



## La Neurociencia Computacional en Churchland y Sejnowski (1992)

**Asier Arias Domínguez**

Universidad Complutense de Madrid

### Resumen

El presente trabajo tiene por objeto la exposición de una obra fundamental –y, desgraciadamente, poco comentada y valorada en el ámbito hispanohablante– dentro de la historia del debate contemporáneo entre simbolistas y conexionistas *de segunda generación* (escuela gestada a partir de la segunda mitad de la década de los ochenta): *The Computational Brain*, de Patricia S. Churchland y Terrence J. Sejnowski. En él no sólo expondremos y comentaremos los contenidos de la misma, sino que presentaremos un marco histórico, teórico, conceptual y filosófico que, entendemos, será de utilidad no sólo de cara a ofrecer una introducción y una visión de conjunto de los elementos críticos del señalado debate, sino, en términos generales, para fomentar una discusión acerca de los mismos dentro del ámbito hispanohablante que incorpore las aportaciones contenidas en esta obra al referido debate.

### Palabras clave

neurociencia cognitiva, neurociencia computacional, conexionismo, representación.

### 1. Introducción

De cara a crear un marco en el que nos sea dable ensayar un sumario pero, en la medida de lo posible, solvente acercamiento a la obra y los temas de los que nos ocuparemos a lo largo de este trabajo, presentaremos en primer término un compendioso cuadro general –*sinóptico*, cabría decir– del ámbito en el que en lo sucesivo habremos de movernos: el de la neurociencia cognitiva, un reciente y promisorio campo del estudio científico de la mente en el que han proliferado valiosas perspectivas y aportaciones desde sus orígenes en la década de los ochenta. Un resolutivo y superficial vistazo a su historia y precursores nos será de ayuda en este sentido.

De entre las diferentes versiones que circulan sobre el nacimiento de la neurociencia cognitiva mencionaremos aquí las dos más frecuentemente citadas: las que implican a David Marr, por una parte, y a Michael S. Gazzaniga y George A. Miller, por otra. Al primero se le atribuye a menudo el mérito de haber creado la base para ulteriores desarrollos en esta área dado que su feraz *Vision*, publicado póstumamente en 1982 –dos años después de la prematura muerte de Marr–, ha ejercido una constante influencia y constituye un perfecto dechado de primevo conato integrador, en el que, para el estudio de la visión y el análisis del procesamiento de la información visual, se atiende a los niveles *computacional* y *algorítmico*, y, asimismo, al nivel *de implementación* –que en el caso de la visión humana tendría que ver con el modo en que ésta es realizada o llevada a cabo en el sistema nervioso por el sistema visual. En la propuesta de Marr tendrían pues cabida planteamientos provenientes de las ciencias de la computación (IA), las ciencias cognitivas (psicología cognitiva), y las neurociencias (neurofisiología). La segunda de las versiones hace referencia más bien

a la acuñación del término y no pasa de constituir una significativa anécdota.<sup>1</sup> Según ella, las primeras intuiciones acerca de qué –y cómo– caería en este campo de estudio provienen de una conversación entre dos grandes figuras de las ciencias de la mente: Gazzaniga y Miller. La conversación habría tenido lugar a finales de la década de los setenta en el asiento trasero de un taxi neoyorquino, camino de una reunión (una cena) organizada por científicos de las universidades Rockefeller y Cornell en la que se plantearía como tema para la discusión la forma de coadyuvar desde diferentes especialidades al esclarecimiento del modo en que el cerebro da lugar a la mente. En ese taxi, pues, se habría acuñado la denominación <<neurociencia cognitiva>> para esta disciplina en ciernes.<sup>2</sup>

Cabe, en líneas tal vez demasiado generales, presentar a la neurociencia cognitiva como un desarrollo o extensión de los programas teóricos y heurísticos de la neurociencia, o, por así decir, como su punto de desembocadura, en el cual confluiría con la psicología. Aquí, en semejante desembocadura, el estudio molecular, anatómico y fisiológico del sistema nervioso deviene análisis del procesamiento de información por parte del mismo, en un campo de investigación interdisciplinar ligado a lo que, desde la neuropsicología, Luria llamó *funciones corticales superiores*<sup>3</sup> — percepción, lenguaje, motricidad, planificación...—; ligado pues a las *funciones mentales* —desde una perspectiva cognitivista-representacional— los patrones conductuales, los neuronales y su interrelación. Se trataría en definitiva de aquella rama de las neurociencias sobre la cual han tenido verdadero influjo los planteamientos desarrollados a partir del Simposio de Hixon (*Los mecanismos cerebrales involucrados en la conducta*, celebrado en Pasadena en 1948 con la participación, entre otros, de, Norbert Wiener, John von Neumann, Warren McCulloch y Karl Lashley) y la Conferencia Dartmouth (*Sobre la teoría de la información*, celebrada en el MIT en 1956 con la participación, entre otros, de John McCarthy, Claude Shannon, Marvin Minsky, Allen Newell, Herbert Simon, Noam Chomsky y George Miller); es decir, de la rama de las neurociencias vinculada al estudio experimental de la cognición desde supuestos a caballo entre la biología y la teoría de la información.

Con lo hasta aquí señalado, la neurociencia cognitiva apenas sería discernible dentro del *paisaje* que *pintan* la mayor parte de los intentos omnímodos de definición de las ciencias cognitivas. Ejemplo de este tipo de intentos globales de definición de <<ciencia cognitiva>> que encajarían con una caracterización somera -pero útil en un primer acercamiento- de <<neurociencia cognitiva>> lo ofrece el bizarro y meritorio -pero, a la sazón, vilipendiado- informe sobre el estado de la/s disciplina/s cognitiva/s<sup>4</sup>

---

<sup>1</sup> No obstante, esta acuñación no deja de ser representativa, al punto que no resulta extraño encontrar que se la mencione como *punto de partida* de la disciplina (así en Brook, A./ Mandik, P. "The Philosophy and Neuroscience Movement". *Analyse & Kritik* 26/2004; pp. 382-397).

<sup>2</sup> Vid., v. g., Gazzaniga, M. S./Ivry, R. B./Mangun, G. R. *Cognitive neuroscience: The biology of mind*. New York, Norton & Comp., 2002; o la página Web de la *Cognitive Neuroscience Society* (<http://cogneurosociety.org/about/cognitive-neuroscience>).

<sup>3</sup> No utilizamos aquí literalmente la noción —es decir, no pretendemos sugerir que la actividad nerviosa cortical sea la única que interesa a la neurociencia cognitiva—, sino que con ella pretendemos llamar la atención sobre el objeto de estudio de la neurociencia cognitiva: los, por así decir, más *mentales* de entre los conglomerados funcionales de actividad nerviosa.

<sup>4</sup> Incluimos aquí los plurales por cuanto, en el momento al que nos referimos, lo que parecía estar en juego era si resultaba posible tender puentes sólidos y directos entre los diversos campos que, en tanto unidad, se hallaban a mediados de los setenta en pleno proceso de consolidación institucional, proceso que se viera plasmado en la edición de la revista *Cognitive science*, en 1977, y la fundación de la *Cognitive Science Society* en 1979. Las reacciones desde la psicología, la antropología, la lingüística, la filosofía, la neurociencia y la IA al informe

encargado en 1978 por la fundación Sloan. En él, las ciencias cognitivas fueron presentadas como un conjunto formado por diversas especialidades aunadas por su dedicación a un objetivo común: la investigación de "*las capacidades de la mente para la representación y la computación, y su representación estructural en el cerebro*".<sup>5</sup> Un primer paso hacia la elucidación de la especificidad de la neurociencia cognitiva -un poco más allá pues de la vaga generalidad en la que venimos moviéndonos- consistiría en definir su objeto. Así, cabe interinamente señalar en este sentido que la neurociencia cognitiva trata de responder a la siguiente cuestión: ¿cómo la mente es efectivamente producida por el cerebro? En su empeño por ofrecer resultados en esta dirección la neurociencia cognitiva recurriría principalmente a métodos y planteamientos provenientes de la neuroanatomía, la neurofisiología, la neurología, la neurocirugía funcional, la neuroimagen (PET, fMRI, TAC), la psicología cognitiva y la modelización computarizada.

Esta caracterización de la neurociencia cognitiva como ciencia dedicada a la investigación del modo en que la actividad cerebral revierte en actividad mental nos vuelve a sumir en la indeterminación: con ella apenas logramos ver el modo en que la neurociencia cognitiva diverge del plan general de las neurociencias. No obstante, la neurociencia cognitiva se alza como disciplina *autónoma*, diferenciada, en tanto ella trata de descubrir, como podemos leer en *The cognitive neurosciences*<sup>6</sup>, los algoritmos que estarían *poniéndose en funcionamiento* con la actividad fisiológica realizada en las estructuras neuronales *implicadas en* la percepción, la cognición, etc. Otro modo de acercarnos a su carácter diferencial nos lo proporciona la comprensión histórica del contexto en el que surgiera, en un momento en que la psicología cognitiva comienza a prestar decisiva atención al cerebro y su funcionamiento en diferentes niveles, propiciando el trabajo conjunto de psicólogos experimentales, neurocientíficos y especialistas en ciencias de la computación. Suele aceptarse que un antecedente fundamental de la confluencia de psicología cognitiva y neurociencias tuvo lugar en los estudios clásicos llevados a cabo durante la década de los cincuenta por Vernon Mountcastle, David Hubel y Torsten Wiesel, los cuales nos legaran la idea de la distribución modular de la corteza. Dos décadas después, el interés mutuo entre ambas disciplinas era ya un hecho inconcuso, y, a finales de los setenta, parecía evidente que un nuevo enfoque en el estudio de la mente-cerebro estaba gestándose, como ponían de relieve los numerosos seminarios a los que psicólogos y neurocientíficos acudían en pos de una sinergia cuya forma concreta habría de ser precisada en lo sucesivo. Por otra parte, los planteamientos representacionales que cobraran destacada pujanza a partir de las postrimerías de la década de los sesenta en psicología -de la mano de Saul Steiner, Michael Posner y Roger Shepar- hallarían en las ciencias de la computación un marco inigualable para el desarrollo de sus perspectivas teórico-experimentales, un marco que, asimismo, ha sido de extraordinaria relevancia en la consolidación de la terminología explicativa de la neurociencia cognitiva, ya que es de las ciencias de la computación de donde proviene su aparato conceptual descriptivo-explicativo de la función cerebral. Como continuador de una tarea emprendida tres décadas antes de él por McCulloch y Pitts dentro de un campo de investigación situado entre el cerebro y el ordenador, David Marr aparece en este sentido como la figura fundamental por lo que al puente entre neurociencia, psicología y ciencias de la computación respecta. Sus trabajos sobre la visión, a los que ya hemos hecho referencia, jalonan decisivamente el campo de la neurociencia

---

encargado por la fundación Sloan dejaron claro que satisfacer por igual en el trazado de los referidos puentes a cada uno de los grupos de especialistas no iba a ser tarea fácil.

<sup>5</sup> Citado por Gardner, H. *La nueva ciencia de la mente. Historia de la revolución cognitiva*. Barcelona, Paidós, 1988; pág. 54.

<sup>6</sup> Gazzaniga, M. (ed.). *The cognitive neurosciences*. Cambridge, MIT Press, 1995.

cognitiva al establecer los tres señalados niveles de análisis (dos computacionales y uno neurológico) de las computaciones llevadas a cabo por el cerebro -tanto es así que, como apuntábamos, no resulta extraño hallar que se presente a Marr como fundador de la disciplina-, una diferenciación por niveles que debería esperar posteriores matizaciones para alcanzar lo que Paul Thagard ha llamado recientemente plausibilidad neurológica -este aspecto de la restricción realista del campo computacional (potencialmente infinito: una enorme cantidad de algoritmos pueden resolver una misma tarea) al substrato neurofisiológico es, como tendremos ocasión de ver, una de las más descoltantes aportaciones de la propuesta desarrollada por Churchland y Sejnowski en el texto *clásico* del que aquí nos ocuparemos.

La referida acuñación de Gazzaniga y Miller alcanzaría oficialidad académica por vez primera en un curso de *Neurociencia Cognitiva* que ambos ofrecieron a comienzos de la década de los ochenta en el Cornell Medical College, con la intención de -tal y como Gazzaniga señala en el *Handbook of cognitive neuroscience* (1984)<sup>7</sup>- mostrar la orientación de ésta hacia la coordinación de los métodos de estudio del cerebro con los de las ciencias cognitivas. Un segundo paso reseñable hacia la oficialidad sería la aparición en 1989 del primer número del *Journal of Cognitive Neuroscience*, sólo un año después de que Gazzaniga lograra obtener de la Fundación James S. McDonnell y el Charitable Pew Trust un importante programa de financiación para la investigación y formación en neurociencia cognitiva.

Igualmente decisiva en esta época fue la publicación del artículo *Perspectives in cognitive neuroscience* (1988), en el que Churchland y Sejnowski llaman la atención sobre las restricciones que la neurobiología ha de imponer a las teorías computacionales; ponen asimismo de manifiesto en él la necesidad de una teorización puramente neurocomputacional -ya que, por contraposición a las pertinentes restricciones neurológicas referidas, no parece que todos los detalles de la cognición vayan a sernos accesibles por precisos que lleguen a ser nuestros conocimientos del nivel neurofisiológico<sup>8</sup>-, además de la de sumar a la investigación de las capacidades psicológicas y las propiedades de las neuronas y las redes neuronales la creación de modelos capaces de explicar el modo en que los patrones de actividad neuronal pueden alcanzar a representar aspectos del mundo, así como el modo en que las redes neuronales pueden aprender y desarrollarse, modelos que, en resumidas cuentas, serán de gran importancia de cara a orientar la experimentación en neurociencias, al tiempo que los resultados de esta experimentación servirán, en un proceso de retroalimentación, para crear mejores modelos. El estudio de uno de los niveles servirá para la corrección, orientación e inspiración del estudio del otro.

Although we need experimental data concerning the properties of neurons and behavioural data about psychological capacities, we also need to find models that explain how patterns of activity in neurons represent surfaces, optical flow, and objects;

---

<sup>7</sup> Este libro, editado por Gazzaniga, aparece -junto a textos como *Cognitive neuroscience: Developments towards a science of synthesis*, de Posner, Pea y Volpe (1982), y *Mind and brain: dialogues in cognitive neuroscience*, editado por LeDoux y Hirst en 1986- como documento fundacional de la disciplina.

<sup>8</sup> A este planteamiento ha llamado Juan Canseco en su artículo "Redes neuronales y conexionismo en las neurociencias" (*Metátesis* N° 3, marzo 2007) el despertar del "sueño de Cajal" (prosiguiendo con esta afortunada terminología de inspiración kantiana propone el autor que en el texto de Churchland y Sejnowski que comentaremos en este trabajo invitan los autores a despertar del "sueño de Boole" -este será el aspecto antisimbolista en el que insistiremos en el apartado final de conclusiones- y del "sueño de Marr", es decir, de la clara delimitación de los niveles a los que aludíamos; cabría añadir que en 1896 Patricia S. Churchland había tratado igualmente de rescatar a la comunidad cognitiva del hechizo del argumento funcionalista de la realizabilidad múltiple -Putnam, 1967-1975/1997 ).

how networks develop and learn, store, and retrieve information; and how networks accomplish sensorimotor and other types of integration. Ideally, modeling and experimental research will have a symbiotic relationship, such that each informs, corrects, and inspires the other.<sup>9</sup>

Churchland y Sejnowski plantean en el primer párrafo de este seminal artículo una idea fundamental en lo que al surgimiento de la neurociencia cognitiva se refiere: “there is now a gathering conviction among scientists that the time is right for a fruitful convergence of research from hitherto isolated fields”.<sup>10</sup> Esos campos que habrían permanecido aislados hasta el momento son los de las neurociencias y las ciencias cognitivas –la vertiente *decididamente computacional* de las mismas–, y el puente entre ellos habrá de construirse mediante modelización informática del procesamiento de información. Además, como subrayarán cuatro años después en la obra de la que aquí nos ocuparemos, inciden en este artículo en un extremo de singular significación: el nivel neuronal de análisis y el cognitivo deben limitarse y apoyarse mutuamente. Citaremos por extenso las conclusiones del artículo, ciertamente esclarecedoras en este sentido:

*It would be convenient if we could understand the nature of cognition without understanding the nature of the brain itself. Unfortunately, it is difficult if not impossible to theorize effectively on these matters in the absence of neurobiological constraints. The primary reason is that computational space is consummately vast, and there are many conceivable solutions to the problem of how a cognitive operation could be accomplished. Neurobiological data provide essential constraints on computational theories, and they consequently provide an efficient means for narrowing the search space. Equally important, the data are also richly suggestive in hints concerning what might really be going on and what computational strategies evolution might have chanced upon. Moreover, it is by no means settled what exactly are the functional categories at the cognitive levels, and theories of lower level function may well be crucial to the discovery of the nature of higher level organization. Accordingly, despite the fact that the brain is experimentally demanding, basic neurobiology is indispensable in the task of discovering the theories that explain how we perform such activities as seeing, thinking, and being aware.*

*On the other hand, the possibility that cognition will be an open book once we understand the details of each and every neuron and its development, connectivity, and response properties is likewise misconceived. Even if we could simulate, synapse for synapse, our entire nervous system, that accomplishment, by itself, would not be the same as understanding how it works. The simulation might be just as much of a mystery as the function of the brain currently is, for it may reveal nothing about the network and systems properties that hold the key to cognitive effects. Even simulations of small network models have capabilities that are difficult to understand. Genuine theorizing about the nature of neurocomputation is therefore essential.*

*Many major questions remain to be answered. Although some problems in vision, learning, attention, and sensorimotor control are yielding, this will be harder to achieve for more complex psychological phenomena such as reasoning and language. Nonetheless, once we understand some fundamental principles of brain function, we may see how to reformulate the outstanding problems and address them in ways that are impossible now to predict. Whatever the outcome, the results are likely to surprise us.*<sup>11</sup>

---

<sup>9</sup> Churchland, P. S./Sejnowski, T. J. “Perspectives in Cognitive Neuroscience”. *Science*, 4 November 1988, Vol. 242, pág. 744.

<sup>10</sup> *Ibíd.* Pág. 741.

<sup>11</sup> *Ibíd.* Pp. 744-745.

Venimos haciendo referencia a la neurociencia cognitiva *en bruto*, latamente; no obstante, de cara a trazar con algo más de precisión las demarcaciones dentro de las cuales habremos de desenvolvemos a lo largo de este trabajo, se hace necesario indicar que la obra de Churchland y Sejnowski de la que vamos a ocuparnos hace especial hincapié en el aspecto más cognitivista (desde una perspectiva conexionista), por así decir, de la neurociencia cognitiva: el de la modelización de redes *realistas* (neurológicamente plausibles<sup>12</sup>) del procesamiento cerebral de información; es por esto que en la misma hablan de neurociencia computacional. A este respecto, la Enciclopedia Routledge de Filosofía, en su artículo sobre el conexionismo, señala (en referencia a la obra de la que aquí nos ocuparemos):

*Computational neuroscience strives for more and more realistic networks with the aim of someday modelling actual neural-network information-processing.*<sup>13</sup>

El contenido elusivo y proverbial de este tipo de *sentencias* puede resultar desorientador, o excesivamente parcial; no obstante, consideramos que, antes de pasar a prestar atención al contenido de *The computational Brain*, servirán de exordio con mayor eficacia que cualquier anotación que pudiéramos sumar a nuestra sucinta introducción:

*A major goal of computational neuroscience is to provide theories as to how the brain works.*<sup>14</sup>

*Computational neuroscience seeks to answer organizational and functional questions about different areas of the brain.*<sup>15</sup>

*The ultimate aim of computational neuroscience is to explain how electrical and chemical signals are used in the brain to represent and process information.*<sup>16</sup>

## 2. The Computational Brain

Trataremos en lo sucesivo de exponer sucintamente los puntos fundamentales con los que en esta obra de Patricia Churchland y Terrence Sejnowski nos encontramos. Podría, ciertamente, resultar una tarea extravagante: nos hallamos ante una obra que compendia los resultados obtenidos e ilustra la dirección seguida por la investigación en neurociencia cognitiva durante su primera década de vida. Así, el sentido en que entendemos que *reseñar* esta obra podría resultar una extraña tarea es el sentido en que comentar un comentario o resumir un –en este caso prolijo– resumen muestran un cariz iterativo y no demasiado prometedor. No obstante, nosotros nos centraremos en los aspectos metodológicos, teóricos y descriptivos generales. Además, la obra cuyas líneas generales nos disponemos a comentar es más que una compilación o repaso de los logros alcanzados a lo largo de una década de modelización computarizada en neurociencia cognitiva: en ella, como veremos, además de una útil visión de conjunto se nos ofrecen potentes intuiciones y propuestas acerca del lugar que la modelización ha de ocupar en la investigación

---

<sup>12</sup> Thagard, P. *La mente. Introducción a las ciencias cognitivas*. Buenos Aires/Madrid, Katz, 2008.

<sup>13</sup> Craig, E (Ed.) *Routledge Encyclopedia of Philosophy. Philosophy of mind*. Londres & New York, Routledge, 1998; pág. 104.

<sup>14</sup> Lytton, W. W. *From computer to brain: foundations of computational neuroscience*. New York, Springer-Verlag, 2002; pág. 11.

<sup>15</sup> *Ibíd.* Pág. 32.

<sup>16</sup> Sejnowski, T. J./Koch, C./Churchland, P. S. "Computational Neuroscience". *Science, New Series*, Vol. 241, No. 4871. (Sep. 9, 1988); pág. 1299.

neurocientífica y del modo en que cabe entender que en el sistema nervioso se *manipulan* representaciones.

### 2.1. Nociones fundamentales y principios metodológicos.

El substrato filosófico de la aproximación a la neurociencia computacional que los autores nos ofrecerán apunta, como cabría esperar, al materialismo eliminativo<sup>17</sup> defendido por los Churchland desde principios de los ochenta.<sup>18</sup> A pesar de que no es éste un extremo desarrollado en *The Computational Brain*, puede fácilmente apreciarse en *declaraciones* que ya en el prefacio traslucen dicha postura: (v. g.: <<*psychological processes are in fact processes of the physical brain*>>). En esta línea, frente al pesimismo misteriano de quienes se empeñan en negar la posibilidad de dar explicación a nuestras nociones psicológicas desde la neurobiología (los autores incluyen aquí a Nagel, Eccles, Searle y Penrose; no mencionan a dos de los principales misterianos: Colin McGinn y David Chalmers –este último no había publicado aún su obra principal y sólo habían salido a la luz con su firma alrededor de una decena de artículos sobre IA y la conciencia), la propuesta de nuestros autores se sitúa, como la de la mayoría de los neurocientíficos (de Llinás, Crick, Damasio, Edelman, Kandel, Zeki o Gazzaniga a Varela), dentro de las más o menos difusas demarcaciones de un emergentismo naturalista según el cual caben *muy muchas* posibilidades de que las propiedades –*mentales*– de alto nivel sean –netamente– explicables en términos de propiedades –neurofisiológicas y computacionales– de bajo nivel. Esta hipótesis de trabajo les parece al menos más provechosa que la adopción de una actitud crítica a priori fundada en argumentos intuitivos o, en la terminología que tan cara le es a Searle, *de sentido común*. No tenemos ningún motivo, más allá de la mera *ideología*, para sostener que lo mental es de algún modo irreducible; es *demasiado pronto* –y seguirá siéndolo mientras no se ensayen más que argumentos filosóficos negativos– para asegurar que los fenómenos psicológicos no pueden explicarse en términos neurobiológicos. Así, queda dicho: *es demasiado tarde* para patéticas actitudes proféticas, dado que no sabemos lo que podremos llegar saber –aunque algunos se empeñen en acudir a imaginativos experimentos mentales, como si de oráculos se tratara, para determinar con taumaturgias artes sortilégicas hasta dónde, y hasta dónde no, pueden llegar las explicaciones de las ciencias de la mente.

La segunda hipótesis de trabajo que subyace al texto que nos ocupa es que el conocimiento del nivel molecular y celular del sistema nervioso es esencial para comprender la mente, pero no suficiente: como es obvio, se hace necesaria una perspectiva sistémica en la que encajar características propias de un nivel superior, el nivel de las redes neuronales. En resumen: las propiedades de las redes, a pesar de su dependencia respecto de las neuronas que las integran, no son necesariamente idénticas a las de éstas. Los datos provenientes del micronivel son pues necesarios para comprender el sistema, pero no suficientes.

La tercera hipótesis de trabajo o principio metodológico que hallamos sustentando la propuesta desarrollada en *The Computational Brain* es una asunción canónica dentro del paradigma cognitivista: en el cerebro se procesan representaciones.

<sup>17</sup> Algunos autores han defendido la idea de que el conexionismo implica o va de la mano del eliminacionismo (Ramsey et al., 1990), mientras otros (v. g. Smolensky, 1988) no encuentran justificación para conclusiones de este tipo en el marco de un adecuado (*proper*) tratamiento del conexionismo.

<sup>18</sup> Vid., v. g., el artículo de referencia “Eliminative Materialism and the Propositional Attitudes”, aparecido en 1981 en el volumen 78 de *Journal of Philosophy*.

Tenemos con lo hasta aquí indicado las bases para comprender el modo en que la neurociencia computacional es presentada por nuestros autores como una disciplina asentada en la neurobiología y las ciencias de la computación y encargada de investigar la forma en que las redes neuronales producen sus complejos efectos de alto nivel. Esta disciplina tendría pues, como señalan los autores, un pie en las neurociencias, otro en las ciencias de la computación, aun un tercero en la psicología experimental y un dedo (*toe*) en la filosofía. Desde este punto de vista, la neurociencia computacional ha de tender en la modelización computarizada de redes neuronales al realismo biológico<sup>19</sup>; en este sentido, el *ordenador* no aparece en el contexto de la neurociencia computacional como una mera herramienta para la investigación, sino que, como tendremos ocasión de ver, Churchland y Sejnowski entienden que los sistemas nerviosos naturales llevan a cabo computaciones y son de hecho computadoras surgidas en el proceso evolutivo, computadoras orgánicas con una arquitectura de procesamiento paralela. De este modo, descubrir los algoritmos, estrategias y principios computacionales del procesamiento neuronal de información equivale a un trabajo de ingeniería inversa en el que de lo que se trata es de descubrir los medios computacionales utilizados por un ingeniero cuya labor se vio constreñida por limitaciones temporales y espaciales (el cráneo), un programador cuya forma de resolver los problemas dados no es siempre necesariamente óptima: la selección natural. Desde este punto de vista biológico, en su artículo sobre computación y cerebro para *The MIT Encyclopedia of Cognitive Sciences* distinguen Churchland (P. S.) y Grush entre computación algorítmica (que coincidiría con la perspectiva cognitivista clásica serial) y computación de procesamiento de información (que coincidiría con la perspectiva conexionista), caracterizando a esta segunda de una manera que invita a contemplar el cerebro del siguiente modo: “as a biological device that processes information from the environment to build complex representations that enable the brain to make predictions and select advantageous behaviors”.<sup>20</sup>

Se hace necesario para cerrar este primer acercamiento a las nociones básicas de la neurociencia computacional incidir en un extremo al que hiciéramos ya referencia en la introducción: nos hallamos ante un entramado de diversos *niveles* (neuronas, conexiones, redes, interacciones de redes en sistemas) que podemos modelar para alcanzar una mejor comprensión de los principios que rigen su funcionamiento. La investigación en el plano neurofisiológico y en el de la modelización estarían, como indicábamos, vinculadas sinérgicamente; pero, además, tanto la modelización como la investigación neurobiológica *en los diferentes niveles* a los que apuntábamos se hallan igualmente integradas en un proceso coevolutivo. Por lo que respecta a los aspectos de la neurociencia computacional en los que habremos de fijarnos, el nivel de las redes será el protagonista –más concretamente, lo será el de los modelos neurobiológicamente realistas de redes neuronales implicadas en tareas y capacidades bien estudiadas por la psicología y la neuropsicología.

---

<sup>19</sup> <<Computational space is undoubtedly vast, and the possible ways to perform any task are probably legion>> (Churchland, P. S./Sejnowski, T. J. *The Computational Brain*. Cambridge, MIT Press, 1992; pág 11), de modo que si pretendemos aprender algo acerca de la cognición humana no nos servirá cualquiera de las formas –computacionalmente viables– de resolver una determinada tarea.

<sup>20</sup> Citamos una versión del artículo (“Computation and the Brain”) previa a la edición de la enciclopedia, editada por Rob Wilson y Frank Keil, de la que pasaría a formar parte.



## 2.2. Neurociencia...

El recorrido a través de la neurociencia computacional que Churchland y Sejnowski nos proponen toma el primer impulso en el primer término del binomio, es decir, en el *componente neurocientífico* de la neurociencia computacional. El repaso que se nos presenta en *The Computational Brain* a los puntos fundamentales de la neurociencia –los cuales son traídos aquí, como cabría esperar, con la intención de propiciar una aproximación computacional al funcionamiento del sistema nervioso– comienza con la siguiente advertencia: el estilo computacional del cerebro y los principios que gobiernan su función escapan al análisis causal y al conductista. Invitándonos a no perder de vista el contenido de tal advertencia, nos abren los autores paso en un recorrido guiado a través de la neurociencia en el que nos orientarán partiendo de la consideración de los diversos niveles de análisis, organización y procesamiento del sistema nervioso. Veamos brevemente a qué se refieren cada uno de los tres.

Por lo que a los primeros –los niveles de análisis– se refiere, nos encontramos ante la distinción de tres niveles propuesta por Marr a la que ya hemos hecho referencia (nivel computacional del análisis abstracto del problema y descomposición en tareas, nivel algorítmico del procedimiento formal para la realización de la tarea y nivel de implementación física). Cabe destacar por lo que a ella respecta la independencia con la que Marr dotó en su caracterización de los niveles superiores a éstos respecto de los anteriores o inferiores<sup>21</sup>. Este punto es un importante foco de polémica, dado que caben diferentes posturas acerca de qué debe entenderse por independencia del nivel algorítmico respecto del de implementación –para algunos es claro que la estrategia vertical descendente propugnada por Marr trae consigo la consecuencia de que la neurobiología queda relegada a un plano marginal o anecdótico cuando de lo que se trata es, curiosamente, de estudiar propiedades de sistemas neurobiológicos; no obstante, hay que señalar que considerar desde este punto de vista el trabajo de Marr equivale a distorsionar su significado y dirección, ya que desde su tesis doctoral puede apreciarse que el nivel neurofisiológico y neurobiológico en general constituyó un importante punto de referencia en su propuesta. La perspectiva de Churchland y Sejnowski al respecto puede resumirse señalando que entienden que la referida independencia puede ser aprovechada para construir máquinas diferentes en cuanto a su implementación pero computacionalmente equivalentes una vez conocemos cómo funciona el cerebro, pero no podría servirnos de guía para descubrir, precisamente, cómo funciona el cerebro.

*The issue of independence of levels marks a major conceptual difference between Marr (1982) and the current generation of researchers studying neural and connectionist models. In contrast to the doctrine of independence, current research suggests that considerations of implementation play a vital role in the kinds of algorithms that are devised and the kind of computational insights available to the scientist. Knowledge of brain architecture, far from being irrelevant to the project, can be essential basis and invaluable catalyst for devising likely and powerful algorithms - algorithms that have a reasonable shot at explaining how in fact the neurons do the job.*<sup>22</sup>

---

<sup>21</sup> Añadamos a esto la siguiente referencia: <<his preferred strategy was top-down rather than bottom-up>>. (Churchland, P. S./Sejnowski, T. J. *The Computational Brain*. Cambridge, MIT Press, 1992; pág 18).

<sup>22</sup> Churchland, P. S./Sejnowski, T. J. *The Computational Brain*. Cambridge, MIT Press, 1992; pág 19. Añadamos que el contenido de este apartado sobre niveles puede encontrarse en gran cantidad de obras colectivas en las que Churchland y Sejnowski han participado (en ocasiones

En lo tocante a los niveles de organización, nos hallamos en un marco en el que prima la anatomía, un marco desde el que nos es dable matizar la anterior distinción de tres niveles de análisis, dado que ésta no nos proporciona unas delineaciones definitivas conforme a las cuales nos resulte posible estructurar y clasificar todos y cada uno de los extremos que vayan apareciéndose al paso. Así, uno y el mismo nivel de organización puede analizarse desde un punto de vista computacional o de implementación.

En cuanto a los niveles de procesamiento, lo fundamental será tener presente el vínculo entre anatomía y representación sin perder de vista que la noción de jerarquía de procesamiento, a pesar de poder responder a ciertas simplificaciones (como *mayor cuanto más lejos de la entrada sensorial*), está fuertemente condicionada por cursos de acción acaecidos en vías de reentrada y proyecciones recíprocas – además de resultar simplemente aproximada en casos como la organización de las áreas sensoriales primarias. Otros aspectos, como el tiempo –y el orden– de procesamiento o la diferenciación de distintas clases de procesamiento para una y la misma tarea, por ejemplo, han de ser tenidos en cuenta y merece la pena hacer mención de ellos en este ínterin, aunque no pasemos de nombrarlos.

### **a) Niveles de organización**

Dado que este primer acercamiento a la neurociencia computacional parte del elemento neurocientífico, los niveles que nos resultarán en este momento interesantes son los segundos, los de organización. Siguiendo una trayectoria descendente (de los sistemas a las moléculas) procuraremos trazar un esbozo de lo que merece la pena tener en cuenta a este respecto.

En la cúspide, en el extremo de las grandes formaciones, tendríamos los sistemas (como el visual o el neuromotor), de los cuales subrayan los autores la importancia de las conexiones recíprocas entre áreas cerebrales, así como el hecho de que a pesar de que una neurona es presentada en los modelos simples de arco reflejo como suficiente para producir la activación de otra, normalmente una cantidad considerable de ellas suele estar involucrada en dicho proceso. En el siguiente peldaño de los niveles de organización nos hallaríamos con los mapas topográficos, de gran relevancia en el contexto de los sistemas sensoriales y motores –un ejemplo habitual en este sentido es el de los famosos homúnculos, o el de los mapas visuales, que en los lóbulos occipitales se organizan topográficamente formando mapas de la retina; del mismo modo pueden contarse como excepciones a este principio de organización el sistema olfativo o, en cierta medida, el cerebelo. El siguiente nivel de organización corresponde a las capas y columnas. Gran cantidad de estructuras cerebrales muestran no sólo una distribución con arreglo a patrones topográficos, sino y al tiempo, una organización en láminas o capas especializadas, tanto en lo relativo a sus aferencias como a sus eferencias (un ejemplo sería el colículo superior, una de las principales formaciones mesencefálicas, cuyas capas más profundas reciben entradas táctiles y auditivas, mientras que las superiores o superficiales las reciben visuales). Las columnas, por su parte, corresponderían a formaciones verticales de células interconectadas a través de capas –recuérdese que la corteza se halla dividida, en su mayor parte, en seis capas formadas por diversos tipos de neuronas. Resulta destacable en este punto el hecho de que los autores apunten que tanto los mapas topográficos como la organización en capas y columnas son casos especiales de un

---

con Koch, vid. op. cit.) –v. g.: Posner, M. I. (Ed.) (1989, pp. 301 y ss.), Schwartz, E. L. (Ed.) (1990, pp. 46 y ss.), Gutfreund, H./Toulouse, G. (Eds.) (1994, pp. 25 y ss.), Gazzaniga, M. S. (Ed.) (2000b, pp. 14 y ss.).

principio general: el aprovechamiento de propiedades geométricas para el diseño de dispositivos de procesamiento de la información. El siguiente nivel de organización es el de las redes locales. De cara a introducir este principio organizativo, son interesantes algunos datos cuantitativos simples: en un centímetro cúbico de tejido neuronal podemos encontrar cien mil neuronas y hasta mil millones de sinapsis. La enorme complejidad a la que esta densidad da lugar ha dificultado considerablemente el trabajo de descifrar los principios que rigen el funcionamiento de las redes neuronales locales, un trabajo que comenzaría a ofrecer frutos gracias al registro discreto de la actividad de células individuales y al apoyo interpretativo en modelización computerizada del que pueden servirse los investigadores a la hora de estudiar el modo en que poblaciones de neuronas pueden representar aspectos del mundo, contribuir a la realización de transformaciones coordinadas o coadyuvar a la ejecución de otras funciones –como en el caso de los modelos de redes para la representación espacial en la corteza parietal desarrollados por Andersen y Mountcastle o para la visual por Zeki o Poggio. Un peldaño más abajo nos encontraríamos ya al nivel de las neuronas, las unidades mínimas de procesamiento en el sistema nervioso, que podemos dividir en excitatorias e inhibitorias (aunque, como es sabido, todas pueden tipificarse con arreglo a diferentes criterios de clasificación: según forma y tamaño, polaridad, tipos de axones y dendritas o mediadores químicos). El avance en la comprensión del funcionamiento de las neuronas ha confutado la interpretación *tradicional* en la que éstas aparecían como dispositivos simples de procesamiento, y así, por ejemplo, la especialización de determinadas dendritas puso ya a mediados de los ochenta a diferentes investigadores<sup>23</sup> en la pista de una nueva concepción de las mismas, según la cual las dendritas podrían ser entendidas como unidades independientes de procesamiento. El penúltimo peldaño lo ocuparían las sinapsis, las vías de comunicación electroquímica entre neuronas –mediadas por la especialización química pre y post sináptica– susceptibles de ser modificadas por la experiencia. Por último, en el nivel molecular, el más elemental, cabe subrayar la importancia de las proteínas de membrana, sensibles al voltaje o activables químicamente, que pueden servir como canales iónicos, los cuales juegan un papel decisivo en las sinapsis.

#### **b) Breve relación de hechos destacados acerca del sistema nervioso desde el punto de vista de la neurocomputación**

Pasaremos a continuación a presentar esquemáticamente –siguiendo el hilo conductor propuesto por los autores– los rasgos estructurales del sistema nervioso relevantes para la *neurocomputación*.

1. La especialización funcional de las diversas estructuras neuronales y regiones del sistema nervioso.
2. La cantidad de neuronas y sinapsis (estimada en cientos de miles de millones y billones, respectivamente).
3. La conectividad: a pesar de la centralidad del fundamental fenómeno de la reentrada –considerada por destacados neurocientíficos, como Gerald Edelman (vid. 1987, 1990, 1992 y 2000), uno de los rasgos decisivos del sistema nervioso–, no todo se conecta en este contexto con todo.
4. La entrada a una neurona es análoga (tiene valores discretos entre 0 y 1) mientras su salida es discreta (se dispara o no) –aunque algunas neuronas puedan tener salidas análogas, los outputs están regidos o determinados por *reglas de umbral*, es decir, por umbrales por encima de los cuales se disparan: que una neurona se

---

<sup>23</sup> Entre ellos Koch y Poggio.

- dispare* depende de que la integración de los inputs que llegan a sus dendritas exceda un determinado umbral.
5. Coordinación temporal o ritmo: por lo que a la percepción se refiere, la escala temporal de las computaciones debe enlazar o hallarse en armonía con la de los eventos externos; en lo que a la motricidad toca dicha escala ha de hacerse corresponder con el tiempo necesario para mover adecuadamente las diferentes partes del cuerpo.
  6. Efectos de célula a célula: por regla general, como apuntábamos, el efecto individual de un input en una célula postsináptica es débil (de entre el 1% y el 5% de lo necesario para alcanzar el umbral por encima del cual se produciría un potencial de acción).
  7. Los diversos tipos de neuronas tienen diversos patrones de disparo –la mayoría de ellas son espontáneamente activas y presentan intervalos aleatorios de disparo en ausencia de inputs, pudiendo llegar a los cincuenta disparos por segundo.
  8. Los campos receptivos son zonas del campo perceptivo desde las cuales les es dable a los estímulos provocar respuestas. Así, por ejemplo, el campo receptivo de las células del sistema visual, que tiende a extenderse en paralelo a la jerarquía de procesamiento, puede adoptar dos configuraciones: una en la que los estímulos que alcancen el centro del campo receptivo de una célula la excite mientras los que permanecen en la periferia la inhiben (on-center/off-surround), y otra en la que todo sucede a la inversa (off-center/on-surround).
  9. Más allá de la concepción clásica del campo receptivo, se ha hallado que estímulos fuera del mismo pueden modular selectivamente las respuestas de la célula.
  10. A la acción de los sistemas específicos cabe sumar la de los llamados centros neuromoduladores de valor (los autores los denominan aquí simplemente <<nonspecific systems>>), como el locus coeruleus, los núcleos de rafe o los tubérculos mamilares.
  11. Sustancias tales como las hormonas pueden alcanzar neuronas, llegando a ellas a través del torrente sanguíneo, y alterar su actividad ejerciendo sobre ellas una *acción a distancia*.
  12. En el cerebro encontramos una ingente cantidad de aferencias paralelas para una función dada. Es esto lo que se ha dado en llamar una *arquitectura paralela*.

### **2.3. ...computacional**

Tras el repaso a los elementos neurocientíficos que resultan de interés para el propósito de nuestro texto, pasan los autores a introducir, con algo más de amplitud que en el caso anterior, los aspectos computacionales que nos serán de utilidad en nuestra aproximación a la neurociencia computacional. Dicha introducción parte de la constatación de la limitación de nuestros conocimientos actuales: “there is a lot we do not yet know about computation”.<sup>24</sup> Y esto, cabe añadir, a pesar de los enormes logros alcanzados en esta área, tanto tecnológica como teóricamente, desde los pioneros trabajos de Turing, von Neumann, Rosenblatt, McCulloch y Pitts hace más de medio siglo. Resistiéndose a ofrecer una definición formal de computación los autores señalan que la ciencia acaba por ofrecer definiciones más que seguirse de ellas, pero avanzan que podemos entender un sistema físico como un sistema computacional cuando sus estados físicos puedan ser comprendidos como representaciones de estados de otros sistemas y cuando las transiciones entre sus estados pueden explicarse como operaciones realizadas sobre esas representaciones.

---

<sup>24</sup> Churchland, P. S./Sejnowski, T. J. *The Computational Brain*. Cambridge, MIT Press, 1992; pág. 61.

En un acercamiento gradual a nuestro objeto, los autores comienzan por definir función, en términos matemáticos de relación entre elementos pertenecientes a conjuntos (no usan estos términos, sino que hablan de pares ordenados), y función computable como aquella en la siguiendo una regla concreta pueden establecerse las correspondencias pertinentes entre los elementos que sean el caso dentro de cada uno de los conjuntos, especificando el modo de obtener el elemento de destino desde el de origen (los autores hablan de mapeo en términos de determinada regla), es decir, la relación entre elementos del conjunto inicial o dominio y el final, rango o codominio (pares ordenados: a cada elemento del dominio corresponde uno y sólo uno del codominio; en estos términos, una función no computable puede definirse como una serie infinita de pares ordenados para la cual no puede en principio establecerse una regla de correspondencia). Un extremo de interés es en este sentido si la actividad del cerebro puede ser caracterizada mediante funciones computables, un extremo que los autores consideran una cuestión empírica.

Dos nociones de álgebra lineal fundamentales para la comprensión de la modelización de redes neuronales son las de *vector* y *espacio de estado*. Vector puede definirse matemáticamente como una serie ordenada de elementos (los autores hablan aquí de series ordenadas de números, y, de cara a operar con ellos, de reglas sensibles al orden). Por lo que a nuestro tema respecta, la clave está en que los sistemas, como el nervioso, por ejemplo, ejecutan funciones en las que se realizan transformaciones de vector a vector (por ejemplo del valor de la velocidad del movimiento de la cabeza al de la del movimiento de los ojos). Por su parte, espacio de estado puede definirse como la articulación geométrica de la anterior noción (con tantos ejes como dimensiones se incluyan), como una serie de todos los vectores posibles. En cuanto a lo que a la neurociencia computacional atañe, a una red de  $n$  unidades corresponde un espacio de estado  $n$ -dimensional en el que registrar su estado de activación en un momento dado; un vector de activación en un momento dado correspondería a un punto de ese espacio, y una secuencia de ellos a una trayectoria en el mismo. Una función puede por tanto relacionar –siempre y cuando podamos encontrar regularidades, cabría señalar– un punto en un espacio de estado con un punto en otro espacio de estado.

Habiendo preparado con las anteriores precisiones el terreno para la discusión acerca de la computación podemos pasar a definir computadora indicando que cabe comprender *algo* como si de una computadora se tratara cuando sus inputs y sus outputs puedan interpretarse sistemáticamente como representaciones de pares ordenados de alguna función que nos interese. Así, tendríamos dos componentes de este criterio: el aspecto objetivo de qué función o funciones describe o describen el comportamiento del sistema y el subjetivo de si nos interesa la función. Con esto, las condiciones para tomar algo como una computadora no podrían sino ser tenidas por relativas –ninguna propiedad “intrínseca” podría guiarnos hacia la comprensión de un dispositivo como si se tratara, *objetivamente*, de una computadora. De todos modos, en determinadas ocasiones proceder a interpretar de este modo el comportamiento de determinados dispositivos resulta más que orientador cuando tratamos de comprender cómo funcionan; así, hallar qué y cómo es computado por una porción de tejido cortical puede revelarnos claves indispensables para la comprensión de su naturaleza y su funcionamiento.

Lo hasta aquí expuesto nos sirve para caracterizar al sistema nervioso como una clase de computadora, una *biocomputadora* con estados físicos representacionales y transiciones causales entre los mismos, es decir, operaciones computacionales realizadas sobre sus estados representacionales: sus estados físicos *representan* y las transiciones de unos estados físicos a otros *ejecutan computaciones*. Por otra parte, girando, por así decir, de la ontología a la epistemología y la metodología, una explicación meramente causal, en la que no se

incluyan elementos funcionales o computacionales, puede resultar insatisfactoria a la hora de analizar ciertos procesos llevados a cabo en el sistema nervioso: necesitamos para comprender su funcionamiento acudir a un nivel representacional, preguntarnos cómo representa, y cómo sus representaciones pueden transformarse en otras representaciones por medio del comportamiento y la interacción neuronal, *detalles* a los que la descripción meramente neurofisiológica no llega a conducirnos.

Pasaremos a continuación a ocuparnos de una serie de principios computacionales que nos serán de utilidad para aproximarnos a la cuestión de la computación en el sistema nervioso.

a) Buscar la solución:

Este título ilustra el principio computacional más simple, un principio a través del cual los autores nos guiarán con ayuda de la noción de *look-up table* (que en ocasiones podemos encontrar traducida como “tabla de consulta”). Una *look-up table* puede ser cualquier suerte de instrumento físico en el que almacenamos respuestas o soluciones para cuestiones específicas, y en el que, como en el caso de una regla de cálculo, siguiendo determinados criterios podemos realizar diferentes transformaciones y operaciones, es decir, computaciones. Indicaremos sucintamente que las limitaciones computacionales que una estrategia del tipo *look-up table* ofrece tienen principalmente que ver con la dificultad de la tarea a realizar, es decir, con la complejidad del problema a resolver. Los datos avalan la opinión de que una aproximación tipo *look-up table* a los procesos computacionales acaecidos – *globalmente*, latamente– en el sistema nervioso no depara buenas perspectivas. No obstante, se trata de un recurso computacional de gran utilidad a la hora de interpretar los datos obtenidos en el estudio del modo en que realmente computa el sistema nervioso: mediante redes.

b) Redes:

Una red neuronal artificial consta de unidades (*basadas* en las neuronas), conexiones (remotos de axones) y pesos (*sinapsis*) de conexión. Una función puede ser ejecutada por una red en tanto determinados valores de input dan lugar a determinadas activaciones de output en función del modo en que se hallen dispuestos los pesos de las conexiones. La primera indicación de nuestros autores será que las redes asocian un vector de entrada con uno de salida mediante la relación proporcionada por determinados pesos de conexión. Será útil en este punto introducir una serie de escuetas referencias y definiciones, de las que serviremos para dar el primer paso hacia una somera comprensión de la noción de red neuronal. Comenzaremos por señalar que la noción de *red neuronal artificial* equivale a la de *modelo computacional inspirado en redes neuronales reales*. Una red neuronal artificial puede entenderse pues como un sistema de procesamiento de información, y definirse como una estructura distribuida de procesamiento formada por neuronas artificiales (las referidas unidades). Uno de los más conocidos modelos de neurona artificial es el modelo clásico de McCulloch y Pitts, en el que la entrada a la unidad de una determinada cantidad de señales puede ser representada por las variables  $x_1, x_2, \dots, x_N$ , relacionadas con los pesos (representados por las variables  $w_{ij}$ , que indicarían el grado en que la neurona  $i$  influencia a la  $j$ ). Podemos hablar respecto del mismo de dos etapas de procesamiento neuronal. En la primera las señales de entrada  $x_j$  y los pesos  $w_{ij}$  son combinados por el sumatorio:

$$y_i = \sum_{j=1}^N w_{ij} x_j$$

(Aquí,  $y_i$  corresponde al estado interno de la neurona).

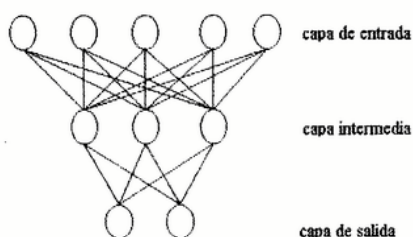
En la segunda se produciría la salida a través de la aplicación de la llamada función de activación:

$$x_i = f(y_i)$$

La salida de la neurona se halla representada en la misma por  $x_i$ . Por su parte,  $f$  corresponde a la función de activación aplicada al estado interno de la neurona.

Existen varios tipos de función de activación (como las lineales, las de umbral y las no lineales; en las primeras la salida es proporcional a la entrada, en las segundas es un valor discreto –dado al exceder un determinado umbral–, y en las terceras, las más interesantes, las que han dado lugar a un mayor número de avances significativos, no es proporcional a la entrada).

Otra noción relevante es en este punto la de la arquitectura de una red.



En una primera aproximación a esta noción cabe subrayar que diferentes tipos de arquitectura resultan útiles para diferentes tipos de tareas o problemas, de modo que el tipo de arquitectura restringe en cierta medida el tipo de problema que puede ser tratado y resuelto por la red.

Podemos asimismo distinguir tipos de redes en función del flujo de las señales. Así, nos hallamos con las redes *feedforward* (o de propagación hacia delante) y las *recurrentes*. Por lo que a las primeras respecta, cabe señalar que su estructura consiste en capas de neuronas en las cuales la salida de cada una de las neuronas de una capa alimenta todas las neuronas de la capa siguiente; lo fundamental aquí es la ausencia de conexiones de retroalimentación. En lo referente a las segundas, indicaremos brevemente que, por contraposición a las anteriores, poseen conexiones de retroalimentación (feedback).

Una propiedad fundamental de las redes neuronales es su capacidad de aprendizaje, propiciada por un proceso iterativo de ajustes en los pesos de conexión; es decir, su capacidad para aprender mediante entrenamiento. Interesa en este punto hacer referencia a un extremo en el que habremos de fijarnos más adelante: el de los algoritmos de aprendizaje, los cuales pueden ser de diversos tipos, como el de aprendizaje por corrección de error (basado en la regla Delta, que busca minimizar la función de error utilizando un gradiente descendente), el competitivo (en el cual las neuronas de una capa compiten entre si por permanecer activas), el hebbiano (en el cual neuronas simultáneamente activas fortalecen su conexión; en caso contrario la misma se debilita) o el de Boltzmann (dado por una regla de aprendizaje estocástico basada en principios de teoría de la información y de la termodinámica). Además, en

función del modo en que una red neuronal se relaciona con su entorno podemos distinguir diversos tipos de aprendizaje, como el supervisado (en el que se indica a la red la respuesta deseada para el patrón de entrada, es decir, se le ofrece una respuesta correcta para cada patrón de entrada), el de refuerzo (variante atenuada de la anterior: la indicación externa no es tan importante como en el caso del aprendizaje supervisado) y el no supervisado (en el que ningún agente externo indica la respuesta deseada para los patrones de entrada).

Una primera aproximación al modo en que la computación en redes de tipo lineal (sumatorio de entradas ponderadas) se encuentra frente a problemas que es incapaz de resolver lo ofrece el ejemplo del reconocimiento visual de objetos. El núcleo del problema sería aquí que cada entrada refleja sólo una parte concreta de la imagen a procesar. En este sentido, las unidades de un sistema de procesamiento paralelo habrían de converger de interacciones locales a una interpretación global correcta. En una red de propagación hacia delante no cabría una constricción output-input que permitiera alcanzar una respuesta mutuamente coherente para las unidades de output dada por la posibilidad de que la información circule entre ellas: la información corre en un sólo sentido y no produce variaciones que partan de las propias salidas. Se hace así necesario un proceso de feedback en el que la información no corra jerárquicamente, sino que siga un curso de *conversación cruzada*, y se facilite la revisión de planes.

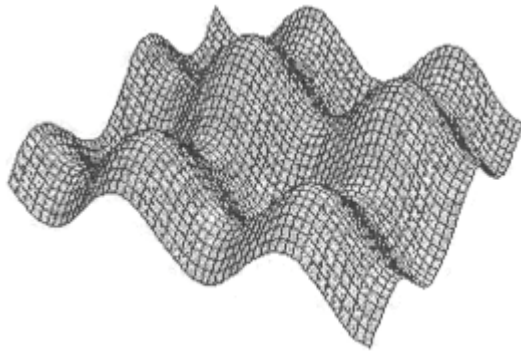
A diferencia de los ordenadores digitales, las redes sobre las que en lo sucesivo discutiremos funcionan sin especificaciones sobre los pasos intermedios. Con todo, al diseñar una red y enfrentarla a una tarea se espera que alcance un estado estable en el que el proceso de ajuste de sus pesos permita afrontar con éxito la tarea en cuestión. ¿Cómo? Los autores nos introducen en este extremo de la mano de las propuestas que Hopfield lanzara durante la primera mitad de los ochenta para la solución del problema del modo de ajustar pesos en una red con feedback (propuestas inspiradas, cabe señalar, en el comportamiento de sistemas físicos, en concreto, del vidrio de espín).

El proceso fundamental llevado a cabo por las redes de Hopfield, en lo que a nuestros intereses presentes se refiere, es el de la relajación (éste es el punto en que se basó Hopfield en una analogía formal con la termodinámica de vidrios de espín). Veamos cómo introduciendo al tiempo algunas de las características fundamentales de este tipo de red. Se trata de un tipo de red denominado *de adaptación probabilística* (o modelos de satisfacción de demandas). La característica fundamental de este tipo de red es que es asíncrona, de modo que las unidades se actualizan o no de forma probabilista, es decir, según cierta probabilidad, durante el ciclo de procesamiento. Se trata de redes cuyo objetivo es alcanzar soluciones –tan aproximadas como resulte posible– a problemas para los cuales se hace necesario tener en cuenta un gran número de demandas o requerimientos simultáneos. De cara a alcanzar dicho objetivo, utilizando retroalimentación (de modo que los estados pueden actualizarse iterativamente tendiendo a la minimización de la *energía de la red\**), parten de un estado dinámico o inestable y tienden a la estabilidad mediante un proceso de relajación, estado que una vez alcanzado puede evaluarse mediante una *función de energía\**. Las unidades en las redes de Hopfield pueden tener dos valores (0 y 1): variando aleatoriamente los valores se aceptan los cambios que reduzcan la energía de la red. “Roughly, the energy is simply the sum of all the weights connecting units which happen to be on, scaled, and inverted in sign, so that the network has the same formal properties as those of a physical system”.<sup>25</sup> Un gráfico tridimensional como el siguiente puede utilizarse para representar la función de energía de una red de Hopfield:

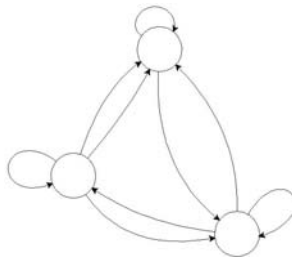
---

<sup>25</sup> Ibíd. Pág. 87.





En un número finito de pasos la red alcanzará un estado estable (tendiendo a mínimos de energía: *atractores*).



Modelo Hopfield.

Dentro de esta de redes de adaptación las máquinas de Boltzmann

caracterización general probabilista caerían igualmente (en las que la distinción entre

hardware y algoritmo se hace difusa ya que la propia configuración física del input codifica el problema computacional, y el algoritmo no pasaría de ser el proceso por el cual el sistema físico llega a la solución), que, al igual que las redes de Hopfield, se caracterizan por la presencia de memorias autoasociativas recurrentes (también denominadas redes retroasociativas). Por lo que al referido ejemplo del reconocimiento visual se refiere, cabe destacar que este tipo de redes son capaces de aprender a reconstruir patrones de entrada, de modo que pueden completar la tarea aun contando sólo con información incompleta o distorsionada, como puede suceder en la presentación de un objeto en varias perspectivas, iluminaciones, etc.

Retomando la cuestión del aprendizaje<sup>26</sup> en las redes neuronales, precisemos un poco la diferenciación entre aprendizaje supervisado y no supervisado, y ampliemos la clasificación. En el caso del primero, el aprendizaje descansa sobre tres basamentos: los inputs, la dinámica interna de la red y el trabajo de evaluación de la disposición de pesos de conexión. Las segundas se apoyan sólo en los dos primeros sin informes acerca de su comportamiento. Con todo, en los dos casos lo fundamental de los algoritmos de aprendizaje será la producción de una configuración de pesos de la que pueda decirse que *representa* algo en el sentido de que al ser activada por un vector de entrada ofrezca la respuesta pertinente. Los autores ponen el acento en que las redes de aprendizaje no supervisado no tienen acceso a feedback externo (como en el caso de la regla de Hebb o la regla de aprendizaje competitivo). Ofrecen en este punto una precisión destinada a deshacer cierta anfibología habitual en la literatura especializada en lo referente al feedback: cuando éste es externo al *organismo* se trata de aprendizaje supervisado; cuando hay una medida interna de error hablamos de aprendizaje *monitoreado*. Tendríamos pues cuatro posibilidades (surgidas de la combinación de supervisión y monitoreo). [Hay que señalar que la referida confusión en la literatura especializada proviene de que el mismo tipo de algoritmo puede

<sup>26</sup> Cabe hacer notar que los autores usan en *The Computational Brain* el término <<aprendizaje>> mostrando ciertas reservas.

utilizarse en las dos estrategias de limitación del error]. De modo que puede haber una medida interna de consistencia o coherencia que puede monitorearse internamente y ser usada para mejorar la representación. Así, por ejemplo, es plausible que en el sistema nervioso una parte del mismo monitoree el aprendizaje de otra. Una conjetura fundamental en este punto es que todo sistema capaz de una autoorganización exitosa, incluidos los sistemas biológicos, tiene una función de objetivo implícito que se optimiza durante el proceso de aprendizaje.

Existen grados en el aprendizaje supervisado en función del tipo de informe proporcionado a la red tras la realización de la tarea. El informe puede simplemente proporcionar datos acerca del éxito en la resolución del problema, especificar con cierto grado de precisión la magnitud del error, o proporcionar indicaciones detalladas acerca de la respuesta dada como contrapuesta a la requerida.

La regla de aprendizaje en las redes originales de Hopfield era de tipo hebbiano. Con ella los problemas que cabía tratar no alcanzaban una alta complejidad estadística. La posibilidad de ampliar este rango limitado de problemas tratables se hallaba en modificaciones en la arquitectura de la red: añadir *unidades ocultas* (que pueden ser lineales o no lineales, las cuales son necesarias para la representación de propiedades de alto nivel). Lo que resulta difícil entonces es encontrar una regla adecuada para el establecimiento de pesos de conexión apropiados para estas unidades ocultas, una regla en virtud de la cual la red pueda lidiar con problemas de alto nivel. A pesar de que este ajustamiento de pesos dependerá de la arquitectura de la red, la búsqueda exhaustiva –un proceso costoso e inviable en la mayoría de los contextos– parece ser la única alternativa. No obstante, para las máquinas de Boltzmann disponemos de un procedimiento con el que la red puede aprender a extraer propiedades de alto nivel. Lo fundamental en este extremo es una propiedad que las máquinas de Boltzmann muestran en su estado de equilibrio: los estados dispondrían en este caso de una distribución de Boltzmann, que para cada estado global proporciona la probabilidad de que ese estado ocurra en equilibrio, lo cual significa que en un estado de equilibrio pueden conocerse las consecuencias de cualquier cambio local de peso de conexión, o, lo que es lo mismo, que para cualquier estado global deseado disponemos de un procedimiento para incrementar la probabilidad de que ese estado ocurra cambiando pesos locales. Importa subrayar que toda la información globalmente relevante para actualizar pesos está disponible localmente. Dada la conectividad y la distribución de Boltzmann en equilibrio, queda garantizado que la información de unidades lejanas se propagará por la red. El aprendizaje en este tipo de red puede ser tanto supervisado como no supervisado. Lo central será, por lo que al ejemplo de la estimulación ambigua respecta, que una red entrenada de este tipo será capaz de, por así decir, reconstruir un patrón de salida a partir de uno de entrada incompleto: será capaz de *representar* el patrón completo. Más aun, resulta interesante subrayar que pueden encontrarse soluciones para el problema del ajuste de pesos en redes con unidades ocultas, y que lo mismo puede decirse para redes con arquitecturas y dinámicas diversas –o, al menos, son problemas cuya posible solución no puede ser negada a priori. El problema del ajuste de pesos será pues, desde este punto de vista, una cuestión de arquitectura, dinámica y procedimientos de ajuste de parámetros.

Por su parte, el aprendizaje competitivo es una forma de aprendizaje autosupervisado en el que las neuronas de salida compiten por la activación en función de la proximidad a un patrón de entrada –resultando que la unidad más cercana al mismo ajusta su peso para hacerse más próxima. A pesar de que este tipo de red ha sido utilizada para explicar la formación de mapas topológicos sensoriales, han resultado por diversos motivos (como la inestabilidad de la configuración de pesos, o el llamado problema de las *neuronas muertas*, que no mencionan los autores) problemáticas y difíciles de manejar.

c) Ejemplos de redes feedforward:

A pesar de la referida ausencia de conexiones de feedback, estas redes pueden tener importantes ventajas en cuanto a velocidad y simplicidad. Por ello merece la pena echar un vistazo al tipo de computaciones que pueden con ellas llevarse a cabo. Asimismo resulta interesante observar sus limitaciones pues ello comporta importantes intuiciones.

El primero de los ejemplos es el proporcionado por la función XOR. Ver cómo puede ejecutarse esta función en una red, en qué tipo de arquitectura, resulta interesante tanto por los fallos que surjan como por lo que ellos significan para la computación del resto de funciones.

Una red arquetípica para esta función consta de una unidad de salida, que puede adoptar los valores 0 y 1, y dos de entrada. Nos hallamos ante tres parámetros: los dos pesos de las entradas a las salidas y el umbral o tendencia de la unidad de salida. Pronto nos hallamos aquí ante un problema: se trata de una función que no resulta linealmente separable y no puede por ello encontrarse un algoritmo de aprendizaje que resuelva el problema de ajustar los pesos en esta red simple –ya que no existe. De modo que es necesaria una unidad oculta. De cara a entrenar la red así reconfigurada se hace necesario un procedimiento de retropropogación o propagación hacia atrás de errores. XOR puede pues ser representado y aprendido en redes feedforward –y, de hecho, una enorme cantidad de funciones puede tratarse en redes de este tipo con la cantidad adecuada de unidades ocultas.<sup>27</sup> Por otra parte, a pesar de lo que pueda parecer a primera vista, este tipo de computación puede resultar relevante desde el punto de vista biológico.

Otro ejemplo de red feedforward entrenada con retropropagación o propagación hacia atrás de errores (por medio de gradiente descendente, pues) es el de una red entrenada para distinguir ecos de sonar correspondientes a minas o rocas, y compuesta por 60 unidades de entrada, 2 de salida y 1-24 ocultas, en la que los pesos se van ajustando mediante la comparación de los valores de output obtenidos y los que debieran obtenerse. Tras el entrenamiento la red acaba por ofrecer valores óptimos o cercanos al óptimo tanto en presencia de estímulos presentados durante el entrenamiento como en presencia de nuevos estímulos.

d) Redes recurrentes:

En la retropropagación en redes feedforward la señal sólo va hacia atrás, por así decir, en el proceso de ajuste de pesos, no en el de la activación de las unidades, y no hay por tanto un flujo de información como input desde unidades de un nivel superior hacia las de otro inferior. De modo que este tipo de redes son sistemas reactivos: responden a un input con un output en función de la señal exterior y la configuración de pesos. Por contra, una red recurrente, con feedback, tendrá una serie de características peculiares: a) incorpora múltiples escalas temporales en las unidades de procesamiento, b) procesa secuencias de inputs extendidos temporalmente, c) genera oscilaciones y ritmos modificables de duraciones variables y d) es capaz resolver ambigüedades. En este tipo de red la entrada no está informacionalmente aislada e inputs internos contribuyen a la activación de las unidades ocultas. Esto dota a la red de unas propiedades temporales que en lo referente a la modelización de redes neurobiológicamente realistas resulta beneficioso por lo que al reconocimiento y respuesta a representaciones temporalmente extendidas toca, dado que los procesos de feedback resultan provechosos para implementar una red, por así decir, temporalmente sensible. La clave está en el papel

---

<sup>27</sup> Redes con cientos de unidades ocultas y cientos de miles de pesos han sido entrenadas con éxito para un amplio rango de problemas.

que el propio tiempo juega en la computación: las propiedades temporales de la red derivan de las propias interacciones físicas en la red, que tendría así una estructura temporal intrínseca: en tanto la señal llega antes a unos lugares de la red que a otros, la señal se transforma, y esto tiene consecuencias en el comportamiento de la red.

e) Redes frente al mundo real:

En primer término cabe indicar que el comportamiento del sistema nervioso, su forma de representar y resolver problemas, puede obedecer a una gran variedad de principios computacionales (incluso al más simple de los comentados: el de look-up table) y a diferentes tipos de redes con diferentes arquitecturas y dinámicas.

Por otra parte, las redes diseñadas por ingenieros se enfrentan a problemas aislados y bien definidos, mientras que el sistema nervioso se halla frente a un mundo complejo y cargado de ambigüedades. Los inputs que le llegan suelen tener muchas más dimensiones que las que interesan a los ingenieros a la hora de diseñar sus redes, y, además, no llegan a él informacionalmente empaquetadas en lotes destinados a diferentes problemas separados y bien definidos. Esto no significa que la creación de redes para desenvolverse ante este tipo de estímulos sea imposible –cabe pensar en redes que incluyan funciones de base radial, con las que puede conseguirse que durante las sesiones de entrenamiento un input afecte sólo a un rango acotado de unidades ocultas, del mismo modo que cabe pensar en la creación de subredes mediante aprendizaje competitivo. Así, la idea de que la modelización no pueda servir para comprender hechos fundamentales acerca de las redes neuronales biológicas –la idea de que la creación de modelos neuronales no es más que una forma en la que la tradición de la caja negra se prolonga en las ciencias de la mente– seguirá sonando a *mero prejuicio* hasta que la imposibilidad empírica resulte abrumadora (cosa muy otra es la que viene sucediendo, como sobra indicar) o aparezca un argumento más *potente* que el de la habitación China (Searle 1980) –nadie parece haber alcanzado a *aplicarlo* con éxito al ámbito conexionista dado que, de hecho, apuntemos interinamente, alberga graves problemas de principio: v. g., atendiendo a su propuesta, ¿de dónde hemos de pensar que surgen –o *caen*– la *semántica* y la “*intencionalidad intrínseca*”?.

En definitiva: no se trata de imitar, sino de hallar la respuesta a la pregunta acerca de qué función está computando una red neuronal real y a la de cómo lo está haciendo. Los modelos computacionales son, en este sentido, una poderosa herramienta cuya sinergia con los estudios tradicionales de neurobiología guiará hacia fructíferas hipótesis y progresivos descubrimientos. En este punto, las constricciones neurobiológicas a la modelización son, como indicábamos más arriba, deseables –los modelos realistas serán más capaces de guiarnos en la investigación de la naturaleza y propiedades de las redes neuronales *reales*–, pero no puede perderse de vista el hecho de que los modelos en los diferentes niveles han de incluir simplificaciones respecto de los niveles inferiores; en este sentido un orientador consejo heurístico podría rezar: haz que tu modelo sea lo suficientemente simple como para que pueda ceñirse a lo que importa, pero lo suficientemente rico como para que pueda incluir cualquier aspecto relevante para las mediciones necesarias.

#### **2.4. Un ejemplo de investigación en neurociencia computacional**

Tal y como indican los autores al final del tercer capítulo del libro, los últimos cuatro están dedicados a presentar ejemplos interesantes y sugerentes de la

metodología de la neurociencia computacional en sus diferentes áreas de estudio. A continuación prestaremos brevemente atención a uno en concreto.<sup>28</sup>

Una cuestión que muchos filósofos (verbigracia Thomas Nagel) han considerado intratable es la de cómo el cerebro representa el mundo. Los autores, en línea con la perspectiva *naturalista*<sup>29</sup> que de Darwin llega a nuestros días, proponen que las ciencias de la mente/cerebro están proporcionando respuestas a esta incógnita. Expondremos de forma resoluta a continuación el modo en que la neurociencia computacional trata un aspecto –que en tanto seres eminentemente visuales nos resulta siempre interesante– de esta capacidad representativa: la visión estereoscópica, un rasgo crucial de la capacidad visual de los primates.

El problema que en este caso debe afrontar el cerebro es el de reconstruir una imagen en tres dimensiones –con distancia y profundidad– a partir de dos imágenes en dos dimensiones –así como de una serie de pistas proporcionadas por el movimiento, etc.

El primer punto a señalar, con nuestros autores, es que la visión incluye muchas facetas diferentes, y que tal vez simular cada una de ellas por separado sólo sirva, en cierto sentido, para aumentar con los resultados alcanzados el grado de realismo de los modelos en futuras investigaciones. No plantean pues la cuestión de la estereovisión en el sistema nervioso como si de un módulo aislado se tratara, pero lo tratan como tal desde el punto de vista de la modelización, con la pretensión de proporcionar desde las redes neuronales simuladas hipótesis acerca de cómo se lleva a cabo esta tarea en las redes reales.

Como antecedentes históricos de la investigación en este campo mencionan los autores que a mediados del siglo XIX David Brewster y Charles Wheatstone propusieron independientemente la hipótesis de que las disparidades entre las imágenes recibidas por cada ojo pueden contar como posibles pistas utilizadas por el sistema visual para producir percepción de la profundidad.

Dado que la distancia media entre los dos ojos es de seis centímetros, la percepción de formas tridimensionales se desvanece a partir de una distancia de cien metros porque los rayos de luz provenientes de más allá de este umbral inciden esencialmente en el mismo punto de las retinas. De modo que a partir de esta distancia son otros los medios –basados igualmente en la disparidad retinal: las pistas claves habrán de ser pues el movimiento relativo, los efectos de perspectiva y la oclusión– que el sistema visual debe usar para extraer diferencias en la percepción de la profundidad.

Una primera noción importante para entender la binocularidad es la de plano de fijación, que podría a grandes rasgos entenderse partiendo de la noción del punto de convergencia al que están dirigidos los ojos. En el plano de fijación la luz proveniente de los objetos incide en la misma localización –relativa– en las dos retinas. Los rayos de luz provenientes de objetos que se hallen detrás o delante del plano de fijación llegarán a diferentes lugares en cada una de las retinas. El área que rodea al plano de fijación se denomina área de fusión o área fusional de Panum (una zona que rodea a la línea curva, llamada horopter, que representa puntos a iguales distancias del observador, puntos desde los cuales la luz llega a las mismas áreas retinales) y *en ella* los objetos aparecen con formas y bordes definidos –enfocados.

La imágenes correspondientes a cada ojo que el cerebro tiene que fusionar son pues algo diferentes, no obstante el cerebro opera como una unidad de procesamiento paralelo de alta capacidad de para producir escenas coherentes y desambiguar (aunque en raras ocasiones puede producirse falsa visión estereoscópica –partiendo

<sup>28</sup> Para el ejemplo que expondremos: vid., especialmente, Churchland/Sejnowski (1992); pp. 199-237.

<sup>29</sup> En sentido amplio, es decir, no exclusivamente *epistemológico*.

de objetos repetidos, como en los ejemplos triviales de los libros de *ojo mágico*). Dicha fusión respondería al problema de qué se enlaza con qué, conocido como el problema de la correspondencia –para la solución del cual los movimientos de los ojos, no sólo los cambios de vergencia sino también los sacádicos, pueden ser de gran ayuda.

Producir estereovisión para el cerebro ha de significar comparar imágenes retinales relativas a planos de fijación variables. Hubel y Wiesel descubrieron en la década de los sesenta que las células de la corteza estriada no responden a los estímulos con uniformidad, sino que unas son altamente monoculares mientras otras responden a estímulos provenientes de los dos ojos, prefiriendo uno u otro. La clave para la visión estereoscópica puede hallarse, muy probablemente, en la arquitectura de procesamiento presentada en las células de las áreas visuales V1 y V2, que reciben pocas aferencias de los núcleos geniculados laterales y se hallan principalmente conectadas a otras células corticales. A las células que muestran patrones de respuesta elevados para estímulos provenientes del área del plano de fijación suele llamárselas *tuned cells* –los autores encuentran que resulta mejor denominación la de células de fijación–; otros dos tipos de células responden a objetos que se hallan delante o detrás del plano de fijación (*near* y *far cells*); otra clase de células podría responder a estímulos correspondientes a las tres referidas posiciones con preferencias por distancias relativas al plano de fijación. Todas estas células pueden entenderse pues como unidades que representan profundidad en relación al plano de fijación. De todos modos, se trata de células con otras propiedades (preferencia por direcciones y orientaciones concretas).

Planteando ya en términos computacionales el problema de la correspondencia cabe preguntar: ¿qué tipo de función puede mapear los puntos de correspondencia para cualquier escena y cómo debe ser computada?

Un primer inconveniente es que los puntos de correspondencia pueden contener discontinuidades y ambigüedades. También la uniformidad en el aspecto de un objeto puede presentar problemas (cuando todo es *igual*, ¿cómo enlazar puntos en base a diferencias?).

El primer cabo del que tirar es que puntos análogos de las dos retinas tienen que hallarse fuertemente correlacionados, dado que recibirán estímulos similares cuya clave por lo que al emparejamiento respecta estará en que la variación entre uno y otro ofrecerá pistas acerca de la profundidad relativa al plano de fijación.

Una posibilidad computacional para dar el primer paso hacia la solución de esta tarea es la búsqueda de compatibilidades entre las dos imágenes. Paul Churchland diseñó el mismo año en que se editara el libro que nos ocupa una red (llamada Fusion-Net) para ofrecer una solución al problema de la búsqueda de compatibilidades; no para el problema completo de la percepción de la profundidad, sino para hallar compatibilidades entre dos imágenes para movimientos oculares dados en un mismo plano de convergencia –la función de búsqueda de compatibilidades puede ser resulta por una serie conectada de redes XOR negativa en la que se implementan *tuned* (que captan disparidad cero), *near* y *far cells* (*units*), que proporcionarían una baja definición para aspectos que se situarían delante y detrás del plano de fijación. Como puede deducirse de lo dicho, Fusion-Net es una red feedforward; además, imita con éxito el hecho psicológico de que delante y detrás del área de fusión la resolución es peor que dentro del mismo. Los autores indican un punto en que la red resulta implausible neurobiológicamente: las unidades de input son al tiempo excitatorias e inhibitorias, mientras que en el sistema nervioso real sólo serían de una u otra clase; además, indican, la inclusión de una dinámica recurrente con conexiones inhibitorias ayudaría al sistema no sólo a solucionar problemas (como la falta de claridad y la caída en errores de fusión) sino también a remedar más de cerca el proceso de la estereovisión en primates.

Marr y Poggio combinaron en 1976 una función de enlazamiento con un proceso cooperativo computando primero compatibilidades para todos los planos de profundidad y usando esto como punto de partida para un algoritmo iterativo que computara correspondencias únicas. La clave de su estrategia era que podían crear una réplica en 3D de objetos en 3D con neuronas dispuestas para cada plano de profundidad –con conexiones excitatorias entre neuronas que respondieran a un mismo plano e inhibitorias entre neuronas que respondieran a diferentes planos de profundidad. El principal de los problemas con esta red, como subrayan los autores, estaría en que al usar un algoritmo cooperativo su procesamiento resulta lento. Por otra parte, comentan los autores –en referencia a una red diseñada por Sejnowski– que redes de aprendizaje competitivo no supervisado pueden aprender a representar la disparidad como invariante.

Para concluir, cabe indicar, con los autores, que la binocularidad es una entre las pistas de las que el sistema nervioso puede servirse para extraer percepción de profundidad, y que, a diferencia de las redes diseñadas hasta el momento, al sistema nervioso llegan estímulos indiferenciados y no informacionalmente empaquetados en función de características discretas –así, señalan, es muy probable que módulos realmente aislados para la resolución de tareas, como en este caso la de extraer percepción de la profundidad, no existan en el sistema nervioso, y que es posible que partir de esta asunción resulte a la larga poco fructífero.

### 3. Conclusiones

Llegada la hora de concluir, y de cara a ofrecer una perspectiva epistemológica amplia que arroje si no luz sí al menos alguna *pista* sobre alguno de los puntos que hemos tratado –acaso un tanto superficialmente– a lo largo del presente trabajo, sería interesante incluir alguna nota de carácter general acerca de un extremo que consideramos central desde el punto de vista filosófico, y que constituye el hilo conductor tanto del texto que hemos comentado como de toda la investigación en ciencias cognitivas desde el cambio de paradigma a principios de los sesenta: el de la mente como instancia manipuladora de representaciones y procesadora de información.

*If the brain is a computer, all we have to explain is how a brain might manipulate representations.*<sup>30</sup>

En una primera aproximación cabe señalar que dicha concepción de la mente, a cuya base encontramos la asunción del supuesto según el cual la misma procesa información, es la concepción canónica cognitivista, según la cual la mente puede de hecho ser tenida por un sistema computacional. Citaremos, para dar el primer paso hacia el tratamiento de estos planteamiento, los siguientes cuatro puntos (tomados de un trabajo en el que Stillings *et al.* señalan algunos de los conceptos fundamentales de la concepción computacional de la mente<sup>31</sup>):

a) El procesamiento de información permite al organismo (o sistema) responder ante variaciones en las condiciones del ambiente. Esas respuestas son típicamente adaptativas o dirigidas a metas; de modo que los procesos de información tienen la cualidad de ser significativos en relación al entorno, es decir, la información en el organismo (o sistema) es algo relacionado con el mundo, como teniendo contenido,

---

<sup>30</sup> McDermott, D. V. *Mind and Mechanism*. Cambridge, MIT Press, 2001; pág. 245.

<sup>31</sup> Stillings, N. A./Weisler, S. E./Chase, C. H./Feinstein, M. H./Gardfield, J. L./Rissland, E. L. *Cognitive Science: An Introduction*. Cambridge, MIT Press, 1995.

significado o sentido. Tales cualidades de contenido son con frecuencia denominadas cualidades *semánticas* o intencionales. Esas cualidades permiten atribuir entendimiento, creencia o conocimiento a cualquier organismo (o sistema) que actúe adaptativamente en respuesta a la información procedente del medio.

b) El procesamiento de información tiene carácter representacional: la información que manipula un sistema computacional debe ser representada de alguna manera. Una buena comprensión de una operación de computación requiere comprensión del modo en que la información está representada en la mente. La representación implica pues símbolos.

c) El procesamiento de información puede ser descrito formalmente, lo cual conduce a la noción de algoritmo, que podemos definir en términos de procesos operados sobre las representaciones –aunque, más generalmente, algoritmo puede definirse como *método* para la solución de problemas definidos. Cabe aclarar que los procesos algorítmicos no operan sobre el dominio que está siendo representado, ni son definidos en términos del significado de la representación. Un algoritmo es un procedimiento o *sistema formal*, definido más en términos de la forma de la representación que de su significado o contenido. Lo que cabe subrayar aquí es que nos hallamos con ellos ante manipulación de patrones de representación.

d) Las nociones referidas son empleadas por los científicos cognitivos para alcanzar un conocimiento científico básico del procesamiento de información. Es decir, los científicos cognitivos pretenden descubrir los principios explicativos del procesamiento de información –lo cual, como resulta obvio, concierne tanto a la IA y las ciencias de la computación como a la psicología experimental y las neurociencias.

Dejando a un lado el campo metodológico –en el que lo que contaría sería el partido que puedan sacar los científicos cognitivos de estos planteamientos y, asimismo, cómo puedan hacerlo–, debemos preguntarnos qué significa ese referido *carácter representacional*, compartido por simbolistas clásicos y conexionistas, que tantas discusiones ha traído consigo –acerca, por ejemplo, de qué tipos de representaciones manipula la mente humana: lingüísticas o proposicionales, pictóricas, modélicas... o todas a la vez, como en el modelo representacional triple de Johnson-Laird– y que nos lleva de la cognición y la computación a la cognición como computación.<sup>32</sup>

Cabe en este sentido destacar en primera instancia que diversas teorías de la representación han sido formuladas a lo largo de la historia de las ciencias cognitivas. Lo importante de todas ellas es su carácter internalista –aspecto que junto con la concepción del sujeto como procesador de información es presentado por Pascual Martínez-Freire<sup>33</sup> como definitorio de la concepción cognitivista de la mente–, es decir, que tratan de explicar mediante la referencia a procesos internos la conducta de los sistemas biológicos o artificiales<sup>34</sup>. Nos topamos pues con una tesis mentalista<sup>35</sup> que

---

<sup>32</sup> Pylyshyn, Z. W. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MIT Press, 1984/1989.

<sup>33</sup> Martínez-Freire, P. “Delimitación de las ciencias cognitivas”. *Anales del seminario de metafísica*, N° Extra 1, Complutense, 1992; pp. 443-452.

<sup>34</sup> Este internalismo no se halla exento de problemas y no han sido pocos quienes han señalado la diferencia entre los estados mentales y los funcionales en tanto estados internos; veámoslo en un elocuente y acre fragmento de un artículo/reseña de González-Castán: <<El funcionalismo se ha dicho lo siguiente: “Ahora que nadie nos ve, ahora que hemos conseguido ser sabios siendo ignorantes, escuchados siendo sordos, vamos a caracterizar lo interno de los estados mentales como si fuera coextensivo, semántica y ontológicamente, con lo interno de los estados funcionales” >>. (González-Castán, O. L. “La locomotora cognitiva y la filosofía de la mente”. *Logos. Anales del seminario de metafísica*, 2003, 36, pp. 269-281; pág. 280).



se opone radicalmente a la tradición conductista –negadora de conceptos intencionales– que en psicología comenzaría a perder peso con el avance del cognitismo en la década de los sesenta. El señalado carácter internalista de las teorías cognitivistas de la representación se dirigiría en la referida explicación de la conducta al esclarecimiento del tipo de procedimientos que subyacen a la forma en que los sistemas biológicos o artificiales interactúan con su medio. Aparecerían así en la concepción cognitivista clásica de representación dos aspectos: el semántico y el mecánico, dado que dichos procesos subyacentes, internos, no son sino procesamiento de información en forma de símbolos, y los símbolos pueden entenderse en este sentido como físicamente implementados y determinantes para el pergeño de la conducta, al tiempo que semánticamente interpretables. Con esto nos hallamos ante otra de las claves del cognitismo clásico: la caracterización de los *sistemas inteligentes* como *sistemas simbólicos*, sistemas adaptativos que al enfrentarse a un medio deben ser capaces de producir conducta coherente con la situación y dependiente de su organización interna. Hasta aquí nada de lo apuntado excedería el marco de la concepción cognitivista clásica –simbolista–, pero (y este es, por lo que toca al texto de Churchland y Sejnowski que hemos estado comentando, un punto clave), con la emergencia del conexionismo y las redes neuronales artificiales (emergencia que hallaría antecedentes en los cuarenta –McCulloch y Pitts, 1943–, aunque resulta justo mencionar aquí a Rosenblatt como verdadero precursor) la noción cognitivista de representación vira de la discreción a la dispersión: en las redes conexionistas la representación se produce no *en* elementos sino *entre* elementos; se trata de representaciones distribuidas por la red, y lo relevante en este sentido es el estado global del sistema. En este sentido, el conexionismo, como rama de la IA y parte del modelo cognitivo, se acerca a la representación de un modo diferente al simbolista clásico, de un modo que podemos entender como *más natural*, en tanto su dechado son las redes neuronales naturales (a partir de las cuales se galibarian *guardando cierta distancia* las artificiales) en las que, hemos de entender, las representaciones no se hallan puntualmente localizadas, sino que dependen de propiedades sistémicas dadas por las relaciones –de pesos– entre elementos del sistema. Ambas perspectivas<sup>36</sup>, la simbólica clásica (Fodor, Pylyshyn) y la conexionista (Rumelhart, McClelland, Hinton) mantendrían un carácter computacional –sus explicaciones sobre los modelos mentales se alzan por igual sobre la idea de *computo mental*–, pero la última rebasaría el campo de la primera rechazando la idea de sistema simbólico y procediendo al transvase del serialismo al procesamiento en paralelo. La actividad representacional deja de atribuirse a símbolos codificados, a un programa almacenado o a un lenguaje interior entendido como manipulación de símbolos, y pasa a ubicarse –a *distribuirse*– en la disposición concreta de fuerzas o pesos de conexión.<sup>37</sup> En este sentido, en el artículo de Churchland y Grush sobre *Computación y Cerebro* para *The MIT Encyclopedia of Cognitive Sciences* que citáramos al final del apartado 2.1 del presente trabajo, caracterizan los autores la representación en sistemas conexionistas como sigue:

<sup>35</sup> Como Pedro Chacón (2006) ha subrayado en *Los límites de una metáfora: la mente como computador* la teoría computacional de la mente es *mentalista*.

<sup>36</sup> Algunos, como Paul Smolensky, han tratado de tender puentes entre simbolismo y conexionismo.

<sup>37</sup> Hay que señalar que en ocasiones nos encontramos con redes en las que las unidades sí que representan aspectos discretos (como en la que McClelland y Rumelhart diseñaron en 1981 para el reconocimiento de palabras), pero se trata de excepciones: las representaciones conexionistas, insistimos, se hallan distribuidas en las unidades del sistema, distribuidas pues por las neuronas de la red.

*Representation is characterized as a pattern of activation across the units in the neural net, where this can be described as a vector, and computations are therefore vector-vector transformations.*<sup>38</sup>

En relación con lo indicado en el apartado anterior precisaremos que tres años antes de que Churchland y Sejnowski colaboraran en la redacción de la obra que nos ha ocupado, Paul M. Churchland se pronunciaba ya respecto de esta nueva concepción de la representación en un marco conexionista (una concepción que presentaba como radicalmente diferente, incluso opuesta, a la simbolista clásica o a la concepción proposicional de la representación)<sup>39</sup>. La noción importante sería en este punto la de transformación de vector a vector, que remitiría de la última cita al apartado anterior, y a la referida obra de Paul M. Churchland, tres años anterior a *The Computational Brain*, en la que el autor ya definía la representación en estos términos (en concreto la representación era allí definida, en el marco de los patrones de actividad neuronal, como puntos y trayectorias en el espacio vectorial, de modo que las computaciones sobre representaciones operarían transformaciones de vector a vector).

Presentado el marco general del modo en que se desenvuelve la noción de representación que hallaríamos a la base de los planteamientos de *The Computational Brain*, nos gustaría en primer lugar subrayar brevemente el sentido en que una teoría representacional y una teoría computacional de la mente presentan convergencias fundamentales. En primer término convendría ofrecer una sucinta –y tal vez por ello sesgada– perspectiva de aquello que con Teoría Representacional de la Mente estaríamos significando. Veámoslo en primera instancia a través del prisma cognitivista clásico –sirviéndonos para ello de un esquema que Clark toma de Fodor según el cual una tal teoría obedecería esencialmente a los dos puntos siguientes:

1. Propositional attitudes pick out computational relations to internal representations.
2. Mental processes are causal processes that involve transitions between internal representations.<sup>40</sup>

Podemos ahora resolver el problema de presentar la confluencia a la que apuntábamos haciendo escueta referencia al apotegma fodoriano: *No hay computación sin representación.*<sup>41</sup> Esto, ciertamente, se colige de las sucintas observaciones generales acerca del significado de computación y computadora que incluimos en el tercer apartado de este trabajo. Cargaremos sobre otro la responsabilidad de incidir en este extremo: *una computadora es un dispositivo que*

---

<sup>38</sup> Citamos, nuevamente, una versión del artículo (*Computation and the Brain*) previa a la edición de la enciclopedia de la que pasaría posteriormente a formar parte.

<sup>39</sup> Nos referimos aquí a la obra de Paul Churchland de 1989 *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MIT Press, 1992 – hay que añadir que recientemente, en *Neurophilosophy at Work* (2007), ha insistido en este punto.

<sup>40</sup> Clark, A. *Mindware. An Introduction to the Philosophy of Cognitive Science*. New York, Oxford University Press, 2001; pág. 43 (Por cierto que en la primera mitad de este libro hallamos frecuentes referencias a *The Computational Brain*).

<sup>41</sup> Contra la práctica habitual, no discutiremos aquí la teoría representacional de Fodor –una de las más conocidas y comentadas, junto con planteamientos como la semántica del rol conceptual de Block.

procesa representaciones de una manera sistemática.<sup>42</sup> Con estas apreciaciones no saldríamos del marco cognitivista clásico –simbolista.

En segundo lugar, y retomando el planteamiento que los autores presentan en el texto del que nos hemos estado ocupando, resulta interesante distinguir entre diferentes posturas en el actual contexto conexionista (como el conexionismo eliminativo, el implementacional –sub-simbólico– y el revisionista de procesamiento de símbolos –la escuela de San Diego), e incluso entre dos formas generales de enfocar el trabajo actual en IA: la subsimbólica –interesada en el trabajo a un nivel análogo al de las neuronas–, y la simbólica –interesada en la modelación de estructuras significativas. Tomando como referencia estas distinciones nos gustaría hacer hincapié en la fuerte influencia de la biología que acusa el examen de la neurociencia computacional presentado en *The Computational Brain*, influencia que sitúa dicho examen y las perspectivas que del mismo se derivan en un marco que apuntaría a los actuales planteamientos sub-simbolistas. De este modo, el problema filosófico de la representación se nos aparece bajo una forma poco habitual en la literatura dedicada al tema: lo que estaría en juego sería qué significa representar cuando las representaciones que sean el caso no tengan nada en absoluto que ver con nuestros estados mentales representacionales, es decir, con los estados mentales que Russel bautizara como actitudes proposicionales (como Churchland y Grush apuntan, la perspectiva conexionista permite abordar formas sub-personales y sub-cognitivas de representación, de función cerebral). Se hace pues evidente que la convergencia entre representacionalismo y computacionalismo no es del mismo carácter aquí que en perspectivas como la que Clark nos presentaba poco más arriba a través de Fodor. Una interesante pregunta sería en este punto si estas micro(sub)-representaciones tendrían intencionalidad. Desde luego que refieren a algo más allá de sí mismas, pero es también claro que este marco tiene muy poco que ver con la forma tradicional de caracterizar la intencionalidad –¿cabría hablar de *intencionalidad derivada*<sup>43</sup> con respecto a ellas?, ¿o podríamos entender, más bien, que deriva *nuestra* intencionalidad de la capacidad que estas micro(sub)-representaciones *sostenidas* por o acaecidas en las redes neuronales tienen de referirse a aspectos y características del mundo?

Por otra parte, no está de más en este apartado final indicar que el énfasis de nuestros autores en las constricciones biológicas (no sólo neurológicas, sino incluso evolutivas) a las que en más de una ocasión hemos aludido no debe hacer que perdamos de vista el modo en que el planteamiento de Churchland y Sejnowski dista de otras propuestas que en la misma década de los noventa pretendieron modificar radicalmente la noción de representación –o incluso borrarla del mapa cognitivista, como brevemente expondremos a continuación. Hablamos de propuestas como la de Varela o la de Clark. El primero habría llegado al extremo de negar la pertinencia de la noción de representación en el contexto de la explicación de los sistemas cognitivos. Apostaba, frente a una concepción estática de la relación cognitiva entre organismo y entorno, por modificar el *realismo idealista* de la concepción tradicional de representación en el ámbito cognitivista por una concepción *enactiva* de la cognición en la cual organismo y entorno se determinan mutuamente dentro del campo de un vínculo entre ambos marcado por la acción. Varela propondría pues excluir del aparato conceptual cognitivo la noción de representación por su carácter realista estático, el cual lleva al error de concebir los actos cognitivos como mediados por características

<sup>42</sup> Crane, T. *La mente mecánica. Introducción filosófica a mentes máquinas y representación mental*. México, D. F., Fondo de Cultura Económica, 2008; pág. 168 (los capítulos III, IV y V de esta obra –pp. 142-328– constituyen una introducción accesible a los problemas filosóficos planteados por la representación mental).

<sup>43</sup> Serale (1992).

fijas tomadas de un mundo que habría en ellos que recuperar mediante el recurso a un sistema representacional igualmente rígido y fijado de antemano, *pre-dado*.<sup>44</sup> Clark, por su parte, con su concepción dinámica de la mente corporizada (conectada con el cuerpo y su actividad y extendida por el entorno) ofrece sobre la noción de representación un planteamiento que, como el de Varela, va más allá del cognitivismo clásico y el conexionismo, pero no desemboca en un antirrepresentacionalismo del tipo del de Varela. En su concepción de la representación nos ofrece Clark (en *Being There: Putting Brain, Body and World Together Again*, de 1997) una interpretación nueva de dicha noción –en el marco de un intento de renovación conexionista y un representacionalismo revisado– en la que el significado de ésta, tal y como era típicamente interpretado en el contexto del conexionista, es sustituido por un enfoque en el que la mente y la cognición aparecen no sólo ligadas al cuerpo y el entorno, sino *extendidas* en ellos (que formarían así parte de la mente y la cognición), y la representación dirigida a acciones particulares –el cuerpo y el entorno se muestran así como elementos constituyentes de la *representacionalidad* mental.

### Referencias

- Brook, A./ Mandik, P. (2004) “The Philosophy and Neuroscience Movement”. *Analyse & Kritik* 6/2004; pp. 382-397
- Canseco, J. (2007) “Redes neuronales y conexionismo en las neurociencias”. *Metábasis* Nº 3, marzo 2007
- Chacón, P. (2006) *Los límites de una metáfora: la mente como computador*. México, D. F., UNAM, 2006
- Churchland, P. M. (1981) “Eliminative Materialism and the Propositional Attitudes”. *Journal of Philosophy*, February 1981, Vol. 78, Nº 2; pp.67-90
- Churchland, P. M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MIT Press, 1992
- Churchland, P. M. (2007) *Neurophilosophy at Work*. Cambridge, Cambridge University Press, 2007
- Churchland, P. S. (1986) *Neurophilosophy. Toward a Unified Science of the Mind/Brain*. Cambridge, MIT Press, 1986
- Churchland, P. S./Sejnowski, T. J. (1988) “Perspectives in Cognitive Neuroscience”. *Science*. 4 November 1988, Vol. 242; pp. 741-745
- Churchland, P. S./Sejnowski, T. J. (1992) *The Computational Brain*. Cambridge, MIT Press, 1992
- Clark, A. (1997) *Being There: Putting Brain, Body and World Together Again*, Cambridge, MIT Press, 1998
- Clark, A. (2001) *Mindware. An introduction to the philosophy of cognitive science*. New York, Oxford University Press, 2001
- Craig, E. (Ed.) (1998) *Routledge Encyclopedia of Philosophy. Philosophy of mind*. Routledge, Londres & New York, 1998
- Crane, T. (1995) *La mente mecánica. Introducción filosófica a mentes máquinas y representación mental*. México, D. F., Fondo de Cultura Económica, 2008
- Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection*. New York, Basic Books, 1987
- Edelman, G. M. (1990) *The Remembered Present. A Biological Theory of Consciousness*. New York, Basic Books, 1990

---

<sup>44</sup> Pueden consultarse, además de *Conocer. Las ciencias cognitivas: Tendencias y perspectivas. Cartografía de las ideas actuales*, de 1988, y *De cuerpo Presente. Las ciencias cognitivas y la experiencia humana*, de 1991 (de la que Varela fuera coautor junto con Thompson y Rosch), los artículos “Francisco Varela y las ciencias cognitivas”, de Cesar Ojeda (*Revista Chilena de Neuro-psiquiatría*, 2001, 39; pp. 286-295), o “El problema de la representación en la filosofía cognitiva”, de Juan C. Vélez (*Contrastes. Revista Internacional de Filosofía*, 2008, vol. XIII; pp. 253-271).

- Edelman, G. M. (1992) *Bright Air, Brilliant Fire. On the Matter of the Mind*. New York, Basic Books, 1992
- Edelman, G. M./Tononi, G. (2000) *El universo de la conciencia. Cómo la materia se convierte en imaginación*. Barcelona, Crítica, 2002
- Gardner, H. (1985) *La nueva ciencia de la mente. Historia de la revolución cognitiva*. Barcelona, Paidós, 1988
- Gazzaniga, M. S. (Ed.) (1984) *Handbook of cognitive neuroscience*. New York, Plenum Press, 1984
- Gazzaniga, M. S. (Ed.) (1995) *The Cognitive Neurosciences*. Cambridge, MIT Press, 1995
- Gazzaniga, M. S. (1996) *Conversations in the Cognitive Neurosciences*. Cambridge, MIT Press, 1997
- Gazzaniga, M. S. (Ed.) (2000a) *The New Cognitive Neurosciences*. Cambridge, MIT Press, 2000
- Gazzaniga, M. S. (Ed.) (2000b) *Cognitive Neuroscience: A Reader*. Massachusetts, Blackwell, 2000
- Gazzaniga, M. S. (Ed.) (2010) *The Cognitive Neurosciences*. Cambridge, MIT Press, 2010
- Gazzaniga, M. S./Ivry, R. B./Mangun, G. R. (2002) *Cognitive Neuroscience: The Biology of Mind*. New York, Norton & Comp., 2002
- González-Castán, O. L. (2003) "La locomotora cognitiva y la filosofía de la mente". *Logos. Anales del seminario de metafísica*, 2003, 36, pp. 269-281
- Gutfreund, H./Toulouse, G. (Eds.) (1994) *Biology and Computation: A Physicist's Choice*. Massachusetts, World Scientific, 1994
- Lytton, W. W. (2002) *From Computer to Brain: Foundations of Computational Neuroscience*. New York, Springer-Verlag, 2002
- Marr, D. (1982) *La visión*. Madrid, Alianza, 1985
- Martínez-Freire, P. (1992) "Delimitación de las ciencias cognitivas". *Anales del seminario de metafísica*, Nº Extra 1, Complutense, 1992; pp. 443-452
- Martínez-Freire, P. (2002) *La nueva filosofía de la mente*. Barcelona, Gedisa, 2002
- McCulloch, W. S./Pitts, W. H. (1943) "A logical calculus of the ideas immanent in nervous activity". *Bulletin of Mathematical Biophysics*, 5, 1943:115-133
- McDermott, D. V. (2001) *Mind and Mechanism*. Cambridge, MIT Press, 2001
- Ojeda, C (2001) "Francisco Varela y las ciencias cognitivas". *Revista Chilena de Neuropsiquiatría*, 39, 2001; pp. 286-295
- Posner, M. I. (Ed.) (1989) *Foundations of Cognitive Science*. Cambridge, MIT Press, 1993
- Putnam, H. W. (1975) *Mind, Language and Reality. Philosophical Papers Volume 2*. Cambridge, Cambridge University Press, 1997
- Pylyshyn, Z. W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MIT Press, 1989
- Ramsey, W./Stich, S./Garon, J. (1990) "Connectionism, Eliminativism and the Future of Folk Psychology". *Philosophical Perspectives* 4, 1990; pp. 499-533
- Schwartz, E. L. (Ed.) (1990) *Computational Neurosciences*. Cambridge, MIT Press, 1990
- Searle, J. R. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3, 1980; pp. 417-457
- Searle, J. R. (1992) *El redescubrimiento de la mente*. Barcelona, Crítica, 1996
- Sejnowski, T. J./Koch, C./Churchland, P. S. (1988) "Computational Neuroscience". *Science. New Series*, Vol. 241, No. 4871. (Sep. 9, 1988); pp. 1299-1306
- Smolensky, P. (1988) "On the Proper Treatment of Connectionism". *Behavioral and Brain Sciences* 11, 1988; pp. 1-23
- Stillings, N. A./Weisler, S. E./Chase, C. H./Feinstein, M. H./Gardfield, J. L./ Rissland, E. L. (1995) *Cognitive Science: An Introduction*. Cambridge, MIT Press, 1995
- Thagard, P. (2005) *La mente. Introducción a las ciencias cognitivas*. Buenos Aires/Madrid, Katz, 2008
- Varela, F. (1988) *Conocer. Las ciencias cognitivas: Tendencias y perspectivas. Cartografía de las ideas actuales*. Barcelona, Gedisa, 1990
- Varela, F./Thompson, E./Rosch, E. (1991) *De cuerpo Presente. Las ciencias cognitivas y la experiencia humana*. Barcelona, Gedisa, 2005
- Vélez, J. C. (2008) "El problema de la representación en la filosofía cognitiva". *Contrastes. Revista Internacional de Filosofía*, vol. XIII, 2008; pp. 253-271