

# *Conexionismo: una útil herramienta para otras ciencias y un problemático modelo para la psicología*

Vicente CABALLERO DE LA TORRE y Francisco José ROBLES RODRÍGUEZ

## **Resumen**

Cuando se habla de “conexionismo” o “teoría conexionista” nos estamos refiriendo a una rama de las Neurociencias y de las Ciencias cognitivas que surgió hace algunas décadas como alternativa minoritaria a la IA (inteligencia artificial). Las redes neuronales artificiales fueron originalmente un intento de simulación abstracta de los sistemas nerviosos biológicos. Sin embargo, no es fácil que haya un modelo completamente satisfactorio. Como conclusión se afirma que la Teoría de redes y la Psicología cognitiva se desarrollan al margen de una serie de cuestiones filosóficamente problemáticas pensando, quizá, que se podrá dar cuenta de ellas cuando el problema de fondo (la relación mente-cerebro) se haya resuelto.

*Palabras clave:* “teoría conexionista”, “neurociencias”, “modelo”, “inteligencia artificial”, “problema mente-cerebro”.

## **Abstract**

When we talk about “conexionist theory” we are talking about a section of Neurosciences and Cognitive Sciences. Artificial Neural Nets born to be an alternative to AI (Artificial Intelligence) and a model of Biological Neural Nets. But, it seems that’s not possible get a model without any problem. The conclusion of the article: the Artificial Neural Nets Theory and the Cognitive Psychology wants to give new explanations without studying all that philosophical problems (mind-brain relationship) because, maybe, these Scientifics believe that those problems can have a solution at the final conclusions of the Cognitive investigations.

*Keywords:* “connexionist theory”, “neurosciences”, “model”, “artificial intelligence”, “mind-brain problem”.

Cuando se habla de “Conexionismo” nos estamos refiriendo a una rama de las Neurociencias y de las Ciencias cognitivas que surgió hace algunas décadas como alternativa minoritaria a los modelos computacionales de la mente tratando de conseguir emular la actividad cognoscitiva del individuo no mediante el cómputo de información sino desde la gestión de insumos por parte de una red que trataría de emular, en la medida de lo posible, la arquitectura del sistema nervioso central (SNC).

Se puede resumir lo anterior diciendo que un sistema de redes neuronales artificiales es un modelo matemático basado en ciertas características de las células neuronales, concretamente en las respuestas ‘todo o nada’ derivadas del potencial de acción, que son interpretadas como valores binarios de ceros y unos – el aumento de la potencia de emulación de las redes cuando hacen uso de valores no booleanos, es decir, cuando comprenden valores intermedios entre el 1 y el 0, ha sido estudiada por uno de nosotros en el artículo *Redes, lógicas no clásicas y neuronas*, referenciado en la Bibliografía.

Las nociones de “cómputo” y de “información” son dejadas de lado para ser sustituidas por las de “conexión” e “insumo”. De la mente al cerebro, del cómputo informacional a la conexión neuronal. Para los psicólogos cognitivos hay que suponer que la mente o “yo activo”, como la llama Popper, es un “programador”<sup>1</sup>: “El cerebro lo posee el yo más bien que a la inversa. El yo está casi siempre activo. Sugiero que la actividad del yo es la única actividad genuina conocida. El yo activo es el programador activo del cerebro (que es el computador); es el ejecutante cuyo instrumento es el cerebro. “Se ha ido cambiando el enfoque explicativo desde la cuestión de cómo funciona la mente a cómo está dispuesto el cerebro pero, durante todo este recorrido, la noción del *algoritmo* (una noción genuinamente cibernética) ha ido vertebrando las distintas propuestas las cuales –a pesar del bloqueo que han sufrido en tanto que modelos teóricos para explicar el SNC– han encontrado múltiples utilidades entre ellas la de herramienta de apoyo de economistas, muchos de ellos de nuestro país, a la hora de lanzar predicciones en el ámbito financiero.

Fue, en efecto, Wiener, el fundador de la Cibernética, quien puso de manifiesto la imperante necesidad de contar con la Psicología –a la que él entendió como *ciencia de la mente*– si se quería avanzar en serio en el campo de la nueva disciplina que da nombre a su obra<sup>2</sup> La esencia misma de la cibernética, como él mismo expone allí, es el *feedback*, es decir, todo aquello que cae bajo el concepto de retroalimentación. En las redes neuronales analógicamente relevantes para explicar el funcionamiento de las redes naturales se produciría también una suerte de “cadenas de retroalimentación en las que no interviene ningún factor humano”<sup>3</sup>

<sup>1</sup> Popper, K. y Eccles, J.: *El yo y su cerebro*. Barcelona, Labor, 1985, p. 135.

<sup>2</sup> Wiener, Norbert: *Cibernética y sociedad*. Buenos Aires, Ed. Sudamericana, 1958.

<sup>3</sup> Wiener, *op. cit.*, p. 135.

Pues bien, dentro de este concepto de retroalimentación puede caer el proceso que, más concretamente, se ha vertido por el término *retropropagación* y al que han dedicado sus trabajos numerosos especialistas entre los cuales cabe destacar a G. E. Hinton y Rumelhart<sup>4</sup>. La retropropagación es el proceso constante de ajuste que una red neuronal debe llevar a cabo para asemejar, en la medida de lo posible, la actividad real de salida con la actividad deseada; ésta, a su vez, se corresponde de una u otra manera con la entrada. La naturaleza de la entrada es indiferente, lo verdaderamente importante son las relaciones entre las entradas y las unidades del siguiente nivel o capa. En efecto, entre la entrada y la salida hay un número considerable de capas intermedias conformadas por las denominadas *unidades ocultas*<sup>5</sup>. La totalidad de *algoritmos de retropropagación* que han propuesto los investigadores que trabajan actualmente en este ámbito –véase la recientísima obra al respecto *Unsupervised Learning*, referenciada en la Bibliografía– consiste en hallar qué unidades ocultas (se entiende por “unidades ocultas” las neuronas artificiales que se encuentran entre la capa de neuronas que reciben los *input* y la capa que da salida a los *output*) hay que estimar como “agentes” (“efectores”, en el *argot*) relevantes de ajustes en sus correspondientes pesos sinápticos con el fin de conseguir la representación más adecuada. Algunos de estos algoritmos prescinden incluso de los mecanismos de retropropagación; en cualquier caso, todos los modelos tienen en común partir de la total ausencia de “instructor externo” tal y como debería ocurrir en el cerebro (nadie va ajustando “desde fuera” los errores en la representación de la realidad “externa” conseguida por nuestro SNC) a no ser que supongamos la existencia de una Causa eficiente extranatural, como la postulada por Malebranche, que se encargara de ajustar nuestras representaciones con la *res extensa*.

---

<sup>4</sup> Este último, junto a McClelland, dio un impulso a la Teoría de redes que la rescató del ostracismo al que había sido sometida por los representantes de la IA– con la publicación de *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*.

<sup>5</sup> Cuya configuración sería la propia de una serie de *subgrafos* (donde el grafo principal sería la totalidad del tejido neuronal) Con respecto a esto último, hemos intentado fundamentar en el trabajo (Caballero de la Torre, 2005, 35) cómo la configuración grafológica de las redes neuronales no supone un argumento a favor de la modularidad de la mente sino que, a lo sumo, tal configuración sería un argumento válido para explicar la imposibilidad de la recuperación de determinada información cuando queda dañado algún componente principal que acaso sirviera de acceso único a la misma. No estaría en lo cierto Fodor cuando supone que los mecanismos computacionales específicos de un dominio dado, a los que denomina *facultades verticales*, son módulos neurológicamente compactos (Fodor, J.: *La modularidad de la mente*. Madrid, Morata, 1986), es decir no hay correspondencia uno a uno entre módulos mentales y regiones cerebrales. Debe decirse que ya Lorente de Nó en sus estudios sobre la corteza cerebral del ratón se opuso a Cajal en lo que refiere a la tesis según la cual la gran complejidad del *cortex* humano constituiría un respaldo de la existencia de correlatos morfológicos para el conjunto de funciones que resumimos con la expresión “intelecto”. Lorente, cuya actividad se desarrolló principalmente en EEUU, desde muy al comienzo de sus trabajos, en su etapa española, se opuso a tal correlación.

Pues bien, es en este principio sobre el que Hinton se apoya a la hora de exponer todos los posibles modelos *sin instructor* que harían posible la adecuación deseada de la salida en el caso de redes naturales. La verdad perceptiva es la adecuación del *output* al *input* y tal logro se debe a la actividad de las unidades ocultas.

Todos estos modelos se basan en procedimientos de uso general que ajustan los pesos de la red y cuyo fin es, precisamente, conseguir una descripción económica, donde se prescindiera de lo “inesencial” pero con información suficiente como para reconstruir las entradas; esta descripción es denominada por Hinton *representación de calidad* basada en una reducción del número de parámetros. Una forma de reducción de éstos consiste en ensayar modelos basados en parámetros lo más generales posibles los cuales se corresponderían con características universales de cualesquiera ambientes o entornos físicos<sup>6</sup>. Estos parámetros pueden clasificarse en tipos en una ordenación que recuerda a la tabla kantiana de las categorías: a) invariantes (*invariant input parameters*) los cuales representarían una cualidad tal como la *identidad* de un objeto y b) transitorios (*transient input parameters*), involucrados en la representación de, por ejemplo, la posición de un objeto. Esta dicotomía hintoniana es criticada por Stone dado que “sería poco realista esperar del mundo físico una neta partición en parámetros invariantes (p. e., la identidad del objeto) y transitorios (p. e. la posición del objeto) [*However, it may be unrealistic to expect the physical world to partition neatly into invariant (e. g., object identity) and transient (e.g., object position) parameter types*]<sup>7</sup>. Por ello propone una tricotomía –invariantes (identidad), medianamente transitorios (v.g.: color) y enteramente transitorios (posición). El modelo descrito por Stone tendría en común, pues, con el de Hinton el postulado de que los “modelos de procesos perceptuales pueden derivarse del análisis de los tipos de cambios espaciales y temporales inmanentes a la estructura del mundo físico [*models of perceptual processes can be derived from an analysis of the types of spatial and temporal changes inmanent in the structure of the physical world.*]<sup>8</sup>. Como indica Stone, la generalidad deseable de los parámetros exige que estos alberguen propiedades transversales u horizontales –captadas por distintas fuentes sensoriales– tales como la orientación tridimensional de los objetos o su forma (accesibles por la vista y el tacto, p. e.)

Debe señalarse que, en lo que concierne al reconocimiento de una serie de impresiones como procedentes de un mismo objeto (es decir, el reconocimiento de su identidad), éste no puede producirse sino es bajo la caída de este objeto bajo un

<sup>6</sup> Hinton, G. & Sejnowsky, T. J.: *Unsupervised Learning: foundations of neural computation*. MIT, 1999.

<sup>7</sup> Hinton & Sejnowski, *op. cit.*, p. 93.

<sup>8</sup> La crítica a los prejuicios fisicalistas propios de estos planteamientos se expondrá más adelante. No obstante puede encontrarse expuesta en sus rasgos fundamentales en el artículo de Robles y Caballero citado en la Bibliografía. Hinton & Sejnowski, *op. cit.*, p. 72.

“universal” a través del cual queda catalogado, reconocido. No obstante cabe indicar, con Putnam, que “las similitudes de ningún modo son sólo constancias del estímulo físico o patrones en el insumo de los órganos sensoriales. Lo que hace similares a los cuchillos, por ejemplo, no es que se parezcan (pues no se parecen), sino que todos están fabricados para cortar o perforar [...] De este modo, cualquier sistema que pueda reconocer los cuchillos como cosas relevantemente similares debe ser capaz de atribuir propósito a los agentes”<sup>9</sup>. Sin embargo, esta *intencionalidad*, con la que Putnam ha tenido que enfrentarse necesariamente, queda fuera de la propuesta computacional de Wiener, el cual presupone que todo *input* perteneciente al mundo físico es, en principio, neutro; Putnam reconoce que el mundo se configura ya en los estratos más elementales de nuestra percepción; esto es una forma tácita de reconocer que los contextos –o “marcos constructivos”, usando la terminología de Minsky– no pueden ser analíticamente descompuestos. Si así se hiciese tal acción resultaría incompatible con su significación contextual misma<sup>10</sup> del mismo modo que el fondo (en la distinción perceptiva entre fondo y figura) desaparece –en tanto que fondo– cuando la atención se centra sobre él y no sobre la figura: en ambos casos se trata de una imposibilidad estructural que hace caer a todo intento de computar un contexto en un círculo vicioso, puesto que la elección de elementos pertinentes de un contexto se hace sobre la previa pre-comprensión del mismo. Putnam da un ejemplo muy interesante sobre este aspecto: la comprobación experimental de que un número muy significativo de físicos e ingenieros son incapaces de observar como movimiento aparente los destellos sucesivos de luz desde posiciones distintas; ven dos o más luces y no una sola luz en movimiento, como sí ocurre con el resto de individuos: “Los propios *inputs* sobre los que se basa nuestro conocimiento están conceptualmente contaminados”<sup>11</sup>. Este problema lo extiende Putnam desde el modelo computacional (la IA tal y como se entiende comúnmente) y el conexionismo sin instructor (redes neuronales artificiales que corrigen por sí solas los patrones)

Sin embargo, a pesar de todas estas objeciones de entrada, se han ido planteando distintos modelos de aprendizaje no supervisado debido a su indudable utilidad práctica en “campos” (usando el término “campo” en un sentido laxo, poco riguroso) muy diversos. Entre dichas aplicaciones, y por su interés filosófico y carácter problemático, sólo vamos a centrarnos en la aplicación al campo de la Neuromimética (modelos que tratan de emular el funcionamiento del SNC). No obstante se hace a continuación un rápido repaso a las otras tres aplicaciones actualmente existentes:

<sup>9</sup> Graubard, S. R. (comp.): *El nuevo debate sobre la Inteligencia Artificial*. Barcelona, Gedisa, 1993, p. 312.

<sup>10</sup> Robles Rodríguez, F. J.: *Para aprehender la Psicología*. Madrid, Siglo XXI de España, 1996, pp. 154-179.

<sup>11</sup> Putnam, H: *Razón, verdad e historia*, Madrid, Tecnos, 1988, p. 64.

*Primera aplicación.* El “campo” de la ingeniería y de la tecnología en sus diferentes ramas se ha interesado por los sistemas de redes neuronales como una herramienta útil para el reconocimiento de patrones que los ordenadores no pueden llevar a cabo. Entre estas aplicaciones tecnológicas se cuentan el reconocimiento de textos manuscritos, reconocimiento del habla, simulación de centrales de producción de energía, detección de explosivos, identificación de blancos de radares, etc.

*Segunda aplicación.* En el “campo” de la Psicología Cognitiva, el interés por los sistemas de redes neuronales se ha centrado fundamentalmente en su uso como posibles metáforas de la arquitectura cognitiva (no necesariamente plausibles biológicamente) en un intento de proponer una forma alternativa a la metáfora simbólico-computacional para explicar aspectos específicos de la conducta humana. Wiener expone muy claramente que las conexiones cibernéticas no son el modelo del cerebro, a su juicio, sino de la mente: “El cerebro, en condiciones normales, no es un equivalente exacto de la computadora, sino más bien el equivalente de un programa de dicha máquina [...] Es más bien el proceso y no la existencia total de la estructura mecánica de la computadora la que corresponde a la vida del individuo”<sup>12</sup>.

*Tercera aplicación.* En Economía, las redes neuronales funcionan en ámbitos tales como la predicción de la quiebra y la estimación de los rendimientos de carteras de valores<sup>13</sup>. Además, la matemática subyacente que permite entender el funcionamiento de todo tipo de redes –la Teoría de Grafos, vinculada al *Analysis situs* leibniano, donde la distancia es irrelevante y lo importante son los vínculos entre distintos nodos– ha sido aplicada a la Dirección y Administración de Empresas<sup>14</sup>. Estos aspectos de la tercera aplicación serán abordados en próximos trabajos.

<sup>12</sup> Wiener: *op. cit.*, pp. 165-175.

<sup>13</sup> “En el ámbito español, es posiblemente la crisis bancaria del periodo 1977-1985, que afectó a un gran número de entidades de crédito, el suceso más estudiado desde la perspectiva de los modelos de predicción de la quiebra [...]. Si bien, los resultados que obtienen son satisfactorios, dado que logran clasificar correctamente a los bancos en un 94% de ocasiones, no ofrecen una comparación de los resultados que se obtendrían en el caso de utilizar otros métodos. No obstante, afirman que han constatado previamente que los resultados que ofrecen las RNAs son superiores [...]. López y Flórez, utilizando aproximadamente la misma base de datos (de la muestra original de 66 bancos, únicamente toman 52, la mitad de los cuales quebraron) observan mejores resultados en las predicciones efectuadas por una RNA [...]. Asimismo, únicamente utilizan 6 ratios contables para clasificar las entidades bancarias en solventes y fracasadas. Otro campo donde las redes neuronales han encontrado una intensiva aplicación es la predicción bursátil [...]. Es posible efectuar predicciones del IBEX-35 a partir de su comportamiento pasado, mostrándose las RNAs como un instrumento de predicción sensiblemente superior” Andrés Sánchez, Jorge de: “Dos aplicaciones empíricas de las redes neuronales artificiales a la clasificación y la predicción financiera en el mercado español”, en *Revista Asturiana de Economía*, número 28, 2003, pp. 61-85.

<sup>14</sup> Las matemáticas convencionales han tratado el espacio conforme a las magnitudes, especial-

*Cuarta aplicación.* En el campo de la biología ‘artificial’ de donde son originarios estos modelos, que pretende desarrollar modelos biológicos plausibles de redes neuronales reales, éstas fueron originalmente un intento de simulación abstracta de los sistemas nerviosos biológicos, formados por un conjunto de unidades llamadas “neuronas” o “nodos” conectadas unas con otras. Estas conexiones pretendían tener una gran semejanza con las dendritas y los axones en los sistemas nerviosos biológicos. Por ejemplo, Hinton no considera que, frente a Wiener, un sistema conexionista sea el modelo de la mente sino el de su “soporte” fisiológico: el cerebro. Para que sean plausibles cualesquiera modelos que se propongan deben cumplir tres condiciones: a) no hacer uso de la retropropagación puesto que el sentido retrógrado no se da entre las neuronas naturales donde las descargas se dan en un solo sentido y, en el caso de tener que hacer uso de retropropagación indirecta –la descarga afecta de nuevo a las unidades por las que ya ha pasado, “dando la vuelta” a todo el recorrido– ésta debe ajustarse a la rapidez propia de las neuronas naturales; b) como es obvio, no requerir de supervisión (ausencia de “instructor externo”) –a no ser que se sostengan consciente o inconscientemente presupuestos metafísicos (ideológicos) ocasionalistas o calvinistas, reformulados muchas veces bajo un fisicalismo reduccionista grosero (al estilo de Laplace); c) que la minimización del costo de codificación (principio de economía) no elimine toda redundancia con el fin de evitar que cuando una sola unidad deje de funcionar se produzcan notables disfunciones en el sistema –algo que, como es obvio, no ocurre en el caso del cerebro.

Con el fin de sortear estos tres escollos se han postulado modelos alternativos que salven la operatividad del sistema en caso de disfunción de una de sus unidades y que, a su vez, respeten el principio de economía, es decir, hagan posible el mínimo costo de codificación. Todos estos modelos tienen en común la característica de regirse en virtud de un *código demográfico* que opera en un espacio segmentado en distintas poblaciones neuronales (grupúsculos de neuronas ratificales) que producen la representación mediante la descripción de una “figura dibujada” a partir de un punto álgido de actividad prominente y de la unión de algunos –*no necesariamente todos*, he aquí la virtud del modelo– puntos constituidos por aquellas neuronas próximas. En caso de representaciones complejas el proceso consistiría en la superación de un valor umbral producido por la aparición de una protuberancia de actividad cuyo origen serían regiones distintas de la red. Al bastar con una sola

---

mente desde la geometría analítica cartesiana, es decir, desde el paradigma del *recinto* y usando el álgebra ordinaria como instrumento de expresión de las magnitudes. El control y la gestión empresarial requieren, en cambio, de una geometría vectorial y no “recintual” que, en combinación con el álgebra booleana y/o procedimientos estadísticos (valor medio, valor dominante, etc.) podría, según, Kaufmann y Aluja dar cuenta de los problemas de carácter altamente combinatorio propios de los menesteres empresariales; los problemas de esta índole son los problemas más importantes que se presentan al hombre según estos autores (Kaufmann y Aluja: *Grafos neuronales para la economía y la administración de empresas*, Madrid, Pirámide, 1995)

protuberancia para representar un complejo cumple sobradamente con el principio de economía que debe regir todo modelo. No obstante es necesario una constante tarea de verificación por parte de la red con el objetivo de conseguir la corrección adecuada entre las distintas partes elementales que constituyen las partes de un complejo, dado que esta corrección es la que proporciona cariz semántico, significativo, a la representación compleja (piénsese en los elementos de un rostro humano, por ejemplo, donde una protuberancia de actividad se corresponde con la boca, otra con la nariz, etc.) Si una región de la red, sin usar todos sus elementos o unidades sino un número suficiente de ellos, es capaz de representar la parte central de una imagen de un rostro y otra región se hará cargo, del mismo modo, de la parte inferior, etc. no por eso queda garantizado que el resultado final (la salida) se ajuste a lo que cabría esperar: un rostro ordenado que se corresponda con el de las unidades de entrada. Para que la salida sea la deseada es necesaria esa constante “tarea” de verificación (auto-supervisión) de la red sobre su propia actividad. Todo esto implica de algún modo algún tipo de propagación retrógrada que corrija el error arrojado por una proto-representación no conforme con el patrón normalizado al que debe de atenerse (el orden de los elementos del rostro, continuando con el mismo ejemplo) hasta conseguir la adecuada representación.

Como puede observarse el modelo más plausible en lo que a su analogía orgánica refiere está también fundamentado sobre la necesidad de retropropagación en la red. El mismo Hinton objetaba a muchos modelos propuestos durante la historia de las redes neuronales el hecho de que en las conexiones biológicas no se da el sentido retrógrado, si bien resolvía esta dificultad apelando a la riqueza cerebral en lo que concierne a la posesión de redes con la capacidad de retrogradar a capas anteriores, de modo que la objeción basada en la *unidireccionalidad* de las neuronas no es insalvable siempre que el tiempo que se emplee sea razonablemente similar al que emplean nuestras neuronas.

Con respecto a todos los modelos cabe subrayar su dependencia con respecto al *cálculo del error*, condición *sine qua non* para la aplicación de propagación retrógrada. Sin embargo, como ya detectó Hinton, el tiempo empleado en el cálculo crece correlativamente –con más rapidez, de hecho, debido al mayor número de veces que se han de corregir los pesos sinápticos– al aumento del tamaño de la red. Dando por hecho la enorme complejidad de la red cerebral esta objeción *material* no resulta baladí.

Pero también nos interesa aquí presentar una objeción de corte *formal* contra estos modelos que requieren de retropropagación para su *correcto* funcionamiento. Esta objeción apelaría sobre todo a la falta de definición (a la *borrosidad*, cabría decir) del estatuto lógico-matemático de la retropropagación misma en tanto que mecanismo de naturaleza algorítmica. Por algoritmo entendemos una lista de instrucciones que especifican una secuencia de operaciones que arrojan un resultado



para un problema específico y cuyas dos características principales serían su naturaleza determinista y su generalidad.

En los años cincuenta, cuando el modelo para abordar la cuestión mente/cerebro era la máquina de Turing, la misma estructura de ésta dejaba en un estado de provisionalidad irresoluble la cuestión acerca de si un computador sería capaz de ejecutar cualquier algoritmo; para que esto fuese posible necesitaríamos un teorema que probase que cualquier algoritmo pudiera reducirse a un “algoritmo-Turing”. La irresolubilidad de la cuestión no radica en la dificultad de hallar el teorema en cuestión sino en la imposibilidad lógica de su consecución debido a la indefinición de la noción misma de algoritmo. De modo que lo máximo que puede conseguirse es un respaldo empírico, a base de éxitos sucesivos, para la reducción de cualquier algoritmo al procedimiento propio de una máquina de Turing, pero no podría haber una prueba matemática definitiva de que los algoritmos de cualquier índole pueden ser susceptibles de cómputo por una máquina de Turing.

Por otro lado, cuando, por ejemplo, un *perceptrón* (red que consigue reproducir imágenes) capta una imagen con un rostro –sea del natural como de una fotografía– y consigue una salida que es esa misma imagen, la naturaleza del *input* como la del *output* es la misma: una imagen en color de un rostro humano que nos resulta tan *material* tanto si se trata del rostro de una persona viva y presente como si se trata del retrato obtenido por las unidades de salida. Nos encontramos ante un doble problema:

*Primer problema.* Si, por un lado, ambas imágenes tienen un significado es porque el intérprete externo – el ser humano que se sitúa frente a los resultados arrojados por el computador– es capaz de dotarlos de algún tipo de significación global. Del mismo modo, volviendo al tema que aquí nos ocupa, una red neuronal puede reproducir fielmente una imagen sin necesidad de instructor externo (de forma no supervisada) pero esa imagen no será sin duda interpretada de la misma forma por un sujeto operatorio –de carne y hueso– acostumbrado a contemplar –v. gr., cualquier occidental al uso– imágenes sin profundidad que representan realidades tridimensionales, como una fotografía en papel o en una pantalla, que por un aborigen que jamás se ha enfrentado a nada semejante en toda su vida, el cual no verá, hasta pasado un cierto tiempo de aprendizaje, otra cosa que manchas. Ahora bien, las manchas cobrarán sentido para él porque alguien le mostrará los vínculos entre las estampas y el mundo que, *fenomenológicamente*, le circunscribe y condiciona.

*Segundo problema.* Por otro lado, en la medida en que se asuman prejuicios fisicalistas y emergentistas a la hora de explicar la emanación de “*eso*” que suele llamarse “mente”, no vemos la forma en que de los procesos acaecidos en las capas de unidades ocultas se siga, como epifenómeno, algo de una naturaleza distinta: una *vivencia* de índole mental. Ponemos una imagen frente a las unidades de entrada, el perceptrón lleva a cabo su trabajo y emite una imagen de una resolución mayor o

menor. Pero ambas imágenes necesitan ser comparadas por un controlador experimental. Este problema no puede solucionarse substituyendo al controlador por ninguna otra maquinaria dado que dicha maniobra no haría otra cosa que desplazar el problema a otro lugar. En efecto, no hay aparato científico que no tenga algún tipo de panel en el cual aparezcan relaciones de identidad sintética (no triviales) las cuales, no obstante, son accesibles de forma operatorio-conductual –como un *pragmata* cualquiera, por ejemplo un cuchillo, si queremos volver al ejemplo anterior de Putnam con respecto a los universales– por un sujeto operatorio, de carne y hueso, que debe emitir algún juicio sobre el valor que debe atribuirse a tales resultados.

Este acceso fenoménico es imprescindible si queremos que las máquinas y el conocimiento –éste, como vemos, aparece así con un perfil muy distinto de aquella otra *cosa* llamada “información”– se divorcien. De modo que *incluso* cuando la conducta de un sujeto operatorio de carne y hueso refiere al estudio de las ciencias *duras* (Física, Química) hay un tipo de operaciones que no puede quedar segregado: la del control de resultados –aunque sí sea posible eliminar las operaciones referidas al acceso al objeto –acceso inviable, en el marco de la ciencia actual, si nuestros sentidos no son potenciados mediante la maquinaria pertinente.

Aunque los Dreyfus no se han hecho cargo de la brecha entre representación mental y proceso físico sí es verdad que apuntan al problema que acabamos de exponer cuando concluyen que una red “ha de aprender de sus propias “experiencias” para hacer asociaciones que se asemejen a las humanas, y no tanto ser enseñada a hacer asociaciones que son específicas de su entrenador; una red debe compartir también nuestro sentido de propiedad de un comportamiento, y esto significa que debe compartir nuestras necesidades, deseos y emociones y poseer un cuerpo parecido al humano con movimientos físicos, habilidades y vulnerabilidad a la violencia igualmente apropiados”<sup>15</sup>. Cabe suponer, tras la lectura del artículo, que los investigadores en redes neuronales consideran dicha dificultad como algo que es cuestión de tiempo; pero, a nuestro juicio, tal “dificultad” sólo podría ser remontada si partimos de una noción de cuerpo fenomenológica, totalmente alejada del prejuicio cartesiano-galileano que lo reduce a pura extensión física.

Y en efecto la solución no puede partir de otro lugar. Uno de nosotros exponía así la importancia del cuerpo operatorio vivido: “El cuerpo detenta y despliega un tipo de espacialidad que, contrariamente a la espacialidad “posicional” cartesiana<sup>16</sup>, se muestra bajo la forma de una “espacialidad de situación” [...] La experiencia del propio cuerpo revela el nexo psíquico que acompaña a toda génesis del sentido, siendo el caso que la estructura de esta instancia primordial exhibe una naturaleza inminente que conjuga la perspectiva y la circunstancia”<sup>17</sup>.

<sup>15</sup> Graubard, *op. cit.*, p. 52.

<sup>16</sup> Recuérdese la explicación de Wiener para solucionar el problema de los universales, donde se otorgaba valor numérico a las características del objeto percibido.

<sup>17</sup> Robles, *op. cit.*, pp. 251-253.

Cierto es que, partiendo de una posición fenomenológica, no resultaría en absoluto un ejercicio de bizantinismo especulativo considerar que organismos con susstratos neuroquímicos distintos de los acotados por la Biología pudieran emitir conducta significativa, es decir, manifestar *psiquismo*. Del mismo modo que partimos de una *epoché* de los prejuicios fisicalistas tampoco ha de concederse mayor crédito a los prejuicios biologists. Ahora bien, lo que sí consideramos como un auténtico desatino es la estructura discreta y elementalista sobre la que descansa el intercambio de señales en las redes neuronales. De hecho, el descenso desde la mente al cerebro –la fuga de científicos desde el marco de la IA y la Psicología cognitiva hacia la investigación en redes neuronales, cuyos motivos tanto internos como externos al devenir de la propia investigación científica describen magníficamente los Dreyfus<sup>18</sup>– no ha evitado que el mismo problema de los contextos, señalado más arriba, reaparezca: “El problema del conocimiento de sentido común, que ha bloqueado el progreso de las técnicas de representación simbólica durante quince años, puede estar asomando en el horizonte de las redes neuronales, aunque los investigadores todavía no lo reconozcan. [...] En situaciones de la vida real buena parte de la inteligencia humana consiste en generalizar modos que son apropiados a un contexto. Si el diseñador restringe la red a una clase predefinida de respuestas apropiadas, la red estará exhibiendo la inteligencia construida dentro de ella por el diseñador para ese contexto, pero no tendrá el sentido común que la habilitará para adaptarse a otros contextos, como lo haría una inteligencia genuinamente humana”<sup>19</sup>. De modo que la única salida es la mera yuxtaposición de redes orientadas cada una de ellas a un tipo de entradas característico: los módulos de Fodor reaparecen en forma de redes y no de factores. Pero el sujeto de conocimiento es, en su enfrentamiento al mundo, una unidad y si ha de enviar ciertas señales a un módulo y otras a otro, etc. se está dando por supuesto que en el nivel de entrada se dan unos mecanismos de discriminación. Pero esos mecanismos, precisamente, sólo son posibles si ese sujeto (en nuestro caso, la red) tiene una pre-comprensión del contexto. Pero, precisamente, lo que se está intentando solucionar con estas hipótesis de la modularidad es el problema de la incapacidad de adaptación y reconocimiento de contextos. Se da de nuevo un círculo vicioso de mano del modularismo, estilo Fodor, círculo que ya aparecía como problema en la IA<sup>20</sup>. Claro está, la pre-comprensión del contexto la da el sujeto operatorio de carne y hueso –el instructor externo– que supervisa y dispone los elementos de la red.

Resumiendo, la tesis filosófica principal de este artículo tiene una forma lógica muy sencilla, basada en el principio de *tercio excluso*: los modelos con instructor funcionan pero son *formalmente* inadecuados (a no ser que demos por hecho la

<sup>18</sup> Graubard, *op. cit.*, p. 48.

<sup>19</sup> Graubard, *op. cit.*, pp. 51-52.

<sup>20</sup> Robles, *op. cit.*, pp. 168-171.

intervención de la armonía preestablecida o de la Divina providencia en un plano fisicalista); los modelos sin instructor son incapaces de dar cuenta de la flexibilidad de las redes neuronales naturales y los que más se acercan – los modelos que usan códigos demográficos– requieren de retropropagación y esto los hace *materialmente* improbables en virtud del factor tiempo.

No puede reconstruirse el plano fenoménico desde el análisis de lo que, supuestamente, podría ocurrir en el SNC a la luz de las investigaciones en redes neuronales artificiales. Una totalidad significativa –holista o gestáltica, si se quieren usar los términos de los Dreyfus– no puede descomponerse en elementos simples (*inputs*), y no sólo por el problema, anteriormente expuesto, del círculo vicioso que supone la necesidad de requerir de un criterio previo a la descomposición para seleccionar los elementos relevantes sino porque un elemento es, por definición, una forma de presencia mientras que la percepción-conducta es el modo de enfrentamiento de un cuerpo vivo y vivido (*Leib*) a un mundo que aparece como simultaneidad de presencia y ausencia; el contacto con la realidad es circunstanciado y pide la acción, nunca ofrece una imagen expuesta, acabada y analíticamente diferenciada de todas sus partes o elementos –salvo en condiciones especiales y mediante artefactos e ingenios contruidos y controlados, no obstante, de forma operatoria. Es descabellado considerar que hay una correspondencia uno a uno entre fenómenos y nómenos, como señala también Putnam.

El genio de Kant jamás llevó al sabio alemán a semejante afirmación (correspondencia uno a uno) pero la forma en que transforma la Metafísica en Doctrina de las costumbres se asienta en una concepción negativa del cuerpo, en tanto que fuente de impulsos y deseos muchas veces contrarios a la moral y otras veces, cuando le son favorables, propicios a la moral pero sin capacidad de otorgar valor ninguno a la acción (recuérdese, el cuerpo nos puede llevar a actuar *conforme al deber* pero jamás fundamentará por sí solo una conducta *por deber*) De este modo Kant reduce, a nuestro juicio, cualquier conocimiento del propio cuerpo a la representación (*qua* conjunto de fenómenos categorizados y científicamente enjuiciados) que hacemos del mismo. Hasta tal punto esto es así que el cuerpo es algo externo al individuo y por esto mismo las impresiones que de él proceden cumplen, según Kant, las dos condiciones a priori de la sensibilidad que atribuye exclusivamente a la sensibilidad “externa”, puesto que ocuparían algún lugar en el *espacio* (corporal) y durarían en el *tiempo*. Por contra, dice Kant, aquello que pertenece a la “interioridad” del individuo sólo cumple con la intuición sensible temporal: un enfado, una pena, una alegría, etc., tan sólo durarían en el tiempo, pero serían ilocalizables en el espacio. Pero ¿es esto realmente así? ¿Pueden localizarse todos los males orgánicos en algún lugar concreto de nuestro propio cuerpo tal y como lo vivimos? O, por contra ¿son los “dolores del alma” enteramente ajenos a las manifestaciones somáticas? Evidentemente, no. ¿Dónde están aquí la claridad y distinción cartesianas? Y, lo que

es más grave: ¿en qué se han influido unos descubrimientos psicológicos de unas escuelas en otras? ¿Cómo podría una inteligencia artificial desarrollar, por ejemplo, un síntoma neurótico como los que describió Freud en su obra? ¿No seguimos acaso en el mismo estado de la cuestión que hace ya más de un siglo?

En conclusión, parece que la Teoría de Redes y la Psicología cognitiva se desarrollan al margen de una serie de cuestiones que se dejan a un lado pensando, quizá, que se podrá dar cuenta de ellas cuando el problema base (la relación mente-cerebro o cerebro-mente) se haya resuelto. La cuestión es que estos temas no son momentos ulteriores del proceso explicativo que se pretende llevar a cabo sino que afectan a las condiciones mismas de tal proceso, volando sus cimientos. Del mismo modo que la geometría de Lobachevsky no es un desarrollo posterior (una “cuestión avanzada”) de la geometría de Euclides sino que se construye desde condiciones diferentes, pudiendo “amoldarse” a un Universo que es distinto al que tradicionalmente se había pensado, dar debida cuenta de todo ese *maremagnum* de “fenómenos de conciencia” difícilmente adjudicables a la sensibilidad “externa” o “interna” requiere de una serie de condiciones (renuncia de prejuicios, abandono de modelos científicos previos, etc.) que deberían ser consideradas por la Psicología y que, desde luego, las Neurociencias y Ciencias cognitivas no respetan.

## Bibliografía

- ANDRÉS SÁNCHEZ, Jorge de: *Dos aplicaciones empíricas de las redes neuronales artificiales a la clasificación y la predicción financiera en el mercado español*, en “Revista asturiana de Economía”, número 28, 2003, pp. 61-85.
- BALLESTEROS JIMÉNEZ, S.: *Psicología general. Un enfoque cognitivo*. Madrid, Editorial Universitas, 1996.
- CABALLERO DE LA TORRE, Vicente: “Operarios y mercancías incorpóreas”, en *Filosofía y cuerpo. Debates en torno al pensamiento de Gustavo Bueno*. Madrid, Ediciones Libertarias, 2005.
- CABALLERO DE LA TORRE, Vicente: “Redes, lógicas no clásicas y neuronas. De los límites de la matematización más allá de la física”, en *Paideia*, publicación periódica de la Sociedad Española de Profesores de Filosofía, 2ª época, año XXVI, 2005, enero-marzo.
- CARPINTERO, Helio: *Historia de la Psicología en España*. Madrid, Eudema, 1994.
- CHRISTOFF, Daniel: *Husserl o el retorno de las cosas*. Madrid, Edad, 1979.
- COASE, Ronald: *La empresa, el mercado y la ley*. Madrid, Alianza, 1994.
- DAMASIO, A. R.: *El error de Descartes*. Barcelona, Crítica, 1996.
- DREYFUS, H. L. y Dreyfus S. E.: “Fabricar una mente versus modelar el cerebro: la inteligencia artificial se divide de nuevo”, en Graubard, S. R. (comp.): *El nuevo debate sobre la Inteligencia Artificial*. Barcelona, Gedisa, 1993.

- ECHEVERRÍA, J.: *Cosmopolitas domésticos*, Barcelona, Anagrama, 1995.
- ECHEVERRÍA, J.: *Los señores del aire. Telépolis y el tercer entorno*. Barcelona, Destino, 1999.
- ELSTER, J.: *Tuercas y tornillos. Una introducción a los conceptos básicos de las ciencias sociales*. Barcelona, Gedisa, 1996.
- FODOR, J.: *La modularidad de la mente*. Madrid, Morata, 1986.
- GRAUBARD, S. R.: *El nuevo debate sobre la Inteligencia Artificial*. Barcelona, Gedisa, 1993.
- HEIDEGGER, Martin: *Kant y el problema de la Metafísica*. México, D.F.: FCE, 1986.
- HEIDEGGER, Martin: *La pregunta por la cosa*. Alfa Argentina S.A, 1975.
- HEIDEGGER, Martin: *Sein und Zeit*. Tübingen: Max Niemeyer Verlag, 2001.
- HINTON, G. y Sejnowsky, T. J.: *Unsupervised Learning: foundations of neural computation*. MIT, 1999.
- IBÁÑEZ, Jesús: (coordinador): *Nuevos avances en investigación social II*. Barcelona, Proyecto A Ediciones, 1998.
- MARTÍNEZ RODRÍGUEZ, F.: *Estudio de las relaciones conciencia-corporalidad desde Merleau-Ponty*. Madrid: Universidad Complutense de Madrid. Colección Tesis Doctorales, n ° 349/91, 1991.
- MERLEAU-PONTY, M.: *Fenomenología de la Percepción*. Barcelona, Península, 1975.
- MOYA, Eugenio: *¿Naturalizar a Kant? Criticismo y modularidad de la mente*. Madrid, Biblioteca Nueva, 2003.
- PASCAL, Blaise: *Obras*. Madrid, Alfaguara, 1981.
- PINILLOS, J. L.: *La mente humana*. Madrid, Temas de Hoy, 1991.
- POPPER, K. Y ECCLES, J.: *El yo y su cerebro*. Barcelona, Labor, 1985.
- POZO, J. I.: *Teorías cognitivas del aprendizaje*. Madrid, Morata, 1994.
- ROBLES RODRÍGUEZ, F. J.: *Para aprehender la Psicología*. Madrid, Siglo XXI de España, 1996.
- ROBLES, F. J. Y CABALLERO V.: “Mentalismo mágico y sociedad telemática”, en *Cuaderno de materiales*, n ° 18, 2003, página 32, [en línea]: <http://filosofia.net/materiales> [consulta: 25/08/05].
- RULLANI, Enzo: «Le capitalisme cognitif: du déjà vu ?» en *Multitudes* n ° 2, mayo, 2000 [en línea]: <http://multitudes.samizdat.net> [consulta: 12/07/05].
- SEARLE, John: *Mentes, cerebros y ciencia*. Madrid, Cátedra, 2001.
- STONE, J. V.: “Learning perceptually salient visual parameters using spatiotemporal smoothness constraints” en G. Hinton y T. J. Sejnowski: *Unsupervised Learning: foundations of neural computation*. MIT, 1999.
- TARDE, Gabriel: *Psychologie économique*. París, Félix Alcan éditeur, 1902.
- VAHINGER, Hans: *The philosophy of ‘as if’. A system of the Theoretical, Practical and Religious Fictions of Mankind*. London: Cox & Wyman Ltd, 1965.

VIRILIO, Paul: *La máquina de visión*. Madrid, Cátedra, 1998.

V.V.A.A.: *Filosofía y cuerpo. Debates en torno al pensamiento de Gustavo Bueno*. Madrid, Ediciones Libertarias, 2005.

WIENER, Norbert: *Cibernética y sociedad*. Buenos Aires, Ed. Sudamericana, 1958.