Why there are Complementary Learning Systems in the Hypocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memora

James L. McClelland, Bruce L. McNaughton, Randall C. O'Reilly

1995

Amnesia retrógrada temporalmente graduada como resultado de lesiones en el lóbulo temporal. Las lesiones hipocampales parecen afectar la consolidación de la memoria, por lo que si el hipocampo se deja intacto un periodo después del evento, este se consolida y la destrucción no lo afectará más adelante. ¿Es la consolidación un reflejo de una propiedad arbitraria del sistema nervioso? ¿Refleja un principio de diseño importante?

Role of hippocampal system in Learning and Memory

Efectos de daños al hipocampo

- Déficit en aprendizaje nuevo (amnesia anterógrada).
- Efecto selectivo en ciertas formas de aprendizaje. Afectación de memoria declarativa (episódica, semántica). Déficit en la adquisición de asociaciones, además de nueva información fáctica.

  En modelos animales, se ha propuesto que el hipocampo es necesario para responder ante *cue configurations*. Se ha propuesto su importancia para el acceso flexible a huellas de memoria, aunque también se propone su rol en las memorias que involucran localización en el ambiente.
- El aprendizaje no declarativo o implícito parece ser inmune a las lesiones hipocampales. No depende del acceso consciente o deliberado a la memoria (por ejemplo, habilidades motoras). Tampoco se afecta la habilidad para leer la estructura de común a ítems, las tareas de repetición de priming. Tampoco parecen afectarse formas de condicionamiento clásico o instrumental.

- Amnesia para la información adquirida poco tiempo antes de la lesión. Parece haber un gradiente de pérdida de memoria cuanto más atrás se va en el tiempo.

  Una teoría que compite pero ha sido desacreditada es la existencia de un sistema dual de almacenamiento de memoria, uno dependiente de hipocampo y otro independiente, cuyo efecto es aditivo. La ejecución peor de los sujetos en tareas más recientes que en tareas más viejas es evidencia en su contra.

One account of the Organization of Memory in the Brain

Se comienza con la presunción de que el cerebro tiene sistemas de aprendizaje complementarios: uno de ellos se basa en la adaptación de conexiones sinápticas entre las neuronas directamente responsables de procesamiento de información y conducta; el otro, en la adaptación de conexiones sinápticas en un sistema especial de memoria que incluye al hipocampo y estructuras relacionadas.

The Neocortical Processing System

Se asume que las tareas cognitivas y conductuales de alto nivel dependen de patrones de activación en el sistema neocortical. Un patrón representando a un input dispara a otro representando a un output. El sistema debe tener una estructura tal que cualquier parte del contenido del patrón objetivo, y cualquier material asociado con él, pueda servir como clave de recuperación.

Se asume que el conocimiento yace en las conexiones de los circuitos neuronales que realizan las tareas que usan la información. Se asume que en cada ocasión que se procesa la información, se realizan pequeños cambios adaptativos en las conexiones de las neuronas involucradas. Con las repeticiones sostenidas, los cambios sinápticos se acumulan. Estos cambios forman la base de la ejecución correcta en tareas que dependen de ese contenido en específico.

The Hippocampal Memory System

Los cambios descritos son pequeños y no permiten el aprendizaje rápido. Éste último depende en cambios sustanciales en las fuerzas de las conexiones en el sistema hipocampal. La información pasa entre el hipocampo y la neocorteza de forma bidireccional. Se cree que la representación cortical se re-representa en forma comprimida en el hipocampo.

En el hipocampo, el evento se representa en un patrón disperso de activación. Cuando tal patrón aparece en el sistema hipocampal, la memoria potencialmente se vuelve estable.

Cambios en la sinapsis del hipocampo incrementan la probabilidad de que un fragmento del patrón evoque el patrón completo.

Reinstatement and Consolidation of Hippocampal Memories in the Neocortical System

Los patrones almacenados en memoria hipocampal pueden ocurrir en situaciones relevantes a la tarea, y en otras como el ensayo, reminiscencia y sueño. Esta reactivación hipocampal lleva a reactivación neocortical. Esto permite controlar respuestas conductuales y ajustar incrementalmente las conexiones neocorticales. Así, el hipocampo no es solo un almacén, sino un maestro. Lo mismo ocurre con otros tipos de memoria. No hay ninguna distinción especial entre ellos.

Los contextos se almacenen de una manera similar: si una información es presentada repetidamente en el mismo contexto, evocar la información evocará el contexto en el que

se aprendió. Pero al repetirse en ambientes variados, ninguno de ellos será pareado con la memoria.

Evidence and Comment

NEOCORTICAL PROCESSING AND LEARNING. No hay disputa en la idea de que el procesamiento de la información ocurre mediante la propagación de la actividad vía conexiones sinápticas.

Hippocampal involvement in some forms of memory. Las conexiones necesarias para llevar información de y hacia el hipocampo existen. Neuronas de áreas CA1 y CA3 del hipocampo disparan de forma selectiva exhibiendo "place fields". Estos place fields representan conjuntos de claves que definen un lugar en el ambiente. En cierto modo, esas neuronas codifican situaciones y no lugares.

Se requiere de un mecanismo de plasticidad sináptica en el hipocampo. La propuesta de tal mecanismo es la potenciación a largo plazo (LTP). LTP puede ser la manifestación experimental de las modificaciones sinápticas que ocurren en el hipocampo en algunas formas de memoria. Estimulación que satura el LTP en el hipocampo produce déficit profundos en el aprendizaje espacial y amnesia retrograda.

La plasticidad en regiones del sistema hipocampal distintas del hipocampo también se involucra en el aprendizaje. Funciones propuestas son la mediación de la comunicación bidireccional entre hipocampo y neocorteza, retención de la información sobre eventos recientes por períodos breves, o una jerarquía de plasticidad, de modo que el aprendizaje del hipocampo es rápido mientras que el de la corteza es lento, y en las regiones parahipocampales ocurre a una tasa intermedia.

Reinstatement of hippocampal memories. Hay evidencia que indica que durante el sueño se presentan *sharp waves*: breves periodos de actividad que permiten el repaso de la información codificada en las sinapsis. Los patrones guardados en el hipocampo pueden completarse durante las *sharp waves*, con lo que dan la oportunidad de que se repasen en la neocorteza. La activación aleatoria puede llevar al repaso en el hipocampo, y se ha mostrado que neuronas activadas selectivamente durante la vigilia se reactivan durante el sueño.

SUMMARY

El déficit tras lesiones hipocampales para aprender asociaciones arbitrarias que involucran conjuntos de pistas sucede debido a que éstos se encontrarían en el destruido hipocampo. Las habilidades sin dañar vendrían de la acumulación gradual de pequeños cambios en poblaciones neuronales relevantes en la neocorteza. La naturaleza temporalmente graduada de la amnesia retrógrada refleja el hecho de que la información inicialmente almacenada en el hipocampo se incorpora a la neocorteza de manera gradual, dado que los cambios en cada repaso son pequeños.

Esta caracterización incluye presunciones, por lo que es en el mejor de los casos un punto de partida para una teoría de la memoria.

KEY QUESTIONS ABOUT THE ORGANIZATION OF MEMORY IN THE BRAIN

¿Por qué el sistema se organiza así? ¿Por qué necesitamos sistema hipocampal si la ejecución depende de cambios en conexiones neocorticales? ¿Por qué los cambios no se hacen ahí en primer lugar? ¿Por qué la incorporación del nuevo material a la neocorteza toma tanto tiempo? ¿Por qué los cambios en las conexiones neocorticales no es más rápido,

inmediatamente después de su almacenamiento en el sistema hipocampal?

Successes and Failures of Coonnectionist Models of Learning and Memory

Estas pregonas se intentan responder dada la información obtenida de modelos de redes neuronales o conexionistas.

Discovery of Shared Structure through Interleaved Learning

Los modelos de sistemas conexionistas son monolíticos en el sentido de que la información se almacena dentro de la fuerza de las conexiones de los nodos individuales. Éstos tienen la ventaja del *interleaved learning*, que significa que el aprendizaje de un ítem particular no se hace en un solo momento, sino que se adquiere gradualmente a través de las exposiciones repetidas a ejemplos de un dominio. Los ajustes hechos a los pesos de las conexiones son pequeños, de modo que la dirección general del ajuste no es gobernada por las características de asociaciones individuales, sino por las regularidades de la estructura común en el ambiente. Esto permite la formación de representaciones del ambiente que capturan la riqueza del mundo real.

El conocimiento de los conceptos se puede representar en *redes semánticas*. En ellas el conocimiento se organiza de manera jerárquica, lo que permite la generalización a nuevas instancias. Estas redes fueron formas populares de representación en los 70s, pero el soporte experimental que tenía la idea de que así organizan las personas el conocimiento fue aparentemente ilusorio. Es difícil saber cuándo una propiedad debe ser considerada general cuando hay excepciones a ella (e.g., todas las aves vuelan, excepto por los pingüinos, avestruces, kiwis...).

Los modelos conexionistas son distintos: según esta aproximación, la generalización depende de un proceso que le asigna a cada concepto una representación interna que captura su similitud a otros conceptos. Esto parece más consistente con la evidencia psicológica. Esta aproximación depende de que una red pueda aprender las relaciones entre los conceptos por medio del aprendizaje intercalado.

Este tipo de red presenta conceptos organizados en módulos y conectados entre sí mediante relaciones del tipo *isa, has, can, is*. Cuando se activa un nodo particular, por ejemplo, uno que represente a una especie de ave, junto con uno que represente la relación *can*, como resultado la red activará los nodos que indiquen aquellas cosas que esa ave puede hacer.

Antes de aprender, los pesos de la red se aleatorizan. Con las exposiciones a proposiciones del ambiente se aprenden pesos adecuados para minimizar la discrepancia entre el output deseado y el obtenido. Esto se puede lograr por aprendizaje intercalado usando un método de descenso de gradiente: al entrenar, cada patrón se presenta varias veces intercalado con otros patrones. Tras cada presentación se calcula el error, y los pesos de cada conexión se ajustan hacia arriba o abajo en una cantidad proporcional a cuánto se reducirá la discrepancia entre lo correcto y lo obtenido con su ajuste. Los cambios en los pesos se escalan mediante una constante ϵ con un valor bajo para que haya cambios pequeños. Con el tiempo, algunos cambios se acumulan, mientras otros se cancelan entre sí.

Con el entrenamiento la red aprende pesos de input a representación interna (capas ocultas), y de representación a output. Se puede ver que con el tiempo las representaciones

de conceptos similares se hacen cada vez más parecidas entre sí (e.g., la representación de "pino" se hace más parecida a la de "abeto" que a la de cualquier otro concepto). Esto sucede no por una similitud intrínseca en los input, sino por la similitud en las respuestas que la red debe aprender cuando se le presentan los varios conceptos.

Cuando a la red se le presenta un input novedoso que comparte características con uno de los inputs previamente aprendidos, el conocimiento puede generalizarse de modo que el nuevo input elicita un patrón de activación similar al de los inputs previos. Esta habilidad de generalización del conocimientos es crucial en el aprendizaje de organismos biológicos. El orden de adquisición de las distinciones conceptuales, desde los grueso a lo fino, es similar a la forma en que los niños aprenden a distinguir la atribución de los conceptos al mundo. Los modelos conexionistas muestran la habilidad de descubrir relaciones apropiadas entre conceptos y de usarlas de manera apropiada, y esto permite reabrir preguntas con respecto a qué tipo de conocimiento puede adquirirse con la experiencia y qué debe tomarse como innato.

Catastrophic Interference

El aprendizaje intercalado, aunque efectivo en algunas tareas, no es ideal en todas. No es funcional para la adquisición rápida de asociaciones arbitrarias. En tales tareas exhiben un fenómeno llamado interferencia catastrófica. Como ilustración se puede pensar en una red que aprende una tarea de relación de palabras AB seguida de una tarea AC. Al volver a probar la ejecución de AB, ésta ha caído hasta cero, como si su conocimiento hubiese sido sobreescrito.

Se ha intentado sortear este problema haciendo que las representaciones de las redes sean más dispersas y compartan menos los patrones de activación. Sin embargo esto resulta en una gran pérdida en la explotación de la estructura compartida: el conocimiento solo se puede generalizar a otros conceptos relacionados si los patrones que los representan se sobrelapan.

La existencia de la amnesia hipocampal y el papel del sistema hipocampal en el aprendizaje y la memoria sugieren que se puede utilizar la interferencia catastrófica como base para entender por qué tenemos un sistema de aprendizaje separado en el hipocampo y por qué el conocimiento almacenado en él se incorpora en la neocorteza de forma gradual.

Incorporating New Material into a Structured System of Knowledge through Interleved Learning

Al intentar introducir conocimiento en un sistema ya estructurado se puede producir una gran cantidad de interferencia con aspectos de lo que ya se sabe. Esta interferencia se puede reducir si la nueva información se agrega gradualmente, de forma intercalada.

Se piensa en introducir en una red ya establecida conocimiento inconsistente (e.g., los pingüinos son aves, pero no vuelan y sí nadan). Se consideran dos casos: uno llamado *focused learning*, en el que el nuevo conocimiento se presenta al sistema de forma repetida, sin intercalar con el resto de la base de datos; y uno llamado *interleaved learning*, en el que la nueva información se añade al set de entrenamiento, de modo que aun hay exposición continua con toda la base de datos. Con el *focused learning* la red aprende el nuevo material sobre pingüinos mucho más rápido. Sin embargo con esto ocurre un efecto negativo en el desempeño en todos los demás conceptos. La red aprende que los pingüinos pueden nadar pero no volar, y comienza a tratar a todos los demás inputs como si tuviesen las mismas

características.

Con *interleaved training* la incorporación del conocimiento de los pingüinos es gradual. Hay menos progreso dado que hay menos exposiciones (se intercalan los ensayos de pingüinos con los ensayos de los otros inputs), y además el progreso con cada ensayo es menor. Pero esto resulta en muy poca interferencia dado que la red ajusta su representación de otros conceptos similares a la vez que incorpora el concepto de pingüino.

Estos efecto aplican de forma general a todo sistema que ajuste los pesos de sus conexiones mediante la experiencia.

Evidencia a favor de esta interpretación se encuentra en un estudio en el que el *focused learning* en cerebros reales mediante la repetición de habilidades motoras llevó a la pérdida de la diferenciación en regiones relevantes de la corteza sensorial: la práctica produjo una severa reducción en la diversidad de las respuestas de las neuronas en estas regiones. Esto vino acompañado de un síndrome llamado *focal dystonia*, que es la afectación de la coordinación sensoriomotora del miembro afectado. Este síndrome puede corregirse mediante terapia que involucra la práctica intercalada.

Se ve a este proceso de incorporación gradual mediante intercalamiento como un reflejo de lo que sucede dentro de la neocorteza durante la consolidación. Squire, Cohen y Nadel dijeron que "sería simplista sugerir que un cambio biológico simple es responsable por una consolidación que dure varios años, como indican los datos de la amnesia anterógrada. En su lugar, este período está lleno de eventos externos y procesos internos. Estos influyen en el destino de la información por consolidar mediante el remodelamiento de la circuitería neuronal subyacente a la representación original"

Three Principles of Connectionist Learning

- El descubrimiento de un conjunto de pesos de conexiones que captura la estructura de un dominio y coloca hechos específicos dentro de esa estructura ocurre mediante un proceso gradual e intercalado.
- Intentos para aprender nueva información rápidamente en una red que ha aprendido un subset de algún dominio lleva a interferencia catastrófica.
- La incorporación de nuevo material sin interferencia puede ocurrir si éste se incorpora gradualmente, intercalado con exposición a ejemplos del dominio ya aprendidos.

Answers to Key Questions

¿Por qué necesitamos un sistema hipocampal si la ejecución depende de la neocorteza? ¿Por qué no se hacen los cambios directamente en ella?

El papel del hipocampo es proveer un medio para el almacenamiento inicial de memorias en una forma que evita la interferencia con el conocimiento ya adquirido en el sistema neocortical.

¿Por qué toma tanto tiempo la incorporación de nuevo material en la neocorteza?

Para permitir que el nuevo conocimiento se intercale con la exposición continua a ejemplares de la estructura de conocimiento existente de modo que el conocimiento nuevo se incorpore en el sistema ya contenido en la neocorteza. Cambios rápidos llevarían a interferencia.

Generality of the Relation between Discovery of Shared Structure and Gradual, Interleaved Training

La experiencia puede llevar al descubrimiento gradual de estructuras por medio del aprendizaje intercalado pero mediante el *focused learning*. Y este desarrollo gradual es el proceso en el centro del desarrollo cognitivo, lingüístico y perceptual.

¿Qué significa descubrir la estructura en un conjunto de inputs y experiencias? ¿Por qué esto requiere de aprendizaje lento?

What is structure?

La estructura es cualquier relación sistemática que exista dentro de los eventos, la cual, si es descubierta puede servir como base para la representación eficiente de eventos novedosos y respuestas a inputs.

En un modelo de redes se podría decir que la estructura es el conjunto de limitaciones que existen en las realizaciones correctas de las proposiciones dado un concepto y un término de relación. En los modelos conexionistas esas limitaciones se capturan en los pesos y en las relaciones de similitud entre conceptos. Otros tipos de estructura pueden ser las redundancias en los patrones visuales, o las similitud en la pronunciación y escritura de palabras, o la correlación entre pares de elementos en un conjunto de patrones. Esas correlaciones pueden usarse para inferir valores de un estímulo novedoso pero incompleto dado que la correlación indica la presencia de redundancia.

Why Discovering Structure Depends on Slow Learning

La primera razón aplica en procedimientos con las siguientes características:

- El procedimiento se aplica a una secuencia de experiencias muestreadas de un ambiente.
- La meta del aprendizaje es derivar una caracterización parametrizada del ambiente que generó esas muestras, y no almacenar las muestras mismas.
- Lo que se almacena no son ejemplos, sino su caracterización parametrizada.
- El proceso de ajuste consiste en la mejora de la medida de adecuación de la caracterización parametrizada.

Este tipo de procedimientos pueden llamarse estocásticos, on-line, o de actualización de parámetros.

El arreglo de pesos de las conexiones en una red es una caracterización parametrizada o un estadístico, una estimación de los pesos adecuados para el ambiente completo del que se muestrea.

Cuanto más pequeña es la tasa de aprendizaje, más precisa es la estimación del valor poblacional del estadístico que la regla de aprendizaje está estimando. Cuando el valor de la tasa de aprendizaje ϵ es 1, entonces cada nueva experiencia reinicia completamente el peso w de las conexiones para reflejar solamente la experiencia actual. Cuanto más pequeña es la tasa de aprendizaje, w depende más del promedio de experiencias previas y pasadas, por lo que representa mejor al verdadero valor poblacional del estadístico.

Este argumento es independiente del estadístico específico que se estima. Solo depende de que cada experiencia representa una muestra probabilística de la población. La precisión incrementará con el tamaño de muestra.

La segunda razón sobre por qué es necesario el aprendizaje lento aplica en casos con una característica adicional:

- El procedimiento ajusta cada parámetro en proporción al estimado de la derivada de la medida de ejecución con respecto a ese parámetro, dados los valores existentes de todos los parámetros.

Tales procedimientos se pueden llamar de descenso de gradiente. Éstos garantizan mejora, pero solo si ajustes ínfimos se hacen a los pesos en cada paso, dado que cuando cambia el peso de una conexión, puede tener un efecto que altera el efecto de otros pesos en el error. Esto es más severo aun en redes multi-capa. Estas redes necesitan repasar varias veces el conjunto entero de patrones. Con cada repaso, los pesos se ajustan solo un poco, pues de lo contrario los cambios en un peso mellarían los cambios en otro.

Arbitrary Associations, Quasi-Regularity, and Memory for Facts and Experiences

Las redes conexionistas también pueden aprender asociaciones arbitrarias, y en ello se desempeñan mejor mediante aprendizaje intercalado que mediante *sequential learning*, lo que es consistente con los hallazgos en humanos: la práctica espaciada es mejor que la práctica concentrada. Las razones son las mismas: se busca mejorar la ejecución en todos los casos, no solamente en el último.

Se puede evitar la necesidad del aprendizaje intercalado mediante la explotación de representaciones que no se sobrelapan, aunque esto es ineficiente cuando se trata de sistemas grandes de asociaciones arbitrarias. Además, las asociaciones arbitrarias son excepción más que regla en los ambientes naturales. Aunque aspectos arbitrarios suelen coexistir con regularidades, por ejemplo, en las excepciones en la pronunciación de palabras. Estos dominios se llaman quasi-regulares. Se propone que los dominios del conocimiento semántico, episódico y enciclopédico son de este tipo.

Para consolidar los contenidos de un evento parcialmente arbitrario, el sistema neocortical tendrá que encontrar un conjunto de pesos de conexión que se ajusten tanto para lo común como para lo idiosincrático. Los aspectos idiosincráticos se tomarán más tiempo en ser consolidados. La tasa de decaimiento de la huella hipocampal es crucial en estos casos: si el decaimiento es muy rápido, mucho del contenido idiosincrático puede no consolidarse. Esta carrera entre decaimiento hipocampal y aprendizaje intercalado da un mecanismo para lo que se ha descrito como la cualidad esquemática de la memoria a largo plazo: aquello arbitrario e idiosincrático tiende a perderse, mientras lo que es común a varios episodios permanece. Aunque no hay nada que evite que un material arbitrario se consolide aun si se encuentra solo una vez, si acaso se repasa con suficiente frecuencia en el sistema neocortical.

Discovery of Structure in Biologically Realistic Systems

Parte de la estructura en arreglos de inputs se puede extraer con reglas simples, como una regla de covarianza. Una de tales reglas es el patrón de intercorrelaciones entre los varios inputs a una neurona o grupo de neuronas. Esta regla le asigna pesos de conexión positivos a los inputs que están máximamente correlacionados con otros inputs de la misma

neurona. El conjunto de items más relacionado entre sí tiene conexiones positivas fuertes con la unidad receptora. El aprendizaje en este modelo debe ser lento, pues de lo contrario muchas de las correlaciones no serían detectadas y quedarían como ruido. Sin embargo, las reglas locales simples que permiten detectar las correlaciones pueden no siempre ser adecuadas para aprender todos los aspectos de una estructura. Por ejemplo, pueden existir estructuras ocultas en el sentido de que no están presentes como una relación directa entre inputs una vez que son adecuadamente re-representadas.

El mayor avance del aprendizaje conexionista es el descubrimiento de procedimientos, mejores que simples reglas correlacionales, mediante los cuales se pueda aprender a formar estas representaciones. El propósito es hacer disponible para cada conexión información sobre el grado en el que el ajuste de esa conexión particular influirá en la discrepancia ente el output y lo esperado.

Sin embargo, el procedimiento para calcular esta discrepancia parece biológicamente irreal. Las señales de activación se propagan en una dirección, y el proceso de determinar los ajustes necesarios en los pesos adecuados depende de una computación que parece corresponder a una transmisión hacia atrás biológicamente implausible. Por ello, este algoritmo se tolera en neurociencias como una forma de conocer los pesos óptimos de la red, pero no como un posible mecanismo usado por sistemas reales.

Una solución viene de la idea de que el aprendizaje en sistemas multi-capa puede explotar la reciprocidad de las proyecciones axónicas que hay entre las regiones de la neocorteza. Esto parece plausible dado que cuando hay conexiones de A a B, también las hay de B a A.

Una segunda aproximación viene de reemplazar la *backpropagation* de la información con una sola señal difusa de reforzamiento. Al compararse ambos tipos de algoritmos (reforzamiento y *backpropagation*) se ha encontrado que ambos pueden descubrir los mismos tipos de campos receptivos que el cerebro. Aunque este tipo de aprendizaje por refuerzo parece requerir de aprendizaje aun más gradual.

No se conocen con certeza los procedimientos usados por el cerebro para descubrir la estructura de conjuntos de patrones. Solo se pretende presentar la idea de que tales procedimientos deben existir, y algunos de los presentados hasta ahora son biológicamente plausibles.

Combining the Hippocampal and the Neocortical Learning Systems: Consolidation and Retrograde Amnesia

La adquisición de información mediante aprendizaje intercalado es muy lenta, y sería insuficiente para ajustarse a las demandas del ambiente. El argumento es que el hipocampo existe para permitir la retención de contenidos específicos a la vez que se evita la interferencia.

Cuando una memoria se almacena en el hipocampo, puede repasarse en la corteza. Estos repasos tienen dos funciones: el repaso puede controlar respuestas conductuales y ajustar las conexiones neocorticales, permitiendo que las memorias se hagan lentamente independientes del hipocampo.

Una pregunta clave es la fuente de los repasos de ejemplares tomados de la estructura existente de conocimiento. Hay varias posibilidades no excluyentes, incluyendo la reactivación debido al ambiente externo y debida a la reminiscencia, además de la espon-

tánea. Hay evidencia de repaso espontáneo durante el sueño. Quizá estos eventos durante el sueño activan patrones en la neocorteza, con lo que estos se vuelven candidatos a la activación durante sueño REM. Esto permitiría que la información vieja y la nueva se activen intercaladamente.

Modeling Temporally Graded Retrograde Amnesia

Se presentan simulaciones de experimentos con lesiones bilaterales del hipocampo. Se toma al hipocampo como una caja negra que ejecuta las operaciones que se han descrito de él sin entrar en más detalles.

Se asume que el aprendizaje hipocampal depende de la saliencia o importancia del evento, que las huellas de memoria del hipocampo se desvanecen con le tiempo, que la probabilidad de repasos neocorticales mediados por el hipocampo disminuye con la fuerza de la huella, y que la probabilidad de repaso en una cantidad dada de tiempo puede ser distinta en distintas tareas y contextos: más probable en contextos relevantes a la tarea.

Una complicación adicional es que el repaso podría fortalecer la representación hipocampal igual que la neocortical, lo que disminuiría el decaimiento de la huella. La supresión de la plasticidad hipocampal durante algunas fases del sueño indica que quizá esa activación no sea auto-reforzarte.

Si los repasos dado un contexto relevante son auto-reforzantes, pero los espontáneos no, esto proveería un mecanismo mediante el cual las memorias que permanecen relevantes tienden a persistir. Aun así, se ignora el auto-reforzamiento por simplicidad.

Se toma al hipocampo como una fuente de datos de entrenamiento para las redes neocorticales. Se asume que las presentaciones generadas por el hipocampo se intercalan con la exposición concurrente a otras experiencias.

Se simula un experimento de condicionamiento aversivo en el que, siguiendo la asociación de un tono con un choque eléctrico, se realizaron lesiones hipocampales tras 1, 7, 14 o 28 días y se evaluó la respuesta de miedo. Cuanto más tiempo había pasado entre el entrenamiento y la lesión, mayor fue la respuesta de miedo.

En la simulación, tras la exposición experimental a una asociación esta se encuentra disponible solo en el hipocampo. Al introducir un nuevo par, el entrenamiento continua normalmente, y se asume inicialmente que la huella hipocampal no decae.

En otro experimento se encontró evidencia de consolidación en una período de 10 semanas en monos. Los monos se entrenaron en discriminaciones binarias. Recibieron lesiones hipocampales bilaterales en momentos distintos y después fueron probados en ejecución. Aquellos animales con más espacio entre la cirugía y el aprendizaje tuvieron una ejecución mejor.

La red simulada fue entrenada con 100 pares de input-output. Se asumió que después de la cirugía no ocurrió ningún repaso adicional basado en el hipocampo.

A Simplified Quantitative Formulation of the Consolidation Process

Por simplicidad se adopta un modelo de dos compartimentos del proceso de almacenamiento y consolidación de memoria.

La formulación asume que cada evento experimentado se almacena en el hipocampo con una fuerza inicial $S_h(0)$. Esta fuerza inicial va de 0 a 1, y la fuerza en el tiempo t

sigue un decaimiento exponencial desde su valor inicial:

$$\Delta S_h(t) = D_h S_h(t) \tag{1}$$

La fuerza inicial $S_h(0)$ y la tasa de decaimiento D_h puede depender de la tarea y las condiciones de los estímulos.

Cuando el hipocampo está off-line, los repasos que sirven a la consolidación ocurren con probabilidad $\rho(t)$ por unidad de tiempo. Esta probabilidad de repaso depende de la fuerza residual de la huella multiplicada por el parámetro de tasa de repaso r_h:

$$\rho(t) = r_h S_h(t) \tag{2}$$

Se asume que la huella de fuerza neocortical incrementa con cada repaso neocortical de la huella. El incremento es proporcional al parámetro de aprendizaje ϵ multiplicado por la diferencia entre la fuerza de huella neocortical actual y el máximo de fuerza de 1.0. La fuerza de huella neocortical también decae a una tasa D_c (esto puede ser pasivo o resultado de la interferencia del almacenamiento de otras huellas). Tomando la probabilidad de repaso en cuenta, el cambio en la fuerza cortical en cada *timestep* está dado por

$$\Delta S_c(t) = C S_h(t)(1 - S_c(t)) - D_c S_c(t) \tag{3}$$

donde C es la tasa de consolidación igual al producto de ϵ y r_h.

Cuando el sistema es sondeado en un contexto particular, la probabilidad de que la huella hipocampal sea repasada de forma suficiente para resultar en conducta adecuada para la tarea está dada por

$$b_h(t) = R_h S_h(t) \tag{4}$$

En esta ecuación R_h refleja la adecuación de la situación como una clave para la huella de memoria hipocampal. La probabilidad de que la huella cortical consolidada para resultar en conducta adecuada para la tarea se asume como

$$b_c(t) = R_c S_c(t) \tag{5}$$

donde R_c refleja la adecuación de la situación como clave de recuperación para la huella neocortical.

La conducta correcta se puede basar en el sistema hipocampal, si este produce un output, o en la neocorteza. Dad esto, la probabilidad $b_{hc}(t)$ de emitir la conducta correcta baasda en cualquiera de los dos sistemas es

$$b_{hc} = b_h(t) + (1 - b_h(t)) b_c(t) \tag{6}$$

Respuestas correctas también pueden surgir de tendencias preexistentes o azar, si las alternativas son discretas. Esto puede introducirse en la fórmula asumiendo que la respuesta correcta se genera en los sistemas hipocampal/cortical con probabilidad b_{hc}, y que en los ensayos restantes los animales se basan en tendencias preexistentes o el azar,

cuya probabilidad de dar la respuesta correcta será denotada por b_p. La probabilidad total de respuestas correctas entonces es:

$$b_t(t) = b_{hc}(t) + (1 - b_{hc}(t))b_p \tag{7}$$

Aunque esta formulación es minimalista, tiene parámetros libres. Sin embargo, el parámetro R_c puede ajustarse a 1 dada la dificultad de separarlo de los efectos del parámetro C de consolidación. De forma similar, R_h se confunde con ${}_h(9)$ y también se puede ajustar en 1. Si el experimento está bien diseñado, habrá una medición separada de b_p, o si hay opciones discretas b_p será simplemente $\frac{1}{n}$.

Esta ecuación fue ajustada a los datos de los experimentos descritos, y mostró un buen ajuste. También tiene un ajuste moderadamente bueno con dos estudios más:

En uno de ellos se hizo que ratas atestiguaran a otra rata comiendo comida con aroma a canela o a chocolate. Después se les realizó una lesión hipocampal en momentos distintos y se midió su preferencia por las dos comidas. Se encontró que cuanta más separación había entre la demostración y la lesión, mayor preferencia hubo por la comida consumida por la rata ejemplo.

En el otro estudio se probó la amnesia retrógrada en humanos mediante una prueba basada en el recuerdo de hechos sobre televisión. Los sujetos experimentales habían recibido terapia electroconvulsiva, que produce amnesia similar a la de lesiones hipocampales. Los pacientes de terapia electroconvulsiva tuvieron un peor desempeño al recordar detalles de un programa viejo de televisión que los sujetos control (aunque se ofrecen explicaciones alternativas). Sin embargo, en este y otros estudios se muestra evidencia de que en los humanos las diferencias en la memoria debidas al hipocampo pueden extenderse hasta los diez años de duración.

Sources of Variation in Hippocampal Decay and Neocortical Learning Rate

Dos factores influyen en la longitud del intervalo de consolidación y el resultado del proceso: la tasa de decaimiento del sistema hipocampal, u la tasa de incorporación de las huella hipocampales en la neocorteza. Si la tasa de decaimiento del hipocampo es alta, el período de consolidación será breve dado que la información se perderá del hipocampo en poco tiempo. Si la tasa de incorporación es alta, el período de consolidación será igualmente breve debido a que la corteza aprenderá rápidamente. Para que el período de consolidación sea largo ambas tasas deben ser bajas, y la evidencia indica que tienden a variar de forma conjunta.

Entre las fuentes de variación en la escala temporal de la consolidación pueden encontrarse diferencias entre especies y entre tareas, o en la saliencia de la información almacenada.

Diferencias entre especies podrían venir de presiones evolutivas distintas: animales con vidas cortas necesitan de aprendizaje rápido.

Otra posibilidad son las diferencias de edad en el aprendizaje neocortical. Los datos obtenidos con humanos vienen de adultos, mientras que los estudios animales suelen usar jóvenes. Quizá la tasa de consolidación disminuye con la edad. Esta diferencia en edad podría ser adaptativa: tiene sentido que en la juventud la tasa de aprendizaje sea muy alta, y que éste se vaya refinando con el tiempo cuando ya hay estructuras y representaciones establecidas.

Aunque bien puede haber variables distintas de la tasa de aprendizaje que estén en juego.

Amnesia infantil. Que la tasa de aprendizaje sea más alta en la infancia puede ayudar a explicar la amnesia infantil. Se ha aportado evidencia que indica que no se puede explicar simplemente por inmadurez del sistema nervioso. El fenómeno puede deberse al cambio inicialmente rápido en las representaciones usadas en el sistema neocortical. Y las memorias serían más difíciles de interpretar si se repasan, puesto que ya no tendrían sentido en un sistema que ha cambiado mucho.

Aprendizaje rápido fuera del sistema hipocampal. Ciertos tipos de memoria permanecen intactos ante lesiones del hipocampo. Este aprendizaje parece depender de la formación de asociaciones de una clave simple y discreta con un estímulo incondicional o con la disponibilidad de reforzamiento.

Según este análisis, aun ese aprendizaje debería ocurrir de forma gradual para prevenir la interferencia. Sin embargo, parece razonable que la evolución provea de mecanismos que permitan sobreponerse a esta consideración en situaciones en las que la adaptación rápida es necesaria para la supervivencia (como el condicionamiento aversivo al sabor). Una alternativa sería la adquisición de conocimiento inicial de manera rápida, para después refinarlo de manera lenta, de una forma análoga a lo que se propone con el envejecimiento. Con esto se esperaría que el aprendizaje de asociaciones simples no fuese afectado por las lesione hipocampales, en contraste a las asociaciones complejas.

General Discussion

Se ha presentado una explicación de los papeles complementarios del hipocampo y la neocorteza en la memoria, y las propiedades computacionales de modelos de aprendizaje y memoria que dan la base para entender por qué la memoria se podría organizar de este modo. Ahora se compara brevemente esta aproximación con otros puntos de vista sobre el papel del sistema hipocampal en la memoria.

Perspectives on Retrograde Amnesia

Este análisis se basa en buena parte en el fenómeno de la amnesia retrógrada temporalmente graduada. Esto pide que el hipocampo sea un sistema relativamente extendido pero no ilimitado. La idea de que el hipocampo tiene un papel en la consolidación ha sido tratada antes por Marr (1971), y se ha enfatizado en el trabajo de Squire (Squire et al., 1975). Squire propuso que la amnesia es un reflejo de un proceso gradual de reorganización, la cual es extendida en las ideas aquí presentes.

La idea de que en la consolidación el hipocampo reproduce las memorias para la neocorteza también puede venir de Marr, quien propuso que el hipocampo almacena las experiencias durante el día, y las retransmite a la neocorteza en la noche.

Tres roles se han sugerido para el hipocampo:

- Como ayuda para la neocorteza para seleccionar una representación para usar en el momento del almacenamiento.
- Como proveedor de una forma crucial de representación no disponible en la neocorteza, y que es necesaria para la ejecución en ciertas tareas de memoria.

- Como ejecutor de un rol explícitamente limitado en el tiempo en la formación de representaciones neocorticales.

Other Points of Comparison

Algunas investigaciones han propuesto que el hipocampo tiene un papel en el aprendizaje de contingencias que involucran conjunción de claves. La explicación presente es similar en tanto que asume que el sistema hipocampal es necesario para la formación rápida de representaciones conjuntivas, pero difiere en tanto que asume que estas representaciones se forman inicialmente en el sistema hipocampal.

Se argumenta que la perspectiva de que el hipocampo tiene una función de codificación conjuntiva, de aprendizaje especial, y la perspectiva actual no necesariamente son excluyentes. Por ejemplo, si se toma a la ubicación espacial como parte del contexto al que es sensible el hipocampo se pueden reconciliar las ideas.

Explicit and Declarative Memory. Los amnésicos humanos tienen dificultad con la memoria explícita, pero también parecen tener problemas con la adquisición de información arbitraria que no vaya acompañada de evocación consciente, por lo que parece que ambos tipos de memoria se encuentran afectados. Aunque se ha dicho que el déficit se encuentra en memoria declarativa, cuyos *contenidos* se pueden recordar conscientemente, sea que esto se haga finalmente o no.

La perspectiva actual sugiere más bien que la memoria afectada es aquella que depende de la formación rápida de asociaciones novedosas, dado que es la que depende del hipocampo.

Flexible use of memory. Se ha sugerido que el hipocampo se especializa en la representación de memorias recientes de un modo que facilita su uso flexible.

Reference versus working memory. Otra propuesta indica que el hipocampo es necesario para la memoria de trabajo, pero no para la de referencia (memoria para aspectos invariantes de una situación). Esto es similar a la perspectiva actual, que señala que el hipocampo es necesario para el almacenamiento rápido de los contenidos de episodios específicos. Sin embargo, la perspectiva actual indica que se puede aprender rápidamente cualquier tipo de asociación, aunque se trate de un aspecto invariante de una tarea, como una aproximación inicial.

Binding. Se ha sugerido que el sistema hipocampal tiene un mecanismo que ata juntos los diversos aspectos de la representación cortical de un episodio específico. Algunas propuestas indican que el hipocampo no almacena la memoria en sí misma, sino solamente una lista de direcciones a las localizaciones de la neocorteza donde realmente se almacena. Esta visión podría coincidir con la presentada en este artículo si se recuerda que se asume que el hipocampo no guarda una copia exacta de la activación cortical, sino una versión comprimida de ella.

Prediction. La última perspectiva indica que el hipocampo es necesario para predecir el futuro con base en el pasado reciente. La predicción basada en la experiencia reciente se ve comprometida dadas lesiones hipocampales. Sin embargo, se ve a esta predicción como un caso especial del aprendizaje asociativo que ocurre en el hipocampo. La predicción puede surgir desde el almacenamiento asociativo y la recuperación posterior mediante la completación de patrones.

Conclusion

Se ha tratado el fenómeno de la consolidación como un reflejo de la incorporación gradual de nuevo conocimiento en sistemas represantacionales neocorticales. Se han señalado los mecanismos computacionales que indican cómo la incorporación de nuevo conocimiento puede llevar a la estructura a adaptarse. El análisis no está completo, aun así, y muchos detalles de la implementación fisiológica son oscuros. Este análisis se enfoca en las dos preguntas clave planteadas en el artículo, pero ha producido muchas otras. Responderlas dependerá de la síntesis de investigación computacional, conductual y neurofisiológica.