

Modelos del cerebro, generación de neuronas y neurotecnología: hacia el desarrollo de prótesis cerebrales y máquinas de aprender

Pablo F. Argibay y Victoria I. Weisz

TEORÍA COMPUTACIONAL DE LA MENTE Y CONEXIONISMO

Mientras que la teoría computacional de la mente sostiene que la mente humana o al menos la actividad cerebral puede ser concebida como un sistema de procesamiento de la información y los pensamientos como formas de procesos computacionales (pensar es computar), el conexionismo parte del presupuesto de que los procesos psicológicos son el emergente de la organización de los circuitos neuronales cerebrales. Independientemente de que los fenómenos sean conscientes o no, el aprendizaje en sentido amplio,¹ es la resultante de la organización cerebral en una red de conexiones neuronales. Este tipo de redes neuronales puede ser modelizado matemáticamente a través de técnicas de neurocomputación.

¿QUÉ SON LAS REDES NEURONALES ARTIFICIALES (ANN)?

- Son modelos matemáticos computacionales.
- Están basadas en la organización del cerebro.
- Presentan interconexión de neuronas artificiales.
- El procesamiento es conexionista (paralelismo-emergentismo).
- Varían su arquitectura en función de la información (plasticidad).
- Son modelos estadísticos no lineales.

La figura 1 ilustra el concepto de ANN en forma simple describiendo una neurona artificial; la información “p” llega a la neurona en forma de input; “w” representa el peso de la conexión sináptica con las señales aferentes de otras neuronas; siguen los propios procesos de cómputos que hace la neurona receptora, consistentes en una sumatoria (Σ) de los productos de los inputs “p” por los pesos sinápticos “w”, sobre los que se efectúa una transformación a través de una función matemática “f”. De este modo el output de una neurona artificial es “a” o sea la función $f(wp+b)$, siendo “b” una variable denominada “bias” reguladora de la actividad neural.

La analogía con una neurona biológica queda graficada en la figura 2 donde se observa una red de neuronas. En el recuadro se destaca una sinapsis.

Históricamente dos hitos en el desarrollo de las ANN fueron por un lado el desarrollo de la primera neurona virtual por los científicos Walter Pitts (1923–1969), lógico matemático, y Warren McCulloch (1898–1969), neurofi-

Figura 1. Neurona artificial.

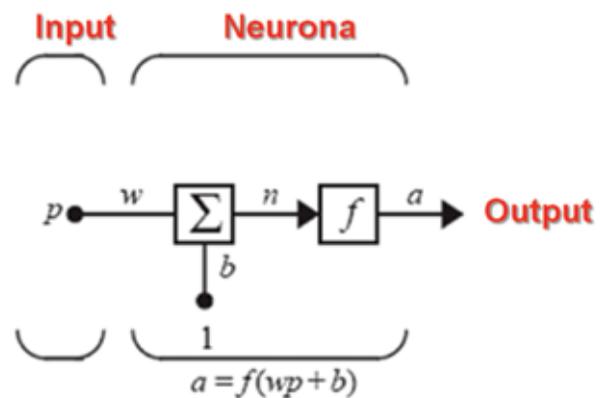
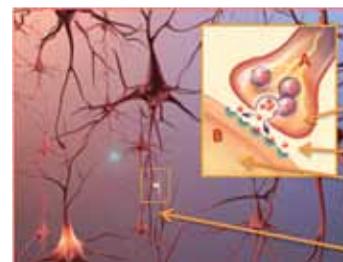


Figura 2. Esquema de las conexiones en un circuito neuronal y su analogía con una ANN (red neuronal artificial).



P= input cuantificado a B

w= fuerza de la sinapsis

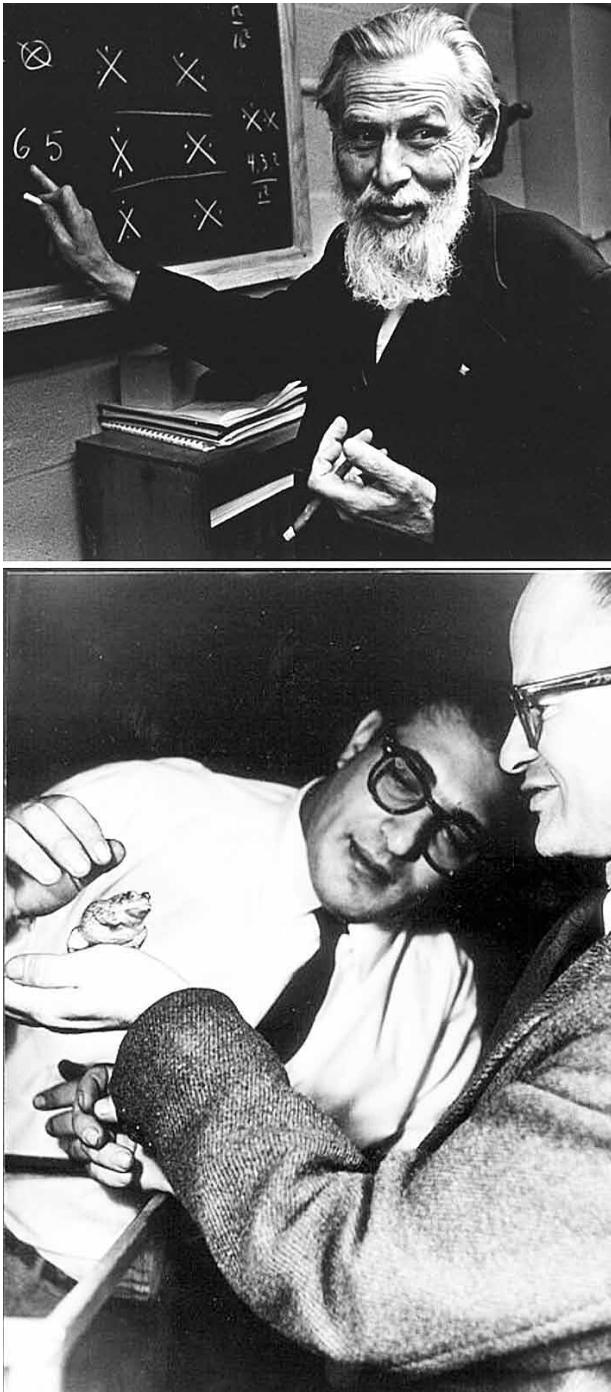
Σ = computación

a= output

1. A los fines de este texto, cualquier cosa aprende cuando cambia su estructura de una manera tal que le genera mejor *performance* para el desempeño futuro en una tarea dada.

siólogo (Fig. 3). McCulloch y Pitts escribieron un trabajo fundacional en el terreno de la inteligencia artificial y las redes neuronales, "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943). En este trabajo propusieron el primer modelo de red neural basado en una unidad neuronal artificial, la neurona de McCulloch y Pitts.

Figura 3. Warren McCulloch (arriba) y Walter Pitts (abajo, a la derecha de Jerry Lettvin).



Otro hito fue el desarrollo en 1957 de una red neuronal, el perceptrón, por Frank Rosenblatt. Esta red es capaz de efectuar tareas de clasificación binaria. Si bien mostró rápidamente sus limitaciones, subsiguientes modificaciones dieron origen a los perceptrones de múltiples capas que se utilizan actualmente en numerosas tareas de clasificación.

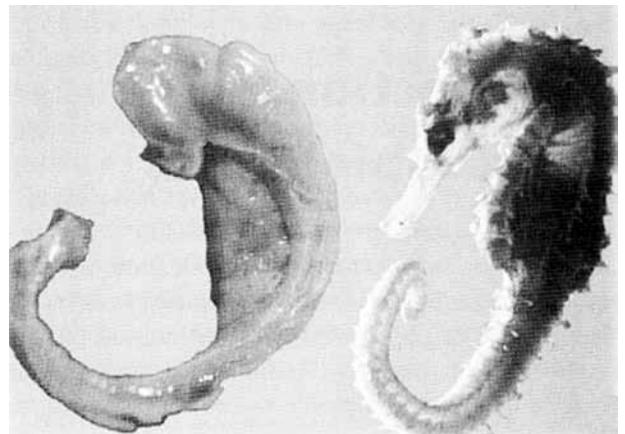
APRENDER ES MODIFICAR LA ARQUITECTURA DE UN CIRCUITO NEURONAL

En 1949 el psicólogo canadiense Donald O. Hebb (1904-1985) publicó un libro tan intuitivo como genial y que modificaría la concepción acerca del aprendizaje. En *The organization of Behavior*, Hebb escribiría: "La idea general es antigua: si dos células o sistemas de células son repetidamente activadas al mismo tiempo, tenderán a estar asociadas de tal manera que la actividad de una facilitará la activación de la otra". Este tipo de modificación estructural en un circuito de neuronas que constituiría un sistema elemental de aprendizaje se denomina aprendizaje hebbiano y es la base de varias reglas de aprendizaje de las redes neuronales artificiales, a la vez que parece ser la base de un tipo de aprendizaje en los sistemas biológicos. En síntesis, para Hebb, los cambios sinápticos son la base del aprendizaje y la memoria. Existe hoy en día bastante evidencia electrofisiológica y molecular para avalar si no totalmente al menos conceptualmente el aprendizaje hebbiano.

EL HIPOCAMPO, UN CIRCUITO CONSERVADO

En el lóbulo temporal del cerebro se esconde una estructura evolutivamente conservada a tal punto de mantener una arquitectura relativamente simple en reptiles, aves y mamíferos. Esta estructura, el hipocampo, denominada así por los anatomistas antiguos a causa de su similitud con el caballito de mar (Fig. 4), está organizada en tres estaciones neuronales: el giro dentado (GD), que recibe

Figura 4. Hipocampo humano y caballito de mar.



sus aferencias de la corteza entorrinal, y las áreas CA3 y CA1 (Fig. 5).

EN EL HIPOCAMPO SE GENERAN NUEVAS NEURONAS

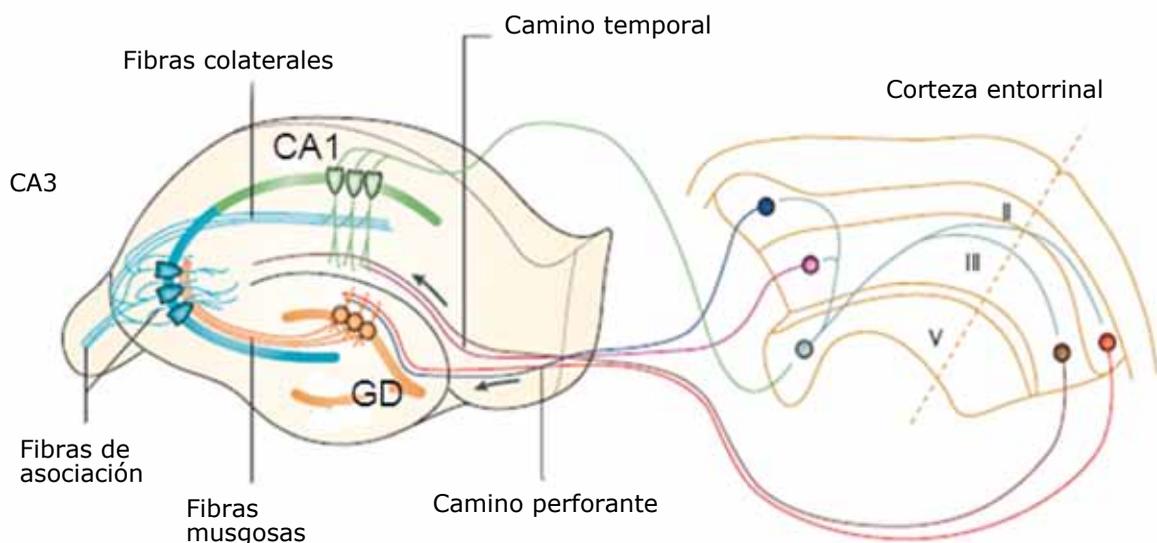
La neurogénesis del adulto es la capacidad que tienen algunas áreas cerebrales de generar nuevas neuronas a lo largo de la vida de un individuo. Una de estas áreas es el hipocampo, donde en la capa subgranular del giro dentado se producen nuevas neuronas que se integrarán a través de sus conexiones con los circuitos existentes. Esta generación, supervivencia e integración de nuevas neuronas dependerá de diversos factores como la edad, el ejercicio físico, el estrés, las hormonas y otros factores. Dada la relación que tiene el hipocampo con la memoria de los individuos se ha asociado la neurogénesis con estos procesos de memoria; sin embargo, aislar el proceso de neurogénesis de otros procesos biológicos hace difícil “disecar” el problema y entender experimentalmente la neurogénesis. Por otra parte, la generación de nuevas neuronas y su integración a los circuitos existentes se relaciona con un problema de cualquier circuito que deba almacenar información en forma estable, pero que a su vez deba aprender a través de modificaciones en su estructura. El denominado dilema “estabilidad-plasticidad”, tiene que ver con ese problema. Imaginemos una computadora que necesite guardar en forma estable su información, pero que, como

consecuencia de diversos procesos, “genere” nuevas conexiones y nuevos componentes. Se encontrará con que debe incorporar nuevo conocimiento y modificar su estructura; todo sin alterar lo aprendido, sin una capacidad de espacios de memoria ilimitados y con pocos fallos.

LOS MODELOS DEL CEREBRO

Una forma de simplificar un problema harto complejo es la creación de un modelo; la física teórica se ha nutrido de la modelización con un éxito extraordinario desde los tiempos del mismo Galileo. Un modelo es “armar” el sistema que se va a modelizar con algunos componentes seleccionados y presuponiendo que funciona de determinada manera. ¿Qué pasaría si el sistema X se copia como x con menos componentes y con el cumplimiento de ciertas reglas de funcionamiento, presupuestas de antemano sobre la base de intuiciones, mediciones y evidencia experimental previa? Ya vimos que las redes neuronales artificiales se basan de alguna manera en el cerebro biológico y funcionan a partir de algunas reglas matemáticas simples. Por otra parte tenemos una red neuronal biológica, el hipocampo, de la cual conocemos arquitectura (número de neuronas y conexiones), mediciones electrofisiológicas, algunos experimentos con animales y la inferencia hecha a partir de las lesiones que presentan algunos seres humanos como consecuencia de infartos cerebrales, tumores y enfermedades degenerativas. Desconocemos muchas de

Figura 5. Esquema de hipocampo de rata donde se esquematizan las estaciones neuronales del giro dentado (GD) y las áreas CA3 y CA1. El GD recibe sus aferencias de la corteza cerebral entorrinal a través de una serie de haces denominados camino perforante. Por otra parte, los axones del GD se conectan con el área CA3 a través de las denominadas fibras musgosas. CA3 es un área rica en conexiones recurrentes sobre sí misma. Las fibras de CA3 se conectan con el área CA1 a través de las denominadas fibras colaterales de Schaffer. Las neuronas de CA1 reciben a su vez información de la corteza entorrinal a través del camino tampo-ammónico, a ella les reenvían sus proyecciones y queda de esta manera cerrado el circuito. El circuito presentado está sobresimplificado pero sirve para entender conceptualmente el circuito hipocámpal (Tomado y modificado de Neves G, Cooke SF, Bliss TV. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nat Rev Neurosci. 2008;9(1):65-75.).



las funciones específicas del hipocampo y si en verdad estas funciones dependen de procesos de cómputo realizados en las neuronas de su circuito. Asumimos algunas funciones, simplificamos la arquitectura, hipotetizamos algunos procesos computacionales de las tres estaciones neuronales (GD, CA3 y CA1) y cargamos nuestro modelo con información. Tenemos, pues, un sistema cognitivo artificial (Fig. 6), copia del hipocampo biológico y al que pondremos en marcha para que luego de un período determinado de funcionamiento nos permita observar qué pasó en términos de la conservación y recuperación de la información cargada en su momento.

La debilidad del modelo es obviamente su simplificación basada en información y presupuestos del diseñador; la potencia radica en que las reglas de funcionamiento son matemáticas y por ende reproducible por otros medios y con capacidad de deducción de otras estructuras matemáticas. En el primer sentido, la reproducción en un medio físico que “entienda” de ecuaciones nos lleva al mundo de lo que se han denominado sistemas neuromórficos o simulaciones de cerebro *in silico* (Fig. 7). La idea es pasar el modelo computacional, (propiedades electrofisiológicas de las células en estudio y su conectividad), a un “chip”.³

Figura 6. Tres formas clásicas de investigar el funcionamiento del cerebro y los procesos cognitivos.²

Experimento, clínica, teoría y simulación computacional son los elementos para la construcción de un modelo cognitivo. A través del experimento, descubrimos la organización del sistema en estudio (molecular, sinapsis, neuronas, módulos, áreas, conexiones). La clínica nos indica cuáles son las consecuencias de la destrucción de determinada área cerebral y las técnicas de diagnóstico por imágenes relacionan la ejecución de tareas con la activación de las áreas cerebrales. La construcción de teorías nos permite explicar a través de abstracciones y, en el mejor de los casos, predecir y deducir nuevas funciones. La simulación computacional nos permite relacionar matemáticamente el funcionamiento de partes seleccionadas del sistema, agregar situaciones y observar cómo se comporta el sistema. El ciclo continúa en el diseño de nuevos experimentos, la construcción de teorías y las nuevas simulaciones.

EL HIPOCAMPO COMO ANN

En un intento de entender la función que la generación de nuevas neuronas podría tener sobre la función del hipocampo, básicamente sobre el almacenamiento y la recuperación de lo que se ha denominado memoria episódica (el tipo de memoria que representa situaciones vividas en términos de cuándo, dónde, qué y cómo), probamos implementar una red neuronal (Fig. 8) que simulara el hipocampo y la ventaja que, en términos de memoria episódica, podría tener ir generando neuronas a lo largo de la “vida” de la red neuronal.

El modelo de hipocampo fue entrenado⁴ con 200 episodios codificados binariamente y se probó en 20 ensayos diferentes. A lo largo de la vida de la red neuronal se incrementó el número de neuronas en el giro dentado en un 30%. La forma de evaluar el modelo consistió en presentarle, luego de su entrenamiento, episodios incompletos (todos conocemos que a partir de una información pequeña, un olor, una melodía reconstruimos un episodio completo de nuestra vida) y evaluar la eficacia de recuperación en términos de relación entre correlaciones de entrada y salida de la red.

Figura 7. Sistemas neuromórficos.

A fines de la década de 1980, Carver Mead desarrolló el Caltech, la primera retina de silicio, dando así inicio a los trabajos con los chips neuromórficos.

Sobre la base de lo que ocurría con el desempeño de las computadoras, que duplicaban su eficacia cada dieciocho meses (según la ley de Moore), en 1990 Mead predijo acertadamente que las computadoras de aquel momento utilizarían diez millones de veces más energía (por unidad de instrucción) comparadas con el cerebro (por activación sináptica). Se propuso achicar esta brecha mediante la construcción de circuitos microelectrónicos basados en el funcionamiento del cerebro.

Mead logró emular el flujo de iones transmembrana de las neuronas mediante el flujo de electrones a través de canales en los transistores. Esta observación no es una sorpresa, ya que en ambos casos activa las mismas fuerzas físicas.

En nuestro caso, hicimos que esos chips fueran programables al introducir un análogo de canales de iones junto con conexiones sinápticas reconfigurables.

2. El concepto de cognición (del latín: cognoscere, "conocer") hace referencia a la facultad de los seres de procesar información a partir de la percepción, el conocimiento adquirido y características subjetivas que permiten valorar y considerar ciertos aspectos en detrimento de otros.

3. El término "chip" se refiere a un circuito integrado en una pastilla pequeña de silicio. Los chips son la base de la moderna electrónica.

4. El entrenamiento se refiere al aprendizaje que la red hace sobre la base de la presentación de ejemplos. La red ajusta sus pesos sinápticos hasta hacer eficaz el aprendizaje. En el modelo presentado, el entrenamiento se refiere a la presentación de los diferentes contextos que van a ser almacenados como memorias.

5. Por correlación se entiende en estadística la relación que existe entre dos variables. En el modelo presentado, cada correlación de entrada o de salida indica qué grado de similitud existe entre el patrón de memoria original y lo presentado (correlación de entrada) o lo rememorado (correlación de salida). Las correlaciones se miden entre 0 y 1; existe correlación perfecta en el caso del 1; cada vez que una variable aumenta o disminuye, la otra la acompaña linealmente. En el caso de una correlación 0, no existe relación demostrable entre las variables.

Se comparó entonces la disimilitud del ingreso (correlación entre episodio incompleto-episodio vivido) con la similitud del egreso (correlación entre episodio recuperado-episodio

vivido). La figura 9 ilustra el concepto de eficacia de la recuperación. La figura 10 ilustra la eficacia del modelo de hipocampo; como se puede observar, aun con una pobre

Figura 8. Simulación del hipocampo a través de una ANN. Se observan los módulos modelizados con la escala correspondiente (número de neuronas), las conexiones y la cantidad de sinapsis que hacen input en cada neurona receptora. La corteza entorrinal emite proyecciones hacia el hipocampo portadoras de la información de todas las cortezas sensoriales, a la vez que recibe la información "construida" en CA1 para el restablecimiento de la información original durante la evocación; el giro dentado remueve las redundancias en la información recibida, es decir, procesa como diferente la información que se parece; CA3 almacena provisoriamente a la vez que construye un episodio por recordar; CA1 recodifica las memorias almacenadas en CA3 y las asocia con la información original en la corteza entorrinal (Weisz VI, Argibay PF. A putative role for neurogenesis in neuro-computational terms: inferences from a hippocampal model. Cognition. 2009;112(2):229-40.).

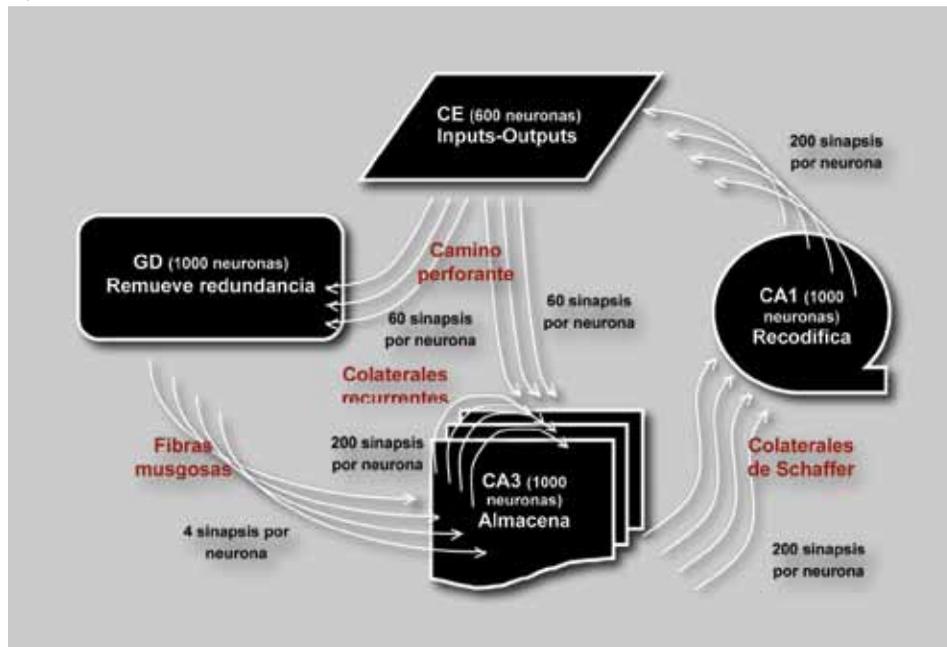
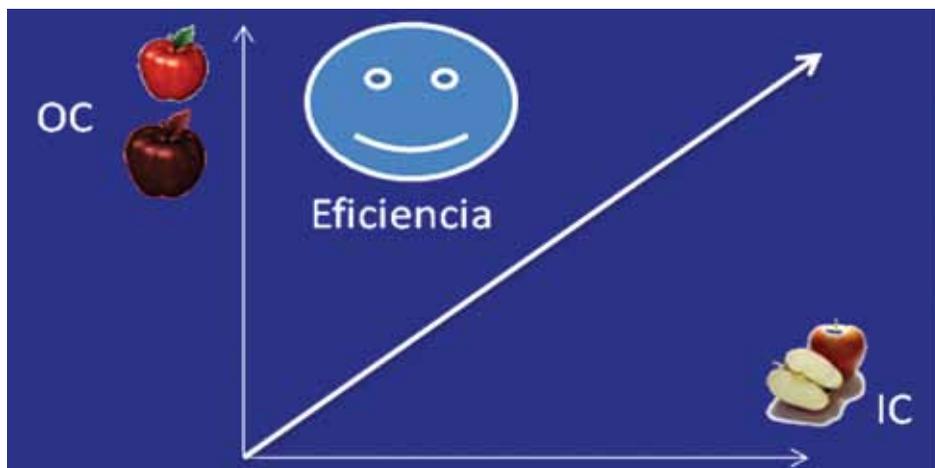


Figura 9. La eficacia de recuperación de la información introducida en el modelo de hipocampo se evaluó relacionando dos correlaciones. La correlación de entrada (IC) representa el grado de similitud entre el episodio incompleto que se presentó a la red luego de su entrenamiento (manzana partida) y el episodio completo almacenado (manzana completa). La correlación de salida (OC) representa cuánto se parece el episodio recuperado al realmente aprendido. Para hablar de eficiencia de recuperación, la correlación de salida debe ser mayor a la correlación de entrada; la relación entre ellas es una medida de cuánto se recupera realmente del episodio original. En la figura, el área de la carita indica la región en la cual debe trabajar la red para decir que es eficiente en la recuperación de información.



correlación de entrada (IC), $r = 0.3$, entre la presentación parcial y el patrón original, la correlación de salida (OC) fue aceptable ($r = 0.55$ entre lo recuperado y el original). La figura 11 ilustra el rendimiento de la red en términos de correlación media de salida (*Mean OC*) versus el número de patrones almacenados. Como se puede observar, hasta 100 patrones almacenados, la correlación media de salida está por encima del 67%, valor medio de la correlación de entrada (*Mean IC*), indicando eficacia de recuperación.

Figura 10. Eficacia del hipocampo implementado como ANN.

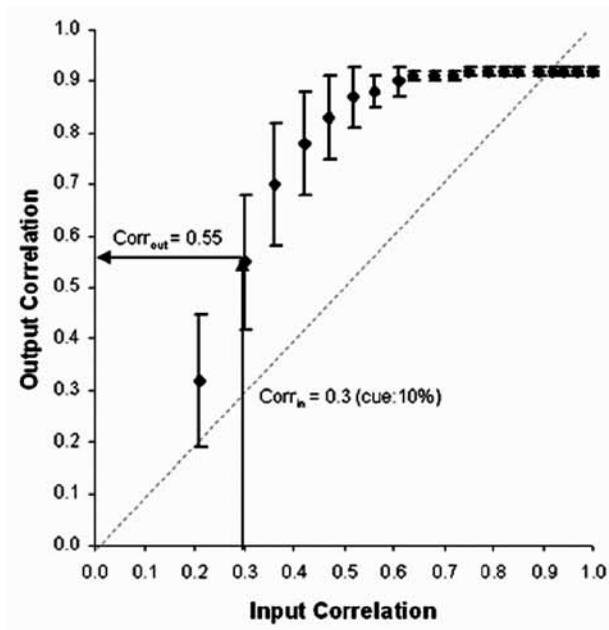
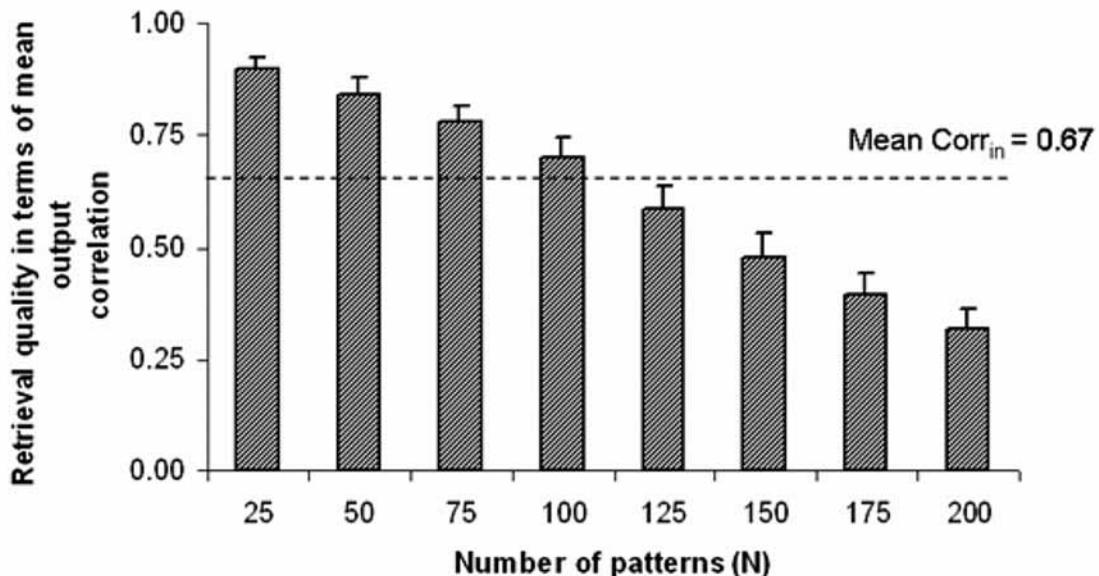


Figura 11. Eficiencia de la red en términos de correlación media de salida versus número de patrones almacenados.



La figura 12 ilustra la influencia de la neurogénesis en términos de recuperación de memoria. Como se puede observar, en A se ve la influencia positiva y estadísticamente significativa sobre la recuperación de las memorias recientes, mientras que en B se observa que la neurogénesis no es eficaz en términos de recuperación de memorias remotas.

HACIA MODELOS MÁS PLAUSIBLES

Si bien los resultados presentados se obtuvieron con una simulación neurocomputacional plausible, esto es, similar al hipocampo biológico, es de destacar que esta plausibilidad se da en términos de arquitectura de la red; es decir, se respetaron las conexiones y se hizo una escala del número de neuronas a valores basados en la literatura acerca de la anatomía hipocampal. Sin embargo, las funciones matemáticas aplicadas son las corrientes en ANN, la conexión entre la corteza entorrinal y el GD se modelizó con una red competitiva, CA3 se implementó con una red asociativa y CA3-CA1 también se simuló con una red competitiva. Este tipo de redes está tendiendo a ser reemplazado por redes de neuronas en las que se simulan las propiedades electrofisiológicas de las neuronas biológicas. Los modelos son más complejos, pero sus resultados se aproximan a lo observado en condiciones experimentales sobre hipocampos in vitro y experimentos animales. En este sentido, el grupo de Fred Gage del Salk Institute en California ha implementado recientemente un modelo de neurogénesis de alta plausibilidad fisiológica observando que en su modelo la neurogénesis actuaría de tres maneras:⁶ las neuronas aún inmaduras realizarían una integración de patrones entre

6. Aimone JB, Wiles J, Gage FH. Computational influence of adult neurogenesis on memory encoding. *Neuron*. 2009;61(2):187-202.

memorias aprendidas al mismo tiempo; la maduración neuronal intervendría en el patrón temporal de las memorias, a la vez que la red global obtenida con las nuevas neuronas integradas en el GD codificarían para nuevas memorias en contextos familiares, tratando diferencialmente los contextos novedosos. De alguna manera nuestros resultados sobre las memorias recuperadas de un patrón temporal reciente y novedoso son compatibles con el modelo de Gage. La posibilidad de modelizar plausiblemente redes neuronales cerebrales abre un camino de resultados impredecibles pero equiparables a los del gigantesco salto realizado por la

física de los últimos 100 años. Sin embargo, la cautela se impone y es justo decir que nuestros modelos son al funcionamiento cerebral, lo que la física de Galileo a las leyes de la naturaleza. El cambio conceptual no es despreciable si tenemos en cuenta que lo logrado a partir de 1604 consistió en una de las mayores rupturas epistemológicas que llega hasta la física de nuestros días y solo fue jaqueada por la física cuántica de principios del siglo XX. La neurociencia del siglo XXI está pariendo a los Galileos y Newtons; con suerte la del siglo XXII engendrará a los Planks, Einsteins y Bohrs de la gran revolución de las ciencias del cerebro.

Figura 12. Influencia de la neurogénesis sobre la recuperación de las memorias almacenadas. A. Memoria reciente (últimas 10 memorias episódicas de 20 ensayos). B. Memoria remota (primeras 10 memorias de 20 ensayos). En síntesis, parecería que, en la simulación computacional presentada, la neurogénesis produciría un aumento en la capacidad hipocampal para rememorar memorias recientes y disminuiría la capacidad de rememorar memorias antiguas.

