

FILOSOFIA Y CIENCIAS COGNITIVAS

Coordinación de:
Pascual F. Martínez-Freire

PHILOSOPHICA MALACITANA
Suplemento nº 3

1995

EL CONEXIONISMO: UN NEXO ENTRE LAS NEUROCIENCIAS Y LAS CIENCIAS COGNITIVAS.

A. Caño¹ y J.L. Luque²

RESUMEN

Se presenta el conexionismo como un marco común para la comunicación entre neurociencias y ciencias cognitivas. En primer lugar, se exponen los rasgos básicos de la nueva *metáfora del cerebro* y cómo diferentes componentes de la cognición, tales como la representación, la memoria y el aprendizaje, son considerados desde la perspectiva conexionista. En segundo lugar, se discute la delimitación del nivel de estudio y de explicación propio de estos modelos. En este contexto, se distinguen dos estrategias: la modelización *inspirada* y la *restringida neuronalmente*, revisándose ambas a la luz de la clásica división de los niveles de explicación para las neurociencias propuesta por Marr. Por último, se presenta la *potenciación a largo plazo* (PLP) como un mecanismo de plasticidad sináptica especialmente interesante, ya que la PLP parece adaptarse a los requisitos de la regla de Hebb, que ha sido implementada con éxito en diferentes modelos conexionistas.

ABSTRACT

Connectionism as a common framework to neurosciences and cognitive sciences is presented. Firstly, major characteristics of the new *brain metaphor* are exposed and several cognition components, such as learning, memory and representation, are considered within the connectionist view. Secondly, the boundaries of the study and explanation level for these models are discussed. In this context, two strategies are distinguished: neurally inspired and neurally constrained modelling, both being considered within the explanation levels for neurosciences proposed by Marr. Lastly, *long term potentiation* (LTP) is introduced as a specially interesting mechanism for synaptic plasticity because LTP satisfies the Hebbian rule requirements, which has been successfully implemented in different connectionist models.

1 Departamento de Psicología Básica y Psicobiología. Universidad de Málaga.

2 Departamento de Psicología Evolutiva y de la Educación. Universidad de Málaga.

1. Introducción.

En los últimos años, viene produciéndose una convergencia de intereses entre diferentes disciplinas en torno al funcionamiento y funciones del cerebro, tradicionalmente objeto de estudio de neurobiólogos y psicólogos. Lingüistas, filósofos, informáticos, bioquímicos, e incluso físicos y matemáticos muestran en la actualidad su interés por estos temas y con frecuencia integran equipos de investigación multidisciplinares sobre los mismos. El nuevo interés de estas disciplinas se ha producido tanto por el impulso de su propio desarrollo interno como por la necesidad de colaboración ante ciertos problemas planteados. Dos buenos ejemplos de esta convergencia interdisciplinar son las disciplinas integradas bajo la denominación de *ciencias cognitivas* o bajo la de *neurociencias*.

Tanto las ciencias cognitivas como las neurociencias han experimentado un espectacular avance en los últimos veinte años, aunque siguiendo caminos distintos. Últimamente, no obstante, se está produciendo un acercamiento entre ambas, motivado por el convencimiento de los neurocientíficos de que el funcionamiento del sistema nervioso no puede entenderse al margen del conocimiento de su capacidad funcional; y por el interés de los científicos de la cognición en la arquitectura y dinámica del sistema nervioso ante una doble evidencia. En primer lugar, la dificultad de sus modelos para simular correctamente el comportamiento cognitivo humano, con sus sesgos y peculiaridades, y, en segundo lugar, la dificultad para construir sistemas de inteligencia artificial tan potentes como el cerebro, incluso para determinadas tareas «simples» como por ejemplo, la percepción o la toma de decisiones.

No hace más de una década, si hubiéramos preguntado a un psicólogo cognitivo qué validez tenían sus hipótesis desde el punto de vista de los conocimientos sobre el funcionamiento cerebral, seguramente su respuesta habría oscilado entre la perplejidad, la indiferencia y hasta posiblemente el rechazo. Parecía entonces imposible que aquellos modelos sobre los diferentes procesos cognitivos, implementados mediante las ya clásicas *reglas de producción*, encontrasen ningún tipo de correlato fácil y directo con el funcionamiento del sistema nervioso. La Psicología se ocupaba del *software* y los neurobiólogos del *hardware*, un dualismo encubierto donde parecía imposible escoger entre dos programas, funcionalmente equivalentes, que simulaban la misma tarea cognitiva. La necesidad de encontrar un espacio común era patente.

Este mutuo interés y acercamiento ha encontrado un marco esperanzador en el desarrollo de los modelos de *procesamiento distribuido y paralelo* (PDP) o *modelos conexionistas* (Hinton y Anderson, 1981; McClelland, Rumelhart y cols., 1986; Rumelhart, McClelland y cols., 1986). Así, en esta encrucijada de interdisciplinariedad, el conexionismo puede situarse como un nuevo puente entre las llamadas ciencias cognitivas y las neurociencias.

No vamos a defender aquí que esta nueva perspectiva, que ha sido calificada por algunos como un auténtico cambio de paradigma (Schneider, 1987), solucione todos los problemas. Pero sí, que la situación ha cambiado al menos cualitativamente. Segura-

mente, las formulaciones pioneras de McCulloch y Pitts (1943) sobre las propiedades lógicas de las redes neuronales no encontraron una psicología lo suficientemente madura. Seguramente, tenía que ser desde el corazón mismo de las ciencias cognitivas, desde donde se planteara este nuevo reto. En cualquier caso, este nuevo lenguaje común ha permitido, por ejemplo, a los neurobiólogos recordar la implausibilidad neuronal, al menos en el momento actual, de cualquier modelo que utilice la *propagación de errores hacia atrás*, o las limitaciones de la velocidad de procesamiento neuronal con respecto a los componentes electrónicos, el conocido «programa de 100 pasos» (Feldman, 1985).

Es en este contexto de diálogo entre disciplinas, donde se inscribe este artículo. Un nuevo contexto más exigente, en el que la opinión de otras disciplinas ha de ser tomada en cuenta, bien para hacer plausibles los modelos de procesamiento psicológico, bien para interpretar la actividad neuronal e incluso orientar la búsqueda de nuevos mecanismos de funcionamiento de nuestro sistema nervioso. No nos cabe la menor duda de que nuestra actividad como científicos se ha complicado aún más, si cabe, como tampoco nos cabe duda de lo apasionante que puede ser el camino por recorrer.

2. Arquitectura conexionista y arquitectura cerebral.

Un breve repaso a los supuestos fundamentales de los modelos conexionistas ayudarán a ver la relación entre éstos y la arquitectura propia del sistema nervioso. El isomorfismo arquitectónico que se aprecia tiene importantes implicaciones en el estudio de los procesos cognitivos básicos y en el abordaje de estos procesos desde una perspectiva neurobiológica. En este apartado vamos a describir muy brevemente la analogía entre las redes neuronales y las auténticas redes de neuronas, haciendo especial énfasis sobre el lugar que ocupan en ellas los conceptos básicos de la psicología cognitiva: la representación, la memoria y el aprendizaje.

2.1. Características generales de los modelos de procesamiento distribuido en paralelo.

Según Rumelhart, Hinton y McClelland (1986) ocho son los principales aspectos o elementos de cualquier modelo de procesamiento distribuido en paralelo:

1. Un conjunto de unidades de procesamiento.
2. Un estado de activación.
3. Una función de salida para cada una de las unidades.
4. Un patrón de conexión entre las unidades.
5. Una regla de propagación para propagar la activación a través de la red de conexiones.
6. Una regla de activación que calcule nuevos niveles de activación para cada unidad a partir de las entradas que recibe de otras unidades y de su estado previo de activación.
7. Una regla de aprendizaje para modificar los patrones de conexión entre las unidades como resultado de la experiencia.
8. Un ambiente dentro del que debe operar el sistema.

La arquitectura común a cualquier modelo conexionista consiste en una multitud de *unidades* conectadas entre sí. Cada una de las unidades constituyen simples procesadores con un *estado de activación* determinado a las que llegan, por medio de las conexiones, la información de otras unidades o del entorno y que mandan información, a su vez, a otras unidades o al entorno. Con el conjunto de entradas que llega a una unidad se realiza algún tipo de operación -por ejemplo aditiva, aunque no necesariamente- y el valor resultante en combinación con el estado de activación del momento producen el nuevo estado de activación, que es utilizado para producir un valor de salida de la unidad. La unidad manda entonces un impulso de *salida*, que es una función de su estado de activación, a otras unidades por medio de las conexiones que las unen. Un aspecto importante es que diferentes conexiones pueden tener diferentes fuerzas, pesos, que son modificables por la experiencia, de modo que la entrada que le llegue a una unidad depende tanto del estado de activación de la unidad que manda la señal como de la fuerza de la conexión entre las dos unidades. Asimismo, los pesos de las conexiones entre unidades pueden ser positivos o negativos, de modo que una entrada particular a una unidad puede colaborar a su estado de activación excitándola o inhibiéndola, dependiendo del signo del peso.

La forma en que el *patrón de conexiones* se ve afectado por la estimulación está regulado asimismo por un *regla de propagación* y el grado en que este patrón cambia o se modifica en función de esta actividad esta regulado por una *regla de aprendizaje*. Todos estos elementos son definidos de distinta forma en función del *ambiente* de la tarea que se pretenda simular. La regla de aprendizaje tiene una gran importancia en la modelización conexionista, ya que de ella depende el ajuste de pesos entre las unidades de la red y por tanto, la dinámica y el funcionamiento de la red. Sin ellas las redes quedarían imposibilitadas para el aprendizaje.

De forma muy simplificada podemos decir que los estados por los que pasa una red son calculados mediante la aplicación de álgebra lineal. Por ejemplo, las unidades de entrada y salida de la red son representadas mediante vectores de activación, mientras que las interconexiones se representan con matrices. Los sucesivos estados de la red se calculan mediante procedimientos de bondad de ajuste entre las restricciones previamente establecidas en la red y las nuevas entradas de información.

2.2. La metáfora del cerebro.

Visto lo anterior, la analogía con la estructura del sistema nervioso resulta inmediata: si a las unidades las llamamos neuronas y a las conexiones entre las unidades sinapsis podríamos estar hablando del sistema nervioso. Las neuronas pueden tener tasas de descarga muy desiguales, que quedarían representadas por el estado de activación diferente de las unidades de la red. Es más, muchas veces la función de activación en cada unidad posee características de umbral, por debajo del cual la unidad no propaga la activación. Esto podría tener igualmente su similitud en el sistema nervioso. Por ejemplo, en el umbral de despolarización necesario en la membrana neuronal para que en la célula se produzca un impulso nervioso. Asimismo, las neuronas se influyen unas a otras de manera muy diferente, dependiendo del tipo de sinapsis particular; ésta puede variar

tanto en relación a su fuerza como a su signo, pudiendo ser tanto excitatoria como inhibitoria. El tipo de sinapsis particular queda representado en los modelos conexionistas por los pesos de las conexiones. Estos pueden variar tanto en su valor absoluto como en su signo. El valor absoluto estaría representando la fuerza de la conexión sináptica; el signo indicaría si la conexión es excitatoria o inhibitoria.

En este tipo de redes, los patrones de activación de las unidades van experimentando cambios en el tiempo, como resultado de la actividad del sistema. La memoria y la representación ocupan un lugar común: el patrón de conexiones entre las diferentes unidades del sistema. El aprendizaje consiste en el cambio de las conexiones entre unidades, y se produce como una consecuencia de la actividad de procesamiento del sistema. La pauta que siguen las conexiones para cambiar sus pesos y, por tanto, la regla por la que varían las relaciones entre las unidades del sistema es la regla de aprendizaje. El énfasis en esta metáfora lleva incluso a que estos modelos tengan un sinónimo en la expresión red neuronal.

De este modo, el desarrollo de los modelos conexionistas abre una vía de comprensión sobre el modo en que habilidades cognitivas complejas pueden emerger de la interacción entre neuronas. Esta analogía ha conducido, como ya hemos dicho, a un abandono de la metáfora del ordenador, dominante en las últimas décadas en psicología cognitiva, por la metáfora del cerebro como un modelo del funcionamiento mental (Rumelhart, Hinton y McClelland, 1986).

3. Nivel de estudio de los modelos conexionistas.

Los modelos conexionistas, heredando los objetivos de las ciencias cognitivas, tanto en la construcción de modelos del funcionamiento cognitivo humano y animal (tradicionalmente objetivo de la psicología cognitiva), como en la de modelos competentes sin los sesgos propios de la cognición humana (tradicionalmente objetivo de la inteligencia artificial), aportan originalidad en cuanto a su arquitectura, inspirada en el sistema nervioso. Es el cambio radical que esta nueva arquitectura proporciona en cuanto a la capacidad de procesamiento, unido a su plausibilidad biológica, lo que sitúa al conexionismo en una posición ventajosa respecto a la orientación simbólica, dominante en las ciencias cognitivas (véase, p. e., Bechtel y Abrahamsen, 1991). Asimismo, su situación, a caballo entre ciencias cognitivas y neurociencias, posibilita abordar los problemas desde perspectivas teóricas y metodológicas diferentes y, lo más importante, facilita un marco y lenguaje común entre disciplinas, permitiendo una comunicación que antes tenía serias dificultades para establecerse.

Sin embargo, su situación, sirviendo a múltiples objetivos e intereses diferentes, aunque sobre un mismo objeto de estudio, sitúan al conexionismo en una posición epistemológica difícil, cuyos objetivos no están siempre claramente definidos (véase p.e., García Madruga, 1992). Dos ejemplos de esto, son los llamados modelos *simplificados* y modelos *realistas* del cerebro, ambos inspirados en la modelización conexionista, pero con objetivos muy diferentes.

3.1. Modelos de cerebro simplificados y realistas.

Si desde el paradigma del procesamiento de información se ha hablado de *metáfora dura* y *metáfora blanda*, según el grado de similaridad postulado entre mente y ordenador, con los modelos conexionistas y la metáfora del cerebro ocurre algo parecido. Por un lado, algunos investigadores defienden que la arquitectura cerebral es una arquitectura útil para construir simulaciones de procesos cognitivos, y que es interesante que un modelo que funcione correctamente posea tal parecido con el cerebro, pero lo más importante es que funcione bien, aunque para ello se tengan que sacrificar detalles de la metáfora. Sin embargo, otros investigadores afirman que es un error desarrollar modelos con redes de unidades y conexiones de las que se sabe positivamente que tienen características distintas a las neuronas y sinapsis, si con ellas pretendemos explicar los procesos mentales producidos por nuestro cerebro.

Estas dos concepciones son frecuentemente conocidas como *inspiradas* y *restringidas* neuronalmente (Morris, 1989) o como modelos de cerebro *simplificados* y *realistas* (Sejnowski, Koch y Churchland, 1988). La aproximación restringida o realista en las redes neuronales es más reciente y va surgiendo como una respuesta que manifiesta el rechazo de los neurobiólogos a la modelización simplificada. En los modelos realistas no se usan reglas de aprendizaje como la propagación hacia atrás o conexiones que puedan pasar de ser excitatorias a ser inhibitorias, al no haberse encontrado fenómenos parecidos en el sistema nervioso. Asimismo, intentan incluir en la modelización características neuronales que frecuentemente no son consideradas en las modelizaciones habituales, como son la actividad de los canales de iones o la actividad de las enzimas en el proceso de transmisión sináptica (Hawkins, 1989). Sin embargo, los modelos simplificados son los más habituales y los que han dado su actual popularidad al conexionismo. Los desarrollados por Rumelhart, McClelland y el grupo de investigación PDP se encuadran entre estos últimos.

El argumento básico de los defensores de los modelos restringidos neuronalmente lo expresa de forma contundente Crick (1989). Según Crick, el escaso interés que los psicólogos, lingüistas y demás interesados en las funciones mentales muestran en el contenido real del cerebro se explica porque para ellos el cerebro es demasiado complicado y difícil de comprender. Según ellos, se posee mucha y no muy clara información respecto a los mecanismos de funcionamiento del cerebro y la labor es aún más complicada si se pretende relacionar esos mecanismos con los procesos cognitivos. Así, si se quieren construir modelos que reflejen con la mayor exactitud los procesos cognitivos superiores resulta necesario desmarcarse de la complicación cerebral y hacer simplificaciones. Éste último sería el argumento básico empleado por los defensores de los modelos simplificados de cerebro. Para Crick, este es un camino equivocado. Si el cerebro es realmente tan complicado, nada claro se puede sacar respecto de su funcionamiento si se estudia desde una aproximación simplificada y de caja negra.

Hawkins (1989) también se sitúa en esta postura y señala varios aspectos por los que una modelización restringida podría resultar de interés. En primer lugar, permitiría

una comprobación a nivel fisiológico de los supuestos de un determinado modelo, haciendo así de la simulación algo más que una modelización formal de procesos. Esto permitiría un enriquecimiento mutuo entre la investigación fisiológica y la investigación cognitiva. La investigación neurofisiológica podría servirse de la modelización conexionista como fuente generadora de hipótesis, en busca de fenómenos sugeridos en un determinado modelo. Por otro lado, en muchos casos sólo la investigación fisiológica puede decidir cuál de los diferentes modelos posibles con resultados similares es el más adecuado, resolviendo de este modo el problema de la equivalencia funcional según la plausibilidad biológica de los modelos. En segundo lugar, el sistema nervioso sigue siendo para muchas tareas el dispositivo más potente que se conoce. Así, no resulta extraño que los modelos de redes neuronales hayan tenido éxito al usar procesamiento en paralelo y distribuido, el que realiza el sistema nervioso, y que este tipo de modelos sea hoy extensamente utilizado en ingeniería para realizar tareas con objetivos que nada tienen que ver con el conocimiento del sistema nervioso. Entonces, ¿por qué no copiar también la estructura de la neurona, del funcionamiento sináptico o la arquitectura concreta de diferentes áreas del cerebro?

La defensa básica de la utilización de modelos de cerebro simplificados se realiza desde la ciencia cognitiva. McClelland, Rumelhart y Hinton (1986) lo dejan claro en el capítulo introductorio a sus dos volúmenes: *«Nosotros somos, después de todo, científicos cognitivos, y los modelos PDP llegan a nosotros por razones psicológicas y de computación.»* (p.11). Se trata de una cuestión de objetivos. No obstante, hay una gran diferencia entre los modeladores de redes neuronales y otros científicos cognitivos: los creadores de redes neuronales, aun las simplificadas, dan una gran importancia al parecido de estas con el sistema nervioso, mostrándose sensibles –hasta el límite que permitan los criterios psicológicos y computacionales– con la información proveniente de él (Rumelhart y McClelland, 1986; Bechtel y Abrahamsen, 1991).

Hay otro argumento para defender la modelización simplificada, en el que no se subordina el conocimiento del sistema nervioso a objetivos computacionales. La idea básica es que el sistema nervioso tiene múltiples niveles de organización, cada uno con sus propiedades, desde niveles muy moleculares hasta niveles más molares, como el estudio de las conexiones entre diferentes núcleos encefálicos. Cada uno de estos niveles presenta un conjunto de propiedades particulares, de forma que los niveles superiores pueden exhibir propiedades que no estén presentes en los niveles más bajos. Estas propiedades podrían ser más fácilmente simulables en modelos simplificados que en modelos realistas (Sejnowski, Koch y Churchland, 1988).

Aunque, en un principio, el panorama expuesto podría llevar a una asociación entre los defensores de los modelos realistas con una perspectiva en la que ocupe una posición preferente el nivel biológico en oposición al computacional en el estudio de los fenómenos cognitivos, esto no sería exacto. En primer lugar, la defensa de modelos simplificados es compatible con un tratamiento prioritario del estudio del sistema nervioso, como hemos expuesto (Sejnowski, Koch y Churchland, 1988). Segundo, incluso si el interés se centra en el estudio de los mecanismos cerebrales de plasticidad sináptica

implicados en los procesos de aprendizaje, para la correcta comprensión de estos mecanismos es necesario el conocimiento de las diferentes formas de aprendizaje, así como su posible implementación en diferentes arquitecturas neuronales (Morris, 1989; 1990). Es necesario, dicho de otra manera, un abordaje complementario del fenómeno a un nivel distinto al biológico, como expondremos en el siguiente apartado.

En un nivel distinto a los modelos simplificados y realistas podríamos situar a los científicos que, utilizando una arquitectura de procesamiento distribuida y en paralelo, tienen como objetivo confeccionar sistemas inteligentes, útiles para algún propósito, con independencia del grado de similitud que tengan tanto con el funcionamiento del cerebro, como con sus funciones. En este punto nos sumamos a las palabras de Crick: *«Construir una máquina que funcione (como un ordenador en paralelo) es un problema de ingeniería. La ingeniería está, con frecuencia, basada en la ciencia, pero su objetivo es diferente. Una buena obra de ingeniería es una máquina que haga algo útil. Comprender el cerebro es, por otro lado, un problema científico. El cerebro ha llegado hasta nosotros como el producto de una larga evolución. No estamos interesados en saber cómo podría funcionar, sino cómo lo hace en realidad»* (Crick, 1989, p. 132). Añadiríamos que los modelos tanto simplificados como realistas tratan, a nuestro juicio, con problemas científicos. Su interés está en estudiar la relación del cerebro con las funciones que de su funcionamiento emergen.

En todo caso, y tanto en la versión simplificada como en la realista, existe una colaboración entre los modeladores de redes neuronales y los neurocientíficos. Esta colaboración funciona en ambas direcciones y para situarla es útil hacerlo en términos de los niveles de explicación que Marr (1982) postuló para las neurociencias.

3.2. Posición de los modelos conexionistas en los niveles de explicación de Marr para las neurociencias.

Marr (1982) defendió la existencia de tres niveles explicativos: computacional, algorítmico e implementacional. El nivel computacional hace referencia al tipo de función que un sistema de procesamiento de información tiene que realizar. En este nivel, un tipo de tarea se caracteriza por su objetivo, la función que realiza, sin importar los procesos. El nivel algorítmico es el que tiene que ver con los procesos. Una misma tarea puede ser resuelta por diferentes procedimientos. Es este el nivel de discusión propio de la psicología (las diferentes teorías explicativas de una tarea como el condicionamiento clásico, por ejemplo). Por último, el nivel implementacional hace referencia al sustrato físico sobre el que los algoritmos se llevan a cabo. En este nivel se situaría la neurobiología. Estos tres niveles gozan de autonomía, aunque no de forma completa. Por ejemplo, aunque hay un nivel propio de discusión psicológico, hay restricciones provenientes del nivel computacional (una teoría debe desempeñar correctamente la función para la que ha sido pensada) e implementacional (debe poder ser ejecutada por nuestro cerebro). Aunque lo normal es que el debate científico se dé dentro de un nivel, la comunicación entre niveles es necesaria, pues hay una relación lógica y causal entre ellos.

Los modelos conexionistas, en su formulación más extendida (Rumelhart, McClelland y cols., 1986) se situarían en el nivel algorítmico, aunque más cercanos al nivel implementacional que otras conceptualizaciones propias del nivel algorítmico, como las que se desarrollaron con anterioridad desde el enfoque computacional de la mente que impulsó la revolución cognitiva. No obstante, los modelos desarrollados por Rumelhart, McClelland y su equipo entrarían dentro de los categorizados como modelos simplificados y, según se haga referencia a modelos simplificados o realistas, el nivel implicado estará más o menos cerca del nivel implementacional. En este sentido, los modelos realistas estarían tan comprometidos con el nivel implementacional como con el algorítmico.

La colaboración entre los distintos niveles era vista por Marr como un elemento imprescindible para el desarrollo de las neurociencias. Pensamos que esta colaboración encuentra un marco óptimo en los modelos conexionistas, que aportan un lenguaje común a los niveles algorítmico e implementacional. Entre los modelos PDP, sumidos en el nivel algorítmico y la neurobiología de la cognición, que representaría el nivel implementacional, se encuentran los modelos realistas, que requieren un planteamiento simultáneo del problema a ambos niveles.

El proceso por el que la comunicación entre los diferentes niveles debería producirse ha sido sugerido por Sejnowski (1986). En primer lugar, es necesario detectar las variables importantes a los diferentes niveles; después, establecer con exactitud el algoritmo por el que esas variables interactúan; por último, debe contrastarse experimentalmente que los mecanismos postulados son los correctos. Esto último implicaría tanto a la investigación en psicología experimental como en neurobiología. La colaboración entre los niveles algorítmicos e implementacional a través de los modelos conexionistas podría ir en las dos direcciones: el nivel algorítmico podría beneficiarse del nivel implementacional y viceversa.

En ocasiones, desde el nivel implementacional se postula la implicación que a nivel computacional tienen determinados fenómenos. Éste es el caso de algunos mecanismos bien conocidos a nivel biológico, como la cantidad de neurotransmisor liberado por una neurona, en función de la entrada de iones de calcio en su botón terminal. Se ha comprobado que un mecanismo tan simple como éste está en la base de determinadas formas de aprendizaje en organismos simples, lo que ha llevado a postular la existencia de un alfabeto celular para el aprendizaje (Hawkins y Kandel, 1984). Es también el caso de la multitud de estudios que relacionan determinadas estructuras del sistema límbico, como el hipocampo, con la memoria. La modelización conexionista podría ser un buen método para poner a prueba esas hipótesis y para conducir a una mayor operacionalización las ideas formuladas a nivel cualitativo (véase p.e., Donegan, Gluck y Thompson, 1989). Asimismo, mediante la utilización de modelos realistas, podrían simularse lesiones específicas imposibles de realizar en el tejido vivo, como la de sinapsis concretas, e incluso canales iónicos.

Por último, a un nivel más general, el desarrollo de modelos conexionistas puede

inspirar la investigación neurobiológica. Los algoritmos desarrollados por estos modelos se basan en determinados supuestos sobre la naturaleza de las neuronas y sinapsis. Con frecuencia se sabe que estos supuestos son falsos. Pero, también, con frecuencia simplemente no se conocen mecanismos similares en el sistema nervioso. En este sentido, los modelos conexionistas estarían proporcionando nuevas líneas de investigación, al sugerir diferentes maneras en que el sistema nervioso podría funcionar, derivando nuevas predicciones.

El nivel algorítmico también puede verse favorecido por el conocimiento proveniente del nivel implementacional. Con frecuencia, los modeladores se encuentran con algoritmos alternativos para una misma función, que pueden ser implementados de maneras deferentes. El conocimiento neurobiológico, en este caso, debe decidir cuál de las diferentes formas de implementación es consistente con el funcionamiento del sistema nervioso, para decidir, de este modo, cuál de los diferentes algoritmos es el adecuado. En esta misma línea, es posible que la capacidad de procesamiento del cerebro dependa críticamente de características peculiares de la transmisión de información sináptica que queden fuera de los modelos habituales (simplificados). En este caso, la información neurobiológica sería muy relevante para la creación de redes neuronales con capacidad explicativa para la cognición humana y animal.

En definitiva, dentro de los niveles que Marr vislumbró para el estudio de las neurociencias, los modelos conexionistas ocupan una posición importante entre los niveles algorítmico e implementacional. De una u otra manera acercan a estos dos niveles y se constituyen en un vehículo de relación entre ambos.

4. Un ejemplo a modo de conclusión.

Hemos preferido concluir con un ejemplo que ilustre el principal argumento de este trabajo, este es, que la nueva perspectiva conexionista está haciendo posible una comunicación productiva entre ciencias cognitivas y neurociencias. Vamos a hacerlo atendiendo a uno de los puntos más críticos en la analogía entre redes neuronales y el sistema nervioso: la regla de aprendizaje. Existen diferentes reglas de aprendizaje utilizadas en la modelización conexionista. Las más frecuentes son la regla de Hebb (Hebb, 1949) y la de Widrow-Hoff (1960), también conocida como regla de Rescorla-Wagner o regla delta (Rescorla y Wagner, 1972; Sutton y Barto, 1981).

Desde la perspectiva neurobiológica, se asume desde los trabajos de Ramón y Cajal (1899) que las sinapsis son la base de la memoria y que la alteración en la eficacia de éstas la base del aprendizaje. Sin embargo, desde la perspectiva neurobiológica, es decir desde el nivel implementacional, la regla de aprendizaje plantea serios problemas. No se conocen mecanismos de plasticidad sináptica en el cerebro con un funcionamiento similar a las reglas de aprendizaje que modifican la fuerza de las conexiones en los modelos conexionistas (nivel algorítmico). No obstante, recientemente se ha descubierto un mecanismo de plasticidad cuyo funcionamiento obedece a los requerimientos de la regla de Hebb, la potenciación a largo plazo (PLP).

Nos interesa, en primer lugar, detenernos en la regla de Hebb, formulada originalmente en 1949: «*Cuando el axón de una célula A está lo suficientemente cerca como para excitar a una célula B, y repetida o persistentemente participa en su excitación, tiene lugar algún proceso de crecimiento o cambio metabólico en una o ambas células, de manera que la eficacia de A como una de las células que excita a B se incrementa*». Aunque esta regla estaba originalmente formulada en términos neurobiológicos, hasta el conocimiento de la PLP no había recibido apoyo empírico, no habiéndose identificado ningún mecanismo similar en el sistema nervioso. Así, su supervivencia hasta nuestros días ha dependido más de la buena aceptación entre psicólogos y estudiosos del aprendizaje del principio funcional que en ella subyace que de su cercanía a la realidad neuronal. De hecho, para los psicólogos, la regla de Hebb es entendida como el fortalecimiento de la conexión entre dos unidades como consecuencia de su activación conjunta. Es, por tanto, el nivel algorítmico el que ha permitido la pervivencia de la regla de Hebb. Así, su extrapolación a las redes neuronales ha sido inmediata, bajo un enunciado ligeramente diferente, adaptado a la terminología conexionista, como por ejemplo: el peso de la conexión entre dos unidades puede incrementar o disminuir en proporción directa al producto de sus activaciones. Esto es representado por la función:

$$\Delta p_{ui} = R a_u a_i,$$

donde el incremento de los pesos de conexión entre las unidades u e i sería igual al producto de una constante que especifica la cantidad de aprendizaje por cada ensayo, por la activación de la unidad u y la activación de la unidad i (Bechtel y Abrahamsen, 1991).

Ya hemos adelantado que recientemente se ha propuesto un candidato a mecanismo de plasticidad sináptica que modifica la fuerza de una sinapsis según un patrón de activación similar al postulado por Hebb: la *potenciación a largo plazo* (Brown, Kairiss y Keenan, 1990). Esta característica de la PLP está estrechamente ligada a un tipo de receptores neuronales del neurotransmisor glutamato (receptores NMDA). La PLP posee, además, un conjunto de características muy sugerentes para ser considerada como un mecanismo de plasticidad sináptica implicado en los procesos de aprendizaje en el cerebro de los vertebrados (Barnes, 1979; Berger, 1984; Barnes y McNaughton, 1985; McNaughton y cols., 1986; Morris y cols., 1986; Skelton y cols., 1987).

Desde el punto de vista de las relaciones entre el nivel algorítmico y el nivel implementacional, no obstante, la más llamativa de estas características es que la producción de PLP depende de que se produzca un potencial de acción en la neurona presináptica al mismo tiempo que la despolarización de la neurona postsináptica¹. Como acabamos de describir, éste es un requerimiento básico de la regla de Hebb. Así, la PLP podría ser el mecanismo de plasticidad sináptica que acercara las redes neuronales a las verdaderas redes de neuronas, allí donde en la metáfora se encuentran más alejadas: en la regla de aprendizaje.

¹ El potencial de acción es el disparo de una neurona (presináptica). En la neurona postsináptica, despolarización es equivalente a excitación y la hiperpolarización a inhibición.

BIBLIOGRAFÍA

- Barnes, C. A. «Memory deficits associated with senescence: A neurophysiological and behavioural study in the rat.» *Journal of Comparative and Physiological Psychology* 93 (1979): 74-104.
- Barnes, C. A., y B. L. McNaughton. «An age comparison of the rates of acquisition and forgetting of spatial information in relation to long-term enhancement of hippocampal synapses.» *Behavioral Neuroscience* 99 (1985): 1040-1048.
- Bechtel, W., y A. Abrahamsen. *Conexionism and the mind*. Oxford: Blackwell, 1991.
- Berger, T. W. «Long-Term potentiation of hippocampal synaptic transmission affects rate of behavioral learning.» *Science* 224 (1984): 627-630.
- Brown, T. H., E. W. Kairiss, y C. L. Keenan. «Hebbian synapses: Biophysical Mechanisms and Algorithms.» *Annual Review of Neuroscience* 13 (1990): 475-511.
- Crick, F. H. C. «The current excitement about neural networks.» *Nature* 337 (1989): 129-132.
- Donegan, N. H., M. A. Gluck, y R. F. Thompson. «Integrating behavioral and biological models of classical conditioning.» *Computational models of learning on simple neural systems*, 109- 156. Eds R. D. Hawkins, y G. H. Bower. London: Academic Press, 1989.
- Feldmann, J. A. «Connectionist models and their application: Introduction.» *Cognitive Science* 6 (1985): 205-254.
- García-Madruga, J. A. «Introducción a la edición española.» En *Introducción al procesamiento distribuido y paralelo*, Comps D. E. Rumelhart, J. L. McClelland, y el grupo de investigación PDP. Madrid: Alianza, 1992.
- Gormezano, I., W. F. Prokasy, y R. F. Thompson. Eds. *Classical Conditioning*. Hillsdale, N.J.: Lawrence, 1987.
- Hawkins, R. D. «A biologically realistic neural network model for higher-order features of classical conditioning.» *Parallel Distributed Processing. Implications for Psychology and Neurobiology*, 214-247. Ed. R. G. M. Morris. Oxford: Oxford University Press, 1989.
- Hawkins, R. D., y E. R. Kandel. «Is There a Cell-Biological Alphabet for Simple Forms of Learning?» *Psychological Review* 91, no. 3 (1984): 375-391.
- Hebb, D. O. *The organization of behavior: a neuropsychological theory*. New York: John Wiley & Sons, 1949.
- Hinton, G. E., y J. A. Anderson, Eds. *Parallel models of associative memory*. Hillsdale, N.J.: Erlbaum, 1981.
- Marr, D. *Vision*. San Francisco: Freeman, 1982.
- McClelland, J. L., D. E. Rumelhart, y G.E. Hinton: «The appeal of PDP.» *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, 3-44. D. E. Rumelhart, J. L. McClelland, y the P.D.P. research group. 1. Foundations. Cambridge: MIT Press, 1986.
- McClelland, J. L., D. E. Rumelhart, y the P.D.P. research group, Comp. *Parallel distributed processing: explorations in the microstructure of cognition*, 2. Psychological and Biological Models. Cambridge: MIT Press, 1986.
- McCulloch, W. S., y W. Pitts. «A logical calculus of the ideas immanent in nervous activity.» *Bulletin of Mathematical Biophysics* 5 (1943): 115-133.
- McNaughton, B. L., C. A. Barnes, G. Rao, J. Baldwin, y M. Rasmussen. «Long-term enhancement of hippocampal synaptic transmission and the acquisition of spatial information.» *Journal of Neuroscience* 6 (1986): 563-571.
- Morris, R. G. M. «Computational neuroscience: modelling the brain.» *Parallel Distributed Processing. Implications for Psychology and Neurobiology*, 203-213. Ed. R. G. M. Morris. Oxford: Oxford University Press, 1989.
- Morris, R. G. M. «Synaptic plasticity, neural architecture and forms of learning.» *Brain organization and memory: cells, systems and circuits*, Eds. J. L. McGaugh, N. M. Weinberger, y G. Lynch. New York: Oxford University Press, 1990.
- Morris, R. G. M., E. Anderson, G. S. Lynch, y M. Baudry. «Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist.» *Nature* 319 (1986): 774-776.
- Ramón y Cajal, S *Textura del sistema nervioso del hombre y de los vertebrados*. Madrid: Imprenta y librería de Nicolás Moya, 1899.
- Rescorla, R. A., y A. R. Wagner. «A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement.» *Classical Conditioning II: Current Research and Theory*, 64-99. Eds. A.H. Black, y W.F. Prokasy. New York: Appleton-Century-Crofts, 1972.
- Rumelhart, D. E., y J. L. McClelland. «PDP models and general issues in cognitive science.» *Parallel Distributed Processing: explorations in the Microstructure of Cognition*, 110-146. D. E. Rumelhart, J. L. McClelland, y the P.D.P. research group. 1. Foundations. Cambridge: MIT Press, 1986.
- Rumelhart, D. E., J. L. McClelland, y the P.D.P. research group, Comp. *Parallel distributed processing: explorations in the microstructure of cognition*, 1. Foundations. Cambridge: MIT Press, 1986.
- Rumelhart, D.E, G.E. Hinton, G.E y J.L McClelland «A general framework for parallel distributed processing» *Parallel distributed processing: explorations in the microstructure of cognition*, 1. Foundations. 45-76. Eds. D.E Rumelhart, D.E. McClelland, y the P.D.P. research group. Cambridge: MIT Press, 1986.

- Schneider, W. «Connectionism: Is it a paradigm shift for psychology?» *Behavior Research Methods, Instruments and Computers* 19 (1987): 73-83.
- Sejnowski, T. E., C. Koch, y P. Churchland. «Computational Neuroscience.» *Science* 241 (1988): 1299-1306.
- Sejnowski, T. J. «Open questions about computation in cerebral cortex.» *Parallel distributed processing: explorations in the microstructure of cognition*. 372-389. Eds J. L. McClelland, D. E. Rumelhart, y the P.D.P. research group. 2. Psychological and Biological Models. Cambridge: MIT Press, 1986.
- Skelton, R. W., A. S. Scarth, D. M. Wilkie, J. J. Miller, y A. G. Phillips. «Long-term increases in dentate granule cell responsivity accompany operant conditioning.» *Journal of Neuroscience* 7 (1987): 3081-3087.
- Sutton, R. S., y A. G. Barto. «Toward a modern theory of adaptive networks: expectation and prediction.» *Psychological Review* 88 (1981): 135-171.
- Widrow, B., y M. Hoff. «Adaptive switching circuits.» IRE WESCON Convention Record, vol. 4: 96-104. New York: IRE, 1960.