la que intentaremos dar respuesta durante este trabajo es: ¿Cuál es la especificidad de la cognición como fenómeno natural pero a su vez diferenciado del resto de los fenómenos que lo sustentan y que lo rodean? O, dicho de otro modo, ¿Cómo se establece esa dicotomía fundacional de la cognición entre sujeto y objeto de conocimiento en una realidad natural continua? ¿Cómo surge la información como un tipo de causalidad diferenciada y propia de la cognición de un universo en el que no existía?

2 LOS CONCEPTOS DE AUTONOMÍA E INFORMACIÓN EN LA FUNDAMENTACIÓN DE LAS CIENCIAS COGNITIVAS

2.1 Información y representación en el funcionalismo cognitivista y evolucionista

El concepto de información (y otros relacionados como el de procesamiento, representación y computación) ha jugado un papel fundamental en la fundamentación y estudio de los fenómenos cognitivos, intencionales y mentales.

El desarrollo de la lógica formal, la teoría de la computación y finalmente el auge de los ordenadores ayudó a afianzar la idea de la posibilidad de un procesamiento mecánico que preservara el valor semántico de los símbolos o estados manipulados (computados) gracias al seguimiento de unas reglas de procesamiento (algoritmos) lógicamente estructuradas. Junto con ello se afianzó la metáfora del ordenador, según la cual la relación entre mente y cerebro es la misma que existe entre software (programas) y hardware (dispositivos físicos) en un ordenador.

En este paradigma de las ciencias cognitivas la especificidad de lo cognitivo ha venido de la mano del carácter representacional de los procesos internos y de su organización funcional en base a las relaciones input-output que determinan. La tesis de la realizabilidad múltiple (los estados funcionales y representacionales que caracterizan a la mente pueden realizarse en múltiples tipos de procesos subyacentes) establece una separación fundamental entre procesos cognitivos y biológicos quedando éstos últimos relegados a meros “detalles” de implementación.

No será hasta el resurgimiento del conexionismo en los años 80 (Rumelhart *et. al.* 1986) que empiece a tambalearse la idea de un procesamiento lineal de símbolos en base a reglas de transformación de tipo lógico-deductivo, dando pie a una concepción distribuida y paralela del procesamiento informacional en los procesos cognitivos. El éxito de los modelos conexionistas inspirados en la estructura de las redes neuronales

supuso un acercamiento a la biología y la “estructura” de implementación que caracteriza a la inteligencia natural. Los “detalles de implementación” introducían constricciones al procesamiento abstracto al tiempo que permitían explicar y reproducir fenómenos cognitivos antes ignorados o relegados a un segundo plano (reconocimiento de patrones, memoria asociativa, mantenimiento de la eficacia global ante daños locales, etc.). Si bien en este paradigma subsimbólico el funcionalismo pierde la transparencia semántica de estados representacionales localizados, el conexionismo no deja de ser una “revolución inacabada” (Clark 1997) ya que aún transformando las expresiones más radicales del funcionalismo representacionista mantiene el núcleo fundacional del mismo: la cognición es fundamentalmente un proceso computacional y representacional descorporalizado y desvinculado de su enraizamiento biológico. Si bien las estructuras biológicas pueden inspirar soluciones computacionales diversas, la cognición sigue siendo un fenómeno que sucede a una escala funcional y desmaterializada, marcada por el carácter representacional de los procesos funcionales. Ya sea de forma distribuida o secuencial, paralela o en serie, simbólica o subsimbólica, la mente es una máquina abstracta de procesamiento de información con valor semántico o representacional.

Para la perspectiva funcionalista el núcleo argumental que marca la especificidad del hecho cognitivo reside pues en el concepto de información semántica o representación. Encontramos dos grandes modelos acerca del origen y fundamentación de la semántica en los procesos informacionales:

1. El modelo de la covarianza o correlación causal sistemática entre estados mentales/neuronales y estados de cosas del mundo (Fodor 1987).
2. El modelo teleosemántico de fijación evolutiva de la función representacional de una estructura en base a su historia selectiva (Millikan 1984).

A pesar de las diferencias sustanciales entre ambos enfoques existen problemas comunes que resultan especialmente relevantes desde una perspectiva naturalizadora:

- a) El valor semántico no puede resolverse desde la perspectiva del sistema mismo (Bickhard 2000) sino que hace falta o bien recurrir a la historia evolutiva del sistema o bien a un observador exterior que asegure la relación representacional de las representaciones. Más concretamente 20

- b) ninguno de los modelos anteriormente mencionados puede dar cuenta satisfactoriamente de la detección y corrección de error por parte del sistema.
- c) En ambos modelos los factores organizativos (mecánicos o de implementación, en la jerga funcionalista) así como los interactivos no juegan ningún papel explicativo relevante. Es precisamente por esta ausencia de relevancia causal de la organización dinámica de los mecanismos cognitivos que el modelo teleosemántico sólo consigue relacionar la cognición y los mecanismos naturales sobre los que se sustenta de una forma débil a través del proceso selectivo-evolutivo de esos mecanismos.

Problemas derivados o relacionados con los anteriores son el “symbol grounding problem” (Harnad 1990), el “frame problem” (Pylyshyn 1987), o el problema de cómo derivar la semántica de la sintaxis (Searle 1980, 1997). En respuesta a éstos y otros problemas derivados del funcionalismo cognitivista surge un nuevo paradigma en el que convergen corporización y dinamicismo. Un nuevo paradigma que, si bien deja de lado el concepto de la información e incluso se enfrenta directamente con él, pondrá las bases para una nueva reformulación del mismo.

2.2 Situaciones, Cuerpos y Ecuaciones Diferenciales

A finales de los años 80 y comienzos de los 90 empiezan a surgir toda una serie de nuevos enfoques en ciencias cognitivas, enfoques que llegan de diversas disciplinas (robótica, psicología, filosofía, vida artificial, etc.) y que, bajo diversas etiquetas (*embodied cognitive science, situated robotics, autonomous robotics, dynamical approach*, etc.) comienzan a cuestionar algunos de los principios y métodos del funcionalismo representacionalista clásico y, sobre todo de los conceptos de información y representación.

2.2.1 De mentes y programas a cuerpos y robots

Uno de los pioneros de esta oleada de nuevos investigadores fue Rodney Brooks (y el grupo de ingeniería robótica del MIT). Según Brooks la IA funcionalista se ha centrado principalmente sobre el razonamiento abstracto (fácilmente sistematizable en base a reglas y a la lógica) sin preocuparse del origen, evolución y corporización de esa inteligencia. Si la evolución ha

trabajado durante 3500 millones años creando todas las formas complejas de adaptabilidad sensoriomotora básica pero sólo en los últimos segundos del reloj de la historia terrestre ha surgido el lenguaje, la lógica, el ajedrez y el razonamiento abstracto... ¿por qué no concentrarse primero en construir robots capaces de adaptarse a entornos reales y construir de abajo arriba formas cada vez más complejas de cognición? En palabras del propio Brooks:

“[evolution] suggests that problem solving behavior, language, expert knowledge and application, and reason, are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surrounding to a degree sufficient to achieve the necessary maintenance of life and reproduction.” (Brooks 1991: 140).

Ahora bien cuando nos enfrentamos al diseño de robots autónomos nos enfrentamos a la tarea de diseñar su arquitectura de control... y aquí es donde la robótica situada (como ha sido bautizada) entra en confrontación directa con los principios analíticos y sintéticos (criterios de diseño y funcionalidad) del paradigma funcionalista. La propuesta de Brooks (la arquitectura de subsunción) rompe con la linealidad percepción-planificación-acción y propone una alternativa: descomposición por actividad (y no por función) e interacción internivelada entre capas de acción. De acuerdo a Brooks la evolución ha operado incrementalmente asegurando primero niveles conductuales de interacción básicos y añadiendo (después) nuevas capas conductuales sobre ellos. Las capas conductuales (*behavioral layers*) actúan como circuitos que, cerrados a través de una interacción recursiva con el entorno, generan conductas estables. La conducta global observada es el resultado de la interacción interna y externa entre estas capas. La atribución de representaciones es un acto interpretativo del observador, pero ni debe ser un principio de diseño, ni resulta un factor causal en la producción de conducta cognitiva. En palabras del propio Brooks “el mundo es la mejor representación de sí mismo” y la conducta inteligente surge en interacción directa con él, no mediada a través de procesos representacionales explícitos internos, tal como postulaba el representacionalismo funcionalista (cuya aplicación robótica apenas había conseguido desarrollar robots capaces de las interacciones sensoriomotoras más básicas).

Al trabajo pionero de Brooks siguió un oleada de investigación y desarrollo en robótica que profundizó en la crítica al cognitivismo funcionalista dando lugar al campo de la robótica autónoma o situada (Maes 1991, Steels 1991, Smithers 1995, 1997, Pfeifer & Scheier 1999) que busca comprender los procesos cognitivos a través de la síntesis de conducta

cognitiva situada en el mundo real (no en entornos puramente simbólicos o entornos físicos simplificados) y en tiempo real (y no en forma de algoritmos atemporales y abstractos).

Resultados de la investigación en otras disciplinas se unen al trabajo en robótica. En el campo de la psicología del desarrollo destaca la teoría corporizada del desarrollo infantil de Thelen (Thelen & Smith 1994), así como su reconstrucción del test de Piaget *A y no B* (Thelen *et.al.* 2001). Estos autores concluyen que el desarrollo psicomotor infantil surge de una interacción entre cuerpo, mente y entorno en el que las constricciones materiales y temporales de la corporalidad material forman un continuo inseparable con los procesos internos formando un todo responsable del desarrollo cognitivo. También destacan los trabajos de percepción activa, en concreto el trabajo de Ballard (1991) sobre la extracción de patrones a través del movimiento sacádico de los ojos haciendo, una vez más, énfasis en la artificialidad del esquema percepción-planificación-acción que el cognitivismo clásico (incluso en su forma conexionista) proyecta sobre los fenómenos cognitivos.

Otra serie de trabajos de investigación, no menos importantes, vienen de la mano del *Artificial Life route to Artificial Intelligence*: es decir, la metodología que genera simulaciones computacionales biológicamente inspiradas de procesos autoorganizados que dan lugar a conducta cognitiva. Entre otros programas de investigación destaca lo que Beer denominó “neuroetología computacional” (Beer 1990, Cliff 1995), también conocido como robótica evolutiva (Husbands *et.al.* 1997, Nolfi & Floreano 2000, Beer 2003): el estudio de redes neuronales diseñadas por evolución artificial en simulaciones de agentes situados. El estudio del funcionamiento de las redes neuronales corporizadas en simulaciones supone un avance significativo frente al conexionismo descorporalizado ya que el funcionamiento de las redes neuronales en interacción recurrente con el entorno transforma drásticamente la naturaleza de los procesos neuronales. Este cambio cualitativo responde al hecho de que la red no es diseñada primero representacionalmente y posteriormente insertada en un dominio interactivo (como sucede con el conexionismo clásico) sino que es evolutivamente diseñada en interacción recurrente con su entorno². Los algoritmos genéticos

² Para evitar que el diseñador imponga sus prejuicios teóricos sobre las formas de organización neuronal y la arquitectura de control se usan métodos de evolución artificial (por lo que el diseño queda en manos de un proceso evolutivo automatizado en el que el programador no

utilizados para la evolución artificial operan navegando el espacio de configuraciones posibles hasta dar con los mecanismos neuronales necesarios para generar la conducta deseada (especificada esta en una función de optimización). El resultado se asemeja más a un bol de spaghetti causal (como lo define Wheeler —1996) que a una serie ordenada de representaciones sobre las que operan transformaciones lógicas. Una maraña causal que atraviesa cerebro, cuerpo y entorno pero que genera globalmente una conducta coherente, aunque difícilmente descomponible en subrutinas funcionalmente relevantes.

La lección fundamental de las perspectivas corporizadas y situadas en ciencias cognitivas puede recogerse bajo el concepto de *funcionalidad emergente* (Steels 1991, Clark 1997). La funcionalidad emergente comprende un aspecto interactivo y otro interno. El interno ya fue explorado por el conexionismo clásico y comprende la distribución funcional de los procesos a través de toda la red neuronal sin que esta funcionalidad pueda ser localizada, es decir, sin que a un componente aislado de la red se le pueda asignar un rol causal funcionalmente relevante. El aspecto interactivo supone una dificultad añadida a las tesis del funcionalismo clásico. Se trata del continuo causal recurrente que une al agente cognitivo con su mundo, un continuo que no permite una fácil demarcación entre mundo y representación (algo que ha empujado a muchos investigadores a abandonar las nociones de información y representación). La funcionalidad cognitiva es, por tanto, una propiedad holista del sistema que no aparece ya como un conjunto de procesos computacionales internos sino como el acoplamiento entre cerebro, cuerpo y mundo unidos por interacciones dinámicas recurrentes que no permiten aislar un momento representacional.

En definitiva las aportaciones irrenunciables de los enfoques corporizados y situados en ciencias cognitivas demuestran que el cuerpo y el entorno son fuentes de autoorganización de la dinámica neuronal y que *conjuntamente* generan la conducta cognitiva en tiempo real e interacción directa que observamos en la naturaleza. Pero... ¿cómo estudiar eficazmente

interviene). Se trata de recoger modelos neuronales del campo de la neurociencia y hacer evolucionar los parámetros de conectividad, las constantes de tiempo de caída del potencial de acción neuronal y otros posibles parámetros neurobiológicos (plasticidad, gas-nets, sistema dopaminérgico, etc), así como algunos parámetros corporales y sensomotores. Se sabe, además, que las redes neuronales recurrentes y dinámicas utilizadas en robótica evolutiva son capaces de, en principio, reproducir el comportamiento de cualquier sistema dinámico (Funahashi & Nakamura 1993): en este sentido son equivalentes en el dominio dinámico a lo que la Máquina Universal de Turing es en el ámbito de la computación.

el resultado emergente de ésta interacción? ¿Una vez abandonados algunos de los principios del cognitivismo clásico es posible aún mantener el marco conceptual computacional? La respuesta para muchos científicos y filósofos de la cognición es negativa. En su lugar proponen adaptar la teoría de sistemas dinámicos (TSD) como marco conceptual con el que dar sentido al conjunto de datos empíricos que surgen de la investigación para ir reconstruyendo un cuerpo teórico nuevo.

2.2.2 Teoría de Sistemas Dinámicos: un pegamento conceptual para cerebro, cuerpo y entorno

De un modo análogo a cómo el funcionalismo representacionista se valió del marco conceptual y metodológico computacional y de la metáfora del ordenador, las perspectivas situadas y corporizadas en ciencias cognitivas han optado generalmente por la TSD como marco conceptual y metodológico.

La TSD es utilizada de formas muy diferentes en este contexto (ver Port & van Gelder 1995 para una colección variada de usos de la TSD en ciencias cognitivas). Sin embargo la *Hipótesis Dinámica* (van Gelder 1998) es quizás la formulación más explícita y el punto de partida de un largo debate con el computacionalismo clásico. La Hipótesis Dinámica afirma que:

"for every kind of cognitive performance exhibited by a natural cognitive agent, there is some quantitative system instantiated by the agent at the highest relevant level of causal organization, so that performances of that kind are behaviors of that system; in addition, causal organization can and should be understood by producing dynamical models, using the theoretical resources of dynamics, and adopting a broadly dynamical perspective" (van Gelder 1998: 622).

Van Gelder desarrolla la HD en dos vertientes diferentes pero fuertemente interrelacionadas: a) la *hipótesis de la naturaleza* es una afirmación sobre la naturaleza dinámica de los agentes cognitivos (una hipótesis ontológica) y b) la *hipótesis del conocimiento* afirma que podemos y debemos entender la cognición de forma dinámica (una hipótesis epistemológico-metodológica). La hipótesis de la naturaleza defiende que los agentes cognitivos instancian una organización causal de tipo dinámico, pero esta organización causal, según van Gelder, está recogida por diferentes variables que no se encuentran, necesariamente, en el nivel mecanístico, es decir, en el nivel biológico operacionalmente manipulable por un observador o experimentador. Los ejemplos dados por van Gelder toman como variables dinámicas magnitudes macroscópicas tales como las preferencias y las acciones. La apuesta por variables tomadas a cualquier nivel de abstracción

resulta quizás el punto más conflictivo de la propuesta de van Gelder en relación a una fundamentación biológica de los procesos cognitivos e ignora una de las potencialidades más grandes de la perspectiva dinamicista: su capacidad de dar cuenta de cómo *emergen* esas variables primitivas abstractas de los mecanismos de nivel biológico y de las interacciones agente-entorno. Este es precisamente el dominio en el que las herramientas dinámicas son las más propicias para un análisis transnivélico. Como bien señala Bechtel: “[d]ynamical and mechanistic approaches are in fact allies” (Bechtel 1998: 629). De hecho el poder de los métodos de análisis dinámicos reside en su capacidad de modelar las interacciones corporizadas agente-entorno y los mecanismos biológicos cuyos componentes estructurales no poseen ninguna función computacional o representacional específica pero que, sin embargo, producen conducta cognitiva. Es decir, la virtud metodológica de la TSD reside en su capacidad de modelar la funcionalidad interna e interactivamente emergente que hemos descrito en la sección anterior.

En concreto uno de los aspectos más interesantes de los sistemas dinámicos (en oposición a los modelos computacionales del funcionalismo clásico) es que permite la emergencia de representaciones. De hecho el enfoque dinámico de la cognición sugiere que no hay sentido alguno en el que un sistema cognitivo sea funcionalmente computacional ya que no hay estados dinámicos del sistema que puedan ser ellos mismos representacionales. Por supuesto que muchos sistemas dinámicos pueden implementarse computacionalmente (como suele ser común en el cálculo numérico realizado en computadoras), pero un estado de la implementación en nada se parece a lo que pudiera ser considerado como un estado funcional-representacional. En todo caso la representación o la funcionalidad aparece como un proceso (una sucesión de estados dinámicos) en un contexto dinámico determinado. Precisamente lo que permite el dinamicismo es evitar tomar representaciones como primitivos teóricos para el modelado de sistemas cognitivos y generar en su lugar procesos emergentes autoorganizados cuya conducta puede categorizarse como cognitiva.

Podemos rescatar las siguientes características que hacen del dinamicismo un marco conceptual preferible al computacional *a la hora de naturalizar la cognición*:

1. Capacidad de modelado de mecanismos biológicos (permite sostener

metodológicamente la tesis de continuidad entre vida y mente)

2. No exige un compromiso con primitivos funcionales-computacionales de tipo representacional. Como consecuencia de ello permite el modelado de funcionalidad novedosa y de procesos de desarrollo.
3. Permite atravesar causalmente mente, cuerpo y mundo por igual (a diferencia del representacionalismo que introduce una separación, un dualismo causal entre representación y mundo) es por ello que la TSD es el marco conceptual más ajustado a los postulados teóricos de la perspectiva corporizada y situada en ciencias cognitivas.
4. Integra el tiempo como componente esencial de la cognición.



El problema del dinamicismo y de la perspectiva corporizada de la cognición es que no consiguen dar cuenta de la especificidad de lo cognitivo. El abandono de las nociones de representación e información como constitutivas del dominio cognitivo y su sustitución por los mecanismos causales dinámicos inespecíficos hace difícil encontrar un criterio que permita justificar la especificidad ontológica de la cognición.

Sin embargo este problema del dinamicismo (su inespecificidad causal en relación a *lo cognitivo* como hecho diferenciado) es el que va a permitir ensamblar las teorías cognitivas con la organización biológica y será precisamente la particularidad de ésta (la organización biológica) la que permita a su vez reconstruir el concepto de información cognitiva y, con ello, redefinir la especificidad de lo cognitivo.

2.3 Autonomía: donde biología y cognición se dan la mano

Cuando observamos la conducta de los seres vivos en la naturaleza, su capacidad de adaptación y supervivencia en entornos cambiantes, atribuimos cierto tipo de agencialidad al organismo; como si la causa última del movimiento, de la conducta, surgiera desde dentro. De hecho la producción de conducta cognitiva presenta una gran asimetría entre factores internos y externos de tal manera que, aunque influenciada por factores externos, la conducta se presenta como dirigida desde dentro del organismo o determinada por su estructura cognitivo-adaptativa. La percepción

internamente evaluada y activamente dirigida, la motivación, la evaluación de las interacciones y sus capacidades interactivas corporales le permiten al organismo definir un mundo de valores, de significaciones, de un entorno físico que resulta en principio neutro y carente de valores de respuesta para un observador absoluto. Pero si nos adentramos en la “interioridad” de los mecanismos de producción de conducta cognitiva en busca del origen de esa conducta no encontraremos un procesador central o un lugar específico al que atribuir voluntad o intencionalidad. Por el contrario encontraremos una red integrada de procesos (químicos, eléctricos, metabólicos, etc.) en constante actividad.

Decimos que los organismos son *autónomos* o presentan cierto grado de *autonomía* en la producción de conducta cognitiva pero no resulta fácil de explicitar o reconstruir ese concepto a partir de los mecanismos que constituyen al organismo como agente.

El término autonomía ha sido elaborado de múltiples maneras en biología, robótica y simulación de conducta adaptativa o ciencias cognitivas: desde noción fundamental para explicar la organización biológica (Varela, 1979) hasta un conjunto de constricciones de ingeniería sobre el diseño y testeo de robots (Maes, 1991).

2.3.1 Autonomía: el origen del término

El concepto de autonomía en ciencias cognitivas fue originalmente introducido por Francisco Varela como concepto fundacional de la biología en su obra “Principles of biological autonomy” (1979) heredado a su vez de los conceptos de clausura operacional y autopoiesis que él mismo, junto a Humberto Maturana, desarrollaron a principios de la década de los 70). El concepto de autonomía de Varela, es mucho más abstracto y abarca un abanico más amplio que el de su aplicación en el campo de las ciencias cognitivas (especialmente el campo de la robótica). Varela se sitúa en la tradición de la cibernética y la teoría de sistemas en la que inserta la noción de autonomía como un tipo de organización³ sistémica.

Para Varela la autonomía un tipo de dinámica automantenida y autorregulada que es producida por una red recursiva de procesos que generan estructuras invariantes características del sistema y las mantienen

³ El término organización hace aquí referencia al tipo de relación entre los componentes de un sistema que los hace miembros de una clase.

frente a perturbaciones internas y externas. La autonomía es una organización dinámica que define su propia identidad, es decir, su unidad como sistema distinguible de los procesos que la rodean. Esta noción abstracta de autonomía se realiza a diferentes escalas y dominios biológicos. Es precisamente la autonomía de cada dominio lo que define su especificidad; es decir, lo diferencia del resto de dominios. Como ejemplo paradigmático la vida se define como un tipo especial de autonomía: la *autopoiesis* o autonomía en el dominio físico. En concreto los sistemas autopoieticos son una red de producción de componentes que componen la propia red y que generan además un borde físico (la membrana) que los separa de su entorno. La *conducta cognitiva* es el resultado de un nivel superior de autonomía: la del sistema nervioso. La organización dinámica y en red del sistema nervioso establece una serie patrones invariantes de correlaciones sensomotoras definiendo al organismo como unidad móvil en el espacio (Varela 1979, Varela & Bourgine 1992, Varela 1992). De este modo la autonomía instaura una perspectiva, unas coordenadas fenomenológicas desde las que la interacción de un sistema (automantenido y autodefinido) con su entorno devienen intencionales, cargadas de valor para el sistema mismo que establece la interacción (al margen de un observador externo que asegure una correlación semántica o una historia evolutiva cuyos efectos no se encuentren ya integrados en la organización del organismo, en la estructura que determina su acción).

La autonomía es, por tanto para Varela el resultado de una organización interna, de una relación específica entre los componentes de un sistema que actúan recursivamente entre ellos generando un holismo causal automantenido, una identidad.

Pero la perspectiva autónoma de Varela ha sido recientemente criticada por su énfasis en la clausura y el rol secundario que la interacción sistema-entorno juega en la definición y la constitución de un sistema autónomo. La introducción de la teoría de sistemas complejos y la termodinámica de procesos irreversibles por autores como Bickhard (2000), Christensen & Hooker (2000), Collier (2002), Kauffman (2000), Ruiz-Mirazo (2001), Ruiz-Mirazo & Moreno (2002, 2004) y Ruiz-Mirazo, *et. al.* (2003) han enriquecido notablemente el concepto de autonomía y su aplicación tanto a la biología como a las ciencias cognitivas. Para todos estos autores la autonomía debe definirse en términos de sistema termodinamicamente

abierto y recursivamente automantenido en condiciones alejadas del equilibrio. Para estos autores la vertiente interactiva de la autonomía es esencial en la definición: la condición alejada del equilibrio y termodinámicamente abierta exige asegurar el flujo necesario de materia y energía para el automantenimiento del sistema; para que éste sea autónomo la fijación de ese flujo termodinámico debe ser internamente generada, es decir, encauzada a través de constricciones producidas por el propio sistema.

2.3.2 La autonomía básica como fuente de normatividad funcional

Los desarrollos posteriores de la noción de autonomía que hemos mencionado brevemente en la sección anterior han permitido redefinir un marco conceptual que dé cuenta de la normatividad y la funcionalidad en procesos naturales (ya sean éstos de carácter metabólico, adaptativo o cognitivo). Ésta es sin duda una de las aportaciones más notables de la perspectiva autónoma a la tarea de naturalizar la cognición. En esta sección veremos cómo una noción termodinámicamente abierta de autonomía permite una fundamentación naturalizada de los conceptos de funcionalidad y normatividad desde una perspectiva radicalmente diferente a la del cognitivismo clásico (el funcionalismo). Ello exige primero definir más concretamente la noción de autonomía a partir de su expresión más fundamental, la termodinámica.

Empecemos por la etimología del término. La palabra autonomía viene del griego *auto-nomos* (ley propia). Podemos, por tanto, ofrecer una primera caracterización intuitiva de los sistemas autónomos como aquellos que producen sus propias leyes⁴. Pero esta noción intuitiva requiere una noción previa de identidad sobre la que predicar el “propio” de la expresión “leyes propias”. Por eso los sistemas autónomos son primariamente aquellos cuya organización básica es la de una entidad que se construye y se mantiene a sí misma en el tiempo y en el espacio. Su ser es un proceso de producción recursiva de su estructura constitutiva: una recursividad que genera una identidad (una estructura automantenida y diferenciada de su entorno). Éstos es lo que vamos a denominar *autonomía básica* siguiendo a Ruiz-Mirazo & Moreno (2000). Más concretamente (y sin recurrir a una noción previa de auto, propio, o identidad) la autonomía básica es el tipo de organización (de

⁴ Hablando propiamente un sistema autónomo nunca será capaz de crear nuevas leyes, sin embargo sí será capaz de crear constricciones que gobiernen su conducta.

relación entre componentes y procesos) por la que un sistema alejado del equilibrio y termodinámicamente abierto genera adaptativamente constricciones internas y externas para modular el flujo de materia y energía necesario para mantener la estructura del sistema. Un sistema autónomo surge cuando en un conjunto continuo de procesos se genera un bucle retroalimentado que genera una circularidad causal automantenida. Los sistemas autónomos son por tanto conjuntos de procesos causalmente integrados en los que la estabilidad de un proceso depende del funcionamiento específico de otros procesos para cuya estabilidad a su vez es necesario el proceso anterior (Rosen 1966, 1967, 1971, Varela *et. al.* 1974, Kauffman 1986). De este modo la autonomía de un organismo se convierte en la fuente de una funcionalidad definida por los propios procesos que la componen (una especie de teleología mecánica): una dependencia entre procesos que genera una causalidad recursiva interpretable en términos de valor para el automantenimiento. La normatividad viene marcada por la red de dependencias dinámicas que se establece entre los componentes y procesos del sistema.

La naturaleza termodinámicamente abierta y la estructura disipativa de los sistemas autónomos básicos permite ya diferenciar dos tipos fundamentales de contribución al automantenimiento del sistema, y por tanto dos tipos dependencias funcionales características:

- a) *Las constructivas*: definidas por la generación de constricciones internas que controlen el flujo y la organización necesaria de materia y energía para el automantenimiento del sistema. En este sentido los sistemas autónomos (autopoiéticos) pueden ser entendidos como una red de producción de componentes (o mantenimiento energético de estructuras) que producen (o mantienen) los componentes y procesos que constituyen la propia red (Maturana & Varela 1980). La expresión más significativa de este aspecto de la autonomía en los seres vivos lo constituye el metabolismo como red autocatalítica.
- b) *Las interactivas*: definidas por las necesidades termodinámicas globales de la red que definen el dominio de constricciones modulatorias necesarias para el mantenimiento de la condiciones de contorno del sistema. Esto exige gestionar apropiadamente el intercambio de energía y materia con el entorno incluyendo las posibilidades de transformar activamente estas condiciones del entorno ya sea a través del desplazamiento como del

cambio de concentración dentro de la membrana, pasando por la ingestión o el bombeo activo de sustancias al entorno⁵. La interacción recurrente sobre sus condiciones de automantenimiento diferencia a los sistemas autónomos del resto de las estructuras disipativas que mantienen su organización sólo bajo ciertas condiciones de contorno que no pueden controlar (es el caso de las células de Benard, los huracanes o los láseres). Un sistema autónomo permanece constitutivamente abierto a su entorno integrando su mundo en su propia realización.

Podemos por tanto definir la *clausura constructiva* como la satisfacción de las constricciones constructiva (endergónicas y exergónicas) y la *clausura interactiva* como la satisfacción de las constricciones interactivas necesarias para el automantenimiento.

En general la autonomía, en cualquiera de los niveles biológicos en los que aparece, siempre presenta una dialéctica recíproca entre los procesos constructivos internos y las interacciones necesarias para mantenerlos. Lo característico de los sistemas autónomos es que la dinámica interna es más cohesiva e integrada (más compleja) que la dinámica interactiva que sostiene, produciendo así una asimetría de control inclinada hacia los procesos internos.

Lo que define la funcionalidad de los sistemas autónomos es la satisfacción de las condiciones de clausura (Collier, 2002) de los procesos interactivos y los constructivos. Un proceso (interno o interactivo) es funcional si contribuye al automantenimiento global del sistema. Al mismo tiempo esta funcionalidad deviene *normativa*⁶ a través de la *presuposición dinámica* de ese proceso en la organización del sistema (Christensen & Hooker 2002) ya

⁵ Podemos adelantar que la forma que toma esta organización autónoma de la funcionalidad en relación a las interacciones con el entorno (la funcionalidad interactiva) evolucionó en una teleosemántica cognitiva que emerge de la autonomía del sistema nervioso y su inserción en un cuerpo biológico (que es a su vez un sistema termodinámicamente autónomo). Este será el objeto de análisis del siguiente capítulo.

⁶ En este contexto la normatividad se refiere al valor atribuido a un proceso u objeto; es decir, adaptada o inadaptada a una interacción o a la estructura de un organismo, verdadero o falso a un estado cognitivo, belleza o fealdad de una obra de arte, etc. La normatividad reta las perspectivas fisicalistas de la comprensión del mundo porque introduce una asimetría de valor (verdadero/falso, bueno/malo, adaptativo/maladaptivo) en la descripción de la naturaleza. Una asimetría que no se encuentra presente en ninguna de las propiedades fundamentales de la materia (tal y como es descrita por las ciencias físicas). Pero, a pesar de ser extraña para la física, la normatividad es un componente esencial de las ciencias biológicas, cognitivas y sociales: el hecho de que una interacción sea adecuada, adaptativa, acertada, etc. es un juicio de valor al que el científico interesado en biología o ciencias cognitivas debe realizar cotidianamente. Un juicio que debe ser justificado en términos naturalistas, a saber, en término de la organización causal del sistema y desde una serie de preferencias valorativas propias del observador ajenas al sistema bajo estudio.

que los procesos funcionales interactivos y constructivos son *condición de posibilidad* del sistema autónomo (como sistema alejado del equilibrio). La naturaleza de los seres vivos como sistemas alejados del equilibrio que mantienen sus condiciones de existencia de forma recursiva, permite definir la funcionalidad normativa (cómo debería de funcionar) de sus procesos constitutivos en relación a su contribución al automantenimiento. De este modo podemos salvar la dificultad de una justificación del funcionamiento *propio* de un componente del sistema (como opuesto a su funcionamiento real en un momento dado). En concreto la dificultad de justificar el funcionamiento propio o normativo de un sistema (o un componente del mismo) radica en la dificultad de encontrar un fundamento naturalizado del que derivar este funcionamiento propio que no sea su funcionamiento *de facto*. Pero si derivamos el funcionamiento propio (normativo) del funcionamiento *de facto* no puede darse el caso de que algo funcione incorrectamente. Desde la perspectiva autónoma podemos salvar esta dificultad ya que el funcionamiento propio se deriva de las necesidades que el sistema como un todo tiene sobre el funcionamiento de una de sus partes. Es decir podemos especificar el funcionamiento normativo de una estructura como aquel funcionamiento que permite que el sistema en el que se inserta siga automanteniéndose (obviamente esta solución sólo tiene sentido en el marco de los sistemas alejados del equilibrio cuyo componentes han de permanecer activos para sostener la estructura del sistema). Pongamos un ejemplo: el hecho de que el automantenimiento de un mamífero dependa crucialmente de una forma concreta de funcionamiento del sistema circulatorio permite justificar un juicio normativo sobre el funcionamiento del corazón. El bombeo activo de sangre a una velocidad determinada X define el funcionamiento propio normativo del corazón, ya que de otro modo el organismo perecería (haciendo desaparecer al propio corazón). Podemos así hablar de un funcionamiento correcto e incorrecto, adaptado o maladaptado de un sistema o de alguna de sus partes ya que un funcionamiento diferente al correcto conlleva una destrucción del sistema (o de la parte) de la que predicamos la funcionalidad. Es por tanto en relación a sus capacidades para el automantenimiento alejado del equilibrio que los seres vivos generan su propia normatividad y sus propios valores de forma *autónoma*.

El cambio de perspectiva que implica la noción de autonomía queda especialmente resaltada en la oposición entre dos tipos principales de

explicación causal que rivalizan o definen niveles diferentes de comprensión de los sistemas complejos. Los modelos construidos sobre la noción de algoritmo y procesamiento lineal o computación y los construidos sobre la noción de red. Los algoritmos procesan como árboles o procedimientos lineales de ejecución de funciones que contienen (como mucho) bucles repetitivos de procesamiento, pero que actúan siempre de forma secuencial. La funcionalidad, en este esquema, viene determinada por el lugar que ocupa una subrutina en el algoritmo completo. En una red, en cambio, la arquitectura es interconectada y recurrente y la actividad se determina por la estabilidad de toda la red. En una red la funcionalidad viene determinada por las transiciones de estados globales (*eigenbehaviors*) definidas por las condiciones de estabilidad de la red. La perspectiva autónoma se sitúa en la tradición reticular del estudio de los sistemas complejos y se preocupa especialmente de la realizabilidad termodinámica de esa red. La estructura en red del metabolismo y las redes neuronales permite además una naturalización más inmediata que la del funcionalismo, cuya tesis de realizabilidad múltiple desconecta la explicación funcionalista de la estructura del sistema que la instancia.

2.3.3 Características comunes de las perspectivas autónomas en ciencias cognitivas

Lejos de un programa de investigación establecido y unitario el concepto de autonomía ha suscitado toda una serie de trabajos de investigación en el ámbito de las ciencias cognitivas (desde las neurociencias a la robótica, pasando por la filosofía de la mente y la inteligencia artificial)⁷. Sin embargo

⁷ Entre los trabajos de investigación que han reclamado la autonomía como concepto central a la hora de estudiar procesos cognitivos destaca el campo de la robótica autónoma. El término *robótica autónoma* ha sido usado a partir de los años 90 para referirse a una serie de constricciones ingenieriles en la construcción y experimentación en robótica y ha pasado así a convertirse en una etiqueta de uso común en la investigación robótica en ciencias cognitivas e ingeniería aplicada. Estas constricciones incluyen condiciones como la ausencia de control remoto en los robots, la ausencia de intervención humana en las tareas que el robot ha de resolver por sí mismo (autónomamente) o la más exigente condición de respuesta en tiempo real y en entornos físicos reales (no simulados ni simplificados). Cercana en inspiración y desarrollo a la robótica situada (Brooks 1991), la robótica autónoma defiende la fisicalidad, la corporización, el situacionismo (*situatedness*) y el dinamicismo de los procesos cognitivos frente a las perspectivas formalistas, abstractas y virtuales de la inteligencia artificial clásica (en los que los agentes cognitivos operan en entornos virtuales o formales controlados, o en entornos físicos de juguete sin constricciones temporales ni variabilidad ambiental —ruido, imprevistos, etc.). La práctica de la robótica autónoma ha forzado a algunos ingenieros a ir más allá de una lista de constricciones ingenieriles y a desarrollar una noción de autonomía más elaborada que dé cuenta del tipo de proceso interactivo que se establece entre el robot y su entorno físico. Es el caso de ingenieros como Tim Smithers (1995, 1997) o Randall Beer (1995, 1997) que han criticado fuertemente las perspectivas computacionalistas y de procesamiento de información a favor de una perspectiva dinámica, corporizada y situada. En

el término autonomía subsume una serie de conceptos clave para las ciencias cognitivas, una serie de características que permiten definir una perspectiva autónoma de otras estrategias o corrientes científicas y filosóficas en relación a la cognición. Veamos algunas de ellas:

Enraizamiento biológico: es la idea de que la comprensión de la cognición debe realizarse de abajo-arriba (from the bottom-up) en dos vertientes: en términos de la evolución de las capacidades cognitivas en la historia natural y en términos de los mecanismos biológicos que producen conducta cognitiva (redes neuronales, cuerpos, sistema neuroendocrino, nichos ecológicos, etc.).

Autoorganización, complejidad, emergencia: la intuición fundamental es que no existe un procesador central u homúnculo que controle la conducta sino una red distribuida e integrada de procesos recursivos del que emerge un patrón conductual coherente. La noción de autonomía asume un alto grado de complejidad en el sistema, una dependencia holista y no lineal entre los componentes del sistema, lo cual introduce importantes constricciones en el análisis, localización funcional y descomposición de estructuras dentro del sistema.

Interactivismo, corporización, situacionismo y dinamicismo: Desde la perspectiva de la autonomía se considera que la cognición es un proceso cuyo desarrollo y realización no puede comprenderse de forma desacoplada de las interacciones corporizadas que la constituyen. Una perspectiva autónoma asume una dialéctica entre independencia y acoplamiento estructural sistema-entorno: una construcción interactiva del significado y la conducta en la que la corporización y el estar situado se consideran características esenciales de la cognición y son mejor comprendidas por conceptos dinámicos que con los computacionalistas tradicionales; se introducen así constricciones temporales y espaciales como fundamentales para la producción de conducta cognitiva.

Crítica al representacionalismo: el uso del término autonomía está generalmente asociado a una crítica a los fundamentos del funcionalismo representacionalista. Se considera que el representacionalismo supone una explicación heterónoma de los procesos cognitivos, es decir, que la

esta línea se encuentra también al roboticista Eric Prem (1995, 1997) que ha desarrollado el concepto de *autonomía epistémica* como la habilidad del propio sistema de decidir sobre la validez de sus mediciones (percepciones), un proceso que no puede ser reducido a aspectos formales o lógico-sintácticos dada la fisicalidad intrínseca a todo proceso de medición, su naturaleza pre-formal.

representación, en lugar de ser una realidad autónomamente generada por el sistema cognitivo es una explicación que hace un observador externo al margen de los mecanismos dinámicos causalmente efectivos que operan en la producción de conducta cognitiva. La atribución de significados o referentes a estados internos del sistema desde esta perspectiva del observador externo y al margen de la organización global de los mecanismos subyacentes es imponer sobre el agente cognitivo una regularidad que no le es propia. La mente como un espejo que re-presenta un mundo externamente definido rompe con la tesis fundamental de la perspectiva de la autonomía, a saber, que *el mundo cognitivo es interactivamente generado por el propio organismo que recorta un mundo de significados de la amalgama caótica e indefinida de procesos físicos que le rodean*. El punto de partida de la perspectiva autónoma no es, por tanto, el de la representación sino el de una serie de mecanismos biológicos o ingenieriles que sostienen una interacción robusta e internamente guiada, capaz de generar sus propias reglas de interacción, modificarlas y adaptarlas de acuerdo a valores propios del sistema.



A partir de aquí la noción de autonomía no ha sido suficientemente desarrollada en el dominio del SN, generalmente la normatividad y la funcionalidad se consideran transitivas del dominio biológico al cognitivo mientras que algunos autores (especialmente los roboticistas) aíslan la autonomía interactiva de la biológica sobre la que se sustenta. Al mismo tiempo lo que en el terreno del origen de la vida fue resuelto con el surgimiento de la información genética, queda aún por resolver en el dominio cognitivo, a saber el surgimiento de un tipo de causalidad (la informacional) capaz de organizar la enorme complejidad dinámica de los sistemas autónomos. Este es el objeto del siguiente capítulo.

3 AUTONOMÍA E INFORMACIÓN EN EL SISTEMA NERVIOSO: UNA PERSPECTIVA NATURALIZADA.

La cognición es un fenómeno emergente. Minutos después del big-bang no existía el fenómeno cognitivo en el universo (no había nada de lo que pudiéramos predicar capacidades cognitivas); hoy en cambio, atribuimos capacidades cognitivas a diversos sistemas que nos rodean, especialmente a nosotros mismos. ¿Cómo ocurrió esta transición en la historia del universo? ¿Qué tipo de organización característica de la materia lo posibilita? ¿Cómo podemos caracterizarla operacionalmente? En la sección anterior hemos visto cómo la aparición de sistemas autónomos básicos supone la emergencia de una funcionalidad y una normatividad propias del sistema, gracias al surgimiento de estructuras disipativas recursivamente automantenidas. El origen de la cognición y su especificidad radica en una transición parecida e imbricada en la anterior: la aparición del sistema nervioso (SN) en los seres vivos. Las características propias del SN, su organización autónoma y el tipo de causalidad interna que la gobierna (la informacional) marcarán la especificidad de la cognición como fenómeno natural.

Desde la perspectiva naturalizada que venimos defendiendo naturalizar la cognición exige dar cuenta de ella en dos vertientes fundamentales: una histórico-evolutiva (que de cuenta de la emergencia diacrónica de la cognición) y otra organizativa (que de razón de la emergencia sincrónica del fenómeno cognitivo). Para naturalizar diacrónicamente la cognición necesitaremos contextualizar la aparición del SN en el conjunto de la evolución biológica donde la aparición del SN se muestra como una bifurcación esencial que permite el desacoplamiento de la agencialidad del metabolismo y genera una cascada de acontecimientos evolutivos que desembocan en la aparición de la cognición y la conciencia. Como veremos la emergencia de la cognición en la historia natural viene de la mano de la evolución de la funcionalidad interactiva basada en la motilidad y las constricciones organizativas internas que ésta va generando y posibilitando en la organización corporal de los organismos. Por otro lado tenemos que especificar el tipo de organización concreta que permite al

sistema nervioso (corporizado e interactivamente situado) generar el fenómeno cognitivo. Esta especificación de la organización del sistema nervioso debería de satisfacer las constricciones naturalistas que permitan su inserción del discurso científico: por un lado en la forma de concepto capaz de dar coherencia a los datos empíricos de los que disponemos al tiempo que se muestra productiva a la hora de generar nuevas hipótesis de trabajo; por otro como concepto productivo a la hora de sintetizar conducta cognitiva en sistemas artificiales. Este es el horizonte epistemológico en el que se sitúa el presente capítulo.

3.1 El origen evolutivo del SN

3.1.1 El origen de la motilidad y la agencialidad

El surgimiento de la cognición está intrínsecamente unido a la historia evolutiva de la funcionalidad interactiva (la agencialidad) de los organismos como sistemas autónomos y especialmente la motilidad como base de nuevos dominios funcionales interactivos.

Desde el nivel más simple de vida encontramos ya una considerable variedad de formas de adaptabilidad a los cambios del entorno: las bacterias han desarrollado muy variadas formas de vida, basadas en estrategias adaptativas diferentes: cambios morfológicos, modificaciones de sus procesos alimentarios, motilidad, etc. Sin embargo, la aparición de la motilidad tiene un interés particular. La motilidad no reside simplemente en el hecho de que el organismo se mueva por sus propios medios, sino de que lo haga como un todo en función de determinadas diferencias locales de su entorno (es decir, de forma dirigida). Además, en la medida en que para ciertos organismos la supervivencia depende de su capacidad de moverse direccionalmente a *cierta velocidad*, la aparición de lo que llamamos motilidad requiere complejas estructuras moleculares situadas en la membrana (flagelos) capaces de impulsar coordinada y eficazmente al conjunto del cuerpo.

Lo característico de esta estrategia adaptativa es la capacidad de detectar diferencias funcionalmente relevantes relativamente *distantes* y autotransportarse a una velocidad suficiente como para automantenerse a través de la explotación de tales diferencias. Así por ejemplo, una bacteria *E. coli*, es capaz de detectar la presencia de una concentración de azúcar en su entorno y conmutar su movimiento usual de tambaleo por un movimiento

direcciona hacia la concentración (Neihardt 1996).

A pesar de su aspecto llamativamente conductual, la base organizativa de esta forma elemental de interacción “sensomotriz” no es diferente de la que funciona en otros tipos de bacterias con muy diferentes formas de estrategias adaptativas. La motilidad que hemos visto es resultado de la articulación de un modo de conexión entre los mecanismos quimiorreceptores y motores de la célula (unos situados en la zona ‘delantera’, opuesta al flagelo, y los otros en la base o inserción del mismo) mediado por los ciclos o rutas metabólicas que tienen lugar en su interior (Losik & Kaiser 1997, Hoffmeyer 1998). Así, dicha conexión no puede ser entendida más que asociándola estrechamente a los procesos de automantenimiento constructivo del sistema. En los sistemas vivos más pequeños y primitivos todas las diferentes formas de agencialidad adaptativa están basadas en la capacidad de control sobre la plasticidad metabólica. A escala bacteriana los mecanismos que soportan las reacciones adaptativas en tiempo somático son los mismos, ya se trate de una respuesta de movimiento o puramente metabólica. La idea fundamental es que la funcionalidad interactiva y la constructiva comparten la misma infraestructura organizativa. Un organismo pueda adoptar diferentes estrategias adaptativas (las metabólicas y de movimiento) a través de la misma infraestructura organizativa. Así por ejemplo, el procariota *Caulobacter* cuando está en un medio de alta humedad vive fijo al suelo mediante una forma de tipo vegetal, y sin embargo cuando sobreviene un periodo de desecación se reproduce y las nuevas células crecen con un flagelo capaz de transportarlas a otro ambiente de más humedad. El hecho de que el organismo base su modo de vida en la captura del alimento a través del movimiento activo o que lo haga a través de explotar los recursos directamente accesibles, no implica diferencias sustanciales en su modo interno de organización.

3.1.2 Tensión entre motilidad e incremento de tamaño

Esta situación de indistinguibilidad entre los mecanismos que sustentan la funcionalidad interactiva y la constructiva empieza sin embargo a cambiar cuando la evolución produjo organismos de tamaño mucho mayor, como las células eucariotas⁸. La razón es que para la motilidad se necesita que la coordinación entre superficies sensoras y motoras siga siendo rápida; pero

⁸ El volumen medio de una célula eucariota es unas 1000 veces el de una procariota.

ahora la distancia entre ambas es mucho mayor, y también lo es el volumen del organismo que debe ser movido. Si a pesar de ello las células eucariotas han sido capaces de explotar el nicho de la adaptabilidad basada en el movimiento es gracias a cambios sustanciales en su organización interna. Esta organización interna es pues capaz de soportar todos los procesos fundamentales del organismo: mantenimiento metabólico, reproducción, y todas las formas de adaptabilidad, incluida la motilidad. De hecho, la organización interna de las eucariotas (a través de la producción y uso de microtúbulos) constituye la forma más compleja de articular eficientemente, sobre bases puramente metabólicas, todo el abanico de respuestas adaptativas de un organismo individual.

A pesar de estos eficientes y sofisticados mecanismos moleculares, en las eucariotas se plantea un serio conflicto entre la organización simultánea del movimiento y de otras tareas metabólicas, como la reproducción (Buss 1987, Maynard-Smith & Szathmary 1995). La razón básica es que la organización metabólica (cuya función es el mantenimiento material y energético del organismo) se hace cada vez más incompatible con la capacidad de proporcionar una dinámica de configuraciones internas suficientemente versátil y rápida como para coordinar eficazmente las superficies sensoras y motoras, permitiendo movimientos adaptativos rápidos del organismo. A medida que el tamaño del organismo aumenta, los costos energéticos y materiales que implica para la organización metabólica soportar una dinámica suficientemente plástica y rápida con fines de coordinación sensomotriz, devienen incompatibles con su función básica, que es soportar primariamente la construcción y reparación recursiva del organismo (es decir, su automantenimiento).

Así pues, razones estrictamente organizativas impiden que los sistemas de tamaño equivalente a la multicelularidad⁹ puedan desarrollar

⁹ Cabría considerar desde luego, la cuestión de si es posible un crecimiento significativo del tamaño sin pasar a la organización multicelular. Pero esta vía, además de que tampoco solucionaría los problemas de la motilidad, plantea problemas de viabilidad básicos, fundamentalmente por causa de que los mecanismos de distribución están basados en la difusión. Según Bonner (1988) "If one thinks of the rates of different chemical processes occurring within the cell, the distances needed for diffusion, the surface boundaries needed for isolating different chemical components of the motor, and so forth, all of these lead to the conclusion that there is an optimal size with sharp upper and lower limits, which is the size found in nature" (p. 61). En otras palabras, cualquier organismo por encima del tamaño celular de las eucariotas, ha de basar su organización fundamental en la coordinación de muchas células para formar un todo integral (lo cual a su vez va a requerir un proceso de diferenciación entre éstas). Hay alguna excepción, como el alga *Acetabulara mediterranea*, que es un organismo unicelular con un volumen de unos 5 cm cúbicos, viable gracias a que posee mecanismos especiales de transporte interno. Sin embargo, estos tamaños son

formas versátiles de motilidad integral soportadas por la misma organización que a su vez soporta los procesos básicos y universales de automantenimiento constructivo (es decir,, por la organización metabólica). Por eso, como vamos a ver, las formas de motilidad que determinados tipos de multicelulares (los animales) han desarrollado están basadas en una forma de organización radicalmente distinta. En la sección siguiente veremos en qué consiste y cómo pudo aparecer.

3.1.3 La evolución explora nuevas formas de la agencialidad

La potencialidad de diferenciación en las células eucariotas fue un importante factor que posibilitó la aparición de organizaciones multicelulares integradas, con individualidad propia. La aparición de esta nueva forma de vida, como ya ocurriera con los unicelulares eucariotas, abrió un nuevo espacio ecológico a explorar. Pero en este nuevo espacio el ámbito de las estrategias de vida basadas en la motilidad no podría ya ser desarrollado con los mecanismos disponibles hasta entonces, debido al gran incremento de tamaño de los nuevos organismos multicelulares. Este aumento plantea un serio problema para el desarrollo de formas de vida cuya estrategia dependa del movimiento por dos razones: el enorme incremento de las distancias entre superficies sensoras y efectoras (que deben ser conectadas en tiempos cortos para que el organismo pueda reaccionar rápidamente) y la necesidad de modular selectiva y rápidamente la forma de dichas conexiones (de manera que el individuo pueda reaccionar adecuadamente según las variaciones de su entorno). En estas condiciones, si las posibles respuestas adaptativas siguen soportadas directamente por la organización metabólica del organismo, es imposible que éste desarrolle formas de vida basadas en la motilidad. Esta limitación se pone de manifiesto en los animales multicelulares más simples, que sólo son capaces de desarrollar formas de motilidad basadas en mecanismos sensomotores puramente locales. Por ejemplo, en las esponjas¹⁰ las reacciones a los estímulos son locales e independientes (todos los individuos adultos son sésiles). Es obvio que sobre tales bases organizativas no es posible un amplio desarrollo evolutivo de la motilidad en multicelulares.

Sabemos que el problema se resolvió con la aparición del sistema nervioso. Pero, ¿cómo se produjo este cambio tan importante? Probablemente

imposibles en unicelulares con motilidad.

¹⁰ Recordemos que las esponjas pertenecen al filo poríferos que a su vez son parte del reino animal.

este hecho fue resultado de la combinación de dos factores: 1) la aparición (como consecuencia de un proceso de diferenciación celular) de un nuevo tipo de células (las neuronas¹¹) capaces de conectar de manera plástica, rápida y (metabólicamente hablando) económica, superficies sensoras y motoras; y 2) la aparición de algún tipo de metazoo cuyo plan corporal permitiera reclutar a estas células para sostener formas de agencialidad sensomotriz muy sencillas. La potenciación de la conducta motriz que presumiblemente debió de producirse sería la base de un proceso evolutivo en el que se seleccionarían aquellos animales que poseyeran redes de neuronas interconectadas, actuando como soporte de comportamientos sensomotrices funcionales. Esto debió de ocurrir hace unos 600 millones de años, pues poco tiempo después aparecieron los cnidarios (medusas, hidras, etc) que son los primeros animales conocidos que poseen una red de neuronas interconectadas. Esta organización distribuida por todo el cuerpo es la forma más primitiva de red neuronal (aunque todavía no pueda considerarse propiamente un sistema nervioso¹²).

La invención del sistema nervioso parece pues un hecho fundamental para el desarrollo de la motilidad en multicelulares. Dondequiera que veamos movimiento rápido y versátil a escala multicelular, ello es resultado de la existencia en el organismo de una estructura interna formada por neuronas¹³.

¹¹ Las neuronas se diferenciaron como células capaces de formar ramificaciones que pueden interconectarse a través de procesos electroquímicos (polarizando y despolarizando sus membranas o a través de intercambio de sustancias). En la siguiente sección estudiaremos sus características específicas.

¹² En estas primitivas redes neuronales no hay todavía diferenciación entre componentes sensores, motores y asociativos, ni direccionalidad de los impulsos nerviosos. Pero ya tenemos una extensa red de neuronas interconectadas, capaces de producir una dinámica propia de configuraciones de impulsos.

¹³ Sólo vamos a encontrar una clase de organismos multicelulares carente de SN y cuya forma de vida esté basada en la motilidad rápida y versátil. Es el caso de las plantas *Dinee* (las popularmente conocidas como plantas "atrapamoscas") que son capaces de cerrar rápidamente sus hojas cuando un insecto se posa en ellas. Este proceso "sensomotriz" implica que la planta es capaz de realizar una detección y a continuación, un veloz movimiento coordinado de millones de células, sin la participación de nada parecido a un SN. Veamos cómo: cuando la presa entra en contacto con los filamentos de las hojas, el estímulo táctil es transducido en procesos coordinados de cambios en el potencial de membrana, los cuales inducen en las células colindantes cambios que transfieren agua; y estos cambios en la presión del agua son los que provocan el cierre brusco de las hojas (Simons 1981). Este excepcional mecanismo "sensomotriz" en realidad nos permite comprender mejor el carácter fundamental de la forma de organización que representa el sistema nervioso para organizar la motilidad en los multicelulares. Pues, a poco que analicemos el mecanismo anteriormente descrito, veremos que, aunque eficaz en un tipo de operación muy específica, no puede servir de base para formas de conducta motriz que requieran un mínimo de plasticidad adaptativa. Esto se debe a que los mecanismos de la planta atrapamoscas responden de una forma conjunta, uniforme y lineal. La propia estructura del mecanismo impide una variabilidad interna de respuestas. Por consiguiente, aún cuando la evolución ha explorado y encontrado formas diferentes de organizar la motilidad en multicelulares, sólo ha encontrado una (el SN)

Pero además de permitir el desarrollo de la motilidad en organismos multicelulares, la aparición del SN ha abierto el paso a nuevas (cualitativamente diferentes) formas de interacción adaptativa. Ya en los primeros estadios evolutivos del sistema nervioso aparecen formas rudimentarias de aprendizaje, categorización y memoria (Arhem & Liljenstrom 1997). De hecho, tendemos a identificar el significado fundamental de la aparición del sistema nervioso en función del desarrollo de tales características, puesto que empiezan a aparecer como específicamente “cognitivas” (en la medida en que su evolución va a ir aproximando la agencialidad animal a lo que identificamos como nuestro universo cognitivo).

Pero veamos en qué reside esta capacidad tan importante de la organización nerviosa. Un rasgo esencial del sistema nervioso es su capacidad de generar un universo propio que da soporte a formas cada vez más complejas de agencialidad. Pero a su vez, el sistema nervioso forma parte del organismo, en virtud de su integración funcional en el automantenimiento del organismo como un todo. Ahora bien, ¿Cuál es la *forma* particular de contribución al automantenimiento mediante la cual el sistema nervioso se inserta en el organismo? ¿Marca ésta un cambio cualitativo en la organización biológica, tal que permita la emergencia de un nivel o dominio específicamente cognitivo? ¿De qué modo? ¿En virtud de qué principios organizativos y de qué características propias? Analizaremos estas cuestiones en las secciones siguientes.

capaz de servir de soporte a una amplia ocupación del espacio de las formas de vida basadas en la motilidad a la escala de los multicelulares.

3.2 La organización autónoma del sistema nervioso

3.2.1 Desacoplamiento jerárquico y enraizamiento autónomo

El primer factor clave para definir la especificidad del SN va a ser lo que vamos a denominar su *desacoplamiento jerárquico* (Moreno & Lasa 2003). Nos referimos con ello a que el funcionamiento de la dinámica neuronal (en tanto en cuanto esta dinámica contribuye al automantenimiento del organismo) aparece desacoplada y subdeterminada por el conjunto de procesos metabólicos que operan localmente en sus componentes fisiológicos (principalmente las neuronas). Atendiendo detenidamente a la naturaleza de sus componentes y la forma en que se relacionan, podemos entender mejor esta peculiaridad del SN.

1. **Especificidad de sus componentes:** el tipo de células que componen el sistema nervioso (las neuronas) presentan una serie de características propias ausentes en las células que forman otros tejidos o sistemas dentro del organismo. En concreto son destacables las siguientes características:
 - a) capacidad para *propagar* sin disipación (es decir, sin pérdida de variabilidad) el cambio *del potencial de membrana* a lo largo del cuerpo celular y sus extremidades¹⁴,
 - b) *conectividad conductiva* con otras neuronas a través de superficies sinápticas con receptores químicos que permiten la regeneración (*variable*) del cambio de potencial de membrana de las neuronas vecinas,
 - c) *ramificaciones a largas distancias* del cuerpo celular (desde 0.1mm hasta 3 metros),
 - d) *no-linealidad del potencial de respuesta* y
 - e) *plasticidad* (variabilidad) a múltiples escalas temporales y biofísicas en relación a las respuestas a los cambios de potencial y cambios sinápticos: crecimiento dendrítico, variación de concentraciones de neurotransmisores en las sinapsis, cambios moleculares, etc.
2. **Interacción entre componentes:** las características neuronales antes descritas permiten una conectividad recursiva y una velocidad de interacción *selectiva* interneuronal en frecuencias de interacción muy

¹⁴ En los SN desarrollados a esta capacidad hay que sumar la valiosa función de las células gliales que aíslan con mielina la pared celular de las neuronas para aumentar su conductividad.

superiores a las que rigen cualquier otro proceso de "comunicación" en el organismo.

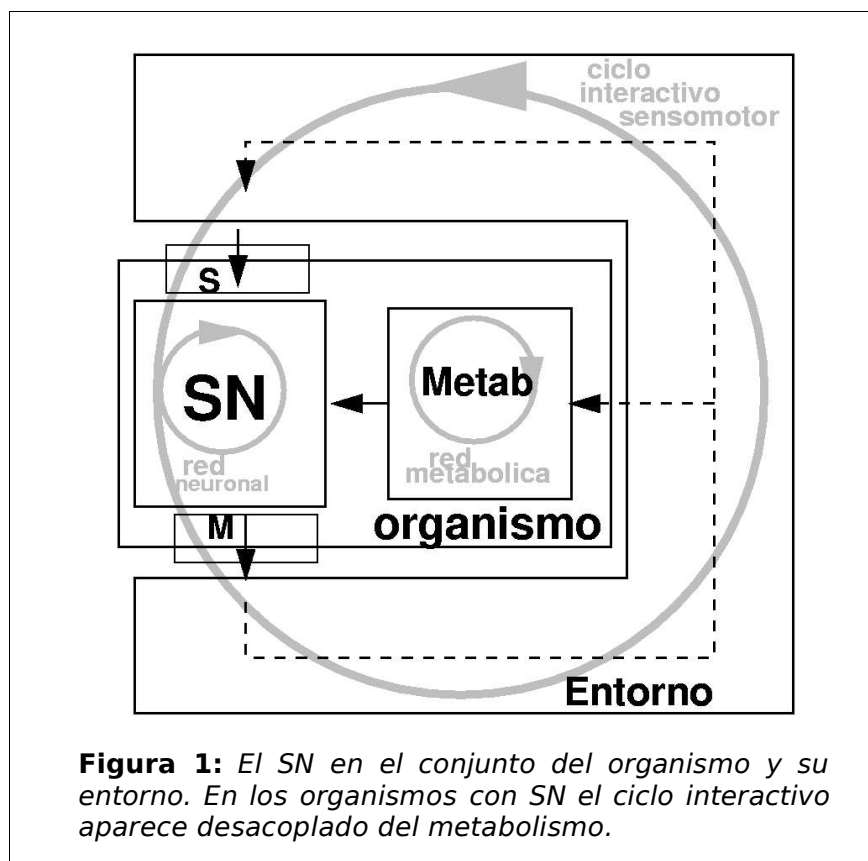
Esto dota a la red interneuronal de unas características especiales en el conjunto del organismo: ningún otro sistema intercelular tiene la capacidad, ni de lejos, que tiene el SN de correlacionar funcionalmente tantas unidades y, al mismo tiempo, de modificar selectivamente los estados de dichas unidades tan rápidamente. La particularidad del SN es por tanto su capacidad de generar una variedad enorme de estados [configuraciones] por unidad de tiempo y de coordinar un número inmenso de transformaciones de estado paralelamente (características todas ellas, como ya veremos, propicias para la generación de procesos autoorganizativos de regulación). Dicho de otro modo, las neuronas son un tipo extremadamente peculiar de células que se han seleccionado por su capacidad para mantener una compleja, plástica y rápida dinámica de interacciones entre sí, minimizando la interferencia con procesos metabólicos locales¹⁵.

Este último factor (la minimización de interferencias con procesos metabólicos locales) va a ser de fundamental importancia para la caracterización del SN, ya que va a definir el aspecto *desacoplado* del SN: es decir, su dinámica va a quedar subdeterminada por los procesos puramente metabólicos (autoconstructivos) de la estructura biológica que lo sustenta (el conjunto de la anatomía neuronal). El desacoplamiento jerárquico del SN apunta a una doble condición del mismo: está localmente construido por procesos metabólicos (aspecto jerárquico: el SN emerge de procesos metabólicos) y, sin embargo, éstos no son suficientes para dar razón de su dinámica funcional (aspecto desacoplado). Otro factor que contribuye a esta subdeterminación de la dinámica neuronal por constricciones metabólicas es su acoplamiento adicional con un entorno sensomotor a través de la interfaz del cuerpo. Su corporización (*embodiment*) y su "estar situado" (*situatedness*) hacen que la dinámica del SN responda también a factores interactivos inscritos en ciclos sensomotores, que no pueden ser reducidos al sustrato metabólico del organismo.

El SN presenta además una dinámica interna reticular, cohesiva y recurrente capaz de mantener patrones invariantes frente a perturbaciones internas y externas, de tal manera que su complejidad dinámica interna es

¹⁵ Cuando decimos que las neuronas han sido seleccionadas por una serie de capacidades que poseen nos referimos a que esas capacidades han abierto un nicho interactivo nuevo (en este caso el de las respuestas adaptativas basadas en la motilidad).

mayor que la dinámica interactiva que establece con su entorno. Esta asimetría de complejidad hace que la dinámica del SN aparezca como autogenerada, o, dicho de otro modo, *autodeterminada* (difícilmente controlable por factores externos a la propia dinámica). En definitiva, si para un momento dado quisiéramos predecir la evolución de la dinámica neuronal,



el conocimiento del estado metabólico de sus componentes aislados no nos serviría de gran cosa. Por el contrario, tendríamos que atender al contexto corporal y ambiental en el que se encuentra el organismo y, sobre todo, a la propia dinámica interna que es recursivamente generada (atendiendo a sus propiedades holistas de red, más allá de las interacciones celulares locales). Sin embargo esta dinámica interna no responde exclusivamente a una regulación propia del SN aislado del organismo. El desacoplamiento metabólico local no implica una independencia global del SN respecto al organismo como un todo. Por el contrario, la dinámica del SN va a estar en última instancia evaluada por su contribución al automantenimiento del organismo. La especificidad y la funcionalidad del SN responden a su capacidad de generar una dinámica propia, automantenida y adaptable, pero siempre en relación a la expansión de las posibilidades interactivas de

automantenimiento del organismo en el que se inserta.

Podemos ahora responder apropiadamente a la cuestión que planteábamos al final de la sección anterior: la *función* del SN en el conjunto del organismo es el control de los ciclos de interacción (de la clausura interactiva necesaria para el automantenimiento del sistema autónomo) a través de una red desacoplada de la dinámica metabólica, pero inserta en una interfaz corporal (extremidades, músculos y sistema perceptivo). Para comprender esto es importante establecer la distinción entre dos ciclos fundamentales de organización de los organismos con SN: el ciclo termodinámico o constructivo y el ciclo sensomotor o interactivo (ver figura 1). Los organismos con SN presentan un flujo material y termodinámico que se realiza entre la ingestión y digestión de alimentos y oxígeno, su secreción y la disipación de calor en el cuerpo, junto a toda la maquinaria autoconstructiva del organismo y la infraestructura necesaria para su mantenimiento (sistema circulatorio, digestivo, respiratorio, etc.). Por otro lado, observamos un ciclo sensomotor de conducta adaptativa regido por una infraestructura (sistema sensomotor y SN) dinámicamente desacoplada de los procesos metabólicos (aunque sustentada en ellos). La funcionalidad del SN viene dada por la forma en la que se conectan ambos procesos a un nivel superior. Ya hemos dicho anteriormente que esta relación funcional exige que la dinámica interactiva controlada por el sistema nervioso satisfaga las condiciones de automantenimiento del organismo, más concretamente, debe satisfacer las condiciones de contorno del flujo termodinámico biológico. Desde el punto de vista del funcionamiento dinámico del SN, la satisfacción de estas condiciones de contorno del flujo termodinámico aparecen como lo que Ashby (1952) denominó *variables esenciales*: un conjunto de variables (temperatura, input nutricional, etc.) que deben ser mantenidas dentro de unos límites de viabilidad. Será en torno al mantenimiento homeostático de estas variables que se construya la dinámica cohesiva del SN y su organización interna (sin que, como veremos más adelante, quede totalmente determinada por estas constricciones de adaptabilidad).



Hemos definido el concepto de *desacoplamiento jerárquico* del SN respecto a la organización básica del organismo como el surgimiento de un dominio dinámico sub-determinado por los procesos biológicos (metabólicos) que lo sustentan como tal. Sin embargo hemos visto cómo al desacoplamiento

dinámico local acompaña un acoplamiento funcional global, por el cual la dinámica neuronal es evaluada en última instancia por la satisfacción de las necesidades de automantenimiento del sistema. La *corporización* del SN (como sistema jerárquicamente desacoplado) se entiende así en un doble sentido de interfaz interactiva con el mundo y de enraizamiento biológico, mediante el cual el metabolismo sostiene constructivamente al SN y éste contribuye a satisfacer las necesidades interactivas de aquél (intercambio termodinámico con el entorno).

3.2.2 Observables constitutivos del SN

La emergencia del SN como nivel desacoplado conlleva la existencia de un conjunto de observables constitutivos de dicho nivel. En la sección anterior hemos visto cómo la especificidad de las neuronas dotaba al SN de unas características organizativas propias y diferenciadas del resto de órganos, sistemas y tejidos del organismo. Sin embargo aún no hemos detallado suficientemente cuáles son las variables que constituyen ese nivel dinámico desacoplado del que hablamos, ni cómo podemos aislarlas para definir con mayor precisión la dinámica de dicho nivel. La delimitación de estos observables es fundamental para construir una teoría propia, específica, de los procesos neuronales.

Si, como señalábamos en el punto anterior, la funcionalidad del SN reside en la satisfacción de las condiciones de clausura interactivas del organismo en su entorno, serán el metabolismo (en su conjunto) y la naturaleza desacoplada del SN quienes impongan las constricciones necesarias para aislar los observables constitutivos del dominio neuronal. En otras palabras, los observables relevantes del funcionamiento del sistema nervioso son aquellos que permiten definir el dominio dinámico causal que explica cómo satisface su función el SN¹⁶. Así pues, entendemos por observables (en el SN) *la clase de distinciones necesarias y suficiente para generar y/o explicar la dinámica del dominio neuronal*. En este sentido, una clase de cambios metabólicos (como el aumento de la concentración de una enzima) en una célula neuronal no es un observable neuronal, a no ser que tenga efectos en la dinámica global del SN. El criterio observacional es pues el que responde a la pregunta: *¿qué conjunto de variables constituyen la*

¹⁶ En el espacio multidimensional definido por todas las variables posibles que pueden abstraerse dentro de los límites anatómicos del SN, el criterio de funcionalidad aísla un hiperplano causalmente relevante en el que tienen lugar los procesos neuronales.

dinámica interna que permite acoplar sensores y efectores en el SN?

Los procesos internos que median entre superficies sensoras y efectoras son enormemente complejos, pero podemos distinguir hoy día entre una serie de variables básicas (observables primarios) que determinan en breves periodos de tiempo las respuestas del organismo, y unos observables secundarios que devienen relevantes en tanto que interactúan con los primeros a una escala temporal más amplia e inducidos por aquellos.

Observables primarios (impulsos o potenciales de acción)

La trayectoria dinámica causal que gobierna la producción de conducta adaptativa viene determinada por los trenes de *impulsos electroquímicos*, también llamados *potenciales de acción*. Estos atraviesan uniformemente el SN uniendo superficies sensoras y efectoras, permitiendo así el establecimiento de correlaciones sensomotoras adaptativas.

Los impulsos neuronales (*neural spikes*) son cambios súbitos de potencial en la membrana de las neuronas que se propagan por el cuerpo neuronal con una amplitud de 100mV y una duración aproximada de 1ms a una velocidad de entre 1 y 100 m/s. Un potencial de acción tiene lugar cuando (debido a los neurotransmisores liberados por las neuronas vecinas) la diferencia de potencial electroquímico en la membrana supera un umbral determinado. En realidad, la continuidad dinámica de los impulsos entre neuronas está asegurada por la conexión sináptica en la que el potencial de acción desencadena la liberación (llamada potencial sináptico) de neurotransmisores, que a su vez desencadena un potencial en los receptores dendríticos de las neuronas adyacentes, haciendo que éstas generen a su vez sus potenciales de acción. Cuando decimos que los potenciales de acción son los observables primarios del SN, nos referimos en realidad a la sucesión de potenciales: de acción – sinápticos – receptores – de acción, y así sucesivamente. Esta concatenación de procesos en red genera el dominio dinámico primario del SN.

La naturaleza electroquímica de los impulsos permite una velocidad de propagación y una velocidad de desencadenamiento y supresión excepcionales, sin que interfieran catastróficamente con otros procesos metabólicos. Ningún otro dominio biofísico presenta estas características. La forma de impulsos que toman los potenciales de acción es un aspecto fundamental de los observables. Sin embargo la naturaleza electroquímica de

los cambios de potencial no implica que la dinámica electroquímica tome la forma de impulsos discretos, activamente generados en forma de potenciales de acción. Sin embargo, si la dinámica electroquímica ha de propagarse sin decaer (es decir, sin pérdida disipativa de la variedad dinámica) es necesario que el impulso se organice activamente y en forma de picos, ya que la amplitud del potencial recibido no puede transmitirse mucho más allá de 1 ó 2 mm por sí mismo. “De hecho, aproximadamente a 1mm a lo largo del axón la amplitud de la señal es sólo un tercio de la que tenía en el lugar de su producción. Para ser transportada con éxito al resto del sistema nervioso central, la señal local ha de ser amplificada; tiene que generar un potencial de acción.” (Kandel *et al* 2000: 29). Por eso, la característica dinámica más significativa de los impulsos reside en su activación no-lineal (todo o nada), una característica que fue ya formulada en la primera ley de Adrian en 1926. La no-linealidad de los impulsos permite una organización de impulsos en términos de frecuencias de disparo y secuencia temporal de los mismos, por lo que constituyen un tipo de procesos idóneos para soportar una dinámica autónoma, plástica, autoorganizada y compleja, como detallaremos más adelante.

Para los efectos de la especificación de este nuevo nivel dinámico todos los procesos moleculares subyacentes son irrelevantes ya que sólo cuenta el todo o nada del impulso. A pesar del esfuerzo metabólico (en términos energéticos) que supone el bombeo de iones a través de la membrana en el proceso de polarización y despolarización, la generación del impulso (el “disparo”) en una neurona determinada no depende exclusivamente de su estado interno, sino del contexto electroquímico, generado a su vez por otras neuronas y ciertas sustancias neuromoduladoras. Es por ello que el impulso es el observable primario del SN como nivel dinámico desacoplado.

Ahora bien... ¿qué aspecto del impulso es el que constituye realmente el observable primario del SN? El todo o nada del potencial de acción y la uniformidad de la amplitud del disparo en las neuronas nos lleva a la conclusión de que las variables relevantes son las relativas a la organización temporal de los impulsos (frecuencias, secuencias, ajustes temporales, etc.) que son además las características pueden conservarse en la transmisión del impulso a través de la neurona. Podemos derivar la organización temporal de los impulsos de dos observables fundamentales: la existencia o no de impulso

y el tiempo en que ocurre. Digamos que no hay variables dinámicamente relevantes *por debajo* de estas magnitudes u observables primarios. Los diversos patrones de impulsos, las combinaciones de ellos y las resonancias que pueden ir generando, constituirán observables de ordenes superiores. Podemos decir por tanto que los potenciales de acción constituyen los *observables primarios*, y las secuencias temporal de los impulsos constituyen *observables de primer orden*. Ulteriormente las redescripciones combinatorias, estadísticas o resonantes que capturen una organización causalmente relevante de orden superior constituirán *observables de órdenes superiores* (dentro del dominio de los observables primarios).

En definitiva, *el número de potenciales de acción y los intervalos de tiempo entre ellos son los observables primarios del SN*, ya que la clase de distinciones internas necesarias y suficientes para generar y/o explicar la actividad nerviosa principal que une dinámicamente las superficies sensoras y efectoras en la producción de conducta adaptativa es precisamente la de la propagación de estos impulsos. Sin embargo para la comprensión de la dinámica de los impulsos es fundamental comprender la conectividad entre las neuronas, es decir, tener en cuenta la arquitectura que sostiene las posibles trayectorias de impulsos a través del SN.

La matriz de conectividad

Los impulsos operan en una red cuya arquitectura está a su vez definida por la *matriz de conectividad* entre las neuronas. La matriz de conectividad es una abstracción del conjunto de conexiones sinápticas del SN. Esta arquitectura básica se configura en el proceso genéticamente instruido de desarrollo ontogenético del organismo adulto. Pero tan pronto empieza a funcionar como soporte de procesos sensomotores interactivos, la dinámica que se induce afecta a la estructura misma de la matriz de conectividad. La transformación de la matriz de conectividad a través del surgimiento de nuevas conexiones neuronales (por crecimiento dendrítico) o el atrofiamiento de las conexiones en desuso es un factor fundamental de los procesos neurocognitivos (Quartz & Sejnowski 1997). Entonces, por un lado la arquitectura de la red aparece como una variable en la escala de desarrollo del organismo, pero desde la escala dinámica de los procesos de impulsos aparece como inerte y actúa como una serie de parámetros de conectividad que determinan las *posibles trayectorias* de un tren de impulsos. Así, la matriz de conectividad aparece como soporte de los observables primarios y

su transformación vendrá marcada principalmente por los observables secundarios que analizaremos a continuación.

Observables secundarios (neuromoduladores)

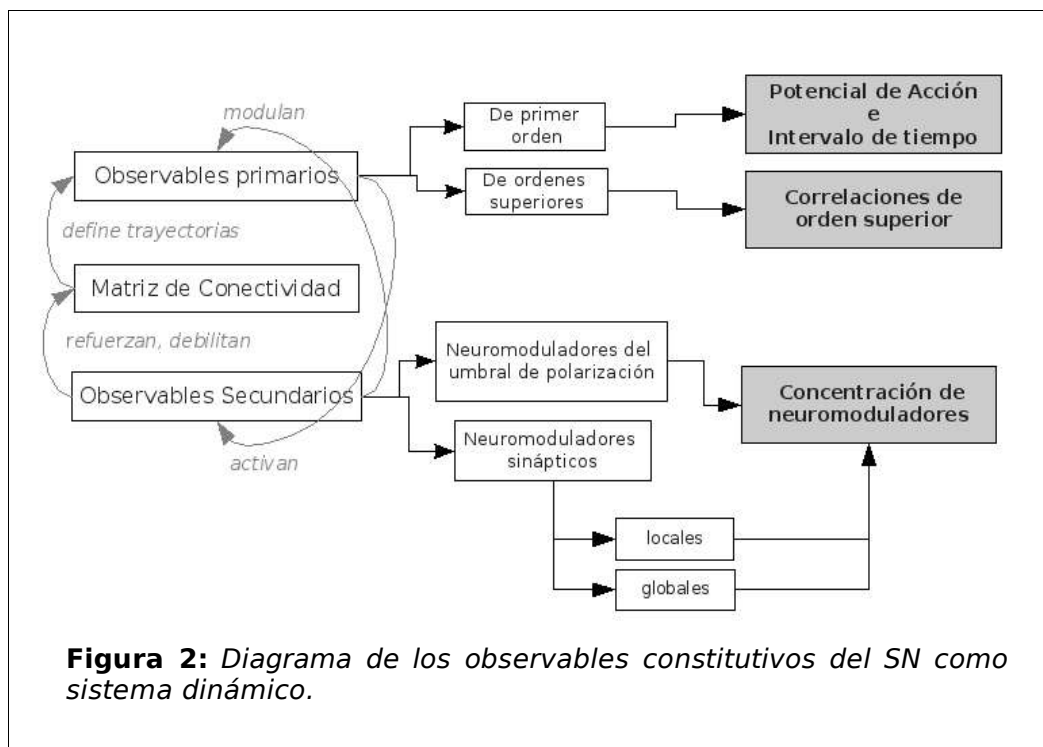
Aunque observacionalmente irreductibles a los impulsos, los observables que llamamos secundarios están dinámicamente determinados por los primarios; es decir, los observables primarios activan los secundarios y, en general, gobiernan su dinámica. En cierto modo, los observables secundarios son la forma en la que las señales actúan sobre sí mismas a escalas temporales y con propiedades modulatorias diferentes a las que puede generar la red de impulsos electroquímicos por sí misma. Aunque la influencia es recíproca y recurrente (entre observables de primer y segundo orden) los impulsos "activan" los cambios de valor de los observables de segundo orden, y éstos devienen observables sólo en la medida en que los impulsos son a su vez observables (y en la medida en que los afectan). Todos estos observables son secundarios porque puede haber una red mínima sólo con impulsos, pero no puede haber una red sólo con modificaciones sinápticas, o con cambios de umbral de despolarización de membrana. Los observables secundarios pueden agruparse bajo el concepto de *neuromoduladores*; dinámicamente hablando pueden considerarse como operadores que actúan sobre el nivel dinámico primario (Marder & Thirumalai 2002) generalmente a escalas temporales superiores a las de los observables primarios¹⁷(ver figura 2).

Distinguimos dos tipos de observables secundarios o

¹⁷ Recientes descubrimientos señalan a que lo que hemos considerado como observables secundarios pueden organizarse en redes que generan una dinámica propia de gran complejidad y que actúan en interacción con los observables primarios más allá de una simple modulación de parámetros. En concreto el reciente descubrimiento de la difusión de los gases neuromoduladores como el NO, CO y H₂S y el análisis de sus efectos en la producción de conducta adaptativa simulada (Philippides *et.al.* 2000) ha generado el concepto de *señales de volumen* que se propagan en redes de gases. Esto podría hacernos pensar en procesos dinámicos internos primarios; sin embargo, las redes de gases actúan de forma paralela, pero inducida por los impulsos neuronales y en espacios localmente constreñidos. Otro ejemplo lo constituyen las complejas interacciones que rigen la relación entre neuronas y células gliales y, especialmente, entre éstas últimas: "(...) una célula glial percibe los potenciales de acción de las neuronas al detectar el ATP que se escapa por un axón activo o bien el que se libera en una sinapsis. La célula glial transmite su mensaje interno vía iones de calcio. Estos, a su vez, activan enzimas que liberan ATP a otras células gliales o activan enzimas que controlan la expresión de genes." (Fields 2004: 12). Esta interacción entre glías crea una red de intercambio de ATP que afecta al proceso de mielinización que aísla a las neuronas pero, también y sobre todo, a las sinapsis, afectando y modulando su plasticidad. Incluso se baraja la hipótesis de un rol fundamental de los circuitos de astrocitos (un tipo de células gliales) en la formación de nuevas conexiones sinápticas y en el reforzamiento de señales interneuronales a través de circuitos locales de astrocitos. La especificidad de los receptores químicos de los astrocitos permitiría incluso hablar de una conectividad por difusión a largas distancias en el cerebro, lo que permitiría activar regiones neuronales distantes (sin conectividad neuronal directa) a través de la comunicación química entre circuitos de astrocitos.

neuromoduladores:

1. **Neuromoduladores que actúan sobre el umbral de polarización de la membrana neuronal:** El umbral de polarización establece el parámetro fundamental que actúa como selector de los impulsos proveniente de las neuronas locales; especifica a su vez las propiedades no-lineales de respuesta de una neurona. Diversos neuromoduladores pueden hacer variar las propiedades de la membrana y el umbral de despolarización.



2. **Neuromoduladores que actúan sobre los neurotransmisores de las sinapsis.** Los neurotransmisores establecen el puente de conectividad del impulso electroquímico entre neuronas en la sinapsis. Muchos neuromoduladores transforman químicamente la actividad de estos neurotransmisores modificando las propiedades sinápticas, modulando así la transmisión de impulsos en el SN. Estas modificaciones pueden ser inducidas localmente (intrínsecas) o externamente (extrínsecas). Los cambios localmente inducidos responden a la actividad relativa entre los impulsos de neuronas adyacentes. Las modificaciones extrínsecas se generan a partir de la segregación (controlada a su vez por neuronas y rutas neuronales específicas) de neurotransmisores globales (como la dopamina o la serotonina) controlados a su vez por circuitos específicos de neuronas dentro del SN (como el sistema dopaminérgico).

La principal función de los observables secundarios es dotar al SN de plasticidad dinámica, es decir, de variabilidad y adaptabilidad de sus procesos dinámicos primarios (los de los potenciales de acción). Como veremos más adelante, esta plasticidad va a ser fundamental para generar una dinámica selectiva a escalas temporales diversas, para generar constricciones autoorganizativas en la dinámica global del SN. Características todas ellas que van a permitir al SN no sólo responder adaptativamente a condiciones del entorno (*adaptividad*) sino cambiar la forma en que el organismo responde a un mismo conjunto de estímulos ampliando la *adaptabilidad* ontogenética del mismo (a través del aprendizaje, la memoria, etc.).

Con estos elementos se puede producir una serie de interacciones recurrentes entre los observables primarios y los secundarios que (sin necesidad de muchas unidades) es capaz de generar una enorme complejidad dinámica. Esta complejidad (que crece exponencialmente con el número de neuronas y neuromoduladores) es tal que, aún cuando su realización biofísica depende de los procesos metabólicos que la sostienen, su dinámica se encuentra principalmente determinada por los procesos recurrentes que la definen (y su acoplamiento adicional al entorno a través de la interfaz corporal sensoriomotora). Es por ello que, desde la perspectiva de la dinámica causal que gobierna el SN, los procesos metabólicos subyacentes son más bien considerados como condición de posibilidad que como determinación de su dinámica. De ahí, una vez más, el desacoplamiento del SN.

3.2.3 Autoorganización y autonomía en el sistema nervioso

Una vez aislados los observables, cabe preguntarse cuál es su comportamiento dinámico y qué principios organizativos los gobiernan. Como ya adelantáramos al principio de este capítulo, tan pronto como, en el curso del desarrollo del organismo, la red alcanza una arquitectura funcionalmente operativa, la dinámica de impulsos se organiza en torno al mantenimiento de ciclos interactivos viables. Desde la perspectiva interna del funcionamiento del SN estos ciclos interactivos viables son una serie de invariantes sensoriomotoras o patrones globales de la dinámica neuronal generados en interacción recurrente con el entorno. Ahora bien, ¿cuales son los principios organizativos de esta dinámica neuronal?

Empezando por el nivel de complejidad más básico, la forma más simple de funcionamiento sensoriomotor que podemos imaginar está soportada sobre todo por una arquitectura innata que fija conexiones sensoriomotoras lineales de tipo arco reflejo en la que los impulsos generados en las superficies sensoras se propagan por rutas predeterminadas hasta desencadenar el proceso motor evolutivamente fijado. Sin embargo, este tipo de circuitos reactivos es muy poco común (incluso en organismos primitivos) ya que son muy sensibles a perturbaciones internas y externas, y es difícil imaginar cómo pueden estar genéticamente especificados patrones de conectividad directamente funcionales, si no es a través de procesos de autoorganización¹⁸. Si, como defienden Elman *et al.* (1996) los únicos niveles genéticamente especificables de la dinámica neuronal son los de constricciones de arquitectura (que no incluyen los detalles de la matriz de conectividad sino tipos de neurotransmisores, presencia o ausencia de recurrencia entre neuronas o rutas neuronales globales) y constricciones cronotópicas (sincronización de procesos de desarrollo) entonces la matriz de conectividad funcional y la dinámica neuronal deben ser el resultado de procesos de autoorganización (Bechtel & Abrahamsen, 2004). De hecho si la conducta de un organismo dependiera de circuitos lineales directamente codificados en el genoma, una pequeña variación en el desarrollo tendría consecuencias catastróficas¹⁹.

Lo más común en la dinámica sensoriomotora, en efecto, es que ésta surja como resultado de la interacción recurrente y en red de todo un conjunto de neuronas, sin que exista un recorrido sensoriomotor lineal y unidireccional que determine la conducta del organismo. Desde esta perspectiva la estimulación perceptiva aparece como un conjunto indefinido de perturbaciones (trenes de impulsos) en el estado dinámico de la red neuronal. Algunos de estos trenes de impulsos son seleccionados o reclutados

¹⁸ Esto se debe a que la información necesaria para especificar circuitos neuronales sobrepasa enormemente la capacidad del genoma. En el caso del ser humano si los circuitos neuronales estuvieran genéticamente preespecificados eso significaría que 10^6 genes (de los cuales sólo un 20-30% participan en la construcción del SN) tendrían que almacenar la información suficiente para codificar 10^{14} sinapsis que, a su vez, pueden tomar valores cuantitativos de un espectro considerablemente amplio. Esa reducción de variabilidad (del conjunto de parámetros y conexiones posibles a las que resultan funcionales para el organismo) sólo puede darse a través de procesos autoorganizativos que seleccionen y establezcan configuraciones específicas. Por lo tanto resulta matemáticamente imposible que la matriz de conectividad final se encuentre genéticamente determinada. Son necesarias otras fuentes de orden para fijarla.

¹⁹ La existencia de estos circuitos reactivos (como la respuesta del reflejo rotuliano) se debe a efectos secundarios del desarrollo del SN o como especificaciones locales de una dinámica neuronal más compleja.

por otro patrón interactivo de orden superior. La estimulación perceptiva es, pues, modulada desde dentro: ciclos sensomotores, internamente generados y regulados, estabilizan un patrón sensorial para integrarlo en el ciclo sensomotor. A su vez, nuevos patrones motores son generados para mantener una serie de invariantes internas. Estos patrones se estabilizan (se seleccionan o activan) en forma de atractores, resonancias, y otras estructuras dinámicas cuando contribuyen al automantenimiento del sistema²⁰. La corporización interactiva del SN hace que este proceso de mantenimiento de invariantes se cierre en ciclos sensomotores con el entorno que se integran funcionalmente para el mantenimiento de los patrones globales. También puede suceder que, como resultado de ciclos acción-percepción-acción, el patrón global neurodinámico cambie y el organismo adopte otro tipo de interacción con el entorno.

En una escala de complejidad aún mayor encontramos que la propia dinámica de la red genera procesos de aprendizaje. La forma más básica de los procesos de aprendizaje funcional es la del reforzamiento de las conexiones que han dado lugar a una conducta de valor positivo para el organismo. En este mecanismo son fundamentales los procesos de activación de los observables de segundo orden que actúan fijando las rutas funcionales de impulsos, generando así un proceso de *selección* interna (Edelman 1987). La forma general de este proceso es que los neuromoduladores extrínsecos (como la dopamina) regulan la estabilidad de aquellos circuitos (o matrices de conectividad) que generan procesos interactivos de valor adaptativo. El aprendizaje actúa pues como la modulación de las condiciones de estabilidad de la dinámica neuronal que sostiene las interacciones con el entorno.

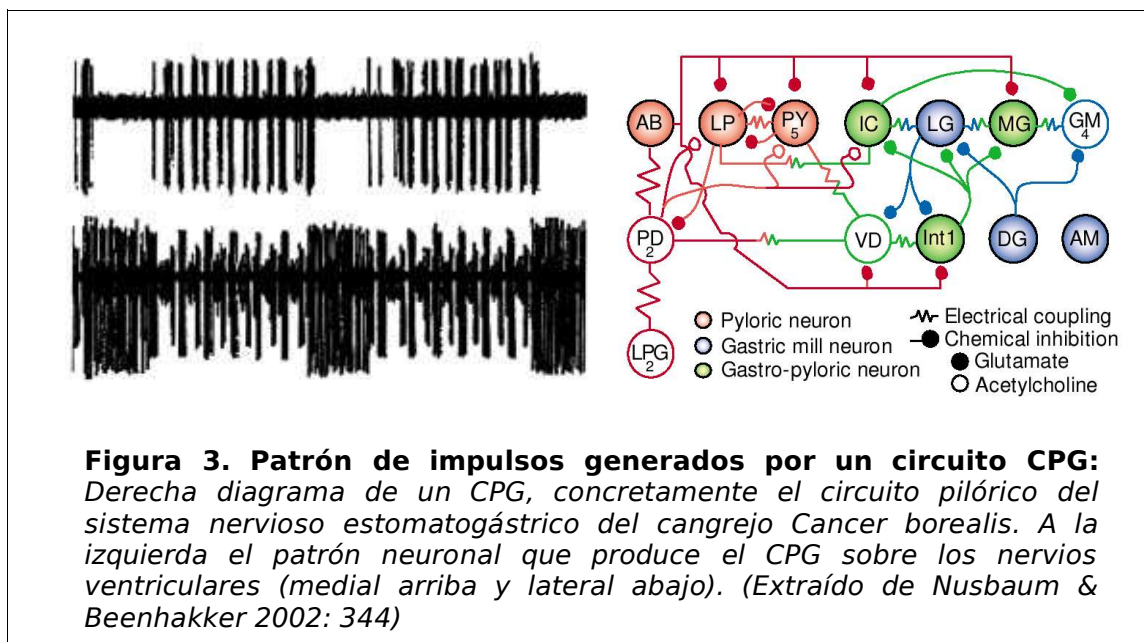
Procesos autoorganizados en el SN

Lo que venimos describiendo son fenómenos típicamente *autoorganizativos* en los que una serie de interacciones locales generan un patrón automantenido a nivel global que es robusto frente a perturbaciones internas y externas. Más explícitamente, entendemos por autoorganización (en sentido general) “el fenómeno por el cual una serie de procesos microscópicos no-lineales generan un patrón de comportamiento

²⁰ De todos modos este criterio regulatorio de adaptabilidad global resulta insuficiente para regular la compleja y variable dinámica interneuronal. Es por ello que ésta aparece autoorganizada y generará progresivamente sus propios mecanismos de regulación. Dicho de otro modo: el automantenimiento del sistema no actúa directamente como variable de control sobre la dinámica neuronal sino que se realiza a través del mantenimiento de otras invariantes globales generadas y sostenidas por procesos autoorganizativos.

macroscópico-colectivo en una situación de alejamiento del equilibrio termodinámico mantenida por la acción continua de un conjunto de constricciones, entre las cuales al menos una es producto del propio fenómeno” (Ruiz-Mirazo 2001). Si bien el concepto de autoorganización ha sido desarrollado principalmente para su aplicación a fenómenos de carácter termodinámico, su aplicación al SN es muy común, especialmente entre algunos dinamicistas (Skarda & Freeman 1987, Kelso 1995, Floreano & Urzelai 2000, Kjeizer 2002, Varela *et. al* 1991, Varela 1997, Thompson & Varela 2001, Pfeifer & Scheier 1999). En el caso del SN el conjunto de fenómenos microscópicos no-lineales lo constituyen los trenes de impulsos y el alejamiento del equilibrio viene marcado por la tendencia a la disipación de los impulsos electroquímicos (que, sin embargo, se mantienen activos gracias a la dinámica global del SN y a la energía metabólica que la sostiene). Las constricciones que mantienen el patrón pueden ser de diversos órdenes (estructuras innatas, neuromoduladores, estabilidad homeostática de las neuronas, inputs de otras regiones neuronales, etc.) de las cuales “al menos una” es generada por el propio circuito neuronal que soporta el patrón autoorganizado. La auto-generación de constricciones en la dinámica neuronal viene generalmente marcada por la naturaleza recurrente de los circuitos neuronales y por la interacción entre observables primarios y secundarios.

Podemos comprender estos procesos como organizadores de la enorme complejidad dinámica que generan los observables. Una complejidad que difícilmente puede ser controlada por constricciones innatas o provenientes del organismo. Los procesos autoorganizados, por el contrario, pueden coordinar espontáneamente la acción de múltiples neuronas; a su vez el patrón global actúa selectivamente sobre el conjunto de la dinámica microscópica, generando un orden global imposible de generar exclusivamente a través de constricciones genéticamente preespecificadas.



Un ejemplo paradigmático de autoorganización dinámica a escala neuronal en circuitos locales es el de los CPG. Los CPG (o *central pattern generator* —ver figura 3) son conjuntos de neuronas que generan patrones multiestables de impulsos que controlan las neuronas motoras (Arshavsky *et.al.* 1997). Estos patrones pueden, incluso, ser generados en ausencia de inputs perceptivos o de otras zonas del SN. La generación de patrones estables por el CPG es crucial para constreñir los grados de libertad del sistema muscular y generar un movimiento coherente y coordinado. Veamos cómo funciona este proceso autoorganizado. Algunas de las neuronas que componen los circuitos de los CPG generan impulsos rítmicos por sí mismas, estas neuronas se acoplan entre sí y con el resto de las neuronas del circuito generando un patrón global estable. Los CPG son capaces de mantener ese patrón frente a diversas perturbaciones, incluyendo lesiones a partes del circuito. Pero los CPG combinan estabilidad y variabilidad, permitiendo la existencia de diversos patrones estables generados por el mismo circuito. Diversos neuromoduladores y la actividad de neuronas adyacentes al CPG son capaces de provocar cambios de patrones, actuando como selectores de atractores de acuerdo a las necesidades motoras del organismo. Además, a pesar de que los CPG de dos organismos de la misma especie generen patrones similares, sus circuitos suelen ser diferentes. Así pues, diferentes circuitos pueden generar los mismos patrones, y el mismo circuito puede generar diferentes patrones estables; cumpliendo así con una característica

común de los sistemas autoorganizados (como bien analiza Kelso 1995: 239—243).

La característica general de los sistemas autoorganizados es que el patrón global es capaz de *seleccionar* algunas fluctuaciones locales, que a su vez refuerzan dicho patrón, generando una estabilidad recursivamente mantenida. Se habla de autoorganización porque el patrón global no es externamente especificado (no es instruido) y porque es el resultado de un conjunto de componentes básicos que interactúan paralelamente, pero donde ninguno de los componentes aisladamente es responsable del surgimiento del patrón global; por tanto no hay algo así como una instrucción localizable. Sin embargo, los procesos autoorganizados dependen de unas condiciones de contorno (*boundary conditions*) y de un conjunto de constricciones que no son parte del propio proceso de autoorganización ni están bajo su control (puede ser el caso de algunas constricciones arquitectónicas de la matriz de conectividad, de ciertos neuromoduladores extrínsecos, señales de otras áreas del SN, etc.). En este sentido los procesos autoorganizados dependen, para su mantenimiento, de ciertas condiciones de contorno estables, sin las cuales el proceso de autoorganización se diluye o cambia de estado.

La organización del SN en procesos dinámicos autoorganizados en diferentes escalas hace que se generen estructuras intermedias (patrones), y que éstas entren a interactuar entre ellas y a organizar la dinámica global del sistema más allá de las interacciones locales de impulsos. Esto nos obliga a distinguir entre niveles dinámicos micro y macro (y entre diversas jerarquías de niveles) y nos permite también hablar de una funcionalidad emergente (es decir, de procesos microscópicos neuronales cuya contribución al automantenimiento del sistema viene marcada por los patrones emergentes que constituyen o en los que participan). Por un lado, la organización de los impulsos neuronales en procesos masivamente paralelos y autoorganizados hace que se diluya la causalidad efectiva de cada uno de los impulsos específicos. Por otro lado, la transformación de patrones globales regenera una causalidad funcionalmente efectiva a un nivel superior. Esto permite hablar de una microdinámica de la que emerge una macrodinámica neuronal funcional²¹.

²¹ La consecuencia principal de este emergentismo causal es la dificultad de una estrategia localizacionista en el estudio de los procesos cognitivos. El localizacionismo opera haciendo una descomposición estructural del sistema y otra funcional y estableciendo un mapeo entre ambos tipos de descomposición para dar lugar a una explicación causal estructura-función en la que los componentes se agregan linealmente para reconstruir los procesos causales que

Ahora bien ¿qué sucede con la dinámica neuronal entendida a escala global de todo el SN? ¿Depende ésta de constricciones externas, o es capaz de mantener autónomamente sus procesos constitutivos? La clave para resolver este problema viene de la mano de la autonomía del sistema nervioso.

La autonomía del SN como clausura operacional

Como señalábamos anteriormente, al contemplar el conjunto del SN desde una perspectiva global (holista), el dibujo que va tomando forma es el de una *dinámica autónoma* de los procesos neuronales (superando las limitaciones de automantenimiento de lo que son los procesos autoorganizativos locales del SN —como el caso de los CPG). La transición entre autoorganización y autonomía viene marcada no sólo por el cambio de escala sino (sobre todo) por la acción recurrente del sistema autoorganizado sobre sus propias constricciones y condiciones de contorno tanto interna como interactivamente. Si un sistema autoorganizado es aquél en el que al menos una de las constricciones que lo constituyen como sistema alejado del equilibrio es autogenerada la dinámica del SN va más allá ya que es un sistema que genera, recursivamente, la mayoría de sus constricciones constitutivas de tal modo que produce un proceso histórico de autodefinition al que llamamos autonomía. Veamos cómo se puede explicitar la noción de autonomía en el SN.

El biólogo y neurocientífico Francisco Varela fue quizás uno de los pioneros en describir el SN como un sistema autónomo y, más concretamente, un sistema operacionalmente cerrado (en su propia terminología sistémica):

"Autonomous systems are mechanistic (dynamic) systems defined as a unity by their organization. We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist." (Varela 1979: 55 —cursiva en el original)

Veamos en qué sentido el SN, como lo venimos definiendo satisface esta definición. Dos condiciones establecen la clausura operacional de un sistema (y por tanto su autonomía). La primera es la dependencia recursiva

gobiernan al sistema (Bechtel & Richardson 1993). Sin embargo, la autoorganización del funcionamiento dinámico que hemos visto no permite llevar a cabo con éxito esta estrategia, ya que la funcionalidad emerge de la interacción no-lineal entre los componentes así como entre su cuerpo y su entorno (Steels 1991, Clark 1997).

en red de los procesos del sistema. La segunda la constitución del sistema como una unidad reconocible en el entorno en el que existe.

La primera condición queda satisfecha por la naturaleza reticular del sistema nervioso, la dinámica recursiva y autoorganizada que la genera, en concreto la dependencia holista del estado de los observables en la producción de conducta funcional. La actividad del SN es la de una red interdependiente de procesos y no la de la interacción lineal entre módulos o bloques de procesamiento aislados. En efecto los actos neurocognitivos exigen la integración de la dinámica de múltiples áreas del cerebro en constante interacción y sincronización: la actividad funcional del SN responde a la actividad neuronal coherente e integrada a gran escala de diferentes partes del cerebro (Tononi & Edelman 1998; Varela *et. al* 2001 y Le Van Quyen 2003).

La segunda característica (la distinción del sistema como una unidad diferenciada de su entorno) ha sido defendida por Varela en términos de la individualidad de un organismo con SN como unidad móvil en el espacio, es decir, el SN como sistema de control de la motilidad del organismo adquiere una diferenciación con su entorno.

Pero el concepto de clausura operacional de Varela va más allá: no se trata solamente de una dependencia holista de los procesos constitutivos de un sistema y de su diferenciación con el entorno sino de que esa dependencia holista se encuentra cerrada sobre sí misma de tal modo que la interacción con el entorno no toma la forma de inputs y outputs sino de perturbaciones cuyo efecto el propio sistema determina: "(...) the nervous system operates as a closed network with no inputs or outputs, (...) its cognitive operation reflects only its organization, and (...) information is imposed on the environment and not picked up from it." (Varela 1979: 238).

Sin embargo este concepto de autonomía del SN es objetable. La interacción con el entorno es necesaria para el mantenimiento de las invariantes internas que definen al sistema. La actividad del SN se encuentra situada e integrada en un entorno interactivo y, si embargo, la definición de autonomía a través del concepto de clausura operacional excluye referencia alguna a un entorno del que sin duda depende la autonomía del sistema. De este modo concepto de clausura operacional dificulta la tarea de integrar interacción y autonomía. Aunque para Varela un sistema operacionalmente cerrado no es un sistema interactivamente cerrado, la interacción con el

entorno aparece siempre como un acoplamiento estructural posterior o adicional a la constitución del sistema como sistema autónomo. Por tanto el concepto de clausura operacional dificulta entender el papel *constitutivo* de la interacción en la propia autonomía del SN²².

La autonomía interactiva del SN

Entendido como un todo, el SN controla sus inputs perceptivos a través del control de la conducta en ciclos acción-percepción-acción (Powers 1973) y es también capaz de modificar su matriz de conectividad gracias a la interacción entre observables primarios y secundarios. El cierre del ciclo sensomotor a través del entorno hace que la dinámica del SN no responda a una mera autoorganización local reticular de los procesos, sino a una integración del entorno en el mantenimiento dinámico de las invariantes del SN. La perspectiva corporizada y situada (que hemos analizado en el segundo capítulo) así como la ecológica (Gibson 1979) muestran cómo la estabilidad (e inestabilidad) de los procesos neuronales está profundamente imbricada en el entorno sensomotor en el que opera, hasta el punto de estructurar la propia arquitectura dinámica interna en un proceso de construcción interactiva (convirtiendo al entorno en una especie de andamio de la dinámica neuronal —Clark 1997). La autonomía del SN está pues abierta a su entorno ya que los procesos interactivos forman parte constitutiva de ese proceso de automantenimiento y autorregulación.

De este modo la centralidad de la dinámica interactiva en la producción de la autonomía neuronal nos obliga a ensanchar la definición de Varela para incluir esta dimensión interactiva como propiamente constitutiva de la autonomía del SN (una dimensión que Varela sólo integra a posteriori en forma de acoplamiento estructural entre el SN y el entorno). Si la interacción con el entorno deviene constitutiva de la autonomía del SN surgen preguntas que parecen poner en peligro la autonomía del SN: ¿podemos distinguir la dinámica del SN del conjunto de la dinámica interactiva SN-cuerpo-entorno? ¿no desaparece, de este modo, la autonomía del SN respecto del entorno? ¿podemos hablar de un sujeto cognitivo diferenciado del entorno en el que opera? La característica crucial que permite salvar estas preguntas de forma satisfactoria reside en que *la neurodinámica interna del SN es más estable y*

²² En un segundo momento de su investigación en ciencias cognitivas Varela *et. al* (1991) desarrollan el concepto de *enacción* y de *ciencias cognitivas enactivas* subrayando el énfasis en la interacción y la corporización pero dejando la noción de autonomía ligeramente de lado. Creemos que el concepto de autonomía puede y debe recuperarse desde una perspectiva interactiva.

compleja que la interacción que establece con el entorno. Algo que no podría ser de otro modo, ya que la dinámica del SN responde en última instancia a las necesidades de automantenimiento del organismo, y la satisfacción de estas necesidades no puede ser impuesta desde fuera. Esta *asimetría de complejidad* permite integrar autonomía e interacción sin pérdida de identidad y ofrece además un criterio más general y preciso de diferenciación entre sistema y entorno. Podemos, pues, definir la autonomía del SN del siguiente modo:

La dinámica del SN es la de un sistema autónomo ya que:

- ◆ La estructura dinámica y alejada del equilibrio de la red es mantenida por:
 - a) la propia red de procesos (cohesivos, recursivos y metaestables) y
 - b) un proceso interactivo con el entorno; y
- ◆ la estructura dinámica de la red se distingue de su entorno por su integración funcional, produciendo una *asimetría de complejidad* en el flujo dinámico tal que los procesos internos son más complejos que los interactivos y permiten mantener la identidad del sistema a través de la diversidad de acomplamientos con el entorno.

De este modo la integración funcional (la dependencia dinámica entre los procesos) y cohesiva del SN es la fuente de un desequilibrio causal entre sistema y entorno que permite denominar agente al primero y lo define como autónomo. Esta cohesión interna de los procesos neuronales permite además recuperar la funcionalidad interactiva cuando ésta se pierde por algún motivo ya que la cohesión se mantiene a pesar de la pérdida momentánea del acomplamiento con el entorno. Un sistema autónomo es por tanto capaz de restablecer y regular una nueva interacción funcional. La interacción con el entorno es funcional y constitutiva y sin embargo mucho más variable e inestable que la interacción entre los observables del sistema. La diferenciación entre sistema y entorno es por tanto el fruto de la permanencia de invariantes internas frente a la variación en la interacción.

Podría uno preguntarse si la autonomía del SN así descrita no se contradice con la necesidad del SN de satisfacer el automantenimiento del organismo que aparecería así imponiendo al SN objetivos heterónomos. La respuesta es que no hay tal contradicción: la autonomía del SN implica una acción complementaria con la autonomía del organismo en el que se inserta, una dialéctica entre condiciones de posibilidad y autorregulación. En

concreto, la autonomía del SN nunca es completa ya que existen toda una serie de constricciones que escapan a su control y que son precisamente las que van a servir de constricciones globales a todo el funcionamiento del SN. Se trata de las constricciones innatas (arquitectónicas y cronotópicas —Elman *et.al.* 1996) y de las señales de valor (*value signals*) que el organismo envía al SN para indicar el valor adaptativo de las interacciones que éste establece. El dolor, el placer, la saciación, el frío, etc., son señales de valor típicas a las que la dinámica del SN responde de forma especialmente sensible y que, de alguna manera, monitorizan la actividad global del SN. Sin embargo, estas constricciones de valor adaptativo *subdeterminan* la actividad del SN, es decir, no son suficientes para explicitar completamente su dinámica. En este sentido, el desacoplamiento jerárquico del sistema nervioso implica el surgimiento de una variabilidad mucho mayor que la que el organismo por sí mismo pueda controlar, y supone también el surgimiento de un dominio regulatorio propio. Es aquí donde podemos hablar de una autonomía (parcial, pero propia) del SN. Por tanto, al hablar de la autonomía del SN no nos estamos refiriendo a un proceso acabado y estático, sino a un proceso abierto y siempre en desarrollo (un *becoming*). Desde los primeros estadios del desarrollo, la dinámica autoorganizada del SN (abierta interactivamente con el entorno) empieza a generar procesos regulatorios propios que superan progresivamente las constricciones genéticamente marcadas. La identidad del SN (su organización regulatoria) es, por tanto, indisociable de su historia de interacciones y de las constricciones que va generando a lo largo de esa historia (Rose 1999). Algunas de estas constricciones autogeneradas, pueden, a su vez, autonomizarse en forma de hábitos automantenidos convirtiéndose así en fuentes de intencionalidad (Di Paolo 2003).

Desde una perspectiva interactiva la autonomía deja de ser, pues, un término absoluto (como en Maturana y Varela o el propio Rosen —y su concepto de sistema cerrado a causalidad eficiente) y deviene una noción gradual y cuantitativa. De hecho cabe esperar que las nociones de complejidad e integración funcional que se están desarrollando en el área de las neurociencias cognitivas (Tononi *et.al.*, 1998) y de la robótica evolutiva (Seth & Edelman 2004) permitirán pronto contar con una definición operacional cuantitativa de autonomía (en el sentido descrito en esta sección) y desarrollar métodos que permitan²³ naturalizar la noción de autonomía en

²³ Un trabajo pionero en esta línea fue ya emprendido por Ashby (1952,1956) a comienzos de los años 50 y desembocó en la ley de variedad requerida según la cual la variabilidad interna de un sistema ha de ser mayor que la de aquellas perturbaciones que ha de compensar para

términos matemáticos aplicables al análisis de redes neuronales. Ésta es sin duda una de las tareas más urgentes de la propuesta naturalizadora que aquí presentamos.



Cabe preguntarse ahora si existe algún principio organizativo propio de esa normatividad y funcionalidad específica de la autonomía del SN. Un principio organizativo añadido a la forma genérica de dinámicas autoorganizadas que hemos descrito en esta sección y a las constricciones innatas que determinan al SN en su origen ontogenético. Cabe preguntarse también qué relación guardaría este principio organizativo interno del SN con su entorno y, finalmente, si esta relación permitiría hablar de un dominio propiamente cognitivo. La respuesta la va a dar el surgimiento de un nuevo tipo de organización dinámica: la información cognitiva.

3.3 Causalidad informacional en procesos neuro-interactivos

Toda la forma de organización que hemos descrito en las secciones anteriores nos dibuja un escenario con un enorme número de variables e interacciones, cuya dinámica concreta no puede ser especificada por la organización básica (metabólica y genética) del organismo. El SN va generando, a través de los ciclos interactivos, principios que operan selectivamente (unas veces a escala local y de forma poco duradera, otras más global y permanente) organizando funcionalmente esa dinámica interactiva. La forma en la que la dinámica neuronal se autoorganiza generando patrones estables de orden superior (invariantes) así como los procesos interactivos entre dichos patrones establecen una serie de estructuras causales de orden superior que soportan las interacciones del organismo y que exigen/permiten la introducción de un nuevo vocabulario.

Hasta ahora hemos venido describiendo la actividad del SN en términos puramente dinámicos. Sin embargo la literatura neurocientífica está cargada de términos que superan el marco descriptivo molecular o bioquímico, introduciendo conceptos como el de *señal*, *código*, *información* o *contenido*. ¿Se puede justificar la introducción de estos conceptos desde las propias necesidad explicativas de la práctica neurocientífica? ¿Tenemos

mantener constantes las variables que controla.

instrumentos y marcos conceptuales apropiados para integrar los niveles mecánicos (aquellos sobre los que podemos operar con instrumentos) y el informacional o semántico?. Con lo que venimos diciendo hasta ahora podemos ya justificar este salto conceptual y definir en qué sentido los potenciales de acción (y algunos neuromoduladores) y, sobre todo, los patrones de orden superior que generan al autoorganizarse pueden denominarse *señales* y adquieren propiedades informacionales.

3.3.1 Los observables como señales

Para que los potenciales de acción constituyan señales (y en general para que cualquier proceso devenga señal) hace falta que se genere un nivel dinámico desacoplado de carácter composicional y recursivo y que, adicionalmente, lo que suceda a dicho nivel sea susceptible de “interpretación” a un nivel superior, es decir, los procesos composicionales han de poseer la potencialidad para una *causalidad formal*. Analicemos por separado estos dos aspectos.

El desacoplamiento jerárquico del SN ha sido extensamente tratado en las secciones anteriores. Recordemos, sin embargo, que la conductividad electroquímica activa de los componentes del SN (las neuronas) organizada en modo de impulsos o potenciales de acción (que conservan a su vez la variabilidad dinámica a modo de frecuencias y distancias temporales) supone un nivel dinámico (una serie de variaciones diferenciables en el tiempo) que se construye tangencialmente a los procesos metabólicos del organismo, aunque sostenido y realizado en ellos. A su vez, la característica de todo o nada de los impulsos nerviosos permite una combinabilidad estable de los mismos que, unida a las características de red de la matriz de conectividad y la acción de los neuromoduladores, genera un dominio composicional (los impulsos pueden organizarse secuencialmente en el tiempo), recurrente (la estructura de red permite circularidad) y recursivo (los impulsos operan sobre sí mismos a través de los neuromoduladores que activan²⁴). A su vez estos observables pueden autoorganizarse en patrones espacio-temporales estables y generar dominios composicionales a niveles superiores.

La propiedad fundamental de las señales es su *potencialidad* para una

²⁴ Por eso, la diferencia fundamental entre los observables del SN y otro tipo de señales, como las hormonas (moléculas que vehiculizan formas de coordinación intercelular) además de una muchísima mayor rapidez de transmisión, es la capacidad de “procesamiento” recursivo (señales operando sobre señales) y que confiere al SN un carácter “cuasi-sintáctico” en relación a la dinámica de los procesos metabólicos.

causalidad formal (es decir,., que la acción causal está ligada a la estructura de las configuraciones y no a elementos energéticos y temporales) desligada de la causalidad energéticamente determinada sobre la que se propagan esas mismas señales. Así por ejemplo, las señales eléctricas que viajan en los cables de redes informáticas pueden causar cambios en las terminales no en virtud de la cantidad de energía eléctrica de la que están constituidas, sino de la secuencia de cambios de amplitud y frecuencia. Del mismo modo, la causalidad motora de los impulsos nerviosos no viene determinada por la cantidad de energía electroquímica que constituye los potenciales de acción, sino por su *forma*, que las células musculares “interpretan” los neurotransmisores segregados por las neuronas. Es en base a esa forma de la actividad nerviosa que las células musculares movilizan (o no) la energía metabólica para producir movimiento. Este tipo de causalidad formal que se hace evidente en la relación entre actividad nerviosa y muscular es característica de la actividad interna del SN.

3.3.2 ¿Cuándo las señales se convierten en información?

La constitución del SN como un dominio autónomo, jerárquicamente desacoplado, constituido por una dinámica recurrente en constante autoorganización y con capacidad para una causación formal de señales emergentes, nos acerca irremediamente al concepto de información. Sin embargo, nos queda por especificar qué es lo que añade el concepto de información a lo que venimos explicando. El factor clave que diferencia a la información como tal de las meras señales va a ser la asignación de contenido (epistémico o instructivo)²⁵.

Hacia una noción naturalizada de información

Cuando, desde una perspectiva naturalizadora hablamos de información en el SN nos referimos a conceptualizar la información en relación al organismo que la genera y la utiliza. Se trata de especificar cómo un proceso físico o biológico deviene informacional no para un observador externo al que se le

²⁵ En cierto sentido podemos considerar a los observables neuronales que hemos caracterizado como señales, como elementos “proto-informacionales”, ya que constituyen un dominio de elementos que actúan exclusivamente en virtud de su forma en relación a la dinámica más básica de los procesos metabólicos. Habría pues dos sentidos del término información: uno, como vehículos de significado en un sentido actual; y otro, como elementos que pueden potencialmente actuar como tales vehículos, aunque no lleven significado. Esta última acepción de información es utilizada en la teoría matemática de la comunicación (Shannon 1948) en la que puede asignarse una cantidad de información a un conjunto de señales independientemente de su significación. En este trabajo usamos el término 'señales' en este segundo sentido y dejamos el término 'información' sólo para ser utilizado en el primero.

presuponen ya capacidades cognitivas, sino en el propio marco del sistema natural en el que ese proceso tiene lugar, jugando un papel causal específico al constituir una nueva forma de organización en dicho sistema. Como veremos más adelante, es en virtud de esta nueva forma de organización interactiva (la informacional) que un sistema deviene cognitivo.

La forma más común en neurociencias de referirse a la información es una forma metodológica o descriptiva, mediante la cual un investigador decide usar el término información para nombrar ciertas regularidades en los fenómenos que observa. El uso de los términos información, código, contenido, significado, etc. en neurociencia teórica está asociado a la correlación entre un estímulo dado y un potencial de acción (o un conjunto de potenciales de acción) en un área específica del SN. Esta correlación causal está generalmente expresada como probabilidad condicional de la ocurrencia de un estímulo, dada una medida de impulsos o una combinación de ellos (Rieke *et. al.* 1997, Dayan & Abbott 2001, Oram *et.al.* 1998). Esta perspectiva nos permite como mucho predecir (*a posteriori* —después de toda una serie de experimentos) la probabilidad de disparo de una neurona en relación a un estímulo determinado; pero en ningún caso *explica* el funcionamiento del SN en relación a esas correlaciones. De hecho, la correlación (así explicada, como probabilidad condicional) no es nunca accesible al propio SN sino sólo al observador que accede al estímulo y sus consecuencias dinámicas de forma separada²⁶, de tal forma que el proceso resulta informacional para el observador pero no necesariamente para el sistema bajo estudio. La perspectiva naturalizadora opta, en cambio, por una concepción ontológica en la que el término información se refiere a un tipo de causalidad específica que permite un tipo de organización característica en el SN. Buscamos dar cuenta del contenido semántico de un proceso informacional desde la organización del organismo²⁷.

La única forma de conseguir que el concepto de información sea naturalizado de esta forma es (siguiendo a Bickhard —2000) buscando una

²⁶ Esta perspectiva permite realizar afirmaciones como la siguiente: “(...) we can ask how the homunculus should best use the spike train data to make a decision about which stimulus in fact occurred. This decision rule would constitute a complete understanding of the neural code (...)” (Rieke *et.al.* 1997: 14). No se trata de inferir “la realidad” a partir de un conjunto de señales, sino de explicar cómo se integran las señales informacionales en la producción de conducta, y en virtud de qué aparecen, para el organismo, asociadas a una intencionalidad o semántica acerca del mundo.

²⁷ De este modo si el concepto de información está generalmente relacionado con la reducción de incertidumbre se trata especificar en qué sentido una señal neuronal reduce la incertidumbre de un estado de cosas para el organismo y no para un observador externo.

noción de información (aunque este autor hablará de representación) que permita detección del error por parte del sistema que maneja la información (y que permita por tanto ajustar expectativa a través de la interacción, es decir, reducción de incertidumbre sobre el resultado de una interacción), Naturalizar la corrección de error exige naturalizar la noción de normatividad (en base a la cual la información pueda ser evaluada como errónea), de tal modo que ésta sea accesible para el organismo mismo. La normatividad se convierte así en el principio en virtud del cual se establece *la evaluación del error* como proceso *causalmente* efectivo en el mantenimiento de la identidad del sistema (y las estructuras dinámicas que lo reconstituyen). De este modo un concepto naturalizado de información pasa por comprender el origen de los valores regulativos de un organismo, y ello a su vez exige dar cuenta del origen y la organización que da lugar a la agencialidad (tal como hicimos al comienzo de este capítulo). Sólo en el contexto de la interacción biológica sistema-entorno, en el que se generan los valores regulatorios de la autonomía del SN a través de los procesos autoorganizativos de selección de señales, puede comprenderse el concepto de información en los sistemas cognitivos naturales. Los procesos informacionales en el SN no son meras correlaciones con estados de cosas en el mundo, ni son meras constricciones impuestas sobre los procesos autoorganizativos sino que emergen de éstos y constituyen un tipo de causalidad específica que surge de su inserción en la autonomía del SN y del organismo en su conjunto.

Información en el SN

A diferencia de la información genética en la ontogenia del organismo la información neuronal es constantemente creada en el SN, y es precisamente esa creación de información a través de la interacción del organismo con su entorno la que va a permitir la progresiva “autodeterminación” del SN. La creación de información es necesaria porque no puede estar genéticamente especificada (Bechtel & Abrahamsen 2004) por lo que la tarea de creación de información a través de procesos autoorganizados es una tarea permanente del SN²⁸.

Nos encontramos por tanto con que los factores clave para una noción

²⁸ A diferencia de la información genética, la información neuronal es generada en tiempo somático. No hay información neuronal genéticamente definida porque la información neuronal se construye en el ámbito del SN. La información genética (via procesos de desarrollo) especifica los mecanismos (entre ellos la arquitectura básica del SN) que van a permitir a su vez la creación de información neuronal (y, naturalmente, la construcción del organismo como sistema metabólico).

naturalizada de información en el SN vienen de la mano de cómo se integra la información causalmente en la organización del SN (en relación a su creación y su capacidad de constricción de la dinámica neuronal) y cómo adquieren estos procesos un contenido semántico (referencial) en relación a la interacción con el entorno y a través de una normatividad internamente generada por el sistema (y que permita, por tanto, la corrección de error). Veamos cómo podemos reconstruir una noción de información que satisfaga estas condiciones desde la perspectiva dinámica y naturalizada que venimos defendiendo.

En primer lugar, podemos diferenciar dos aspectos del contenido informacional. El primero es el instructivo (un aspecto que comparte con la información genética): decimos que el contenido informacional de una estructura o proceso es instructivo si este proceso es capaz de desencadenar-instruir sistemáticamente y en virtud de su “forma” una serie de procesos funcionales (en última instancia, fuera del SN: acciones musculares o secretoras). Es fácil reconocer este aspecto informacional de la dinámica neuronal, por ejemplo, en las señales que hacen cambiar o que desatan un patrón dinámico específico de un CPG y el consiguiente desencadenamiento de una acción motora. De este modo, en el marco conceptual dinámico, *la información son patrones de señales que actúan como parámetros de control de otros procesos autoorganizados*. En este sentido, la dinámica informacional del SN puede ser entendida como el conjunto de procesos que gobiernan los cambios de macroestados generados por los procesos autoorganizativos, y que hacen que el resultado sean acciones evaluables funcionalmente por el organismo y por la propia dinámica regulativa del SN (de tal modo que la relación entre la señal y el proceso que desencadena puede ser corregida-mantenida en virtud de esa evaluación del SN).

Sin embargo, hay otro aspecto más en la información neuronal. Se trata de la dimensión epistémica de la información. Entendemos por ello que los procesos neuronales están organizados en relación a las interacciones con el entorno en virtud de un contenido intencional o semántico de dichos procesos. Es decir, que la organización informacional del sistema nervioso no es un conjunto de “instrucciones” motoras aisladas de las interacciones que el sistema establece con su entorno sino que están vinculadas con los ciclos sensomotores y, sobre todo, con la necesidad de adaptar las respuestas motoras a condiciones específicas del entorno. Dicho de otro modo la

información epistémica integra relaciones dinámicas entre procesos internos e interactivos.

Tentativamente podemos ofrecer la siguiente definición de información epistémica en términos dinámicos:

Un patrón de señales (**S**) se convierte en informacional en relación a **P** acerca de un estado de cosas **F** ssi: **P** selecciona recurrentemente **S**, y **S** participa en el automantenimiento de **P** a través de su correlación con la ocurrencia de **F**.

Donde **S** es una señal o conjunto de señales y **P** es un proceso alejado del equilibrio. En concreto P es una estructura automantenida de orden superior que recluta S o sobre la que S tiene un efecto causal formal (P puede ser un patrón autoorganizado de señales, una invariante del SN, el organismo como sistema autónomo o incluso la especie como estructura colectiva e histórica). Finalmente **F** es un estado de cosas definido por las necesidades de automantenimiento de P.

Veamos un ejemplo menos abstracto que ilustre esta definición de información en el que S es un patrón de actividad nerviosa, P un organismo y F la presencia de comida en el entorno. En este caso el patrón de actividad nerviosa (S) lleva información acerca de la presencia de comida para un organismo (sistema autónomo alejado del equilibrio y recursivamente automantenido) ssi: las acciones que desencadena esa señal en el organismo contribuyen a su automantenimiento *a través de* la presencia de la comida. Esto permite que la información pueda ser falsa: puede darse el caso de que algo que no es comida desencadene esa señal que a su vez desencadena la consiguiente interacción cuyo resultado NO contribuye al automantenimiento del sistema (a través de la presencia de la comida). Además puede darse detección de error si el resultado de la interacción es accesible al organismo, y corrección de error si existen mecanismos de modulación de la conducta en base a la detección del error. Este ejemplo ilustra la forma en que podemos hablar de información epistémica en un organismo sin caer en la falacia del homúnculo, es decir, sin hacer una descripción de correlaciones estímulo-impulso que desplace (indefinidamente) el problema del significado a un sistema que a su vez tenga que *interpretar* el impulso.

La definición que acabamos de presentar permite hablar de procesos

neuronales que operan efectivamente como una causalidad específica (la informacional) para el organismo (ya que éste es capaz de evaluar el contenido semántico del proceso informacional) y no desde la perspectiva de un observador externo que establece el vínculo (en forma de correlación causal o de cualquier otro modo) entre señal y contenido informacional.

La información neuronal aparece en formas diferentes en virtud de la relación entre los aspectos instructivos y epistémicos de la información: a) *Información meramente instructiva (secreto-motora)*: se genera a través de procesos autoorganizativos que se realizan a partir de estructuras genéticamente especificadas (CPG). b) *Información sensomotriz básica (actos reflejos)*: sobre una trayectoria sensomotriz genéticamente especificada se genera información a raíz de una señal perceptiva (el aspecto epistémico y el instructivo son indisociables). c) *Información cognitiva*: es aquella en la que los aspectos epistémicos se pueden desligar de los instructivos. La información cognitiva se genera cuando se fija una relación que no existía anteriormente entre procesos informacionales instructivos y epistémicos. Podemos hablar de dos aspectos organizativamente diferenciados de la información neuronal (el epistémico y el instructivo) porque a diferencia de las trayectorias sensomotoras innatas, los procesos de aprendizaje permiten combinaciones variables de información instructiva y epistémica. Es entonces cuando surge el problema de la verdad, de la adecuación entre información epistémica e instructiva, y surge el dominio cognitivo propiamente dicho.

Analicemos un caso concreto en el que detallar naturalizadamente el marco conceptual que venimos desarrollando.

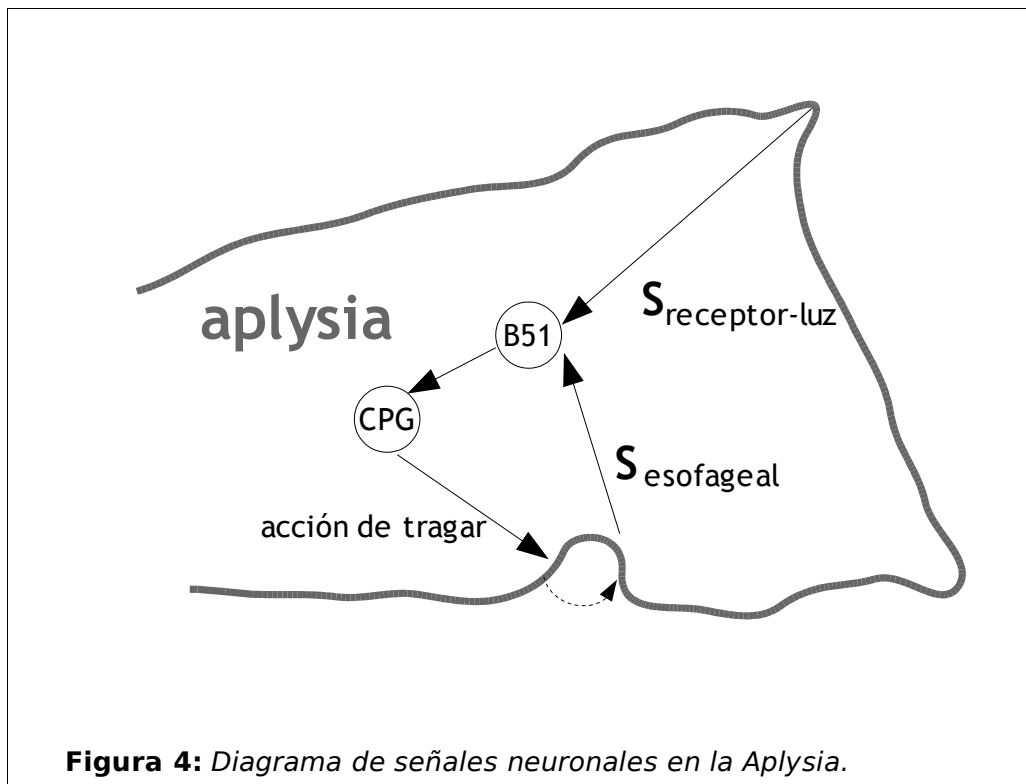
Estudio de Caso (aprendizaje instrumental en la Aplysia)²⁹

Supongamos que la Aplysia es capaz de aprender que a la presencia de un color de luz determinado sigue regularmente comida: es decir, que puede crear un ciclo sensomotor funcional generando un patrón motor específico (el de tragar) ante una serie de patrones perceptivos determinados (los correlacionados con la presencia de luz de ese color). Se trata, en este

²⁹ El caso concreto de aprendizaje instrumental que detallamos a continuación es ficticio. No se ha realizado ningún experimento empírico en relación a la asociación de los estímulos que aquí reproducimos. Sin embargo los mecanismos neurobiológicos que se detallan están extraídos de la reciente literatura neurocientífica que ha analizado los mecanismos neurobiológicos del aprendizaje instrumental en la Aplysia y los circuitos que generan la acción de tragar así como el rol fundamental de la neurona B51: Brembs *et.al.* 2002, Carew 2002 y Kandel 2001.

ejemplo de actividad *cognitiva*, de analizar qué constituye información y cómo se integra en la organización metabólica y neuronal de la Aplysia.

En relación a los procesos neuronales que intervienen en el proceso cognitivo (en el aprendizaje de esta asociación) de la Aplysia podemos diferenciar dos tipos de señales perceptivas diferentes: a) descargas dopaminérgicas del nervio esofageal correlacionadas causalmente con la presencia de ciertos sabores ($S_{\text{ESOFAGEAL}}$) asociados a su vez a compuestos alimenticios para la Aplysia y b) señales perceptivas correlacionadas con la presencia de luz ($S_{\text{RECEPTOR-LUZ}}$). Ahora bien, ninguna de estas correlaciones (por el hecho de ser correlaciones) es de tipo informacional para el sistema porque no hay manera en que el sistema pueda asegurar la correlación entre la señal y su causa: no tiene medios para acceder a la luz o a la comida de forma que no sea la señal misma cuya correlación ha de asegurar. La existencia de estas correlaciones es algo que sólo un observador externo (que tenga acceso tanto al estado de cosas en el mundo —presencia de comida— como a la señal neuronal) puede establecer.



Sin embargo desde la perspectiva que venimos desarrollando es correcto decir que $S_{\text{ESOFAGEAL}}$ lleva información acerca de la comida. ¿En virtud de qué podemos pues hablar de información sin recurrir a la correlación causal observable en entre la señal y el objeto que la causa? En efecto la

Aplysia no puede comprobar esa correlación hacia atrás (en sentido de la causa inmediata de la señal), pero como sistema autónomo su existencia depende de las consecuencias de las interacciones que establece en virtud de esa señal. Si la correlación no se cumple, la Aplysia muere; es decir, si sistemáticamente la Aplysia traga objetos que, sin ser alimenticios, producen $S_{\text{ESOFAGEAL}}$ entonces morirá. Si la correlación se cumple la Aplysia recibirá señales de saciación de sus sistema digestivo indicando la satisfacción de esa correlación. Digamos pues que la Aplysia puede y debe comprobar la correlación hacia el futuro (para su automantenimiento) y es en virtud de esa futura comprobación que la señal deviene informacional para la Aplysia. Cuando esa comprobación se realiza repetidamente se va fijando la semántica.³⁰

Existen sin embargo dos problemas aún por resolver. El primero es que el SN es ciego a esa comprobación semántica, en efecto es el organismo como un todo el que “interpreta” la información y es una señal adicional proveniente del sistema digestivo la que asegura que es información epistémica. Llamemos pues a este tipo de información que lleva la $S_{\text{ESOFAGEAL}}$ información epistémica biológica pero no información cognitiva ya que el proceso informacional no está propiamente inserto en el funcionamiento del SN (sino que necesita un sistema digestivo auxiliar que lo sostenga como proceso informacional). El segundo problema es que $S_{\text{ESOFAGEAL}}$ no genera ninguna acción en sí mismo, es decir, no tiene información instructiva alguna (no produce ningún movimiento)... excepto a través de otros procesos del SN y en la generación de nueva información. Es aquí donde, como veremos a continuación, $S_{\text{ESOFAGEAL}}$ se inserta en los procesos propiamente cognitivos (en los que se produce nueva información) de la Aplysia.

La organización del SN de la Aplysia es tal que cuando la neurona B51 recibe simultánea y repetidamente $S_{\text{ESOFAGEAL}}$ y $S_{\text{RECEPTOR-LUZ}}$ estabiliza una respuesta que produce la acción de tragar. En concreto la B51 modula sus propiedades de membrana (su umbral de activación) de tal manera que para

³⁰ Ahora bien... ¿qué sucede antes de que se fije esa relación semántica la primera vez que una Aplysia come y se genera la señal $S_{\text{ESOFAGEAL}}$? La Aplysia aún no ha tenido oportunidad de comprobar el contenido informacional... ¿Podemos decir entonces que $S_{\text{ESOFAGEAL}}$ lleva información para la Aplysia? Podemos responder de dos maneras diferentes a esta pregunta: en relación a la Aplysia como organismo o en relación a la especie Aplysia. En relación a la especie $S_{\text{ESOFAGEAL}}$ lleva información sobre la comida ya que su contribución al automantenimiento en virtud de su contenido informacional es continuamente comprobado por la especie Aplysia durante su historia evolutiva. Para la Aplysia como organismo individual (a pesar de que no haya habido aún comprobación) la señal sigue teniendo información ya que tendrá efectos para el automantenimiento del sistema.

sucesivas recepciones de la señal $S_{\text{RECEPTOR-LUZ}}$ B51 genera un patrón de impulsos que a su vez modulan los BPM (Buccal Motor Program) producidos por el Ganglio Bucal convirtiéndolos en un patrón que, a su vez, produce la acción de tragar (ver figura 4). De este modo, si se da el caso de que la Aplysia traga en presencia de $S_{\text{RECEPTOR-LUZ}}$ se disparará $S_{\text{ESOFAGEAL}}$ y esto hará que B51 module su umbral de membrana para producir la acción de tragar en sucesivas presencias de $S_{\text{RECEPTOR-LUZ}}$. Esto aumenta la probabilidad de que la próxima vez, en presencia de $S_{\text{RECEPTOR-LUZ}}$ la Aplysia trague y, si, en efecto $S_{\text{RECEPTOR-LUZ}}$ está correlacionado con la presencia de comida ésta desencadenará $S_{\text{ESOFAGEAL}}$ que de nuevo contribuirá a mantener estable el umbral y así sucesivamente. Si la conjunción entre ambos estímulos no se da la estabilidad del valor del umbral de membrana en la B51 (que produce la acción de tragar) se pierde. De este modo la estructura alejada del equilibrio (en este caso el umbral de activación de B51) que sostiene una interacción funcional para el organismo se mantiene si y sólo si ésta funcionalidad puede asegurarse a través de la presencia de comida en el entorno. La comida produce los efectos apropiados en la digestión sosteniendo así la correlación entre la presencia de comida y $S_{\text{ESOFAGEAL}}$, a través de las señales de saciación. A su vez la correlación comida- $S_{\text{ESOFAGEAL}}$ sostiene la correlación comida- $S_{\text{RECEPTOR-LUZ}}$. El cierre del ciclo sensomotor a través de la producción del movimiento de tragar es fundamental en el proceso informacional. Sólo así podemos afirmar que la señal $S_{\text{RECEPTOR-LUZ}}$ lleva información de “comida” para la Aplysia ya que es la acción de tragar y la consecuente $S_{\text{ESOFAGEAL}}$ la que sostiene la estabilidad de la asociación semántica (en realidad el P inmediato de nuestra definición es el patrón asociativo de la B51 —la estabilidad del valor de su umbral de membrana). Nótese que $S_{\text{RECEPTOR-LUZ}}$ “no dice nada” acerca de la presencia de luz (aunque para un observador externo esté claramente correlacionado y “lleve” información sobre la presencia de luz) sino de la comida o, propiamente hablando, de $S_{\text{ESOFAGEAL}}$. ya que presupone dinámicamente que $S_{\text{ESOFAGEAL}}$ seguirá a $S_{\text{RECEPTOR-LUZ}}$ ya que sino los patrones de impulsos que emite la B51 en presencia de $S_{\text{RECEPTOR-LUZ}}$ cesarán. Dicho de otro modo $S_{\text{ESOFAGEAL}}$ es la condición de posibilidad de la estabilidad del umbral de activación de la B51 que genera una serie de patrones dado $S_{\text{RECEPTOR-LUZ}}$. Pero podemos decir que lleva información acerca de la comida para la Aplysia ya que $S_{\text{ESOFAGEAL}}$ lleva a su vez información acerca de la presencia de comida.

Hemos analizado una dependencia entre la estabilidad de los valores de tres de las variables constitutivas del SN de la Aplysia ($S_{\text{RECEPTOR-LUZ}}$, $S_{\text{ESOFAGEAL}}$ y el valor del umbral de activación de B51) con cierto realismo científico. Sin embargo esto constituye sólo una pequeñísima parte de las dependencias entre procesos en el SN de la Aplysia, que apenas presenta complejidad dinámica relevante. Podemos imaginar redes de dependencias dinámicas de este tipo que configuran un universo de significados para la Aplysia, organizados de tal forma que se sostengan en virtud del resultado específico de las interacciones que la propia red de dependencias permite establecer con el entorno. La información aparece así generando unas condiciones de estabilidad particulares dentro del SN, precisamente aquellas que permiten organizar la conducta coordinadamente con los elementos del entorno relevantes para el automantenimiento del sistema y de las propias estructuras que van generándose recursivamente en el propio SN.



El desarrollo de una teoría completa de la información en el SN está fuera del alcance de este trabajo. La definición y los ejemplos tentativos que hemos ofrecido en esta sección deben comprenderse más bien como esfuerzos por señalar la forma y las condiciones que una completa teoría de la información debe satisfacer desde la perspectiva naturalista. En cualquier caso las consideraciones que hemos realizado en el proceso son suficientes para sostener la hipótesis fundamental sobre la especificidad de la cognición que presentaremos en el siguiente capítulo y que constituye la aportación principal de este trabajo de investigación. Recapitulemos brevemente algunas de estas consideraciones.

A partir de una semántica o referencialidad corporizada en las necesidades adaptativas del organismo y guiada por señales de valor evolutivamente fijadas, va generándose toda una red de procesos autoorganizados en la que unos niveles van seleccionando procesos de niveles inferiores en una jerarquía anillada de procesos informacionales. De este modo los grados de libertad de un sistema (el SN) capaz de un inmenso número de transformaciones de estado (a través de una enorme variabilidad recurrente entre diversos tipos de observables) se van organizando informacionalmente a lo largo del desarrollo ontogenético. Los procesos de aprendizaje, sostenidos sobre la normatividad básica de automantenimiento del organismo, van estabilizando una serie de relaciones epistémicas que, a

su vez, estabilizan otros procesos de órdenes superiores, creando una red de relaciones epistémicas siempre revisables, ya que su estabilidad depende de la satisfacción de ciertas correlaciones sensomotoras acopladas al entorno. La organización informacional cognitiva es, pues, aquella en la que los procesos neuronales se instruyen a través de señales seleccionadas recursivamente mediante su inserción en ciclos sensoriomotores que dependen de estructuras específicas del entorno definidas por las capacidades y necesidades evaluativas del organismo.

4 NATURALIZAR LA COGNICIÓN

4.1 La emergencia de la cognición desde lo biológico

Hasta ahora hemos procurado desarrollar nuestro análisis en un nivel esencialmente biológico, tratando de describir e interpretar conceptualmente la naturaleza de los procesos interactivos tal y como se van configurando en los organismos con SN. Hemos por eso mismo restringido nuestro análisis a las formas más básicas (en otras palabras, las evolutivamente más primitivas) para que la labor de análisis conceptual mantenga el mayor grado posible de apoyatura en estudios empíricos de aquellos dominios donde se conocen bastante bien los mecanismos de funcionamiento. Ello nos permite hacer comparaciones con otras formas evolutivamente más elementales de agencialidad y definir una transición evolutiva fundamental marcada por la aparición del SN como sistema dinámico que controla los ciclos de interacción sensomotora con el entorno, informacionalmente organizados y jerárquicamente desacoplados de los procesos metabólicos constructivos.

Tal como vimos en la sección 2.1 para el funcionalismo clásico (ya sea en su vertiente computacional o evolutiva) la especificidad de la cognición reside en el carácter representacional de los primitivos teóricos que la constituyen. Las ciencias cognitivas así entendidas no sólo impiden el surgimiento de nuevas representaciones como parte del fenómeno de la cognición sino que desplazan la naturaleza específica de la representación como correlación con estados de cosas en el mundo a un observador al que se le presuponen ya capacidades cognitivas. Esta forma desnaturalizada de representación impide comprender la información como una forma específica de organizar la causalidad en los procesos naturales (más concretamente en los organismo con SN). Estas dificultades reaparecen en el uso del concepto de información en neurociencia. De este modo el funcionalismo representacionalista establece un dualismo anti-naturalista entre mecanismo y función que difícilmente permite un reconciliación emergentista entre biología y cognición.

Los intentos de corporizar los procesos cognitivos en agentes situados

han supuesto un avance significativo frente al funcionalismo. Sin embargo las ciencias cognitivas corporizadas se han quedado a medio camino; la corporización sensomotora acoplada al entrono es una característica necesaria pero no suficiente para dar razón del fenómeno cognitivo. De hecho el dinamicismo no puede diferenciar entre lo que es cognitivo y lo que no, no puede dar razón de la cognición como fenómeno específico. Desde la perspectiva dinamicista no hay una diferencia cualitativa entre dos péndulos acoplados y la conducta cognitiva de un organismo en interacción con su entorno.

Nos encontramos, pues, ante un panorama en el que o bien el dominio cognitivo se define estableciendo un corte funcionalista difícilmente naturalizable en la organización misma de los mecanismo que conforman el sistema cognitivo (una especie de dualismo) o bien se postula una continuidad dinámica tan estrecha entre mundo y agente que impide dar razón de la especificidad de la cognición como fenómeno propio. Frente a estas dos alternativas paradigmáticas ofrecemos (a modo de hipótesis) una propuesta fundacional del fenómeno cognitivo como emergente de la biología.

4.2 Hipótesis de la Corporización Biológica de la Cognición

La Hipótesis de la Corporización Biológica (HCB) establece que:

*La cognición es un fenómeno que emerge de la **corporización biológica** en la que se inserta y que por tanto no puede ser comprendido (analizado ni sintetizado) al margen de la organización dinámica neuronal e interactiva autónoma: jerárquicamente desacoplada pero agencialmente enraizada en las necesidades termodinámicas del organismo como sistema autónomo alejado del equilibrio.*

De acuerdo a la HCB la corporización no es sólo un enraizamiento físico y en tiempo real de una inteligencia que se había considerado abstracta (como defienden los partidarios del acercamiento corporizado a la cognición) sino la forma en la que las condiciones de posibilidad de la cognición se encarnan en la organización biológica misma (y en su historia evolutiva)

definiendo una dinámica específica. Es por ello que la hipótesis de la corporización biológica supera las tesis de las ciencias cognitivas corporizadas y exige una comprensión más radical de la corporización que, más allá de la inserción del fenómeno cognitivo en un cuerpo a modo de armazón o interfaz con el entorno, lo inserta en las condiciones biofísicas de alejamiento del equilibrio y autonomía (y por tanto en las relaciones interactivas con el entorno necesarias para su automantenimiento como fundamento último de su normatividad). Sólo sobre esta hipótesis puede articularse un dominio propiamente cognitivo de forma naturalizada.

4.3 La especificidad de la cognición

Desde esta hipótesis podemos finalmente reconstruir la especificidad de la cognición (tal como señalábamos al inicio de este trabajo) de forma naturalizada. Para ello podemos recoger cuatro principios organizativos fundamentales del SN que hemos venido detallando a lo largo de este capítulo. La especificidad de la cognición viene marcada por cuatro marcas características de la cognición:

1. **Desacoplamiento jerárquico del SN.** El principio del desacoplamiento nos permite comprender la cognición como fenómeno dinámicamente diferenciado del resto de los procesos biológicos de un organismo y define un dominio dinámicamente específico sin impedir su integración en el organismo.
2. **Corporización sensomotora y metabólica.** A su vez, el SN no puede entenderse sino enraizado corporalmente, es decir,, como una organización causalmente dependiente (para su construcción y mantenimiento, por un lado; para su normatividad básica, por otro) del resto del cuerpo. Las necesidades metabólicas del organismo juegan un papel fundamental en la organización de la dinámica neuronal, especialmente en relación a las señales de valor que la organización metabólica utiliza para constreñir y monitorizar la actividad nerviosa. Pero a su vez, el SN es fundamental para el mantenimiento viable de la organización metabólica (básica) del animal.

3. **Autonomía regulativa.** El principio de autonomía regulativa del SN señala la especificidad del dominio neuronal como dominio autónomo capaz de generar una normatividad propia en forma de constricciones de estabilidad históricamente generadas en la ontogenia del organismo.
4. **Organización informacional.** Finalmente, la *cognición* aparece como una forma de organización específica de esa dinámica autónoma y biológicamente corporizada. La información aparece no como un primitivo teórico desnaturalizado, sino como una forma específica de autoorganización de la interacción con el entorno que establece relaciones epistémicas e instructivas entre el sistema y su entorno, y se configura como el concepto fundamental para entender los procesos cognitivos como constitutivos de un dominio específico.

4.4 Sujeto, entorno y mundo desde la HCB.

El tratamiento de un sistema como autónomo no se limita a constatar una asimetría de complejidad sino que introduce un cambio gestáltico en la forma de comprender y naturalizar la relación sistema-entorno, el concepto de información y, en definitiva, el fenómeno cognitivo en su conjunto. Esto se debe a que el concepto de autonomía remite, en última instancia, a la constitución de un sujeto y de una identidad y apunta a la forma en la que la creación de esta subjetividad genera y define interactivamente un mundo de significados: una estructura normativa de la interacción informacionalmente organizada.

Desde la perspectiva de un observador el entorno de un agente es una amalgama de variables que interactúan entre sí pero que aparecen, en principio, neutras e indiferenciables en relación al agente. Sin embargo un agente genera un orden determinado en ese entorno: un mundo en el que opera. En concreto el mundo primario en el que un agente cognitivo se encuentra situado es el definido por la forma en que el SN media las condiciones de contorno de su autonomía biológica (en sentido termodinámico), a saber, las relaciones entre variables de su entorno que resultan relevantes para el automantenimiento activo de su estructura (temperatura, comida, etc.) y aquellas que potencialmente pudieran desestabilizarla (depredadores, colisiones, etc.). Por lo tanto, a través de su inserción en la autonomía del organismo el SN se encuentra en un mundo

que su propia corporalidad ha definido a través de las relaciones que ese cuerpo (esa estructura alejada del equilibrio) necesariamente ha de establecer para su automantenimiento: es lo que llamamos un nicho ecológico. Pero los nichos ecológicos no son espacios que existan independientemente de los organismos que los habitan, sino que organismo y nicho se (co)definen mutuamente ya que el organismo recorta, sobre un entorno indefinido, el conjunto de relaciones que constituyen su nicho y, a su vez, las condiciones de contorno que genera un nicho sobre el organismo sostienen la estabilidad de éste.

La definición de un mundo a través de un conjunto de necesidades interactivas para su automantenimiento (o autopoiesis) es lo que ha impulsado a algunos autores (Maturana & Varela 1974, Stewart 1996) a considerar que todo ser vivo es un sistema cognitivo. Sin embargo la relación entre un cuerpo biológico y su nicho ecológico y un SN con su mundo es de una naturaleza lo suficientemente diferente como para aconsejar guardar el término cognición para la relación que se establece entre un SN corporizado y su mundo, especialmente a la luz de la autonomía que la propia dinámica y estructura del SN genera a través de su ontogenia interactiva. Analicemos más detalladamente este proceso y su relación con la autonomía metabólica o constructiva.

Como hemos visto la relación del SN con su entorno se origina en el control de la interacción motriz jerárquicamente desacoplada del metabolismo pero funcionalmente integrada en el mantenimiento del sistema a través de la satisfacción de la clausura interactiva. Originariamente, pues, la relación entre SN y entorno está marcada por el mantenimiento interactivo de ciertas variables esenciales (las que definen las condiciones de automantenimiento del sistema) bajo constricciones de viabilidad. Sin embargo el SN no tiene un acceso directo al estado de estas variables ni ese acceso puede garantizar (por sí mismo) una regulación completa de la interacción funcional con el entorno. Tanto las señales de valor como las constricciones genéticamente heredadas limitan la variabilidad de las interacciones que el SN puede sostener con el entorno. Pero los grados de libertad del SN como sistema dinámico son tan grandes que exigen la generación de constricciones propias, la creación de mecanismos internos de regulación a través de la interacción directa con el entorno. El proceso de desarrollo y de aprendizaje de un SN es el proceso de estabilización y

complejización de una serie de invariantes internas que actúan como vicariantes (sustitutivas) de las variables esenciales ampliando el dominio interactivo más allá de lo constructivo. Estas invariantes permiten establecer interacciones más estables y estructuradas con el entorno. Esta interacción desacoplada genera pues, a través de un proceso de desarrollo y una historia interactiva, un nuevo dominio cognitivo fruto de la autonomía que el propio SN progresivamente genera.

Inevitablemente, un nuevo dominio autónomo, el del SN en este caso, genera a su vez un nuevo dominio normativo fruto de la necesidad de regular la enorme complejidad dinámica que alberga a través de constricciones autogeneradas. A su vez estas constricciones se estabilizan en estructuras invariantes internas dando lugar al dominio funcional definido por la contribución al automantenimiento de estas invariantes. En este sentido, el SN genera una normatividad y una funcionalidad propias, más allá de la meramente adaptativa (marcada por las más inmediatas necesidades de automantenimiento del organismo en el que se inserta). Esta normatividad autogenerada es la fuente del contenido informacional de los procesos neurocognitivos.

La información no aparece por tanto como respuesta a la necesidad de modelar un entorno ya que en las interacciones agente-entorno no se está regulando o controlando el entorno. Es tentador pensar que el control de las condiciones de contorno y por tanto del entorno es una función primaria del SN y que, por tanto³¹, modelar el entorno es la forma fundamental de funcionamiento y de relación entre SN y entorno. Sin embargo las variables de control son las variables esenciales aunque sólo a través de las invariantes internas que el SN genera. Lo que el SN controla, por lo tanto, son sus propias invariantes a través de la interacción sensomotora, generando en el proceso una subjetividad y un mundo (un espacio de interacciones estables y estructuradas). El mundo cognitivo no es por tanto un entorno absoluto sujeto a representación isomórfica, un mundo dado y previo al sujeto que lo conoce, sino un dominio de interacciones recortado sobre un fondo indiferenciado en el que cabe un número ilimitado (aunque no arbitrario) de configuraciones. El problema del error y de la verdad en ciencias cognitivas no responde primariamente a un problema de adecuación entre estados internos y externos sino a un problema de ajuste entre estabilidad interna e interacción

³¹ Como bien señalaron Conant y Ashby (1970): "Every good regulator of a system must be a model of that system".

necesaria para mantenerla.

En términos operacionales la HCB predice que para saber lo que un sistema está informando hace falta saber más acerca del propio sistema, de su organización interna y de la forma en que ésta está interactivamente abierta a su entorno de forma biológicamente corporizada, que de la correlación entre la estructura del entorno en sí (o tal y como se le pueda aparecer a un observador que pretenda objetividad en su medición o descripción del entorno) y los estados internos del agente. De este modo la in-formación se acerca más a un proceso que a una entidad isomórfica a un mundo predefinido, información es al acto de in-formar (o formar desde dentro) un mundo estable de interacciones que reclutan señales de entorno para sostener su propia estabilidad. Una estabilidad que se encuentra a su vez sostenida sobre una jerarquía anillada de procesos subyacentes (biológicos y termodinámicos) sin los cuales el fenómeno cognitivo es imposible de naturalizar.

REFERENCIAS Y BIBLIOGRAFÍA BÁSICA

- Arhem, P. & Liljenstrom, H. (1997) On the Coevolution of Cognition and Consciousness. *J Th Biology*, **187** 601-612
- Arshavsky, Y., Deliagina, T. & Orlovsky, G. (1997) Pattern Generation. *Current Opinion in Neurobiology*, **7**: 781—789.
- Ashby, W.R. (1952). *Design for a Brain. The origin of adaptive behaviour*. Chapman and Hall, 1978 edition.
- Ashby, W.R. (1956) *An introduction to Cybernetics*. Chapman and Hall. versión online: <http://pespmc1.vub.ac.be/books/IntroCyb.pdf>
- Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48:57–86
- Bechtel, W. (1998) Dynamicists versus computationalists: Whither mechanists? *BBS* **21**(5): 629.
- Bechtel, W. & Abrahamsen, A. (2004). Mechanistic explanation and the nature-nurture controversy. *Bulletin d'Histoire et d'Epistémologie des Sciences de la Vie*, in press.
- Bechtel, W. & Richardson, R. (1993). *Discovering Complexity. Decomposition and Localization as strategies in scientific research*. Princeton University Press.
- Beer, R. D. (1990). *Intelligence as adaptive behavior: An experiment in computational neuroethology*. Academic Press, San Diego.
- Beer, R. D. (2003) The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, **11**(4): 209—243.
- Bickhard, M. H. (2000). Information and Representation in Autonomous Agents. *Journal of Cognitive Systems Research*, **1**(2):65—75.
- Bonner, J. T. (1988) *The evolution of complexity by means of natural selection*, Princeton University Press. NJ.
- Bonner, J. T. (1988) *The evolution of complexity by means of natural selection*, Princeton University Press. NJ.
- Brembs, B., Lorenzetti, F.D., Reyes, F.D., Baxter, D.A. & Byrne, J.H. (2002) Operant Reward Learning in *Aplysia*: Neuronal Correlates and Mechanisms. *Science* **296**:1706—1709.
- Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, **47**:139—160.
- Buss, L. (1987) *The Evolution of Individuality*: Princeton University Press. Princeton.

- Conant, R. C. & Ashby, W. R. (1970), Every good regulator of a system must be a model of that system. *Intern. J. of Systems Science*, **1**(2):89—97.
- Campbell, D.T. (1974) Evolutionary Epistemology. En P.A. Schilpp (ed.) *The Philosophy of Karl Popper*, pp. 413-463.
- Carew, T.J. (2002) Understanding the consequences. *Nature* **417**: 803—806.
- Christensen, W. D. & Hooker, C. A. (1998). From Cell to Scientist: Toward an organisational theory of life and mind. In J. Bigelow (Ed.) *Our Cultural Heritage*. (275-326). Australian Academy of Humanities, University House, Canberra, Australia.
- Christensen, W.D. & Bickhard, M.H. (2002), The Process Dynamics of Normative Function, *Monist.*, **85**, (1). 3–28.
- Clark, A. (1997) *Being There: putting, body and world together again*. MIT Press, Cambridge, MA.
- Cliff, D. (1995) *Computational neuroethology*. In Arbib, M.A. (de.) *The handbook of brain theory and neural networks*. MIT Press, pp. 626–630.
- Dayan & Abbott (2001) *Theoretical Neuroscience*. MIT Press, Cambridge, MA.
- Dennett, D. (1996) *Kinds of Minds*. Weidenfeld and Nicholson, London (Traducción española (2000) Debate. Barcelona)
- Di Paolo, E. (2003). Organismically inspired robotics. In Murase, K. and Asakura, T., (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*, pages 19—42. Advanced Knowledge International, Adelaide, Australia.
- Edelman, G.M. (1992) *Bright, Air, Brilliant Fire*. Basic Books.
- Edelman., G.M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books.
- Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.
- Fields, R.D. (2004) Células de la glía. *Investigación y Ciencia*. Junio 2004: 7–14.
- Floreano, D. & Urzelai, J. (2000) Evolutionary robots with on-line self-organization and behavioral fitness Source *Neural Networks* Volume 13 , Issue 4-5 pages: 431 – 443
- Fodor. J.A. (1987) *Psychosemantics*. MIT Press, Cambridge, MA.
- Funahashi KI and Nakamura N. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, **6**: 801–806.
- Gelder, T.van (1998) The dynamical hypothesis in cognitive science. *BBS* **21**(5):615–628.
- Gerhart, J & Kirschner, M. (1997) *Cells, Embryos and Evolution*. Blackwell Science
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Hillsdale NJ: Lawrence

Erlbaum.

Gould, S. (1989) *Wonderful Life: The Burgess Shale and the Nature of History*. W. Norton. NY

Harnad, S (1990). The symbol grounding problem. *Physica D*, **42**: 335—346, 1990.

Hoffmeyer, J. (1998) Life: The Invention of Externalism. In Farre, G. and Oksala, T. (eds.) *Emergency, Coplexity, Hierarchy, Organization*. Selected and Edited Papers from the ECHO III Conference. Acta Polytechnica Scandinavica . 91, Espoo, pp. 187–196.

Hooker, C & Christensen, W. (1999) The organisation of knowledge: Beyond Campbell's evolutionary epistemology, *Philosophy of Science*, 66: 237—249.[Proceedings, PSA 1998]

Husbands, P., Harvey, I., Cliff, D., & Miller, G. (1997) Artificial Evolution: A New Path for Artificial Intelligence? *Brain and Cognition*, **34**:130—159.

Kandel, E., Schwartz, J. & Jessell, T. (2000) *Principles of Neural Science*. Mc Graw Hill, 2000, 4th Edition.

Kandel, E.R. (2001) The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses. *Science* **294**: 1030—1038.

Kauffman, S. (1986) Autocatalytic Sets of Proteins. *Journal of Theoretical Biology* **119**: 1—24.

Kim, J. (1988) What is naturalized epistemology? En Martin Alcoff, L., (ed.) *Epistemology: The Big Questions*, Blackwell, 1999. pp.265—281.

Keijzer, F. (2002) Representation in dynamical and embodied cognition. *Cognitive Systems Research* **3**: 275—288.

Kelso, J.S.A. (1995) *Dynamic Patterns. The Self-Organization of Brain and Behavior*. MIT Press.

Le Van Quyen, M. (2003) Disentangling the dynamic core: a research program for neurodynamics at the large-scale. *Biological Research* 36: 67—88.

Lorentz, K. (1982:1941) Kant's doctrine of the a priori in the light of contemporary biology. In Plotkin, H.C. (ed) *Learning, development and culture: Essays in evolutionary epistemology*. New York: Wiley, pp. 121—143

Losik, R. & Kaiser, D. (1997) Why and how bacteria communicate. *Scientific American* **276**: 68—73

Maes, P. (ed). (1991) *Designing Autonomous Agents. Theory and Practice from Biology to Engineering and back*. MIT Press, Cambridge, MA.

Mandik, P. & Clark, A. (2002), Selective Representing and World-Making. *Minds and Machines* **12**: 383—395.

Marder, E. & Thirumalai, V. (2002) Cellular, synaptic and network effects of

- neuromodulation. *Neural Networks* **15**: 479—493.
- Maturana, H. & Varela, F. (1980) *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: Reidel.
- Maturana, H. y Varela, F. (1990) *El árbol del conocimiento*. Barcelona: Debate
- Maynard-Smith, J. & Szathmary, E. (1995) *The Major Transitions in Evolution*. Freeman.
- Millikan, R.G. (1984) *Language, Thought and Other Biological Categories*. MIT press, Cambridge MA.
- Millikan, R.G. (2004): The varieties of Meaning. MIT Press. Disponible en <http://vm.uconn.edu/~wwwhttp://vm.uconn.edu/~wwwphil/contents.htmwphil/contents.html>
- Moreno, A., Ibáñez, J. & Umerez, J. (1997) Cognition and Life. The Autonomy of Cognition *Brain & Cognition* **34**(1):107—129.
- Moreno, A. & Lasa, A. (2003) From Basic Adaptivity to Early Mind *Evolution and Cognition* **9** (1):12—30.
- Moreno, A. (2002) El problema de la relación entre autonomía e información en la estructura de la organización biológica. *Ludus Vitalis*, 10 (17): 123—147
- Moreno, A. & Barandiaran, X. (2004) A Naturalized Account of the Inside-Outside Dichotomy *Philosophica* (**73**)
- Moreno, A. & Fernández, J. (1992) From records to self-description: the role played by RNA in early evolutive systems. *Acta Biotheoretica* **40** pp. 1-9
- Moreno, A. & Fernández, J. 1990) Structural limits for evolutive capacities in molecular complex systems. *Biology Forum*, 83: 335—347
- Moreno, A. & Lasa, A. (2003) From Basic Adaptivity to Early Mind *Evolution and Cognition* **9** (1): 12—30
- Moreno, A. & Ruiz-Mirazo, K. (2002) Key issues regarding the origin, nature and evolution of complexity in nature: Information as a central concept to understand biological organization. *Emergence*, 4.1/4.2: 63—76
- Moreno, A.(1998) Information, Causality and Self-Reference in Natural and Artificial Systems In Dubois D.M. (ed) *Computing Anticipatory Systems* Woodbury, NY: American Institute of Physics,. 202-206
- Morowitz, H. (2002) *The Emergence of Everything. How the world became complex*. Oxford UP. Oxford.
- Neihardt, F. (ed.) (1996) *Escherichia coli and salmonella: Cellular and molecular biology*. American Society for Microbiology. Washington.
- Nolfi, S. & Floreano, D. (2000) *The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press.

- Oram, M.W., Földiák, P., Perret D.I. & Sengpiel, F. (1998) The 'Ideal Homunculus': decoding neural population signals. *Trends in Neuroscience* **21**(6): 259—265.
- Pfeifer, R. & Scheier, C. (1999). *Understanding Intelligence*. MIT Press.
- Philippides, A., Husbands, P. O'Shea, M. (2000) Four-Dimensional Neuronal Signaling by Nitric Oxide: A Computational Analysis. *Journal of Neuroscience* **20**(3): 1199—1207
- Philippides, A., Husbands, P. O'Shea, M. (2000) Four-Dimensional Neuronal Signaling by Nitric Oxide: A Computational Analysis. *Journal of Neuroscience* **20**(3): 1199—1207
- Port, R. & van Gelder, T. (1995) *Mind as motion: Explorations in the dynamics of cognition*. MIT Press, 1995
- Powers, T. (1973) *Behavior: the control of perception*. Aldine Publishing Company, Chicago.
- Pylyshyn, Z. (ed.) (1987) *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Ablex, Norwood, NJ.
- Quartz, S.R. & Sejnowski, T. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, **20**: 537—555.
- Quine, W. (1969) Epistemology Naturalized. En *Ontological Relativity and Other Essays*. Columbia University Press.
- Rieke, F., Warland, D., van Steveninck, R. & Bialek, W. (1997) *Spikes. Exploring the Neural Code*. MIT Press. Cambridge, Massachusetts.
- Rose, S. (1999) Précis of Lifelines: Biology, freedom, determinism. *BBS* **22**:871—921.
- Rosen, R. (1966) A note on replication in (M,R)-systems. *Bulletin of Mathematical Biophysics* **28**: 149—151.
- Rosen, R. (1967) Further comments on replication in (M,R)-systems. *Bulletin of Mathematical Biophysics* **29**: 91—94.
- Rosen, R. (1971) Some realizations of (M,R)-systems and their interpretation. *Bulletin of Mathematical Biophysics* **33**: 303—319.
- Ruiz-Mirazo, K. (2001) Condiciones físicas para la aparición de sistemas con capacidades evolutivas abiertas. [Ph. D. Dissertation] San Sebastián: University of the Basque Country.
- Ruiz-Mirazo, K. & Moreno, A. (2000): Searching for the roots of autonomy: The natural and artificial paradigms revisited. *CC AI:Comunication and Cognition-Artificial Intelligence*, **17**(3-4): 209—228.
- Ruiz-Mirazo, K., Pereto, J. & Moreno, A. (2003) A Universal Definition of Life: Autonomy and Open-ended Evolution. *Origins of Life*. Accepted for publication.
- Ruiz-Mirazo, K. & Moreno, A. (2004). Basic Autonomy as a Fundamental Step in the Synthesis of Life. *Artificial Life*, **10**:235—259.

- Rumelhart, D.E., McClelland, J.L., and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vo.1: Foundations*. MIT Press, Cambridge, Mass.
- Searle, J. (1997). *The Mystery of Consciousness*. New York, New York Review Press.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* **1**: 417–24.
- Seth, A.K. & Edelman, G.M. (2004). Environment and behavior influence the complexity of evolved neural networks. *Adaptive Behavior*, **12**(1):5–21.
- Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal* **27**(July): 379—423 & **27**(October):623—656.
- Simons, P.J. (1981). The role of electricity in plant movements. *New Phytologist* **87**, 11-37.
- Skarda, C.A., & Freeman, W.J., (1987) How brains make chaos in order to make sense of the world. *BBS* **10**:161–195.
- Smithers, T. (1995). Are autonomous systems information processing systems? In Steels, L. and Brooks, R., (Eds.), *The artificial life route to artificial intelligence: Building situated embodied agents*. New Haven. Erlbaum.
- Smithers, T. (1997) Autonomy in Robots and Other Agents. *Brain and Cognition* **34**:88—106.
- Steels, L. (1991) Towards a theory of emergent functionality in Meyer, J. Wilson, R. (eds.) *Simulation of Adaptive Behavior*. MIT Press, Cambridge MA, pp. 451–461.
- Stewart, J.(1996) Cognition=Life. Implication for higher-level cognition. *Behavioral Processes* **35**:311–326.
- Thelen, E. & Smith, L. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, Mass.
- Thelen, E., Schöner, G., Scheier, C. & Smith, L., B. (2001) The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching. *Behavioral and Brain Sciences* **24**:1–18.
- Thompson, E. & Varela, F. (2001) Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences* **5**(10): 418—425.
- Tononi, G., Edelman, G., & Sporns, O. (1998). Complexity and coherency: integrating information in the brain. *Behavioural and Brain Sciences*, **2**(12):474—484.
- Uexküll, J von (1982: 1940) The theory of meaning. *Semiotica* 42(1):pp 25-82
- van Gelder, T. (1998) The dynamical hypothesis in cognitive science. *BBS* **21**(5):615–628.
- Varela, F. (1979) *Principles of biological autonomy*. Elsevier, New York
- Varela, F. (1989) *Connaître Les sciences cognitives. Tendances et perspectives*, Paris: Seuil

- Varela, F. (1997) Patterns of Life: Intertwining Identity and Cognition. *Brain & Cognition* 34 (1) 72-87.
- Varela, F., Coutinho, A., Dupire, H., Vaz, N. (1987) Cognitive Networks: Immune, Neural, and Otherwise. en A. Perelson Ed. *Theoretical Immunology*, Redwood City CA: Addison Wesley, vol. II 359—376.
- Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001) The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience* 2:229—239.
- Varela, F., Maturana, H. & Uribe, R (1974) Autopoiesis: The Organization of Living Systems, its characterization and a model. *BioSystems* 5:187—196.
- Varela, F., Thompson, E, & Rosch, E. (1991) *The embodied mind. Cognitive Science and Human Experience*, Cambridge MA: MIT Press (Traducción castellana en Gedisa, 1992).
- Wheeler, M. (1996) From Robots to Rothko. En Boden, M. (ed) *The Philosophy of Artificial Life*, Oxford University Press, 1996.



Este texto ha sido escrito con OpenOffice bajo un sistema **Debian GNU/Linux**, su realización hubiera sido imposible sin la comunidad del software libre.



La presente versión terminó de escribirse el 23 de Junio del 2004 con el tipo de letra Bitstream Vera Sans y Bitstream Charter.

