

Aprendizaje causal y recuperación de la información
Perspectivas teóricas

**N. Javier Vila
Juan M. Rosas
(Eds.)**

**Aprendizaje causal y
recuperación de la información**
Perspectivas teóricas



CONACYT

Consejo Nacional de Ciencia y Tecnología



Colección Universitas

© los autores

© del lunar 2005

c/ Cruz de la Magdalena 8

23004 Jaén

Tfs. 696 84 53 58 - 606 30 67 06

© Fotocomposición: ediciones del lunar

Imprime: Gráficas La Paz de Torredonjimeno S. L.

Tf. 953 57 10 87

D. L.: J - 236 - 2005

I. S. B. N.: 84 - 95331 - 29 - 2

www.dellunar.com

Presentación

Este libro nace a partir de un symposium titulado *Extinción y recuperación de la información en aprendizaje causal: perspectivas teóricas* organizado en el marco del XV Congreso de la Sociedad Española de Psicología Comparada, celebrado en Barcelona en septiembre del año 2003 y en el que la mayoría de los grupos hispanos dedicados a este tema presentaron los últimos avances de sus investigaciones y reflexiones teóricas.

Hijo natural de ese symposium, el libro que vas a comenzar a leer recoge las comunicaciones presentadas en aquella reunión mejoradas y actualizadas por sus autores hasta conformar una perspectiva bastante exacta del estado del estudio de los mecanismos y contenidos de la recuperación de la información en aprendizaje causal en el ámbito hispanoamericano.

En el primer capítulo Maldonado, Herrera, Catena, Cándido y Perales, de la Universidad de Granada, analizan los contenidos del aprendizaje causal, revisando y ampliando el modelo de revisión de creencias a partir de los resultados recientes que sugieren que el sujeto almacena tanto el juicio causal como los eventos específicos a los que se expone.

En el segundo capítulo Luque, Cobos y López, de la Universidad de Málaga, analizan el fenómeno de la interferencia entre señales, presentando una explicación del mismo basada en modelos de razonamiento causal que destacan la influencia de los procesos de orden superior en el aprendizaje y la recuperación de la información.

En el tercer capítulo Vadillo y Matute, de la Universidad de Deusto, realizan una revisión teórica de los modelos asociativos e inferenciales de aprendizaje causal aplicados particularmente a la extinción y a la recuperación de la información tras la extinción. Estos autores concluyen destacando la necesidad de ampliar el marco teórico donde se desarrollan estos estudios, tomando en cuenta el papel de los procesos de razonamiento en la recuperación de la información después de la extinción.

En el cuarto capítulo Rosas, García, Abad y Callejas, de la Universidad de Jaén, analizan el papel del contexto en la recuperación de la información tras la interferencia. Tras proponer una definición operacional de contexto, analizan los factores que hacen que la recuperación sea dependiente del contexto, concluyendo que el factor fundamental es la atención que los sujetos presten al contexto durante el aprendizaje, atención que viene regulada por la ambigüedad de la información presentada.

Finalmente, en el quinto capítulo Vila y Alvarado, de la Universidad Nacional Autónoma de México, analizan las diferencias en la interferencia retroactiva entre animales humanos y no humanos, destacando sus diferencias cualitativas y la necesidad de tomar en cuenta el lenguaje para poder explicar la recuperación de la información en aprendizaje causal humano.

En conjunto, el libro recoge los avances teóricos recientes más importantes en el estudio de la recuperación del aprendizaje causal dentro del ámbito hispano, quizá el ámbito

donde estos avances han sido más espectaculares y donde más consolidada se encuentra la investigación en estos aspectos del aprendizaje. El lector curioso y el experto interesado encontrarán una revisión actualizada y concisa de los principales problemas a los que se enfrenta hoy el estudio del aprendizaje causal y la recuperación de la información, así como una serie de guías sobre la dirección que han de tomar sus soluciones.

Queremos agradecer al Consejo Nacional de Ciencia y Tecnología mexicano (CONACYT) por su apoyo financiero para la realización de este trabajo a través del proyecto 40345. Iguamente, agradecemos a todos los autores su disponibilidad para la participación en esta aventura que no hubiera sido posible sin ellos. Confiamos en que todos se encuentren tan satisfechos del resultado como lo estamos nosotros.

Juan M. Rosas y Javier Vila

Procesamiento de la información en el aprendizaje causal: ¿qué se aprende?

Antonio Maldonado, Amparo Herrera, Andrés Catena, Antonio Cándido
y José César Perales
Universidad de Granada, España

El aprendizaje causal se ha convertido en uno de los temas de estudio fundamentales de la psicología del aprendizaje actual, desde una doble perspectiva: en primer lugar, porque ha puesto de manifiesto nuevos fenómenos, muchos de ellos específicamente humanos; pero sobre todo, porque ha permitido el desarrollo de nuevos modelos explicativos ampliando sus límites de aplicación. Todo ello hace de la psicología del aprendizaje actual un área científica más interesante y explicativa del comportamiento humano.

En el fenómeno de aprendizaje causal se demuestra que aprender que una “causa produce un efecto” requiere un proceso previo de detección de la relación de covariación existente entre ambos eventos para poder realizar posteriormente una atribución causal en forma de juicio o creencia. La investigación reciente sugiere que en esta dualidad es necesario considerar la participación de varios mecanismos en el proceso de inferencia causal, así como la implicación de la memoria para almacenar los contenidos del aprendizaje resultante, que en el caso del aprendizaje causal serían creencias causales sobre nuestro entorno. Esta multiplicidad de mecanismos permite entender que se hayan sugerido dos tipos de modelos explicativos, incluso en el aprendizaje animal, que creemos que resultaría más conveniente considerar como complementarios en lugar de excluyentes.

La mayoría de los modelos se han centrado en los mecanismos que permiten explicar el proceso de adquisición, normalmente un mecanismo asociativo como ocurre en el modelo de Rescorla y Wagner (1972), modulado según otros modelos por el mecanismo atencional (Pearce y Hall, 1980). Modelos más recientes (como el de Bouton, véase Rosas, Vila, Lugo y López, 2001, para su aplicación al aprendizaje causal) enfatizaron la necesidad de atender a los contenidos del aprendizaje en la memoria, para poder explicar fenómenos más relacionados con la extinción o la expresión de lo aprendido, por ejemplo como resultado del cambio de contextos.

Centrándonos en el aprendizaje causal, y dado que en nuestro ambiente la causalidad casi siempre es probabilística, una medida que permite estimar objetivamente el grado de relación entre una causa y un efecto es la conocida *regla ΔP* que establece el grado de relación en función de la probabilidad del efecto dada la presencia o ausencia de la causa:

$$(1) \quad \Delta P = P(E|C) - P(E|\tilde{C}) = \frac{a}{a+b} - \frac{c}{c+d}$$

El supuesto básico es que el individuo calcula la relación causal a partir de las frecuencias de cada tipo de suceso posible entre una causa y un efecto. Así, es posible que ambos aparezcan juntos (ensayos *tipo a*), que sólo aparezca la causa o el efecto (ensayos *tipo b*)

o *tipo c*) o que no ocurra ninguno (*tipo d*). El hecho de que (1) se haya demostrado repetidamente que los humanos tenemos en cuenta todos los sucesos que nos ocurren para estimar el grado de relación entre causas y efectos, y no cualquiera de ellos aisladamente, y (2) sobre todo, que esta regla permite predecir de manera bastante precisa los juicios causales emitidos en múltiples condiciones experimentales, condujo a proponer modelos inferenciales, en los que se sugiere que las personas actúan como estadísticos intuitivos que detectan y analizan los sucesos que les ocurren en su propio ambiente para calcular la contingencia entre dichos sucesos y estimar el grado de relación entre ellos (Cheng y Novick, 1992; Perales, Catena y Maldonado, 2002). Es importante señalar que en este tipo de modelos el resultado del mecanismo estadístico se asume como el propio contenido del aprendizaje ya que el juicio sería el cálculo realizado.

Ahora bien, el interés en el fenómeno de aprendizaje causal se derivó principalmente de investigaciones relacionadas con el aprendizaje animal (véase Dickinson, 2001; Miller y Matute, 1996). La investigación de aprendizaje animal sirvió de guía para la experimentación con humanos utilizando procedimientos similares a la búsqueda de efectos análogos. El hallazgo de que en ambos casos se producen efectos tales como curvas de adquisición y extinción, fenómenos de competencia entre causas en función de la validez predictiva, ensombrecimiento y bloqueo, entre otros (véase, Allan, 1993; Miller y Matute, 1996) llevó a postular que un mismo mecanismo asociativo podría explicar el aprendizaje asociativo animal y aprendizaje causal en humanos. En la mayoría de los casos, la regla utilizada se derivó del algoritmo de Rescorla y Wagner (1972) en la que el mecanismo asociativo depende de dos capas de nodos, de forma que el incremento del peso de la conexión entre una unidad de *input i* (el nodo que representa la causa) y una unidad de *output j* (el nodo del efecto) se computa como:

$$(2) \quad \Delta V_n = \alpha_i \beta_j (\lambda - \Sigma V_{n-1})$$

donde ΔV_n representa el incremento de la fuerza asociativa entre las dos representaciones mentales en el ensayo n . α_i y β_j son parámetros de actualización que dependen de la relevancia de la causa y el efecto, respectivamente, λ determina el nivel máximo de aprendizaje alcanzable (valor de fuerza asociativa asintótico) y ΣV_{n-1} es la fuerza asociativa acumulada hasta el ensayo $n-1$ por todos los estímulos presentes en el ensayo n .

En el caso de una tarea de aprendizaje causal, el grado de relación percibida entre la causa (o clave predictora) y el efecto (o resultado) sería una función directa de la fuerza asociativa acumulada entre los nodos i y j , que representan a dichos estímulos. Por tanto, de nuevo nos encontramos con modelos basados en la actuación de un solo mecanismo, de forma que el juicio causal sería el traslado automático del resultado del mismo, es decir, debe reflejar la fuerza asociativa calculada entre la causa y el efecto hasta ese preciso ensayo.

Después de más de dos décadas, la controversia entre los defensores de los modelos asociativos y los de los modelos basados en reglas estadísticas aún no se ha resuelto, y en

general los dos tipos de algoritmos, o sus sucesivas modificaciones, tienden a predecir efectos asintóticos muy similares (Perales, Shanks y Castro, 2005). Sin embargo, en los últimos años, la aparición de nuevos fenómenos en el aprendizaje causal en humanos ha ampliado la investigación teórica y aplicada del aprendizaje. Efectos como la reevaluación retrospectiva de las relaciones causales (Dickinson, 2001), la influencia de la direccionalidad causal en los fenómenos de validez predictiva (Waldman, 2000) o el efecto de la frecuencia del juicio durante el aprendizaje de causalidad (Catena, Maldonado y Cándido, 1998) han generado un nuevo tipo de modelos que distinguen entre la actuación de un mecanismo basado en el mero cálculo de la contingencia, ya sea mediante un algoritmo basado en reglas o un algoritmo asociativo, y los procesos de integración y atribución de más alto nivel que subyacen a la emisión de un juicio de causalidad. Estos últimos procesos implicarían la activación de nuevos mecanismos, relacionados ahora más con el contenido, entendido por ejemplo como “poder causal” (Cheng, 1997) o “modelos mentales” (Walkman y Matignon, 1998), que con los mecanismos básicos de aprendizaje (De Houwer y Beckers, 2002).

Uno de los modelos que tiene en cuenta esta dicotomía fue presentado por Catena et al. (1998). Según la propuesta original del modelo, el aprendizaje causal depende de un proceso de revisión de creencias basado en la acción serial de dos mecanismos (véase la Figura 1). En primer lugar, antes de la emisión de un juicio, un mecanismo básico de aprendizaje sería el encargado de calcular la contingencia establecida entre dos sucesos (la causa y el efecto) a partir de las frecuencias de cada tipo de ensayo, almacenadas en la memoria de trabajo. Esto implica un mecanismo básico de detección de los sucesos que ocurre en el medio ambiente sobre cuya frecuencia el mecanismo de cálculo establece el grado de relación objetiva.

Ahora bien, el problema es que se ha demostrado repetidamente que no somos tan objetivos. De hecho, uno de los supuestos básicos del modelo de revisión de creencias parte del hecho de que los individuos valoran el peso relativo de sus experiencias de forma desigual. Se ha demostrado experimentalmente que se asigna más valor a las experiencias positivas y directamente confirmatorias (ensayos tipo *a*) que a las experiencias negativas o disconfirmatorias (ensayos tipo *b* o *c*). Además, las experiencias de tipo *d* o indirectamente confirmatorias, resultan más difíciles de evaluar y tienen menor influencia en nuestros juicios. Este fenómeno de ponderación diferencial ha sido demostrado, tanto de forma indirecta mediante el análisis cuantitativo de la proporción de cambio en el juicio causal como resultado de un determinado tipo de ensayo (Kao y Wasserman, 1993), como con técnicas más directas, es decir, pidiendo a los sujetos que estimaran la cantidad de influencia de cada tipo de ensayos en su propio juicio de causalidad (Maldonado, Catena, Cándido y García, 1999). Aunque ese valor diferencial puede depender del carácter biológico de los eventos (Miller y Matute, 1996), también depende de la experiencia previa con esos eventos. Maldonado et al. (1999) encontraron que, antes de tener ningún tipo de experiencia de aprendizaje, los sujetos tienden a evaluar la influencia de cada tipo de ensayo en consonancia con lo demostrado en investigaciones previas, $a > b > c > d$;

pero una experiencia previa no-contingente (y en parte también la negativa), tiende a producir un cambio en la ponderación de los tipos de ensayo, sobre todo rebajando sensiblemente el valor de las evidencias positivas confirmatorias (ensayos tipo a), lo que retomaremos posteriormente para revisar el modelo.

Por todo lo anterior y para permitir además la posibilidad de explicar los efectos de adquisición y extinción, el modelo de revisión de creencias propone que el mecanismo de cálculo utilizaría la llamada *regla ΔD ponderada*, que permite el cálculo de contingencia incluso cuando sólo hay un único ensayo o una única experiencia:

$$(3) \quad NuevaEvidencia = \frac{w_1 a + w_2 b + w_3 c + w_4 d}{a + b + c + d}$$

donde, a , b , c y d representan la frecuencia de cada tipo de ensayo, como se explicó anteriormente, y w representa el peso asignado a cada uno de ellos, con la restricción de que $a > b > c > d$, para ajustarlo al peso real que los individuos conceden a cada tipo de ensayo, como ha sido demostrado consistentemente (Kao y Wasserman, 1994; Maldonado et al., 1999). Es importante señalar que investigaciones recientes sobre juicios de causalidad han demostrado que una *regla ΔD ponderada* es mucho más predictiva que ninguna otra, incluida una *regla ΔP* con o sin ponderación (Perales et al., 2005).

En la explicación del aprendizaje causal, todo lo anterior sugiere la necesidad de un mecanismo que permita no sólo detectar sino también calcular el grado de relación objetiva entre la causa y el efecto, y hasta aquí este modelo sólo difiere de los anteriores en que el cálculo se haría mediante una regla diferente. Ahora bien, según el modelo de revisión de creencias, una vez estimada la evidencia medio-ambiental sobre el grado de relación objetiva entre la causa y el efecto, entraría en funcionamiento el mecanismo de integración de información (Figura 1). Este mecanismo superior jerárquicamente sería el responsable de la atribución de causalidad, mediante la actualización del juicio en ese ensayo n , integrando dicha *Nueva Evidencia*, obtenida desde el juicio anterior con las creencias causales previas, representadas, normalmente, por el último juicio emitido (J_{n-k}). Dicha integración se realiza mediante la fórmula:

$$(4) \quad J^n = J^{n-k} + \beta (NuevaEvidencia - J^{n-k})$$

donde J sería el juicio en el ensayo n (o $n-k$), k representa el número de ensayos desde el último juicio emitido, β es un parámetro de revisión que depende de las características de la tarea y que permite un aprendizaje gradual, y *Nueva Evidencia* se refiere al cómputo –realizado por el mecanismo de cálculo previamente descrito– de la información presentada entre el ensayo $n-k$ y el ensayo actual, n , según la fórmula anterior. Este segundo mecanismo estaría relacionado con procesos implicados ya no en la detección o cálculo de contingencias, sino en la atribución de causalidad. Por tanto, este mecanismo sería el responsable del juicio final del individuo en esa determinada situación.

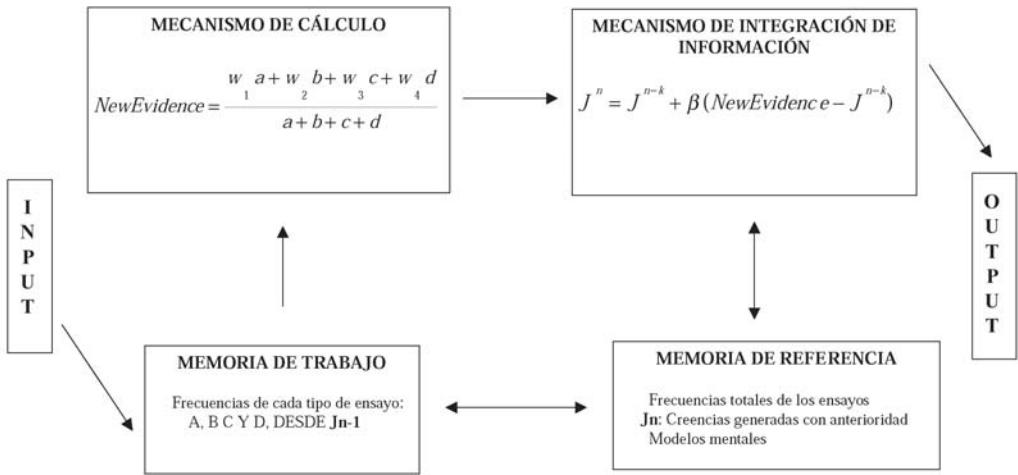


Figura 1. El modelo de revisión de creencias (*Belief revision model*).

Este modelo permitió explicar el efecto de la frecuencia de juicio de una forma clara y elegante y mucho más ajustada que los modelos estadísticos o asociativos que se utilizan normalmente para explicar el aprendizaje causal en situaciones diferentes que incluían tanto experiencias ensayo a ensayo, como incluso cuando se presentan tablas de contingencia, (véase Catena et al, 1998, 2002 y 2004, para simulaciones de modelos estadísticos, asociativos y del modelo de revisión de creencias). Este efecto se explica por el hecho de que cada vez que se aplica el mecanismo de actualización, el punto de enganche para realizar el ajuste producido por la nueva evidencia es siempre el último juicio emitido (J_{n-k}), independientemente de que se haya llegado a éste realizando el juicio cada ensayo o cada cierto número de ensayos. A partir de entonces, este modelo ha permitido extender la investigaciones a nuevos fenómenos como el efecto asimétrico sobre la detección de relaciones generativas y preventivas de una experiencia previa no-contingente (Maldonado et al., 1999, 2004), o más recientemente el efecto también asimétrico de la atención sobre el aprendizaje causal que llega a producir una “ceguera inatencional” de las relaciones negativas, es decir la incapacidad de detectar la existencia de relaciones negativas entre una causa y un efecto (Maldonado et al., 2005).

Centrándonos en el efecto de la frecuencia del juicio, investigaciones posteriores han replicado ese mismo efecto en condiciones diferentes (Matute, Vegas y De Marez, 2002; Vila, 2000). Dicha investigación demostró que tras un entrenamiento que consistía en una primera fase de adquisición, es decir donde existía una relación positiva entre la causa y el efecto, y una posterior de extinción, en donde la relación era negativa, si sólo se pide el juicio al final del entrenamiento (modo global), los juicios reflejan la media de todos los ensayos (es decir promedian ambas fases); mientras que si se piden ensayo a ensayo (o cada cierto número de ensayos), los juicios se van ajustando a cada fase del entrenamien-

to. Además, el efecto de la frecuencia parece depender del tipo de pregunta que se le hace al individuo (Matute et al., 2002) y el paso del tiempo tiene un efecto diferencial, puesto que no cambia el juicio en la condición global (es decir cuando se pide un único juicio al final del entrenamiento), pero sí cuando se ha pedido ensayo a ensayo, en cuyo caso el juicio se hace teniendo en cuenta toda la experiencia anterior, es decir como en el modo global (Vila, 2000).

Podría argüirse que los dos modos de respuesta ponen en marcha mecanismos de detección de la contingencia distintos (asociativo para el modo global y estadístico para el modo ensayo-a-ensayo). Esa posibilidad, poco parsimoniosa, puede descartarse en virtud de la evidencia empírica de la que disponemos, sobre todo porque el efecto de la frecuencia del juicio no aparece súbitamente a partir de una cierta frecuencia de juicio, sino gradualmente conforme dicha frecuencia se va incrementando (Catena et al., 1998). Además, se ha demostrado también que dicho efecto no se acompaña de un deterioro general en la detección global de la contingencia, sino que sólo la modula (Catena, Perales y Maldonado, 2004). Por tanto, es probable que, si se mantienen inalterados otros factores como la carga cognitiva o el modo de presentación de la información, el mecanismo básico de cómputo de la contingencia (ya sea éste estadístico o asociativo) sea siempre el mismo, independientemente del modo de respuesta. Lo que debe variar no es el mecanismo básico de aprendizaje, sino el funcionamiento de mecanismos jerárquicamente superiores que integran la información facilitada por dicho mecanismo para la emisión del juicio (véase también Catena et al. 2002, para una discusión más completa y Collins y Shanks, 2002, para un argumento similar).

Todos estos nuevos resultados ponen de manifiesto la necesidad de modelos basados no sólo en el mecanismo de adquisición, sino que sean capaces de explicar cómo se procesa el contenido de dicho aprendizaje. En ese sentido, el modelo de revisión de creencias establecía que el resultado final del aprendizaje era una “creencia” derivada de la actuación de un segundo tipo de mecanismos (atencionales, de memoria, etc.). Datos recientes nos han llevado a introducir nuevos elementos que permiten extender dicho modelo y adecuarse a lo que realmente ocurre cuando realizamos un aprendizaje.

En investigaciones recientes se ha analizado el efecto de la atención o de la emoción en el aprendizaje causal. En estos estudios sólo se pedía el juicio de causalidad, sino que una vez terminada la tarea y sin que el individuo hubiera sido previamente advertido, se le pedía que estimara el número de veces que había ocurrido cada tipo de ensayo (*a*, *b*, *c* o *d*). Entre otras manipulaciones, la más importante fue variar el número de ensayos en función del tipo de contingencia. En la Tabla 1 se puede observar en primer lugar el nivel de contingencia que fue muy positiva (+75), nula (0) o muy negativa (-75) y a continuación los juicios de causalidad emitidos por los sujetos al final del entrenamiento. Como se puede observar viendo los juicios de causalidad en la Tabla 1, los individuos detectan los diferentes tipos de contingencias objetivas establecidas y son más exactos en el caso de la contingencia positiva que la negativa, lo que es fácil explicar a partir de la influencia diferencial de cada tipo de ensayo, sobre todo de los ensayos tipo *a*, que hacen que sea más difícil

Tabla 1

Nivel de contingencia, juicios de causalidad emitidos por los participantes y número de ensayos de cada tipo empleados para establecer los niveles de contingencia

Contingencia	Juicio	Tipo a	Tipo b	Tipo c	Tipo d
+75	57	14	2	2	14
0	9	2	2	14	14
-75	-41	2	14	14	2

detectar relaciones negativas. Por último, en esa tabla se presenta cómo se estableció objetivamente la contingencia en cada grupo, en función del número de cada tipo de ensayo.

En la Figura 2 se puede observar hasta qué punto los sujetos son sensibles a las diferencias de ocurrencia de cada tipo de ensayos y cómo son bastante exactos en sus estimaciones, a pesar de que nunca fueron avisados o instruidos para que prestaran atención a aquellos. En el caso de la contingencia positiva, distinguen adecuadamente que el número de ensayos tipo *a* es sensiblemente superior a los *b* y *c*, y parecen tener algún problema para recordar con suficiente precisión el número de ensayos tipo *d*, probablemente influido porque también son a los que se les concede menor peso a la hora de la estimación causal. En el caso de la contingencia negativa el contenido es el contrario, ya que ahora se recuerda que el mayor número de ensayos eran de tipo *b* y *c*, muy superior a los de *a*, y de nuevo los peor detectados son los *d*. Ahora bien, es especialmente importante señalar cómo en el caso de la contingencia nula, que estaba construida usando una mayor proporción de ensayos tipo *c* y *d*, los que normalmente son más difíciles de recordar, los sujetos perciben perfectamente esa diferencia y el recuerdo se ajusta a la ocurrencia real de cada uno de los tipos de ensayo, con un alto nivel de precisión. Este último resultado deja muy claro que el recuerdo no se produce como una inferencia resultado del juicio, porque en ese caso lo más sencillo sería asignar el mismo número de ensayos de cada tipo (*a*, *b*, *c* y *d*), tal y como normalmente se define la contingencia nula. El hecho de que en cada caso se recuerde con suficiente precisión cada tipo de ensayo presentado durante el entrenamiento, obliga a matizar cualquier modelo explicativo del aprendizaje causal.

En función de los resultados anteriores podemos entender mejor el contenido del aprendizaje causal ya que se demuestra que tras una experiencia de aprendizaje, los sujetos no sólo establecen una creencia causal (reflejada en el juicio de causalidad), sino que guardan un recuerdo bastante exacto de todo lo que les ha sucedido, lo que permite entender algunos de los efectos previos que resultaban difícil de explicar. Por ejemplo, dado que el sujeto guarda en memoria lo sucedido y dada la capacidad humana para entender y seguir instrucciones, parece obvio que si obligamos a los sujetos mediante el tipo de pregunta a emitir el juicio en función de un determinado tipo de recuerdo (por ejemplo, tener en cuenta todos los ensayos o sólo valorar el final o el principio del entrenamiento), los juicios pueden cambiar porque la “nueva evidencia” se calcula sobre el recuerdo concreto de ese

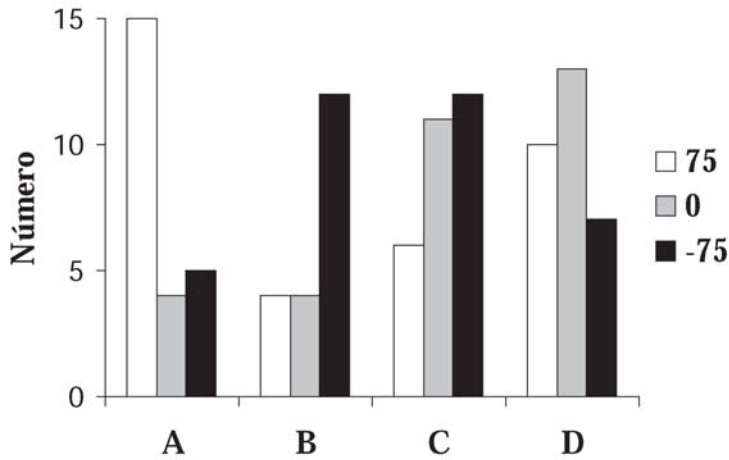


Figura 2. Estimaciones de los participantes con respecto al número de ensayos de cada tipo (*a*, *b*, *c* y *d*) cuando son expuestos a distintos niveles de contingencia (0.75, 0 y -0.75).

tipo de experiencia. Asimismo, el paso del tiempo tendría el efecto de vaciar la memoria de trabajo por la necesidad de procesar otro tipo de información. Por tanto, si volvemos a pedir el juicio tras un largo periodo de tiempo, éste se hará de modo global, es decir, reflejará el recuerdo de todos los eventos almacenados en la memoria de referencia, que es lo que se traslada a la memoria de trabajo cuando se vuelve a pedir el juicio, como han demostrado repetidamente los trabajos de Vila y colaboradores.

En cualquier caso, todo este conjunto de resultados enfatiza una vez más la imposibilidad de un modelo basado en un mecanismo único, sea asociativo o estadístico, que traslade el resultado de su cálculo directamente a juicios, para explicar un fenómeno tan complejo como el aprendizaje causal en humanos. De ahí la necesidad de desarrollar modelos más complejos para adecuarse a esa realidad, como sugiere el modelo de revisión de creencias y otros modelos relacionados, modelos que postulan que el juicio causal depende de algo más que del cálculo de la mera contingencia entre una causa y un efecto, sea por la necesidad de estimar el poder causal (Cheng, 1998), sea por la existencia de un “modelo mental” previo que se revisa en función de la experiencia (Walkman y Matignon, 1998).

Conclusión

En este capítulo hemos intentado mostrar en primer lugar la necesidad de modelos explicativos del aprendizaje causal en humanos que van más allá de la actuación de un único mecanismo, sea asociativo o estadístico. En segundo lugar, hemos intentado presentar cómo el modelo de revisión de creencias propone una arquitectura cognitiva basada en la acción serial de dos mecanismos que permiten explicar dicho aprendizaje causal, que implica siempre la detección de relaciones causales y la atribución de causalidad en nuestro medio ambiente.

Es importante señalar que aunque a lo largo de esta exposición hemos descrito el mecanismo de integración en su actuación sobre la información ya computada, el modelo asume que, antes de iniciar la tarea, este mecanismo de orden superior es sensible a las expectativas y demandas de la propia tarea, derivadas de las instrucciones y la experiencia previa con esa o tareas parecidas, lo que permite al individuo construir un modelo mental de la misma. En función de dicho modelo se produciría un efecto triple sobre el mecanismo de computación de la contingencia.

En primer lugar, como se ha demostrado repetidamente incluso en investigaciones derivadas de este modelo (Perales et al., 2004), es posible que en determinadas situaciones el mecanismo computacional utilice un cálculo incondicionado, cuando sólo se evalúa una causa de un efecto o de diferentes efectos; mientras que, cuando se utilizan claves múltiples de un mismo efecto se use un cálculo condicional (Cheng, 1988; Waldman y Matignon, 1988; véase también Maldonado et al., 2005 para el caso del modelo de revisión de creencias y su explicación de fenómenos con claves múltiples); en segundo lugar, investigaciones recientes han demostrado que las instrucciones pueden modificar el tipo de procesamiento, automático o controlado, y el nivel de atención utilizado durante la tarea (Maldonado et al., 2005); por último, como se demostró en trabajos anteriores (Maldonado et al., 1999), dicho mecanismo debe determinar los pesos relativos y el valor asignado a cada tipo de ensayo, que pueden modificarse en función de la experiencia previa con la tarea, entre otros factores que también necesitan ser investigados.

Por aclarar el funcionamiento de alguno de los puntos anteriores, fijémonos en la importancia de determinar el cálculo condicionado o incondicionado de la información. Este supuesto permite entender y explicar desde un modelo computacional los efectos de validez relativa, ensombrecimiento, bloqueo y reevaluación que han sido comúnmente causa de crítica del mecanismo estadístico, de forma similar a como lo hace el mecanismo asociativo. Pero también este supuesto permite entender otros efectos nuevos, como la importancia de la percepción de la aditividad y maximalidad de la naturaleza de las causas, inducida mediante instrucciones, para que aparezca o no el fenómeno de bloqueo (Beckers, De Houwer y Miller, 2004), lo que demuestra la importancia del modelo mental sobre la naturaleza de las causas para entender los efectos de competencia entre claves en humanos. Por otra parte, también se ha demostrado que los efectos de validez relativa cuando existen múltiples claves y una única consecuencia dependen no sólo del número de claves, sino también de la direccionalidad causal (Walkman, 2000), de forma que la contingencia puede calcularse incondicionalmente en el caso de una causa única con efectos múltiples, pero condicionada en el caso de causas múltiples de un efecto único; esto mismo se confirma por la influencia de la direccionalidad causal en el fenómeno de aprendizaje mediado (Perales, Catena y Maldonado, 2004). Por tanto, cómo y cuándo se determina dicho cálculo de la contingencia, es hoy día uno de los retos abiertos en el aprendizaje de causalidad y abre nuevas perspectivas de investigación futura.

Ahora bien, una vez iniciada la tarea, el modelo de revisión de creencias asume, como explicamos antes, que cada vez que se solicita un juicio, primero actuaría el mecanismo de

cálculo y luego el mecanismo de integración ajusta el juicio en función de dicha información o evidencia nueva (véase la Figura 1) y de los demás factores cuya influencia ha sido recientemente demostrada y que implican tanto factores cognitivos como emocionales (Waldman y Martingon, 1998; Perales et al., 2004).

Por último, una vez emitido el juicio se produce un vaciado de la memoria de trabajo y tanto ese juicio (J_n) como cualquier modificación del modelo mental construido inicialmente se guardan en la memoria de referencia y representarían las creencias causales del individuo en esa situación o situaciones similares. La aportación nueva es que también se guarda en dicha memoria la frecuencia total de cada tipo de ensayo (véase Maldonado et al., 2005), para ser recuperado por la memoria de trabajo cada vez que sea necesario, bien sea porque se produzca una nueva información o porque se solicite un nuevo juicio.

Esta arquitectura cognitiva permite explicar la mayoría de los efectos encontrados en el aprendizaje causal y además predice la influencia de otros nuevos factores cognitivos y emocionales cuyo estudio permitirá ampliar nuestra comprensión del aprendizaje causal y del propio comportamiento humano.

Nota de los autores

Este trabajo ha sido realizado dentro del proyecto BSO2003-03723 del Ministerio de Ciencia y Tecnología, España. La correspondencia correspondiente a este trabajo debe dirigirse a Antonio Maldonado, Departamento de Psicología Experimental, Facultad de Psicología, Universidad de Granada, Campus de Cartuja s/n. Granada 18071. España. E-mail: anmald@ugr.es.

Referencias

- Allan, L. (1993). Human contingency judgment: Rule based or associative? *Psychological Bulletin*, 114, 435-48
- Beckers, T., De Houwer, J. y Miller, R. R. (2004, May). Outcome additivity and outcome maximality as independent modulators of blocking in human and rat causal learning. Communication. *1 Special Interest Meeting on Human Contingency Learning*. Belgium.
- Catena, A., Maldonado, A. y Cándido, A. (1998). The effect of the frequency of judgement and the type of trial on covariation learning. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 481-495.
- Catena, A, Maldonado, A., Megías, J. L. y Frese, B. (2002). The frequency of judgement effect: Belief revision and serial processing of causal information. *Quarterly Journal of Experimental Psychology*, 55B, 267-281
- Catena, A., Perales, J. C. y Maldonado, A. (2004). Judgment frequency effects in generative and preventive causal learning. *Psicológica*, 25, 67-85
- Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405.
- Cheng, P. y Novick, L. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365-382.
- Collins, D. y Shanks, D. R. (2002). *Memory & Cognition*, 30, 1138-1147.
- De Houwer, J. y Beckers, T. (2002) A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology*, 55B, 289-310.
- Dickinson, A. (2001) Causal learning: an associative analysis. *Quarterly Journal of Experimental Psychology*, 54B, 3-25.
- Kao, S. F. y Wasserman, E. A. (1993). Assessment of information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 1363-86

- Maldonado, A., Catena, A., Cándido, A. y García, I. (1999). The belief revision model: Asymmetrical effects of no contingency on human covariation learning. *Animal Learning & Behavior*, 27, 168-180.
- Maldonado, A., Catena, A., Perales, J. C. y Herrera, A. (2004). El efecto de la frecuencia de juicio y de la experiencia previa no contingente en el aprendizaje causal. *Cognitiva*, 16, 75-93
- Maldonado, A., Herrera, A., Jiménez, I., Perales, J. C. y Catena, A. (2005). Inattention blindness for negative relationships in human causal learning. *Quarterly Journal of Experimental Psychology (A)*. En prensa.
- Matute, H., Vegas, S. y De Marez, P. (2002). Flexible use of recent information in predictive and causal judgments. *Journal of Experimental Psychology: Learning, Memory and cognition*, 28, 714-25.
- Miller, R. R. y Matute, H. (1996). Animal analogues of causal judgement. *The Psychology of Learning and Motivation*, 34, 133-166.
- Pearce, J. y Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Perales, J. C., Catena, A. y Maldonado, A. (2002). Aprendizaje de relaciones de contingencia y causalidad: hacia un análisis integral del aprendizaje causal desde una perspectiva computacional. *Cognitiva*, 14, 15-41.
- Perales, J. C., Catena, A. y Maldonado, A. (2004). Inferring non-observed correlations from causal scenarios: The role of causal knowledge. *Learning and Motivation*, 35, 115-135.
- Perales, J. C., Shanks, D. R. y Castro, L. (2005, *remitido*). Formal models of causal learning: A review and synthesis.
- Price, P. C. y Yates, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1639-55.
- Rescorla, R. y Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. En A. H. Black y W. R. Prokasy (Eds.), *Classical conditioning, vol. 2: Current research and theory*. New York: Appleton-Century-Crofts.
- Rosas, J. M., Vila, N. J., Lugo, M. y López, L. (2001). Combined effect of context change and retention interval on interference in a contingency judgment task. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 153-164.
- Vila, J. (2000). Extinción e Inhibición en Juicios de Causalidad. *Psicológica*, 21, 257-273.
- Waldmann, M. R. (2000). Competition between causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 53-76.
- Waldmann, M. R. and Martignon, L. (1998). A Bayesian network model of causal learning. *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: LEA.

Procesos de razonamiento causal en la interferencia entre señales

David Luque, Pedro L. Cobos y Francisco J. López
Universidad de Málaga, España

Introducción

De modo general, llamamos interferencia al fenómeno por el cual el hecho de aprender una información en un momento determinado dificulta la recuperación de otra información aprendida en otro momento distinto. Este fenómeno, estudiado desde diferentes aproximaciones dentro de la psicología cognitiva (e.g., la computación neuronal, véase Lewandowski, 1994; o el aprendizaje verbal, véase Slamecka y Ceraso, 1960), ha recibido un gran interés desde el ámbito de la psicología del aprendizaje animal desde sus orígenes (Pavlov, 1927). En los últimos años se ha venido acumulando una evidencia empírica importante a favor de que determinadas teorías asociativas, surgidas originariamente en el campo del condicionamiento animal para explicar estos fenómenos de interferencia, ofrecen una buena explicación de fenómenos de interferencia equivalentes en el aprendizaje causal humano (e.g., la teoría de la recuperación de Bouton, 1993). En particular, la evidencia se ha venido acumulando en el terreno de los fenómenos tanto de interferencia entre diferentes resultados de una misma señal, como de interferencia entre diferentes señales de un mismo resultado (Castro, Ortega y Matute, 2002; Escobar, Pineño y Matute, 2002; Matute y Pineño, 1998a, 1998b; Miller y Escobar, 2002; Ortega y Matute, 2000; Pineño y Matute, 2000; Pineño, Ortega y Matute, 2000; Rosas, Vila, Lugo y López, 2001; Vila y Rosas, 2001a, b).

Tanto en la interferencia entre resultados como en la interferencia entre señales, y de manera consistente con esas teorías, se ha mostrado la relevancia que tiene el contexto en el que tiene lugar el aprendizaje y la recuperación de la información. La evidencia de esta relevancia del contexto es amplia y se ha mostrado de modos diferentes, como en el efecto de renovación (Matute y Pineño, 1998a; Vila y Rosas, 2001a), el efecto de recuperación espontánea de la respuesta (Pineño et al., 2000; Rosas, et al., 2001) o el de reinstauración (Pineño et al., 2000; Vila y Rosas, 2001b). El hecho de que estos efectos contextuales se hayan mostrado en ambos tipos de interferencia ha llevado a plantear modelos en los que se intenta dar cuenta de ellos a partir del mismo mecanismo asociativo. Los modelos planteados para explicar ambos tipos de fenómenos suponen que la interferencia se da cuando se expone al sujeto a una nueva relación que comparte un elemento (ya sea éste la señal o el resultado) con una relación señal-resultado previamente aprendida. Posteriormente, en la recuperación de la información, el contexto en el que se da la recuperación actuaría de clave que primaría la expresión de la asociación adquirida en ese contexto (Matute y Pineño, 1998b; Miller y Escobar, 2002). Esta explicación, deudora en gran parte del modelo de recuperación de Bouton (1993), surgió originariamente para explicar los efectos contextuales únicamente en la interferencia entre resultados (véanse en este

volumen Rosas, García, Abad y Callejas; Vadillo y Matute; Vila y Alvarado), extiende este modelo para que pueda dar cuenta de los resultados de interferencia entre señales (podemos ver un diseño de interferencia entre señales en la Tabla 1).

Objetivos del presente capítulo. En el presente trabajo vamos a plantear un modelo alternativo al asociativo para explicar el fenómeno de interferencia entre señales y el papel que el contexto desempeña en el mismo. Este modelo se basa en la participación de los que tradicionalmente se han llamado procesos de orden superior. Concretamente, planteamos que organizar el conocimiento usando para ello los contextos forma parte del proceso por el cual los sujetos intentan buscar una explicación en términos causales de la evidencia a la que son expuestos. Este proceso de búsqueda de explicaciones estaría a su vez guiado por el conocimiento causal previo. Ejemplos de modelos que comparten estas intuiciones, a los que a partir de ahora llamaremos modelos de razonamiento causal, son los modelos basados en la importancia de los mecanismos causales (e.g., Ahn y Kalish, 2000), los modelos de redes Bayesianas causales (e.g., Glymour, 2001, 2003), la teoría del modelo causal (e.g., Waldmann y Holyoak, 1992) y la teoría de los modelos mentales adaptada para el aprendizaje causal (Goldvarg y Johnson-Laird, 2001). Aunque un gran número de factores diferencian a estos modelos, todos consideran que la inferencia causal representa una búsqueda de un modelo causalmente legal de la evidencia y que esta búsqueda está restringida o guiada desde el conocimiento previo.

La extensión de estos modelos a la interferencia entre señales parte de varias consideraciones que hacen pertinente su propuesta. Primero, es importante señalar el apoyo empírico creciente que están recibiendo los modelos de razonamiento causal en situaciones de aprendizaje cuando dichas situaciones son susceptibles de recibir una interpretación causal (e.g., Ahn, Kalish, Medin y Gelman 1995; Dennis y Ahn, 2001; Goldvarg y Johnson-Laird, 2001; Gopnik, Glymour, Sobel, Schulz, Kushnir, Danks, 2004; Lagnado y Sloman, 2004; López, Cobos y Caño, en prensa; Sobel, Tenenbaum y Gopnik, 2004; Sloman y Lagnado, 2005; Tangen y Allan, 2004; Waldmann, 2000, 2001; Waldmann y Holyoak, 1992)

Aunque los procesos de orden superior a los que hacemos referencia han mostrado su influencia en la inducción causal en numerosas ocasiones y de modos muy diferentes, para los objetivos del presente trabajo debemos destacar los resultados que muestran la sensibilidad al rol causal de los eventos y la capacidad de los sujetos de inferir factores intervinientes pero no explícitos. En relación a la sensibilidad al rol causal de los eventos, estudios recientes muestran cómo los sujetos se comportan de modo diferente dependiendo del papel causal que tomen las señales y los resultados (López et al., en prensa; Tangen y Allan, 2004; Waldmann y Holyoak, 1992; Waldmann, 2000, 2001). Es decir, los sujetos diferencian situaciones predictivas (en las que el aprendizaje se realiza de causas a efectos) de diagnósticas (en las que el aprendizaje se realiza de efectos a causas). El hecho de discriminar entre los diferentes roles causales que puedan desempeñar las señales y los resultados permite a los sujetos poder razonar de causas a efectos, que es el orden tempo-

Tabla 1
Diseño de interferencia entre señales

	Fase I	Fase II	Test
Grupo experimental	A→1, C→3	B→1, C→3	A?
Grupo control	A→1, C→3	B→2, C→3	A?

ral de aparición consistente con nuestro conocimiento general acerca de las relaciones causales. Esta sensibilidad al rol causal se ha mostrado incluso en situaciones en las que ambas condiciones (predictiva y diagnóstica) eran estimularmente idénticas.

Como hemos comentado, los modelos de razonamiento causal también plantean que los sujetos son capaces de inferir causas ocultas a partir de patrones de evidencia explícita. Al igual que en el caso de la sensibilidad de los sujetos al rol causal de los eventos, estudios recientes avalan esta capacidad de inferencia (Danks y McKenzie, estudio no publicado citado en Glymour, 2003; Gopnik et al, 2004; Kushnir, Gopnik, Schulz, y Danks, 2003). Estos resultados muestran cómo cuando la única posibilidad de formar un modelo causal legal de los eventos implica la existencia de una causa no observada, los sujetos son capaces de inferirla a partir de la información de las co-ocurrencias de los eventos observables.

Todos estos resultados resultan difíciles de reconciliar con los mecanismos asociativos propuestos hasta ahora para explicar el aprendizaje causal. Así, de un modo general se ha mostrado la relevancia de los modelos de razonamiento causal en situaciones de aprendizaje y uso del conocimiento causal. Además, centrándonos en los procesos que nos interesan en el presente trabajo, se han descrito resultados en la literatura que apoyan la validez de los mecanismos específicos que estarían implicados en la explicación de la interferencia entre señales que propondremos en el siguiente apartado, es decir, la sensibilidad al rol causal de los eventos y la capacidad de inferir causas no explícitas.

Otra de las consideraciones que hacen pertinente la aplicación de los modelos de razonamiento causal a la interferencia entre señales se basa en un análisis de los estudios previos en los que se ha observado el fenómeno. Los estudios en los que se ha mostrado interferencia entre señales implicaban tareas que pueden ser interpretadas como diagnósticas, es decir, las señales que eran presentadas a los sujetos servían como indicadores para averiguar la causa o causas de las mismas¹; no se ha mostrado en la literatura interferencia entre señales en tareas en las que la interpretación causal fuese predictiva.

¹Dos son las tareas en las que se ha obtenido interferencia entre señales en humanos. La tarea de los marcianos (e.g., Matute y Pineño, 1998a), en la que las señales eran luces que indicaban la activación de un campo de fuerza por parte de unos malvados alienígenas, y la tarea de la radio espía, en la que las señales eran unas luces que informaban a los sujetos de la presencia o no de minas en la carretera (Pineño et al., 2000). Decimos que estas tareas pueden ser interpretadas como diagnósticas ya que el encendido final de las luces que hacen de señales es fácil de representar como efecto de los eventos que actúan como resultados (ya sea la activación del escudo de fuerza o la presencia de minas en la carretera), siendo difícil imaginar algún otro tipo de modelo que explique la situación, como por ejemplo que una causa común tuviese como efecto tanto las señales como los resultados.

Esta situación concuerda con nuestras predicciones, ya que nuestra hipótesis basada en el razonamiento causal prevé, como veremos más adelante, un efecto de interferencia entre señales en tareas diagnósticas, pero no en tareas predictivas.

Por último, nuestra hipótesis explicativa de la interferencia entre señales pretende dar un sentido computacional a la codificación del contexto en la interferencia entre señales. En la interferencia entre resultados el modelo de Bouton (1993) justificaba la codificación del contexto cuando una sola señal se asociaba con dos resultados contradictorios basándose en un sentido adaptativo: eliminar la ambigüedad en tal situación conflictiva. Para poder predecir lo que va a ocurrir después de la señal, es necesario primar una relación u otra, de ahí el papel del contexto (véase Rosas et al., en este mismo volumen). Sin embargo, el uso de los contextos no se puede justificar del mismo modo en el caso de la interferencia entre señales. En este caso no existe la necesidad de eliminar la ambigüedad de las señales. Los modelos asociativos propuestos para explicar la interferencia entre señales no proponen ninguna hipótesis acerca del posible sentido de la codificación del contexto en estas situaciones, sino que su nivel de análisis se centra en aspectos mucho más descriptivos (véase Miller y Escobar, 2002). Como veremos, un análisis detallado desde los modelos de razonamiento causal nos permite entender el sentido computacional de la interferencia entre señales así como la función que desempeña el contexto.

Interferencia entre señales según los modelos de razonamiento causal

Como hemos comentado, los modelos de razonamiento causal tienen en común el otorgar una especial relevancia al conocimiento causal previo en el aprendizaje del conocimiento causal. La influencia del conocimiento previo puede concretarse de diferentes modos, ya sea mediante la influencia del conocimiento sobre los mecanismos causales implicados en una situación concreta (Ahn et al., 1995; Hagmayer y Waldmann, 2002), o mediante la influencia de conocimiento causal no específico acerca del funcionamiento de las relaciones de causalidad en general. Acerca de la influencia de este último tipo de conocimiento, destacan las aportaciones de Michael Waldmann y sus colaboradores (e.g., Waldmann, 1996; Waldmann & Holyoak, 1992). Su *teoría del modelo causal* refleja el efecto de las asunciones a priori acerca de las características de las relaciones de causalidad, centrándose en el efecto del conocimiento acerca del orden temporal de las direcciones causales, es decir, el conocimiento de que las causas anteceden a los efectos y no al contrario. De este modo, los modelos de razonamiento causal entre los que se incluye la *teoría del modelo causal* mantienen, en oposición a las teorías asociativas, que los sujetos son sensibles al estatus causal de las señales y los resultados (qué eventos, las señales o los resultados, son causas o efectos), y este hecho explica las diferencias encontradas entre tareas de aprendizaje diagnóstico y tareas de aprendizaje predictivo (e.g., Waldmann y Holyoak, 1992; Waldmann, 2000, 2001; pero véase Cobos, López, Caño, Almaraz y Shanks, 2002, para un punto de vista alternativo).

Otro punto en común de las teorías del razonamiento causal es suponer que el sujeto busca un modelo causal que explique la evidencia a la que es expuesto. Existe evidencia

acerca de cómo los sujetos tratan de explicar las situaciones conflictivas a partir de explicaciones causales en las que es necesario inferir causas que intervienen en la situación de aprendizaje, pero que permanecen ocultas a lo largo de la tarea (Danks y McKenzie, estudio no publicado citado en Glymour, 2003; Gopnik et al., 2004; Kushnir et al., 2003). De igual modo, Johnson-Laird ha mostrado cómo cuando los sujetos se enfrentan a información que es contradictoria con la información previa que tenían, no se conforman con primar la información nueva o antigua, sino que buscan explicar el cambio de situación en términos causales, búsqueda que implica la inferencia de nuevos elementos causales no explícitos en la situación (Johnson-Laird, Girotto y Legrenzi, 2004; Johnson-Laird, Legrenzi, Girotto y Legrenzi, 2000).

Al analizar la situación de interferencia entre señales teniendo en cuenta ambas asunciones (sensibilidad a la direccionalidad causal de la tarea y capacidad de inferir causas ocultas para construir un modelo causal aceptable que explique la evidencia) llegamos a conclusiones interesantes. Dos situaciones pueden tener lugar: que el rol causal de las señales sea el de causas (situación predictiva) o que el rol causal de las señales sea el de efectos (situación diagnóstica). Vamos a analizar en primer lugar las predicciones de los modelos de razonamiento causal en la situación diagnóstica, situación equivalente a los escenarios causales en los que se ha encontrado interferencia entre señales en la literatura (e.g., Matute y Pineño, 1998a; Pineño et al., 2000), al menos según nuestra interpretación. Observando la Tabla 1 llegamos a la conclusión de que, en dirección diagnóstica, la estructura causal de los datos difiere según la condición experimental. En el caso del grupo experimental, la información que se recibe durante la primera fase es que la causa 1 produce sistemáticamente el efecto A, mientras que en la segunda fase la causa 1 deja de producir el efecto A para producir sistemáticamente el efecto B. En cambio, el grupo de control pasa por una primera fase en la que la causa 1 produce sistemáticamente el efecto A y posteriormente pasa por una segunda fase en la que la causa 2 produce sistemáticamente el efecto B. Desde el punto de vista de las teorías de razonamiento causal, la información que recibe el grupo experimental es conflictiva de cara a la construcción de un modelo causal, porque las personas parten del conocimiento de que si dos efectos sólo se explican por la ocurrencia de una causa común entonces, dada la causa, la ocurrencia de un efecto debe ser independiente a la del otro. Sin embargo, dada la correlación negativa existente entre los efectos A y B y el modo en que se distribuyen sus probabilidades en las diferentes fases temporales, la única explicación posible es que debe haber alguna causa oculta que correlaciona con las distintas fases temporales y los contextos en los que se dan y que interactúa con la causa 1 para producir bien A, bien B. Esta causa oculta impone límites sobre la capacidad de 1 para producir A y B. Como la fase de test suele formar parte de la fase 2 y suele tener lugar en el mismo contexto, cuando se presenta el efecto A, los participantes tienen sus dudas de que la causa 1 haya ocurrido porque, si las condiciones no han cambiado, la causa 1 ya no produce A sino B. Ello explica que la respuesta ante A sea menor en el grupo experimental que en el grupo de control. En este último nada de esto se aplicaría porque la información recibida en cada fase no plantea

ningún tipo de conflicto. Es decir, no hace falta apelar a causas ocultas que correlacionen con las distintas fases y que limiten la capacidad de las causas 1 y 2 para producir los efectos A y B, respectivamente.

Sin embargo, si la situación fuera predictiva, las predicciones que habría que deducir de los modelos de razonamiento causal serían distintas. En este caso, la información que se recibe en el grupo experimental durante la primera fase es que la causa A produce sistemáticamente el efecto 1, mientras que en la segunda fase es otra causa, la causa B, la que provoca el efecto 1. La información que recibe el grupo experimental ya no es conflictiva y no es necesario apelar a causas ocultas para explicar los eventos. En el grupo de control la situación no cambia con respecto a la situación diagnóstica (véase la Tabla 1), con lo que tampoco es necesaria la inferencia de estas causas ocultas. De este análisis se deriva que en la fase de test los sujetos no tendrán problemas para expresar la relación $A \rightarrow 1$ ni en el grupo control ni en el grupo experimental, con lo que no se obtendrá el efecto de interferencia entre señales.

Primeras evidencias empíricas

Recientemente hemos realizado en nuestro laboratorio un experimento con el fin de analizar la validez de los modelos de razonamiento causal en la explicación de la interferencia entre señales. En este experimento manipulamos ortogonalmente dos factores, la condición experimental y la dirección causal de la tarea. Por un lado, los sujetos podían ser asignados al grupo experimental o al grupo control del diseño de interferencia entre señales que podemos ver en la Tabla 1. Además manipulamos la dirección causal de la tarea, en un grupo la tarea se pasó en dirección predictiva y en otro la tarea se realizó en dirección diagnóstica. La diferencia fundamental existente entre el grupo predictivo y el grupo diagnóstico era el rol causal que adoptaban las señales y los resultados. Mientras que en el grupo diagnóstico las señales eran los efectos y los resultados eran las causas, en el grupo predictivo las señales eran las causas y los resultados los efectos. En ambos experimentos los contextos se distribuían de modo que se producía un cambio contextual de la fase 1 a la fase 2, manteniéndose el contexto de la fase 2 durante la fase de test (ver Matute y Pineño, 1998a, para un procedimiento similar). Las predicciones de los modelos asociativos actuales son idénticas para ambos experimentos: interferencia entre señales en ambos casos. La interferencia se daría porque las dos fases comparten un elemento (el resultado); cuando la fase de test se da en el contexto de la segunda fase, la asociación $B \rightarrow 1$ aprendida durante la segunda fase se ve primada ante $A \rightarrow 1$, implicando un descenso en la expresión de $A \rightarrow 1$, esto es, interferencia. El efecto de interferencia sería independiente del papel causal de las señales y los resultados, ya que esta información es irrelevante para los mecanismos asociativos implicados.

Las predicciones que se derivan de los modelos de razonamiento causal son otras bien distintas. Como vimos en el apartado anterior, si tenemos en cuenta los planteamientos que proponen estos modelos sólo se debería producir interferencia en situaciones diagnósticas, ya que es en estas situaciones donde es necesario suponer causas ocultas para

explicar la correlación negativa de los efectos A y B. Sin embargo, en situaciones predictivas la evidencia se puede explicar mediante un modelo causalmente legal sin la necesidad de suponer causas ocultas que interactúen, imponiendo restricciones a las condiciones bajo las que las causas visibles producen los efectos observados.

Para realizar estos experimentos tuvimos que desarrollar una nueva situación de aprendizaje en la que se pudiese invertir el orden causal sin que por ello perdiese verosimilitud el escenario causal. El nuevo escenario causal está inspirado en el escenario de la radio espía (para ver una descripción detallada del escenario de la radio espía véase Pineño et al., 2000). En nuestro escenario los sujetos debían suponer que eran miembros de la Cruz Roja realizando una labor humanitaria en una región muy pobre. Los habitantes de esta región, movidos por el hambre, habían consumido unas plantas a las que no estaban acostumbrados, una o varias de las cuales eran venenosas. Su tarea consistía en administrar un antídoto a los habitantes de las diferentes tribus de la región que hubieran consumido este tipo de planta. La administración del antídoto debía de hacerse con cuidado, ya que otra de las plantas que podían haber consumido, llamada planta “extraña”, intoxicaba al paciente en interacción con el antídoto. Además, en los grupos control había otro tipo de planta que podían consumir los pacientes, la planta sana, que no era venenosa y que no interactuaba con el antídoto. Resumiendo, los sujetos debían administrar antídoto al paciente si creían que había tomado una planta venenosa y no administrárselo si la planta que creían que había tomado el paciente era la extraña, siendo irrelevante si se administraba antídoto o no si el paciente había tomado la planta sana. El consumo de cada tipo de planta provocaba otro efecto además de los que hemos comentado: la alteración de uno entre tres reactivos disponibles, siendo cada reactivo de un color diferente. Para cada paciente se alteraba un solo reactivo, quedando los otros dos inalterados. Para la correcta administración del antídoto, los sujetos debían usar la información de cuál de los reactivos se alteraba para cada paciente para saber qué tipo de planta había consumido. El antídoto sólo podía administrarse durante el tiempo en que la señal estuviese presente en la pantalla, que era 3’5 segundos. La máxima cantidad de antídoto que podía darse por cada paciente era de 100 unidades. Los diferentes contextos consistían en las diferentes tribus de procedencia de los pacientes.

Como hemos señalado, la situación podía ser diagnóstica o predictiva. En la situación diagnóstica los sujetos debían decidir la cantidad de antídoto a administrar a partir de la información de cuál de los tres reactivos se alteraba. De este modo, en dirección diagnóstica la señal en cada ensayo consistía en el color del reactivo alterado por cada paciente, mientras que el resultado era la planta que había consumido.

En la situación predictiva, sin embargo, los sujetos debían decidir la cantidad de antídoto a administrar a partir de una foto de la planta consumida por cada paciente. En esta ocasión la señal a partir de la cual debían administrar más o menos antídoto era la planta consumida, mientras que el resultado era el reactivo alterado.

Al igual que en la tarea de la radio espía, los sujetos recibían información sobre su nivel de ejecución en forma de puntos ganados o perdidos. Había tres posibles consecuencias

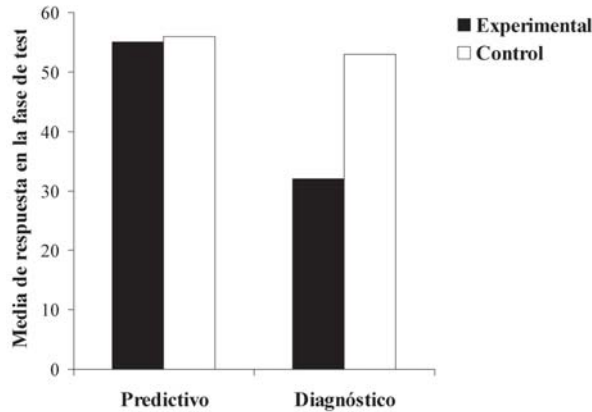


Figura 1. Resultados. El efecto de interferencia se observa sólo en el grupo diagnóstico.

de su respuesta en cada ensayo. Si el paciente había consumido la planta venenosa (en los ensayos A→1 y en los ensayos B→1), los sujetos ganaban tantos puntos como dosis de antídoto habían administrado; si el paciente había consumido la planta extraña (en los ensayos C→3), los sujetos perdían igual número de puntos que la cantidad de antídoto administrada; si el paciente había consumido la planta sana (en los ensayos B→2), no podían ganar ni perder puntos.

De este modo, para obtener interferencia la cantidad de antídoto administrado ante la señal A en la fase de test debería ser menor en la condición experimental que en la condición de control. Si esta disminución de la respuesta se da independientemente de la dirección causal de la tarea, los resultados serían consistentes con los modelos asociativos propuestos hasta ahora para explicar estos fenómenos. Sin embargo, si la disminución de la respuesta en el grupo experimental comparada con el control es menor en la dirección predictiva que en la dirección diagnóstica, las predicciones apoyarían las hipótesis que desde los modelos de razonamiento causal estamos proponiendo en el presente capítulo.

Podemos ver los resultados del experimento en la Figura 1. Como podemos observar, la disminución de la respuesta ante la señal A en la fase de test de la condición experimental, comparada con la condición de control, fue mayor en la dirección diagnóstica que en la predictiva. El análisis estadístico de los resultados se llevó a cabo mediante un ANOVA de dos factores inter-sujeto [dirección causal (grupo predictivo, grupo diagnóstico) x condición experimental (grupo experimental, grupo control)], mostrándose una interacción significativa entre ambos factores, $F(1, 78) = 4.67$, $MSE = 495.31$, $p = 0.03$. El análisis simple del efecto de interferencia en cada dirección causal nos muestra cómo las diferencias entre el grupo experimental y el de control fueron significativas en la dirección diagnóstica $F(1, 35) = 6.69$, $MSE = 631.73$, $p = 0.01$; mientras que no lo fueron en la dirección predictiva $F(1, 41) < 1$, $MSE = 385.51$, $p = 0.99$. Estos resultados nos indican un claro efecto de interferencia sólo en la dirección diagnóstica.

Discusión

El experimento que acabamos de describir fue diseñado para evaluar la validez de los modelos de razonamiento causal a la hora de explicar el efecto de interferencia entre señales. Para ello manipulamos la dirección causal de la tarea manteniendo el mismo diseño de interferencia entre señales tanto en dirección predictiva como en dirección diagnóstica (véase la Tabla 1). Los resultados (interferencia sólo en dirección diagnóstica) son los previstos por la extensión de los modelos de razonamiento causal que proponemos para la interferencia entre señales. Estos resultados replican los que ya obtuvimos con anterioridad en una versión previa del mismo experimento (Luque, Cobos, López y Caño, 2004). Una explicación alternativa a la que proponemos se basa en que la manipulación de la dirección causal correlaciona con el uso de distintos estímulos como señales y resultados. Cuando la tarea estaba diseñada en dirección diagnóstica, las señales eran los colores de los reactivos alterados y los efectos las fotos y nombres de las plantas, mientras que en dirección predictiva las señales eran las fotos de las diferentes plantas y los resultados el color del reactivo alterado. Fruto de estas diferencias se puede proponer una explicación alternativa acorde con los modelos asociativos. Es posible que las diferencias en la configuración estimular de las señales conlleve una mayor dificultad de discriminación en la dirección predictiva que en la dirección diagnóstica. Si este fuera el caso, sería esperable encontrar una curva de adquisición menos acelerada en ese grupo así como una peor ejecución al final de la primera fase. Sin embargo, no encontramos ninguna diferencia entre las curvas de aprendizaje del grupo predictivo y el grupo diagnóstico, siendo la forma de las mismas equivalentes. Tampoco encontramos ninguna diferencia entre los grupos en cuanto al nivel de ejecución en los últimos ensayos de la primera fase. Es decir, los sujetos de ambos grupos aprendieron las relaciones programadas de forma equivalente. El resultado de este análisis resta credibilidad a esta posible explicación de las asimetrías entre la condición predictiva y la diagnóstica.

Existe otra posible explicación alternativa a la que proponemos. Las diferencias entre la condición predictiva y la condición diagnóstica pueden explicarse sin otorgar un papel especialmente relevante a los contextos. Esto es debido a los valores de contingencia diferentes de la relación diana $A \rightarrow I$ si calculamos ΔP condicional siempre en dirección causa \rightarrow efecto (independientemente del orden temporal de los acontecimientos). Esta ΔP condicional es menor en la condición diagnóstica que en la condición predictiva, de ahí la menor respuesta ante A en la fase de test en el grupo diagnóstico que en el grupo predictivo. Hemos de señalar que esta posible explicación también se basa en el uso del conocimiento causal previo, diferenciándose de la que proponemos en este capítulo en que el papel desempeñado por el contexto pasa a ser irrelevante. Aunque esta posible explicación pueda ser atractiva por su parsimonia, no podría dar cuenta de los diferentes efectos contextuales que se han mostrado en la literatura sobre interferencia entre señales. Nuestra hipótesis sí podría explicar los diferentes efectos contextuales como son la renovación, la reinstauración o la recuperación espontánea de la respuesta. En todos estos efectos se encuentra que la respuesta ante A en la fase de juicio se recupera (se debilita el efecto de

interferencia) si el contexto de la fase de juicio es diferente al contexto de la fase 2. Dada esta situación, los modelos de razonamiento causal realizan las mismas predicciones que los modelos asociativos, siempre que la situación sea diagnóstica. En estas situaciones, los sujetos supondrían la existencia de un factor que está mediando la expresión de un efecto u otro dada la causa y que covaría con los contextos. En la fase 2 este factor facilita la aparición del efecto B más que el efecto A, de ahí que si se pregunta por A en el contexto de la fase 2 los sujetos duden de que se haya producido por la causa 1. Ahora bien, si se pregunta por A en cualquier otro contexto, los sujetos no tienen por qué dudar del conocimiento adquirido durante la fase 1, lo que implicaría la desaparición de la interferencia entre señales.

Hemos de señalar que el papel que proponemos al uso contexto está en consonancia con otras evidencias que muestran cómo los sujetos interpretan activamente la información a la que son expuestos, con el objetivo de crear un modelo causal que explique la evidencia. Johnson-Laird desde el campo del razonamiento deductivo describe cómo los sujetos cuando se enfrentan a situaciones inconsistentes, en gran parte similares a un diseño de interferencia, tratan de averiguar cuál de las premisas “falla” y después intentan explicar la situación. Estas situaciones consisten en casos en los que los sujetos se encuentran con premisas incompatibles con las premisas previamente aprendidas; por ejemplo, cuando alguien que creías puntual llega tarde a una cita. Los sujetos buscan la coherencia apelando a otro tipo de factores causales no explícitos en un principio (Johnson-Laird, et al., 2000; Johnson-Laird et al., 2004), como por ejemplo, “habrá tenido un problema con el coche”. Este proceso, denominado “proceso de generación de explicaciones” forma parte fundamental de la aplicación de la teoría de los modelos mentales a casos en los que los sujetos se enfrentan a situaciones conflictivas, siendo semejante a los procesos que creemos que están teniendo lugar en las situaciones de interferencia.

Conclusiones

A lo largo de este capítulo hemos ofrecido una explicación alternativa a la asociativa para el fenómeno de interferencia entre señales y para el papel que el contexto desempeña en el mismo. Esta explicación está basada en los modelos de razonamiento causal, modelos que destacan la influencia de los procesos de orden superior en el aprendizaje y recuperación de las relaciones de causalidad. En línea con estos modelos, proponemos que el fenómeno de interferencia es consecuencia de la búsqueda activa que realizan los sujetos para formar un modelo causal aceptable de la situación, búsqueda activa guiada o restringida por la influencia del conocimiento causal previo. Aunque el modo en que los procesos de orden superior influyen en el aprendizaje y recuperación de las relaciones causales puede ser variado, dos procesos tienen especial relevancia para la explicación que proponemos de la interferencia entre señales: nos referimos al hecho de que los sujetos sean sensibles al rol causal de los eventos y a su capacidad para inferir causas ocultas que intervengan en la expresión de los eventos explícitos.

Las primeras evidencias empíricas parecen apoyar estas hipótesis. Las asimetrías encontradas entre los grupos predictivo y diagnóstico en el experimento descrito no pueden ser explicadas por los modelos asociativos propuestos hasta ahora, siendo los resultados acordes con las predicciones derivadas de los modelos de razonamiento causal. Estos resultados arrojan dudas acerca de la capacidad explicativa de los modelos planteados hasta ahora para dar cuenta del fenómeno de interferencia entre señales.

Aunque no disponemos de resultados directos que nos permitan asegurar que los sujetos estén infiriendo causas ocultas, estos resultados sí se pueden añadir a la evidencia ya existente en la que se muestra que los sujetos son sensibles al rol causal de los eventos (e.g., López et al., en prensa; Waldmann, 2000, 2001). Hay que señalar, en el caso de la sensibilidad al rol causal de los eventos, la existencia de estudios en los que los resultados avalan a los modelos asociativos frente a los modelos de razonamiento causal (e.g., Cobos et al., 2002). No es el objetivo del presente capítulo analizar las razones por las que se han obtenido datos contradictorios cuando se ha evaluado la sensibilidad al rol causal de los sujetos, solo señalaremos, en la línea de López et al. (en prensa) o Waldmann y Walker (en prensa) que ambos tipos de procesos parecen coexistir, siendo variables como la relevancia percibida del conocimiento causal o la demanda de recursos cognitivos las que permiten la puesta en marcha de procesos de orden superior.

El modelo teórico propuesto puede ser extendido a fenómenos similares explicados hasta ahora mediante modelos asociativos. Por ejemplo, hemos expuesto en el apartado anterior cómo la inferencia de factores causales asociados al contexto puede dar cuenta de ciertos efectos contextuales mostrados en la literatura, como por ejemplo el efecto de renovación o reinstauración. Asimismo un análisis equivalente al realizado en este capítulo sobre la interferencia entre señales podría realizarse en el caso de la interferencia entre resultados. Al igual que la interferencia entre señales, en la interferencia entre resultados se produce una situación que no puede ser representada mediante un modelo causal legal sin suponer la participación de causas intervinientes. Al igual que en la interferencia entre señales, en la interferencia entre resultados se ha probado la importancia que tiene el contexto en la codificación y recuperación de la información. Por tanto, creemos que resulta pertinente comprobar si nuestra hipótesis basada en los modelos de razonamiento causal puede ser extendida a los casos de interferencia entre resultados, tarea en la que nos encontramos inmersos en estos momentos.

Nota de los autores

Este trabajo ha sido financiado por la ayuda para proyectos de investigación HUM 0105, concedida por la Junta de Andalucía, y por la ayuda para la formación de personal docente e investigador en las universidades andaluzas N° 4381, concedida a DLR por la misma institución. La correspondencia correspondiente a este trabajo debe dirigirse a Francisco López, Departamento de Psicología Básica, Facultad de Psicología, Universidad de Málaga, Campus de Teatinos. Málaga29071. España. E-mail: frjlopez@uma.es.

Referencias

- Ahn, W., Kalish, C. W., Medin, D. L. y Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition*, *54*, 299-352.
- Ahn, W. y Kalish, C. (2000). The role of mechanism beliefs in causal reasoning. En R. Wilson, y F. Keil (Eds.) *Cognition and explanation*, (p. 199-226) Cambridge, MA: MIT Press.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*, 80-99.
- Castro, L., Ortega, N. y Matute, H. (2002). Proactive interference in human predictive learning. *International Journal of Comparative Psychology*, *15*, 55-68.
- Cobos, P. L., López, F. J., Caño, A., Almaraz, J. y Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 331-346.
- Dennis, M. J. y Ahn, W. (2001). Primacy in causal strength judgments: The effect of initial evidence for generative versus inhibitory relationships. *Memory & Cognition*, *29*, 152-164.
- Escobar, M., Pineño, O. y Matute, H. (2002). A comparison between elemental and compound training of cues in retrospective reevaluation. *Animal Learning & Behavior*, *30*, 228-238.
- Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T. y Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1-31.
- Glymour, C. (2001). *The mind's arrows*. Cambridge, MA: MIT Press.
- Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Science*, *7*, 43-48.
- Goldvarg, Y. y Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565-610.
- Hagmayer, Y. y Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*, 1128-1137.
- Johnson-Laird, P. N., Legrenzi, P., Girotto, V. y Legrenzi, M. Z. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531.
- Johnson-Laird, P. N., Girotto, V. y Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*, 640-661.
- Kushnir, T., Gopnik, A., Schulz, L. y Danks, D. (2003). Inferring hidden causes. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 699-703). Boston: Cognitive Science Society.
- Lagnado, D. A. y Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 856-876.
- Lewandowsky, S. (1994). On the relation between catastrophic interference and generalization in connectionist networks. *Journal of Biological Systems*, *2*, 307-333.
- López, F. J., Cobos, P. L. y Caño, A. (*en prensa*). Associative and causal reasoning accounts of causal induction: symmetries and asymmetries in predictive and diagnostic inferences. *Memory and Cognition*.
- Luque, D., Cobos, P. L., López, F. J. y Caño, A. (2004). Interference between effects but not between causes in designs of interference between cues: importance of previous causal knowledge. *16th Meeting of the Spanish Society of Comparative Psychology/ 12th Biennial Meeting of the International Society of Comparative Psychology*. Oviedo (España).
- Matute, H. y Pineño, O. (1998a). Stimulus competition in the absence of compound conditioning. *Animal Learning & Behavior*, *26*, 3-14.
- Matute, H. y Pineño, O. (1998b). Cue competition in the absence of compound training: Its relation to paradigms of interference between outcomes. En D. L. Medin (Ed.), *The Psychology of Learning and Motivation (Vol.38)* (pp. 45-81). San Diego: Academic Press.
- Miller, R. R. y Escobar, M. (2002). Associative inference between cues and outcomes presented together and presented apart: An integration. *Behavioural Processes*, *57*, 163-185.
- Ortega, O. y Matute, H. (2000). Interference between elementally trained stimuli can take place in one trial. *Learning and Motivation*, *31*, 323-344.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Clarendon Press.

- Pineño, O. y Matute, H. (2000). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology*, *13*, 16-33.
- Pineño, O., Ortega, N. y Matute, H. (2000). The relative activation of the associations modulates interference between elementally trained cues. *Learning and Motivation*, *31*, 128-152.
- Rosas, J. M., Vila, N. J., Lugo, M. y López, L. (2001). Combined effect of context change and retention interval on interference in causality judgment. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 153-164.
- Slamecka, N. J. y Ceraso, J. (1960). Retroactive and proactive inhibition of verbal learning. *Psychological Bulletin*, *57*, 449-475.
- Sloman, S. A. y Lagnado D. A. (2005). Do we “do”? *Cognitive Science*, *29*, 5-39.
- Sobel, D., Tenenbaum, J. B. y Gopnik, A. (2004). Children’s causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303-333.
- Tangen, J. M., y Allan, L. G. (2004). Cue-interaction and judgments of causality: Contributions of causal and associative processes. *Memory and Cognition*, *32*, 107-124.
- Vila, N. J. y Rosas, J. M. (2001a). Reinstatement of acquisition performance by presentation of the outcome after extinction in causality judgments. *Behavioural Processes*, *56*, 147-154.
- Vila, N. J. y Rosas, J. M. (2001b). Renewal and spontaneous recovery after extinction in a causal learning task. *Mexican Journal of Behavior Analysis*, *27*, 79-96.
- Waldmann, M. R. (1996). Knowledge-based causal induction. En D. R. Shanks, K. J. Holyoak, y D. L. Medin (Eds.), *The Psychology of Learning and Motivation, Vol. 34: Causal learning* (pp. 47-88). San Diego: Academic Press.
- Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 53-76.
- Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: evidence from an overshadowing paradigm. *Psychonomic Bulletin and Review*, *8*, 600-608.
- Waldmann, M. R. y Hagmayer, Y. (*en prensa*). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Waldmann, M. R., y Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222-236.
- Waldmann, M. R. y Walker, J. M. (*en prensa*). Competence and performance in causal learning. *Learning & Behavior*.

Nuevas perspectivas teóricas para el estudio de la extinción en juicios de causalidad

Miguel A. Vadillo y Helena Matute

Universidad de Deusto, España

Introducción: de Rescorla-Wagner a Bouton

Aunque la mayor parte de las teorías científicas pasa a la posteridad por su éxito en la explicación de los hechos conocidos y en la predicción de nuevos fenómenos, no menos importantes son todas aquellas hipótesis, teorías y modelos cuyo mayor mérito reside en impulsar la investigación y el desarrollo de una disciplina, aunque sea a causa de los errores e incorrecciones que puedan contener. Tal vez en psicología, una ciencia aún joven, no abundan teorías del primer tipo (y es difícil saber si las hay en las ciencias con más solera), pero sin duda hemos tenido y seguimos teniendo muchas de las segundas. Y si en el campo de la psicología del aprendizaje tuviéramos que elegir el error más fructífero, no sería de extrañar que más de uno se decidiera por la teoría del condicionamiento animal propuesta por Rescorla y Wagner (1972). Surgido tras el renacer del estudio del condicionamiento en los 60, el modelo de Rescorla y Wagner permitió explicar por medio de una fórmula sencilla e intuitivamente atractiva algunos de los fenómenos por aquel entonces recién descubiertos que tenían en jaque a la visión tradicional del aprendizaje animal que se había mantenido desde los tiempos de Pavlov, fenómenos tales como el bloqueo (Kamin, 1968), la validez relativa (Wagner, Logan, Haberlandt y Price, 1968) o la sensibilidad de la respuesta condicionada a la contingencia EC-EI (Rescorla, 1968). De la misma forma, el modelo realizaba predicciones completamente novedosas, tales como la sumación o la sobre-expectación (Kremer, 1978; Rescorla, 1970). Bajo su influencia se establecieron las líneas actuales de investigación (para una revisión, véase Dickinson, 1980), se desarrollaron la mayor parte de las teorías modernas del aprendizaje (Mackintosh, 1975; Pearce, 1987; Pearce y Hall, 1980; Wagner, 1981), y lo que es más importante para el presente trabajo, se extendió la perspectiva asociativa al estudio de la inducción causal en la especie humana (Alloy y Abramson, 1979; Dickinson, Shanks y Evenden, 1984; López, Cobos, Caño y Shanks, 1998).

En el condicionamiento animal, este modelo supone que cada vez que un estímulo condicionado (EC) va seguido de un estímulo incondicionado (EI) inesperado se fortalece una asociación entre las representaciones mentales de los dos estímulos. Por el contrario, si el EC no va seguido de un EI que se espera, la asociación entre las representaciones del EC y el EI se debilita. La respuesta condicionada (RC) que un animal da ante la presentación de un EC estaría determinada directamente por la fuerza de la asociación EC-EI. Así, si un animal ve que en varias ocasiones la aparición de una luz (EC) va seguida de una descarga (EI) que no espera, se desarrollará una asociación luz-descarga que hará que el animal acabe por tener miedo a la luz (RC) y la intensidad de este miedo dependerá, según el modelo de Rescorla y Wagner (1972), de la fuerza de esta asociación.

Las similitudes encontradas entre el condicionamiento animal y el aprendizaje causal humano (Dickinson y cols., 1984) hicieron que se propusiera la hipótesis de que las personas aprenden las relaciones de causa-efecto de la misma forma que los animales aprenden las relaciones EC-EI: cuando una clave va seguida de una consecuencia que no se espera se fortalece una asociación clave-consecuencia y las personas pueden basarse en la fuerza de esta asociación a la hora de juzgar si la clave y la consecuencia guardan entre sí una relación causal. Esta idea según la cual el conocimiento causal de las personas se basa en asociaciones de representaciones mentales es el centro de toda una familia de modelos de aprendizaje causal conocidos como modelos asociativos (Dickinson y Burke, 1996; Van Hamme y Wasserman, 1994).

A pesar de lo atractivo del modelo de Rescorla y Wagner (1972), algunos de sus errores obligaron a los investigadores a proponer ideas alternativas que a veces fueron más allá de la mera revisión o matización. De entre sus muchas limitaciones (para una revisión, véase Miller, Barnet y Grahame, 1995) tal vez la más importante sea la equivalencia que establecía entre aprendizaje y ejecución, que le obligaba a asumir que cuando los organismos no mostraban evidencia de haber aprendido era porque, de hecho, no habían aprendido nada. Así, mientras que el modelo de Rescorla y Wagner (1972) explicaba la competición de claves (esto es, el bajo nivel de RC que provoca un EC que se ha entrenado junto con un segundo EC que predice mejor el EI) como un fallo en el desarrollo de una asociación entre el EC crítico y el EI, la investigación posterior mostró que los animales sí aprenden que existe una relación entre el EC y el EI en estos paradigmas, aunque este aprendizaje no se exprese conductualmente si no se dan ciertas condiciones (por ejemplo, Batsell, 1997; Cole, Denniston y Miller, 1996; Miller y Grahame, 1991; Miller, Jagielo y Spear, 1993; Pineño, Urushihara y Miller, 2005). El impacto de todos estos estudios difícilmente puede exagerarse, ya que mostraban que el mayor triunfo del modelo de Rescorla y Wagner, la explicación de los fenómenos de competición de claves, era puramente ilusorio y se realizaba por medio de un mecanismo poco plausible.

Más importante para los propósitos del presente trabajo es la explicación insatisfactoria que se ofrecía de la extinción desde el modelo de Rescorla y Wagner (1972). Según este modelo, el decremento en la respuesta condicionada que se observa cuando tras una fase de adquisición (es decir, de emparejamientos del EC con el EI) el EC comienza a presentarse sistemáticamente sin el EI se debe a un debilitamiento o “desaprendizaje” de la asociación EC-EI. De nuevo se trataba de una explicación forzada por la equivalencia asumida entre aprendizaje y ejecución: si la respuesta se reduce durante la extinción, este descenso necesariamente se debe a la desaparición de la asociación que dio lugar a la respuesta durante la adquisición. No hacía falta recurrir a nuevos estudios para ver que se trataba de una explicación errónea. Los propios experimentos de Pavlov (1927) muestran que la respuesta extinguida reaparecía si se dejaba pasar cierto tiempo tras la extinción, fenómeno conocido como recuperación espontánea (véase también Rosas y Bouton, 1996), o si el EC se presentaba junto con un estímulo novedoso, lo que se conoce con el nombre de desinhibición externa. En ambos casos queda claro que la extinción no puede

deberse a la desaparición de una asociación. Otros fenómenos, como la renovación de la respuesta (la recuperación de la respuesta condicionada que tiene lugar si se presenta el EC en un contexto diferente del de la extinción; véase Bouton y Bolles, 1979; Nakajima, Tanaka, Urushihara e Imada, 2000) o la reinstauración (la recuperación de la respuesta condicionada debida a la presentación el EI aislado tras la fase de extinción; véase Rescorla y Heth, 1975) muestran también que los organismos siguen manteniendo un registro de la relación EC-EI tras la extinción. La idea de que la adaptación a las nuevas contingencias del entorno conlleva la destrucción o el deterioro de la información aprendida anteriormente no es un supuesto exclusivo del modelo Rescorla-Wagner, sino que afecta en mayor o menor grado a gran parte de los modelos de aprendizaje, incluyendo a las reglas de propagación hacia atrás del error utilizadas actualmente en los modelos conexionistas de procesos cognitivos (Hetherington y Seidenberg, 1989; Lewandowsky, 1991; McCloskey y Cohen, 1989; Ratcliff, 1990).

En el caso de las teorías del condicionamiento animal, los fenómenos de recuperación espontánea, desinhibición externa, renovación y reinstauración provocaron el desarrollo de nuevas teorías asociativas que entendían la extinción no como una destrucción del conocimiento adquirido durante la fase de adquisición, sino como la formación de una nueva asociación que podía o no interferir con la expresión de la asociación EC-EI dependiendo de en qué circunstancias se presentara el EC. Para Bouton (1993; véase también Bouton, 1997) el procedimiento de extinción hacía que el EC se convirtiera en un predictor ambiguo del EI, puesto que al principio el EC predice el EI y después deja de hacerlo. Los animales resolverían esta ambigüedad prestando atención al contexto en busca de elementos que les puedan ayudar a decidir si el EC irá o no seguido del EI en un ensayo determinado. En concreto, Bouton supone que cuando un EC deja de ir seguido del EI, los animales empiezan a codificar el contexto, de modo que aprenden que la presentación del EC en ese contexto no predice la aparición del EI. Según Bouton, esta codificación del contexto se realiza únicamente cuando una clave se convierte en ambigua (esto es, el animal no tiene en cuenta el contexto a menos que perciba que la clave es un predictor ambiguo y que el contexto puede ayudarle a resolver esta ambigüedad), lo que implica que no tiene lugar durante la fase de adquisición (en la que el EC predice sistemáticamente el EI). Esto explica que no sea necesario volver a la situación original de los emparejamientos EC-EI para que se recupere la respuesta condicionada. Como el contexto sólo se codifica en la fase de extinción, lo que se aprende en esa fase se manifiesta únicamente en ese contexto y basta con que se salga del contexto en el que tuvo lugar la extinción para que la respuesta reaparezca. Si asumimos que el propio paso del tiempo o la presentación de estímulos ajenos a la fase de extinción (por ejemplo un estímulo neutro novedoso, o el propio EI utilizado durante la fase de adquisición) pueden hacer que el animal perciba que el contexto ha cambiado con respecto a la fase de extinción, este mecanismo nos permite explicar no sólo la renovación, sino también la recuperación espontánea, la reinstauración y la desinhibición externa.

La aparición de este modelo supuso un avance tal con respecto a los modelos anteriores de condicionamiento clásico que su aplicación a los fenómenos de aprendizaje causal en humanos apenas se hizo esperar. La investigación sobre juicios de causalidad realizada desde esta perspectiva teórica ha sido extremadamente fructífera (para una revisión, véase Vila, Alvarado, Jara y Flores, 2003) y ha venido, en muchos casos, a apoyar la validez del modelo de Bouton (1993, 1997). Este éxito nos ha invitado a nosotros mismos en varias ocasiones a adoptar este modelo como guía para nuestros trabajos empíricos (Matute, Vegas y De Marez, 2002; Vadillo, Vegas y Matute, 2004). Sin embargo, también se han encontrado algunos resultados que demuestran sus limitaciones y que nos obligan a revisar la validez de este marco teórico en su aplicación a determinados fenómenos del aprendizaje causal. A lo largo de este capítulo mostraremos algunos ejemplos de fenómenos que podrían explicarse desde el modelo de Bouton y nos fijaremos después en las limitaciones del mismo a la luz de algunos resultados recientes. Finalmente, propondremos algunas perspectivas teóricas complementarias que podrían ayudarnos a superar estas limitaciones, y en general a avanzar en la comprensión del aprendizaje causal humano.

Recuperación de la respuesta en el aprendizaje humano: alcance y límites del modelo de Bouton

Los experimentos realizados con el paradigma de adquisición-extinción y con otros paradigmas relacionados (como la inhibición latente o la inversión de la discriminación) en el ámbito del aprendizaje de relaciones causales en humanos no han hecho sino confirmar la insuficiencia de los modelos asociativos similares al de Rescorla y Wagner (1972) para dar cuenta de los resultados. Hasta donde alcanza nuestro conocimiento, todos estos fenómenos (recuperación espontánea, renovación y reinstauración) se han replicado en humanos, con la posible excepción de la inhibición externa (García-Gutiérrez y Rosas, 2003a, 2003b; Paredes-Olay y Rosas, 1999; Romero, Vila y Rosas, 2003; Rosas, Vila, Lugo y López, 2001; Vadillo y cols., 2004; Vila, Romero y Rosas, 2002; Vila y Rosas, 2001). Pero, lo que es más importante, a la luz de la evidencia obtenida en experimentos de aprendizaje causal con humanos, se advierte que la actuación de las personas en situaciones con claves contradictorias es mucho más flexible incluso de lo que los experimentos con animales muestran a primera vista, y que la recuperación espontánea, la renovación y la reinstauración son sólo algunas de las manipulaciones que producen una recuperación de la respuesta (en este caso, de los juicios de causalidad) en el paradigma de la extinción. Como veremos a continuación, algunos de estos fenómenos pueden explicarse adaptando algunas ideas del modelo de Bouton (1993, 1997) con algunos supuestos adicionales. Otros, por el contrario, suponen un claro desafío a cualquier intento de ser explicados desde esta perspectiva teórica.

Resultados explicables desde el modelo de Bouton. Uno de los hallazgos que amplía el espectro de resultados que hay que tener en cuenta a la hora de abordar la explicación de la extinción en los juicios de causalidad es el efecto que tiene la frecuencia con

que se solicitan los juicios. Dicho efecto consiste, en líneas muy generales, en que los juicios de los participantes se hacen más sensibles a la información más reciente si dichos juicios se solicitan con cierta frecuencia durante el entrenamiento que si se solicitan únicamente tras finalizar el entrenamiento (véanse Catena, Maldonado y Cándido, 1998; Catena, Maldonado, Mejías y Frese, 2002; Collins y Shanks, 2002; Matute y cols., 2002; Vadillo y cols., 2004). En un paradigma de adquisición-extinción esto significa que los juicios de los participantes son al final menores (más sensibles a los ensayos de extinción) si se pide a los participantes que den un juicio sobre la relación percibida entre clave y resultado en todos los ensayos de entrenamiento que si únicamente se les pide un juicio tras la fase de extinción. La importancia del fenómeno en cuestión es doble. Por una parte, supone un desafío para un gran número de modelos de aprendizaje causal (incluyendo tanto a modelos asociativos como a modelos estadísticos o de reglas) que no otorgan importancia alguna a un factor tan aparentemente trivial como la frecuencia del juicio. Y por otra parte, este efecto de la frecuencia del juicio ha dado lugar a la aparición de una nueva generación de modelos de aprendizaje llamados modelos de revisión de creencias (Catena y cols., 1998, 2002; Hogart y Einhorn, 1992).

Aunque sea difícil entender este efecto desde los modelos asociativos que tradicionalmente se han invocado a la hora de dar cuenta de los juicios de causalidad (p.ej., Dickinson y Burke, 1996; Rescorla y Wagner, 1972; Van Hamme y Wasserman, 1994), puede resultar relativamente sencillo explicarlo complementando el modelo de Bouton (1993, 1997) con algunas ideas adicionales. Según Bouton, que se observe o no poca respuesta condicionada tras un tratamiento de extinción dependerá de la semejanza entre el contexto (físico o temporal) en que tuvo lugar la extinción y el contexto en que se realiza la prueba. Desde esta perspectiva, podríamos entender que en un experimento de aprendizaje causal, el hecho de solicitar un juicio por primera vez tras un entrenamiento de adquisición-extinción introduce un cambio con respecto al contexto de extinción que hace que los ensayos de extinción tengan un menor impacto en la respuesta. Por el contrario, si los participantes deben dar estos juicios durante todo el entrenamiento, y no sólo en el ensayo final de prueba, entonces la presentación de la pregunta con la que se recogen los juicios no producirá cambio alguno con respecto a la fase de extinción (ya que esa pregunta se presenta también durante la extinción), lo cual hará que los ensayos de extinción tengan un mayor impacto en los juicios y se observe en consecuencia un mayor grado de extinción (para un desarrollo más detallado de esta idea, véanse Matute y cols., 2002; Vadillo y cols., 2004).

Fenómenos problemáticos para el modelo de Bouton. Otras manipulaciones muestran tener efectos que plantean problemas más graves a esta perspectiva. Entre ellas se encuentra la influencia que diversos tipos de instrucciones pueden ejercer sobre los juicios de los participantes. Por ejemplo, Matute y colaboradores (2002) mostraron que la extinción de los juicios de causalidad desaparece si entre la extinción y el ensayo de prueba se introducen instrucciones que invitan al participante a tener en cuenta tanto los ensayos de

adquisición como los de extinción a la hora de elaborar el juicio. En principio, y si se toma en cuenta sólo esta evidencia, parece que el modelo de Bouton podría adaptarse fácilmente para dar cuenta de estos resultados, asumiendo simplemente que la presentación de esa pantalla en la que se presentan las instrucciones supone una ruptura con el contexto físico de la extinción que haría que dicha fase de extinción tuviera menos influencia sobre el juicio del ensayo de prueba.

Un primer problema de esta explicación reside en que, de acuerdo con ella, siempre que las instrucciones anteriores a la prueba supongan una interrupción entre la fase de extinción y la prueba, su contenido semántico debería ser irrelevante. Sin embargo, tal y como cabría esperar, el contenido de esas instrucciones afecta claramente a los juicios que dan los participantes. Sin ir más lejos, si en lugar de presentar unas instrucciones que inviten a tener en cuenta tanto la fase de adquisición como la de extinción se presentan instrucciones que inviten a tener únicamente en cuenta la información recibida en la fase de extinción, se observarán juicios bajos y no una recuperación de la respuesta (Collins & Shanks, 2002; Vila, Alvarado, Jara, Flores y Flores, 2002). Vemos pues que el contenido informativo de las instrucciones puede tener mayor impacto que la interrupción o el cambio de contexto que pueda provocar la introducción de esa pantalla entre la extinción y la prueba.

Un segundo problema del modelo de Bouton para explicar el efecto de las instrucciones es que estas pueden tener el mismo efecto si se presentan entre la fase de extinción y la prueba que si se presentan al principio del entrenamiento, antes de la fase de adquisición (Alvarado y Vila, 2004). Es decir, la respuesta de los participantes se recuperará no sólo cuando se les da instrucciones para tener en cuenta toda la información recibida justo antes del ensayo de prueba, sino también cuando se les insta a comportarse de esa manera desde el principio del entrenamiento. Sin embargo, este resultado no es fácil de explicar desde el modelo de Bouton, ya que en la ausencia de un cambio de contexto entre la fase de extinción y la fase de prueba, siempre debería esperarse una respuesta más sensible a la fase de extinción.

Otra manipulación que muestra un problema semejante es el efecto del tipo de pregunta sobre el progreso de la extinción en las tareas de juicios de causalidad. Matute y colaboradores (2002) encontraron que la extinción era notablemente más acusada y acelerada si en cada ensayo los participantes tenían que responder a una pregunta de tipo predictivo sobre la relación entre la clave y el resultado (valorar en qué medida cabía esperar el resultado si se había presentado la clave) que si tenían que responder a una pregunta de tipo causal (valorar hasta qué punto era plausible que la clave fuera la causa del resultado). Es decir, el impacto de lo aprendido en la fase más reciente (en este caso, en la fase de extinción) fue mayor sobre los juicios de predicción que sobre los juicios de causalidad. Este resultado difícilmente puede explicarse desde una perspectiva boutoniana sin introducir numerosos supuestos adicionales. Por una parte, no se trata de una recuperación de la respuesta tras la extinción, sino de una resistencia a la extinción que surge cuando se utiliza una pregunta de tipo causal para solicitar los juicios de los participantes. Por otra

parte, en ningún momento se da una interrupción o una alteración del contexto dentro de la fase de extinción o entre la extinción y la prueba que pueda invitar a tal nivel alto de respuestas. Al igual que en el caso de la manipulación de las instrucciones, el efecto del tipo de pregunta sólo puede entenderse atendiendo al significado de las preguntas utilizadas. En un procedimiento de adquisición y extinción, la utilidad de la clave para predecir el resultado cambia en cada fase, lo que hace que tenga sentido fijarse únicamente en la información más reciente si se trata de hacer una predicción. Sin embargo, una relación causal entre dos eventos se caracteriza por ser una propiedad estable, lo que puede llevar a los participantes a no sobrestimar la importancia de ninguna de las fases de entrenamiento a la hora de elaborar sus juicios de causalidad. El hecho de que los juicios de causalidad se vean alterados en función del tipo de pregunta utilizado supone un problema no sólo para modelos asociativos concretos, como el de Bouton, sino que desafía la propia idea de que los juicios de causalidad puedan entenderse atendiendo única y exclusivamente a procesos asociativos de cualquier naturaleza (Matute, Arcediano y Miller, 1996; Vaddillo, Miller, y Matute, en prensa; pero véase también Cobos, Caño, López, Luque y Almaraz, 2000).

Así pues, si bien parte de la evidencia obtenida en los experimentos de aprendizaje causal en humanos puede entenderse dentro del marco propuesto por Bouton (1993, 1997) sin necesidad de introducir grandes modificaciones, lo cierto es que algunos estudios muestran resultados que no pueden entenderse desde esta perspectiva sin cambios radicales. Para dar cuenta de estos resultados problemáticos sería interesante adoptar perspectivas teóricas nuevas que nos permitieran contemplar estos resultados desde una óptica diferente. Lamentablemente, en los estudios centrados en la adquisición y extinción apenas se han mencionado propuestas teóricas diferentes de la asociativa y cuando se ha hecho ha sido superficialmente. Sin embargo, otros paradigmas de aprendizaje causal, principalmente los de competición de estímulos, han destacado por su fecundidad a la hora de dar lugar a explicaciones teóricas alternativas. Tal vez, como se mostrará a continuación, una inspección atenta de estas propuestas nos permita detectar claves que podrían ser aplicadas a la explicación de los fenómenos de adquisición y extinción mencionados hasta ahora.

Los modelos inferenciales de aprendizaje causal

Aunque hasta aquí nos hemos centrado en las limitaciones de las teorías asociativas para explicar los fenómenos de adquisición y extinción, lo cierto es que han surgido problemas similares en el estudio de otros fenómenos que tradicionalmente se habían abordado desde la perspectiva asociativa. Como ya se ha mencionado en la introducción, el modelo asociativo de Rescorla y Wagner (1972) surgió principalmente para dar cuenta de diversos fenómenos de condicionamiento animal que posteriormente recibieron el nombre genérico de fenómenos de competición de claves (Kamin, 1968; Wagner y cols., 1968; Wasserman, 1974). Cuando se vio que estos fenómenos aparecían también en los experimentos de aprendizaje causal humano (Dickinson y cols., 1984), surgió pronto la hipótesis

de que los mismos procesos asociativos podrían subyacer tanto al condicionamiento animal como a los juicios de causalidad. Sin embargo, durante los últimos años se ha acumulado cierta cantidad de resultados que cuestionan esta idea de que los fenómenos de competición de claves que se observan en el aprendizaje causal humano puedan entenderse como resultado de un mecanismo asociativo similar al propuesto por Rescorla y Wagner (1972).

El estudio de De Houwer, Beckers y Glautier (2002) sobre el bloqueo en aprendizaje causal fue pionero en algunos sentidos. En esta serie de experimentos, los participantes debían aprender la relación que existía entre el lanzamiento de varias armas en un vídeo juego y la destrucción de un tanque en un campo de minas que aparecía en la pantalla del ordenador. Durante la primera fase se mostraba que el disparo de un arma, A, iba seguido de la destrucción de un tanque (A+). Durante la segunda fase, en algunos ensayos este arma se disparaba junto con una segunda, T, todo ello seguido de la destrucción del tanque (AT+) y, de la misma forma, en otros ensayos se disparaban otras dos armas, K y L, que también iban seguidas de la destrucción del tanque (KL+). En esta preparación, puede decirse que se ha obtenido bloqueo si los participantes juzgan que el grado de relación entre el arma T y la destrucción del tanque es menor que la relación entre el arma L y la destrucción del tanque. Dado que tanto T como L han sido emparejados con la explosión del tanque el mismo número de veces, las diferencias en los juicios sobre una y otra tienen que deberse a que T se ha entrenado junto con otra clave, A, que ya permitía predecir el resultado. En otras palabras, la relación entre T y el resultado se ve debilitada debido a la competición de otra clave, A, por convertirse en predictora de ese mismo resultado.

La principal aportación de De Houwer y colaboradores (2002) consiste en haber demostrado que la cantidad de bloqueo observada cambia radicalmente dependiendo de la información que se da sobre las claves (A, T, K y L) y sobre los resultados (la explosión del tanque, “+”). En concreto, si se les dice a los participantes que la intensidad de la explosión del tanque podría haber sido mayor en todos los ensayos, el bloqueo es más fuerte. Según los autores, es difícil ver cómo podría explicarse este fenómeno desde modelos como el de Rescorla y Wagner (1972), ya que para estos modelos el procesamiento de la información está guiado únicamente por las propiedades concretas de los datos, independientemente de cualquier información de carácter abstracto sobre claves y resultados. Sin embargo, los resultados se explican fácilmente si se asume que los participantes han abordado la tarea como si se tratara de un problema lógico. Desde esta perspectiva, los participantes realizarían un razonamiento del siguiente tipo: “Si T es una causa de la explosión del tanque, entonces disparar a la vez A y T debería dar lugar a una explosión más intensa que disparar sólo A. Sin embargo, esto no sucede y no puede deberse a un efecto techo (puesto que me han dicho que la explosión podría haber sido más intensa). Luego, T no es una causa de la explosión.”

Por tanto, esta nueva perspectiva consiste en asumir que los procesos de competición de claves no se deben a la acción de un mecanismo asociativo, sino a procesos inferenciales semejantes a los que tienen lugar cuando las personas razonan deductivamente. Aun-

que la idea de que el aprendizaje causal está guiado por procesos de razonamiento no es del todo novedosa (véanse por ejemplo Cheng, 1997; Cheng y Novick, 1992; Perales, Catena y Maldonado, 2002; Waldmann, 2001; Waldmann y Holyoak, 1992), ha recibido una atención considerable desde el trabajo de De Houwer y colaboradores (2002) y ha dado lugar a un gran volumen de estudios (Beckers, De Houwer, Pineño y Miller, 2005; De Houwer, 2002; De Houwer y Beckers, 2002a, 2002b, 2003; Lovibond, Been, Mitchell, Bouton y Frohardt, 2003; Mitchell y Lovibond, 2002; Vandorpe, De Houwer y Beckers, en prensa).

Aunque poco mencionado, un aspecto interesante de este nuevo marco teórico es que no resulta incompatible con la perspectiva asociativa (Beckers y cols., 2005; López, Cobos y Caño, en prensa; Perales, Catena y Maldonado, 2004; Sloman, 1996; Vadillo y cols., en prensa). Los procesos inferenciales que según De Houwer y colaboradores (2002) son responsables de los fenómenos de competición de claves bien podrían realizarse manejando información que previamente se ha aprendido por medio de un mecanismo asociativo. Los juicios de causalidad pueden estar basados en un argumento lógico con unas premisas y unas consecuencias, pero algunas de las premisas tienen que extraerse de la propia experiencia (por ejemplo, en el argumento lógico propuesto más arriba hace falta saber que tanto A como AT van seguidos del resultado) y esto podría lograrse por medio de un mecanismo asociativo que extrajera del entorno la información relevante sobre la relación clave-resultado.

También en el caso de los fenómenos de adquisición y extinción es posible asumir la intervención tanto de mecanismos asociativos como de mecanismos inferenciales para llegar a una explicación de los fenómenos presentados en la sección anterior. Algunas de las ideas del modelo de condicionamiento animal de Bouton (1993, 1997) resultan extremadamente interesantes y podrían mantenerse a la luz de los datos: el supuesto de que la relación entre claves y resultados se aprende por medio de procesos asociativos, el almacenamiento independiente de las asociaciones excitatorias y las asociaciones inhibitorias, la recuperación selectiva de las asociaciones a la hora de producir una respuesta observable. Lo que no puede mantenerse a la hora de intentar extrapolar estas ideas al área del aprendizaje causal humano es que los únicos factores que determinan qué asociaciones intervienen en un momento determinado sean elementos relacionados con el contexto. A veces podrá ser el contexto el que determine que se exprese una asociación excitatoria o una inhibitoria y sin duda esto introduce buen grado de flexibilidad y ajuste a las demandas cambiantes del entorno. Pero en otros casos, la recuperación de una u otra asociación (o de ambas) puede depender de factores que nada tienen que ver con el contexto o con las propiedades estimulares del material que se presenta a los participantes. Las expectativas con las que lleguen los participantes al experimento, su conocimiento general sobre las relaciones causales y predictivas, las preguntas que les hagamos y la información que se les dé sobre el propósito de la tarea experimental son también elementos que pueden intervenir a la hora de modular la respuesta.

Aunque en principio estas ideas puedan parecer un tanto ajenas a las hipótesis que tradicionalmente se han manejado en el ámbito de la psicología del aprendizaje, en última instancia no hacen referencia sino a un problema ampliamente discutido en el seno de nuestra disciplina: la distinción entre aprendizaje y ejecución. Muchos modelos de aprendizaje asociativo se han centrado casi exclusivamente en determinar cuáles son los algoritmos por los que los organismos aprenden, sin prestar mucha atención a todos aquellos mecanismos encargados de convertir eso que se ha aprendido en una respuesta observable (por ejemplo, Dickinson y Burke, 1996; Mackintosh, 1975; Pearce y Hall, 1980; Rescorla y Wagner, 1972; Van Hamme y Wasserman, 1994; Wagner, 1981). Esta situación ha invitado a asumir que existe una relación directa entre aprendizaje y conducta: si se observa una respuesta es que se ha aprendido algo y si no se observa respuesta alguna es que o bien no se ha aprendido nada o se ha destruido o cancelado lo que se hubiera aprendido. La constatación de que este supuesto tan atrevido es falso (Batsell, 1997; Bouton y Bolles, 1979; Cole y cols., 1996; Miller y cols., 1993; Miller y Grahame, 1991; Pineño y cols., 2005; Rescorla y Heth, 1975) ha obligado a desarrollar modelos que establecen una clara diferencia entre aprendizaje y ejecución. El de Bouton (1993, 1997) es uno de tales modelos que tratan de enfatizar la diferencia entre lo que el organismo aprende y lo que expresa su conducta. Lo que aquí proponemos es que para poder explicar muchos de los fenómenos observados en el aprendizaje causal humano es necesario profundizar en esta línea postulando nuevos mecanismos que puedan intervenir entre el aprendizaje y la conducta.

La necesidad de confrontación teórica en el estudio del aprendizaje causal

No quisiéramos terminar este estudio sin retomar nuestro punto de partida. Decíamos en la introducción que la función de las teorías científicas no es sólo describir o explicar los fenómenos que observamos, sino también estimular la realización de nuevos estudios y descubrimientos, y sobre todo provocar un debate fructífero entre investigadores. En general, la psicología del aprendizaje causal ha sido un lugar de encuentro de teorías muy diversas, encuentro que ha dado lugar a un intenso debate sobre los fenómenos que tratamos de estudiar y sobre cuál debe ser la manera de abordar su explicación. La confrontación de las perspectivas asociativa (Dickinson y Burke, 1996; Rescorla y Wagner, 1972; Van Hamme y Wasserman, 1994), estadística (Allan, 1980; Cheng, 1997; Cheng y Novick, 1992) e inferencial (De Houwer y cols., 2002) ha sido intensa durante los últimos 15 años y ha supuesto una gran aportación en cuanto a propuestas y descubrimientos científicos. Sin embargo, esta discusión no ha parecido afectar a los estudios sobre extinción e interferencia hasta fecha muy reciente. Con la excepción de unos pocos trabajos (por ejemplo, Luque, Cobos, López y Caño, 2004; Maldonado, Catena, Cándido y García, 1999), la mayor parte de las investigaciones se han inspirado principalmente en ideas procedentes del modelo de Bouton (1993, 1997; véanse, sin ir más lejos, nuestros propios trabajos, Matute y cols., 2002; Vadillo y cols., 2004). Aunque esta orientación teórica ha dado sus frutos y nos ha permitido llegar a una comprensión más profunda del aprendizaje causal

humano, la uniformidad en el marco teórico de referencia ha tenido también algunas consecuencias desafortunadas. La confrontación de diferentes teorías es el principal estímulo para recabar nuevo material empírico y reflexionar sobre él. Tal vez la escasa confrontación teórica en el estudio de la extinción en juicios de causalidad sea la responsable de que los estudios sobre este fenómeno sean menos numerosos y también menos citados que los estudios sobre otros fenómenos como, por ejemplo, la competición de claves.

Otra consecuencia no menos problemática es que la adhesión excesiva a un modelo teórico por parte de una comunidad científica puede llevar a que las nuevas interpretaciones se entiendan mal. Por ejemplo, la propuesta de Matute y colaboradores (2002; véase también Vadillo y cols., 2004) de que una recuperación de la respuesta tras la extinción podría deberse a una integración de toda la información recibida durante la adquisición y la extinción a menudo ha sido criticada por otros investigadores acostumbrados a ver la recuperación como una reactivación de lo aprendido en la fase de adquisición. Desde una perspectiva boutoniana, lo que se aprende en la fase de extinción es una asociación entre la clave y el resultado con fuerza negativa. Así pues, si la respuesta se extingue completamente, pero en una fase posterior de prueba se observa que la respuesta es de nuevo mayor que cero, esto tiene que querer decir que lo que se aprendió en la primera fase es lo que ha resultado activado en la fase de prueba. Si se activara lo que se aprendió en ambas fases, entonces la asociación aprendida en la segunda fase (negativa) debería cancelar completamente la que se aprendió en la primera (positiva) y no se observaría respuesta alguna.

Aunque este análisis es factible, una respuesta superior a cero también puede interpretarse como una activación simultánea de lo aprendido en ambas fases de entrenamiento. Por ejemplo, podría asumirse de forma muy intuitiva y natural que durante la primera fase los organismos aprenden que la clave va seguida de la consecuencia y durante la segunda que la clave ya no va seguida de la consecuencia. Si lo que se aprende en ambas fases (la tendencia a responder y la tendencia a no responder) está igualmente activo en un momento determinado, lo que se debería observar no es una ausencia de respuesta, sino una respuesta intermedia o un juicio cercano al 50%. Esta idea general puede implementarse perfectamente en un modelo asociativo, como hacen Matute y colaboradores (2002), pero también es compatible con una visión de carácter estadístico: si los juicios se basaran sólo en la primera fase deberían ser máximos porque la probabilidad del resultado en presencia de la clave durante esa fase es de un 100%; si se basaran sólo en la segunda fase los juicios deberían de ser iguales a cero, porque conforme a lo visto en la segunda fase la probabilidad del resultado en presencia de la clave es cero; y, finalmente, si se basaran en ambas fases, los juicios deberían ser intermedios porque si se toman los ensayos de ambas fases la probabilidad del resultado en presencia de la clave es de un 50%.

Así pues, que la recuperación de la respuesta condicionada o de los juicios de causalidad tras la extinción se entienda como una activación de lo aprendido en la adquisición o como una activación simultánea de lo aprendido en ambas fases, adquisición y extinción,

depende de las ideas previas que uno pueda tener sobre la naturaleza de lo que se aprende durante la extinción. Esto no es negativo per se, pero puede serlo si impide que los investigadores aborden los problemas desde todas las perspectivas posibles.

En algunas ocasiones también se ha señalado que la aparente recuperación de la respuesta que se observa en algunos procedimientos podría deberse a la incertidumbre de los participantes (Matute y cols., 2002; Vadillo y cols., 2004). Sin embargo, esta posible explicación apenas se ha desarrollado en la literatura y permanece más como una explicación alternativa a excluir que como una hipótesis digna de estudio en sí misma. De nuevo esta reticencia a considerar con seriedad una perspectiva alternativa es causa y efecto de la adhesión excesiva a uno o unos pocos modelos.

La revisión aquí presentada sugiere que para avanzar en nuestra tarea es necesario prestar cierta consideración a los procesos de razonamiento que pudieran estar interviniendo en los fenómenos de extinción e interferencia. Independientemente de que se acepte o no la perspectiva general que hemos adoptado aquí, esperamos que la adopción de nuevos marcos teóricos nos ayude no sólo a comprender mejor algunos de los fenómenos que desafían a los modelos actuales, sino también a realizar experimentos novedosos o a reconocer la importancia de procesos que hasta ahora habían permanecido en el olvido.

Nota de los Autores

Este trabajo ha sido financiado por la ayuda para proyectos de investigación PI-2000-12, concedida a H.M. por el Departamento de Educación, Universidades e Investigación del Gobierno Vasco, y por la beca de formación de investigadores BFI01.31, concedida a M.A.V. por la misma institución. Los autores desean dar las gracias a Fernando Blanco por sus valiosos comentarios a una versión previa de este capítulo. La correspondencia relacionada con este trabajo deberá enviarse a Helena Matute, Departamento de Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, España. E-mail: matute@fice.deusto.es.

Referencias

- Allan, L.G. (1980). A note on measurement of contingency between two binary variables in judgement tasks. *Bulletin of the Psychonomic Society*, 15, 147-149.
- Alloy, L. B., y Abramson, L. Y. (1979). Judgements of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 441-485.
- Alvarado, A. y Vila, N. J. (2004). Efectos de las instrucciones en una tarea causal de interferencia con dos consecuencias. *XVI Congreso de la Sociedad Española de Psicología Comparada*. Universidad de Oviedo.
- Batsell, W. R. (1997). Retention of context blocking in taste-aversion learning. *Physiology & Behavior*, 61, 437-446.
- Beckers, T., De Houwer, J., Pineño, O. y Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 238-249.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian conditioning. *Psychological Bulletin*, 114, 80-99.

- Bouton, M. E. (1997). Signals for whether versus when an event will occur. En M. E. Bouton y M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 385-409). Washington, DC: American Psychological Association.
- Bouton, M. E. y Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*, 445-466.
- Catena, A., Maldonado, A. y Cándido, A. (1998). The effect of the frequency of judgment and the type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 481-495.
- Catena, A., Maldonado, A., Megías, J. L. y Frese, B. (2002). Judgement frequency, belief revision, and serial processing of causal information. *Quarterly Journal of Experimental Psychology, 55B*, 267-281.
- Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review, 104*, 367-405.
- Cheng, P. W. y Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365-382.
- Cobos, P. L., Caño, A., López, F. J., Luque, J. L. y Almaraz, J. (2000). Does the type of judgement required modulate cue competition? *Quarterly Journal of Experimental Psychology, 53B*, 193-207.
- Cole, R. P., Denniston, J. C. y Miller, R. R. (1996). Reminder-induced attenuation of the effect of relative stimulus validity. *Animal Learning & Behavior, 24*, 256-265.
- Collins, D. J. y Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition, 30*, 1138-1147.
- De Houwer, J. (2002). Forward blocking depends on retrospective inferences about the presence of the blocked cue during the elemental phase. *Memory & Cognition, 30*, 24-33.
- De Houwer, J. y Beckers, T. (2002a). Higher-order retrospective reevaluation in human causal learning. *Quarterly Journal of Experimental Psychology, 55B*, 137-151.
- De Houwer, J. y Beckers, T. (2002b). Second-order backward blocking and unovershadowing in human causal learning. *Experimental Psychology, 49*, 27-33.
- De Houwer, J. y Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology, 56B*, 345-357.
- De Houwer, J., Beckers, T., y Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology, 55A*, 965-985.
- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.
- Dickinson, A. y Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology, 49B*, 60-80.
- Dickinson, A., Shanks, D. R. y Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A*, 29-50.
- García-Gutiérrez, A. y Rosas, J. M. (2003a). Context change as the mechanism of reinstatement in causal learning. *Journal of Experimental Psychology: Animal Behavior Processes, 29*, 292-310.
- García-Gutiérrez, A. y Rosas, J. M. (2003b). Recuperación de la relación clave-consecuencia por el cambio de contexto después de la interferencia en aprendizaje causal. *Psicológica, 24*, 243-269.
- Hetherington, P. A. y Seidenberg, M. S. (1989). Is there "catastrophic interference" in connectionist networks? *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 26-33). Hillsdale, NJ: Erlbaum.
- Hogart, R. M. y Einhorn, H. J. (1992). Order effects in belief updating: The Belief-Adjustment model. *Cognitive Psychology, 24*, 1-55.
- Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. En M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9-31). Miami, FL: University of Miami Press.
- Kremer, E. F. (1978). The Rescorla-Wagner model: Losses in associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 4*, 22-36.

- Lewandowsky, S. (1991). Gradual unlearning and catastrophic interference: A comparison of distributed architectures. En W. E. Hockley y S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 445-476). Hillsdale, NJ: Lawrence Erlbaum Associates.
- López, F. J., Cobos, P. L. y Caño, A. (en prensa). Associative and causal reasoning accounts of causal induction: Symmetries and asymmetries in predictive and diagnostic inferences. *Memory & Cognition*.
- López, F. J., Cobos, P. L., Caño, A. y Shanks, D. R. (1998). The rational analysis of human causal and probability judgment. En M. Oaksford y N. Chater (Eds.), *Rational models of cognition* (pp. 314-352). Oxford: Oxford University Press.
- Lovibond, P. F., Been, S. L., Mitchell, C. J., Bouton, M. E. y Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition*, 31, 133-142.
- Luque, D., Cobos, P. L., López, F. J. y Caño, A. (2004). Interferencia entre efectos y no entre causas en diseños de interferencia entre señales: Importancia del conocimiento causal previo. *XVI Congreso de la Sociedad Española de Psicología Comparada*. Universidad de Oviedo.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Maldonado, A., Catena, A., Cándido, A. y García, I. (1999). The belief revision model: Asymmetrical effects of no contingency on human covariation learning. *Animal Learning & Behavior*, 27, 168-180.
- Matute, H., Arcediano, F. y Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 182-196.
- Matute, H., Vegas, S. y De Marez, P. J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 714-725.
- McCloskey, M. y Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. En G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 24 (pp. 109-165). San Diego, CA: Academic Press.
- Miller, J. S., Jagielo, J. A., y Spear, N. E. (1993). The influence of retention interval on the US preexposure effect: Changes in contextual blocking over time. *Learning and Motivation*, 24, 376-394.
- Miller, R. R., Barnet, R. C. y Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117, 363-386.
- Miller, R. R. y Grahame, N. J. (1991). Expression of learning. En L. Dachowski y C. F. Flaherty (Eds.), *Current topics in animal learning: Brain, emotion, and cognition* (pp. 95-117). Hillsdale, NJ: Erlbaum.
- Mitchell, C. J. y Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *Quarterly Journal of Experimental Psychology*, 55B, 311-329.
- Nakajima, S., Tanaka, S., Urushihara, K. e Imada, H. (2000). Renewal of extinguished lever-press responses upon return to the training context. *Learning and Motivation*, 31, 416-431.
- Paredes-Olay, C. y Rosas, J. M. (1999). Within-subjects extinction and renewal in predictive judgments. *Psicológica*, 20, 195-210.
- Pavlov, I. P. (1927). *Conditioned reflexes*. London: Oxford University Press.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61-73.
- Pearce, J. M. y Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Perales, J. C., Catena, A. y Maldonado, A. (2002). Aprendizaje de relaciones de contingencia y causalidad: Hacia un análisis integral del aprendizaje causal desde una perspectiva computacional. *Cognitiva*, 14, 15-41.
- Perales, J. C., Catena, A. y Maldonado, A. (2004). Inferring non-observed correlations from causal scenarios: The role of causal knowledge. *Learning and Motivation*, 35, 115-135.

- Pineño, O., Urushihara, K. y Miller, R. R. (2005). Spontaneous recovery from forward and backward blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 172-183.
- Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, 97, 285-308.
- Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, 1-5.
- Rescorla, R. A. (1970). Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation*, 1, 372-381.
- Rescorla, R. A. y Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, 1, 88-96.
- Rescorla, R. A., y Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. En A. H. Black y W. F. Prokasy (eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appelton-Century-Crofts.
- Romero, M. A., Vila, N. J. y Rosas, J. M. (2003). Time and context effects after discrimination reversal in human beings. *Psicológica*, 24, 169-184.
- Rosas, J. M. y Bouton, M. E. (1996). Spontaneous recovery after extinction of a conditioned taste aversion. *Animal Learning & Behavior*, 24, 341-348.
- Rosas, J. M., Vila, N. J., Lugo, M. y López, L. (2001). Combined effect of context change and retention interval on interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 153-164.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Vadillo, M. A., Miller, R. R. y Matute, H. (*en prensa*). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior*.
- Vadillo, M. A., Vegas, S. y Matute, H. (2004). Frequency of judgment as a context-like determinant of predictive judgments. *Memory & Cognition*, 32, 1065-1075.
- Van Hamme, L. J. y Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, 25, 127-151.
- Vandorpe, S., De Houwer, J. y Beckers, T. (*en prensa*). Further evidence for the role of inferential reasoning in forward blocking. *Memory & Cognition*.
- Vila, N. J., Alvarado, A., Jara, E. y Flores, J. (2003). La extinción experimental en aprendizaje causal. En J. Vila, J. Nieto y J. M. Rosas (Eds.), *Investigación contemporánea en aprendizaje asociativo* (pp. 159-174). Jaén, España: Del lunar.
- Vila, N. J., Alvarado, A., Jara, E., Flores, J. y Flores, S. (2002). Efectos de las instrucciones de prueba y del intervalo de retención en la recencia e integración de fases en una tarea de aprendizaje causal con interferencia. *XIV Congreso de la Sociedad Española de Psicología Comparada*. Universidad de Sevilla.
- Vila, N. J., Romero, M. y Rosas, J. M. (2002). Retroactive interference after discrimination reversal decreases following temporal and physical context changes in human subjects. *Behavioural Processes*, 59, 47-54.
- Vila, N. J. y Rosas, J. M. (2001). Reinstatement of acquisition performance by the presentation of the outcome after extinction in causality judgment. *Behavioural Processes*, 56, 147-154.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. En N. E. Spear y R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.
- Wagner, A. R., Logan, F. A., Haberlandt, K. y Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, 76, 171-180.
- Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, 8, 600-608.

- Waldmann, M. R. y Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222-236.
- Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, 103, 284-297.

Contexto y recuperación de la información: ¿qué hace que la recuperación de la información sea dependiente del contexto?

Juan M. Rosas, Ana García-Gutiérrez, María J. F. Abad y José E. Callejas-Aguilera
Universidad de Jaén, España

Tal y como aparece reflejado en todos los capítulos que integran este volumen, el contexto donde la información se aprende parece ser un factor determinante en su recuperación posterior. En este capítulo se analiza el papel que juega el contexto en la recuperación y olvido de la información a través de los resultados más recientes de nuestro laboratorio. El término contexto se ha empleado de forma amplia y hasta cierto punto confusa en la literatura. Los mismos estímulos que se han usado como contexto en algunos experimentos, en otros han jugado el papel de estímulos puntuales directamente relacionados con las consecuencias. Así, en la primera parte de este capítulo se analiza el término contexto y sus diferentes acepciones, proponiendo finalmente una definición operacional del mismo. Después se pasa a analizar el papel del contexto en la recuperación de la información, resaltando los escasos efectos que el cambio de contexto parece tener sobre la información consistente y centrándonos en el efecto que tiene sobre la información interferente. En la segunda parte del capítulo se presentan resultados que sugieren que el cambio de contexto es el factor fundamental que produce el olvido de la información interferente, sea este cambio causado por el simple paso del tiempo (recuperación espontánea), por un cambio explícito en los estímulos contextuales (renovación) o por la presentación de una consecuencia en el contexto de prueba (reinstauración). En la parte final del capítulo se analiza qué hace a la información dependiente del contexto. En la literatura se ha propuesto que el cambio de contexto afecta a un tipo de información específico, bien información inhibitoria, bien información interferente aprendida en segundo lugar. En este capítulo se presentan resultados que sugieren que el que la información se haga dependiente del contexto no depende de que aquella tenga unas características específicas, sino de que la situación lleve a los participantes a prestar atención al contexto.

Definición del contexto

El interés por el estudio de la influencia del contexto en el aprendizaje y la memoria es común a distintas áreas de la psicología. Una gran variedad de investigaciones muestran que el contexto afecta al aprendizaje, a la memoria y a la realización de respuestas previamente aprendidas. Este interés es lógico si tenemos en cuenta que tanto el aprendizaje como el recuerdo o reconocimiento de algo experimentado ocurre inevitablemente en un lugar y en unas condiciones específicas, es decir en un contexto. Por ejemplo, imaginemos una situación en la que le presentamos a una rata un estímulo condicionado, EC (v. gr., un tono) que predice la aparición de un estímulo incondicionado, EI (v. gr., comida). Después de una serie de emparejamientos EC-EI, la mera presentación del EC termina por provo-

car en la rata una conducta de búsqueda del alimento en el lugar en el que éste se dispensa. Esta situación tan simple de aprendizaje es incomprendible si no tenemos en cuenta que los emparejamientos de los estímulos se producen siempre en un contexto particular: una caja de Skinner con paredes y suelo de metal (contexto externo), o un estado biológico del organismo, como puede ser el hambre (contexto interno). En la rata, el episodio de memoria de la relación EC-EI puede incluir a su vez información acerca del contexto en el que dicha relación se ha adquirido, igual que en el caso de un ser humano el recuerdo de una lista de palabras aprendidas en un laboratorio puede incluir información del lugar donde ésta se aprendió e incluso de las características del experimentador que nos la presentó.

Por tanto, el contexto es un marco de referencia en el que se produce un acontecimiento relevante para un organismo. En muchas ocasiones, el contexto puede actuar prediciendo el acontecimiento (v. gr., si nos encontramos en el Reino Unido podemos suponer que se hablará inglés), pero también puede dar al acontecimiento un significado influyendo en lo que aprendemos, la forma en que lo hacemos y en cómo recordemos experiencias pasadas. Por ejemplo, si nos encontramos sentados en un parque, el significado de la palabra “banco” que se activará será el de “asiento, con respaldo o sin él, en que pueden sentarse varias personas”. No obstante, si lo que se describe es una situación en la que se está realizando una transferencia bancaria, el significado de esta palabra será “establecimiento público de crédito, constituido en sociedad por acciones”. Lo que entendamos por “banco” dependerá del contexto concreto en el que se nos presente dicha palabra.

A la hora de aproximarnos a una definición del contexto en una situación experimental de aprendizaje, podemos establecerla conceptualmente desde un punto de vista estructural o desde una perspectiva funcional. Estructuralmente, el contexto se refiere a todos los aspectos de un ambiente experimental que están presentes junto con el EC o el material que se ha de aprender. Entre estos aspectos se incluye el contexto asociativo o cognitivo de lo que se ha aprendido antes, así como el contexto ambiental definido por la localización, tiempo y características específicas de la situación experimental. En este tipo de contexto se incluye tanto el ambiente externo que proporciona el lugar de aprendizaje (estímulos sensoriales), como el ambiente interno (v. gr., estímulos interoceptivos, estado emocional de sujeto, ritmo circadiano).

Desde el punto de vista funcional, el contexto se usa para hacer referencia a cualquier estímulo que influye en el control que ejercen otros estímulos (v. gr., EC). Teniendo en mente la definición funcional del contexto, se hace preciso distinguir entre contexto incidental y contexto intencional (Baddeley, 1992). La diferencia básica entre estas dos formas de contexto es el tipo de procesamiento que reciben. En la mayoría de los estudios que versan sobre la memoria animal y humana el contexto cumple un papel incidental, definido como el conjunto de estímulos ambientales que está presente durante la ejecución o el procesamiento de la tarea experimental, pero que es secundario a éstos. Sin embargo, en algunas investigaciones el contexto toma un papel distinto, adopta una función princi-

pal. Se entrena a los sujetos para que realicen distintas tareas en contextos diferentes, por lo que los estímulos contextuales se convierten en estímulos discriminativos y por tanto entran a formar parte de dicho entrenamiento. De esta forma se da al contexto un papel significativo y recibe un aprendizaje intencional.

Si el contexto puede actuar como un estímulo discriminativo como acabamos de describir, ¿qué es lo que diferencia a los estímulos contextuales de los definidos como estímulos objetivo (v. gr., el E.C.)? Básicamente, la diferencia radica en dos aspectos. En primer lugar, el contexto suele mantenerse constante mientras que los estímulos objetivo se presentan bien de manera puntual o en varios momentos durante el tiempo en el que perdura el contexto. En segundo lugar, los estímulos contextuales suelen ser peores predictores de las consecuencias que los estímulos objetivo. No obstante, una pregunta que podemos hacernos es ¿qué ocurre con los estímulos contextuales cuando un estímulo objetivo pierde su poder predictor (v. gr., si lo asociamos con consecuencias distintas)? ¿Pueden someterse entonces a un procesamiento o aprendizaje intencional para poder resolver el conflicto predictivo generado por el cambio en las consecuencias de la clave? En este capítulo nos centraremos precisamente en esas situaciones, analizando cómo el contexto pasa de cumplir un papel incidental a jugar un papel intencional en la recuperación de la memoria.

Para evitar este tipo de ambigüedades en la definición del contexto en este capítulo y en toda la investigación que hemos desarrollado en nuestro laboratorio hemos adoptado una definición del contexto desde el punto de vista estructural. Así, cuando hablemos de contexto nos referiremos a todos aquellos estímulos internos y externos que proporciona la propia situación experimental y que rodean a los estímulos objetivo sobre los que el sujeto tiene que realizar su aprendizaje. En principio estos estímulos serán incidentales a la tarea que el sujeto tiene que aprender, sin embargo las características de las tareas que proponemos y las explicaciones teóricas que barajamos del comportamiento de los sujetos podrían terminar convirtiendo contextos incidentales en contextos intencionales, entendiéndose por esto que pasan a recibir una atención controlada. Nuestro objetivo de hecho será analizar los factores que determinan que un contexto incidental se transforme en un contexto intencional.

Tomemos como ejemplo un experimento realizado por García-Gutiérrez y Rosas (2003c). En este experimento, como en la mayoría de los que referiremos aquí se utilizó una tarea de aprendizaje predictivo en seres humanos en el que los participantes se entrenaban en la predicción de las consecuencias gastrointestinales de distintos alimentos que supuestos clientes habían ingerido en distintos restaurantes. Durante la primera parte de la tarea los participantes recibían ensayos en los que un alimento determinado iba seguido por una consecuencia concreta en un restaurante (v. gr., ajos-diarrea en la chocita canadiense). El restaurante jugaba un papel de contexto incidental, como se comprueba por el hecho de que la relación predictiva clave-consecuencia es juzgada por el sujeto de forma idéntica cuando se pregunta por ella en dos contextos distintos (véase la parte izquierda de la Figura 1). Sin embargo, cuando después se establece una relación entre esa misma clave

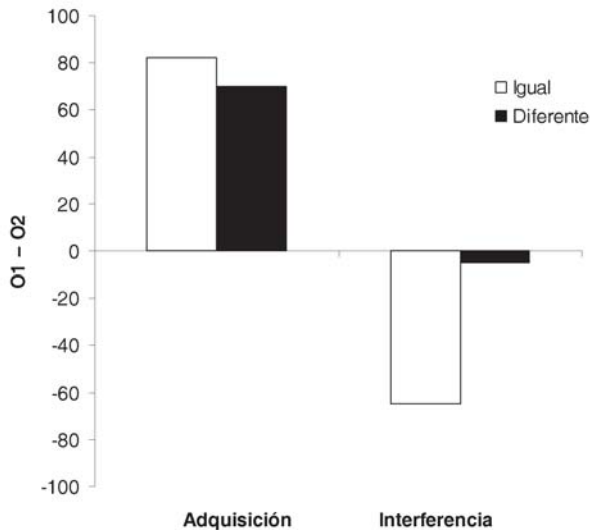


Figura 1. Diferencias medias entre las probabilidades adscritas a la relación entre una clave (X) y dos consecuencias distintas (O1 y O2) en las pruebas realizadas después de la adquisición de la relación X-O1 (izquierda) y después del tratamiento de interferencia en el que la clave que antes se relacionaba con O1 ahora se relaciona con O2 (derecha) en el contexto donde se aprendieron secuencialmente las dos relaciones (Igual) y en un contexto diferente (Diferente). Diferencias positivas implican mayor recuperación de la relación X-O1 que de la relación X-O2, y viceversa las diferencias negativas. Basado en García-Gutiérrez y Rosas (2003c).

y una consecuencia distinta (v. gr., ajos-estreñimiento), la recuperación de la información en una prueba posterior depende del contexto donde se realice la prueba: cuando coincide con el contexto donde se aprendieron las dos asociaciones el participante predice que la clave va a ir seguida por la segunda consecuencia, más que por la primera, mientras que la recuperación de esta información interferente se atenúa en un contexto diferente (véase la parte derecha de la Figura 1). El hecho de que la información se hiciera dependiente del contexto durante la prueba nos sugiere que el contexto podría haber comenzado a jugar un papel intencional desde el momento en el que se cambiaron las consecuencias de la clave. Las razones de ese cambio en el papel del contexto son el objetivo teórico principal de este capítulo, pero antes de adentrarnos en este análisis realizaremos una revisión del papel del contexto en la recuperación de la memoria.

Cambio de contexto y olvido de la información interferente

Buena parte de las teorías que tratan de explicar el olvido como un “fallo” reversible de la memoria resaltan la influencia del contexto en la recuperación de la información. La principal asunción de estas teorías es que la información nunca llega a borrarse de la memoria, no pudiendo recordar cierta información porque su acceso ha quedado temporalmente restringido (v. gr., Bouton, 1993).

Como factor fundamental que dificulta el acceso a la información debemos considerar el efecto de *interferencia* por la competición entre informaciones similares, incluyendo la adición de nuevo significado a una información ya aprendida. De acuerdo con Bouton (1993), el olvido se hace presente sólo cuando la información consistente adquiere un significado nuevo que por lo general contradice lo aprendido anteriormente. En este momento la memoria de la información interferente se encuentra más accesible en la memoria que la información interferida o adquirida originalmente, aunque ésta es susceptible de recuperarse si se dan las circunstancias adecuadas. Dentro de las circunstancias que favorecen la recuperación de la información consistente se han discutido dos posibilidades: el efecto del cambio de contexto después de la interferencia y el intervalo de retención transcurrido entre la fase de interferencia y la prueba de recuerdo.

Los estudios dedicados a explorar los efectos del paso del tiempo sobre la recuperación de la información en situaciones de interferencia comenzaron en el laboratorio de Pavlov (1927). Este autor encontró que la interposición de un intervalo de tiempo entre la adquisición de la información interferente y la prueba conllevaba la *recuperación espontánea* de la primera información adquirida. Tras la extinción de la relación EC-EI por la presentación del EC en ausencia de consecuencia, la RC se recupera espontáneamente tras un periodo de tiempo sin presentar el EC. Este efecto ha sido consistentemente replicado en animales (v. gr., Brooks y Bouton, 1993; Robbins, 1990; Rosas y Bouton, 1996) y en seres humanos (v. gr., Romero, Vila y Rosas, 2003; Rosas, Vila, Lugo y López, 2001; Vila y Rosas, 2001b).

El cambio de contexto después de la interferencia tiene resultados similares a los del paso del tiempo, produciéndose en este caso un efecto de *renovación* de la información interferida. Procedimentalmente la renovación de la información interferida puede conseguirse cuando la adquisición se desarrolla en un contexto (A), la interferencia en un contexto distinto (B) y se regresa al contexto de adquisición durante la prueba (renovación ABA, v. gr., Paredes-Olay y Rosas, 1999; Rosas y Bouton, 1997b, 1998), cuando adquisición, extinción y prueba se realizan en tres contextos distintos (renovación ABC, v. gr., Bouton y Swartzentruber, 1986) y finalmente cuando la adquisición y la extinción tienen lugar en el mismo contexto y éste se cambia durante la prueba, tal y como ocurre en el ejemplo recogido en la Figura 1 (renovación AAB, v. gr., Bouton y Ricker, 1994; García-Gutiérrez y Rosas, 2003c). Estos tres procedimientos conllevan un mejor recuerdo de la información original en detrimento de la información interferente. En otras palabras, la información interferente muestra una dependencia contextual que no se observa con la información que no ha sido interferida (véase el efecto nulo del cambio de contexto en la parte izquierda de la Figura 1).

La literatura en renovación y recuperación espontánea muestra un fuerte paralelismo entre ambos efectos, aquellas situaciones en las que puede producirse renovación son las mismas en las que aparece recuperación espontánea (v. gr., Rosas y Alonso, 1997a, b) y los factores que atenúan la renovación, como el uso de claves de recuperación de la interferencia, atenúan también la recuperación espontánea (Brooks y Bouton, 1993, 1994;

Brooks y Bowker, 2001; Brooks, Palmatier, Garcia y Jonson, 1999). Este paralelismo ha llevado a sugerir que ambos efectos podrían estar mediados por el mismo mecanismo subyacente en una aproximación que considera que el paso del tiempo es una forma particular de cambio de contexto (v. gr., Bouton, 1993; Bouton, Nelson y Rosas, 1999; Rosas y Bouton, 1997a, 1998; Rosas et al., 2001).

Bouton (1994) extendió la explicación del cambio de contexto a otro fenómeno de recuperación de la información tras la interferencia, la *reinstauración*. Este fenómeno se define como el efecto de recuperación de la información original tras la interferencia que produce la mera presentación de la consecuencia entre la interferencia y la prueba (v. gr., Rescorla y Heth, 1975; Vila y Rosas, 2001a). Las interpretaciones tradicionales de este efecto suponían la formación de un aprendizaje nuevo durante la presentación de la consecuencia en ausencia de la clave. Así, Rescorla y Heth (1975) sostenían que parte de la relación condicionada clave-consecuencia sobrevive a la presentación de la clave sola durante la extinción. Durante la exposición a la consecuencia se establecería una asociación nueva entre la consecuencia y el contexto donde se presenta que se sumaría a la relación residual clave-consecuencia dando como resultado la reinstauración de la RC durante la prueba. Desde otra perspectiva Westbrook, Iordanova, McNally, Richardson y Harris (2002) sostienen que el cambio en el significado de la clave durante la extinción lleva a que la clave se asocie con el contexto donde la extinción tiene lugar. Esta asociación lleva a que cuando se presente el contexto durante la fase de reinstauración se recupere la representación de la clave en la memoria y esta representación se asocie con la consecuencia, restableciendo la asociación clave-consecuencia y reinstaurando la RC durante la prueba. Bouton (1994) realiza una sugerencia distinta. Conociendo que la extinción y la interferencia son dependientes del contexto donde se adquieren, sostiene que la presentación de la consecuencia en el contexto podría llevar a que éste se transformara en un contexto diferente por su asociación con la consecuencia, con el resultado de que la prueba se realice en un contexto distinto al contexto en el que tiene lugar la extinción, provocando el olvido de la información interferente y la recuperación de la información original. El fenómeno de reinstauración se interpretaría entonces como un caso especial de renovación AAB.

García-Gutiérrez y Rosas (2003a) realizaron una serie de experimentos en los que trataron de contrastar la idea de que la reinstauración podía ser un caso especial de renovación. Estos autores utilizan la tarea de aprendizaje predictivo que describíamos más arriba en la que se establecía una relación secuencial entre una clave (un alimento) y dos consecuencias distintas (diarrea y estreñimiento) en una situación de interferencia (García-Gutiérrez y Rosas, 2003b, c, d). Tras este entrenamiento de interferencia encontraron que la presentación antes de la prueba de la primera consecuencia con la que se asoció la clave producía una reinstauración de la relación entre la clave y la primera consecuencia, en detrimento de la segunda, extendiendo los resultados de reinstauración a una situación de interferencia retroactiva diferente de la extinción (véase también Brooks, Hale, Nelson y Bouton, 1995). Estos resultados podrían interpretarse desde cualquiera de las teorías

Tabla 1

Extracto del diseño del Experimento 4 de García-Gutiérrez y Rosas (2003a)

Grupo	Adquisición	Interferencia	Reinstauración	Prueba
A:O1			A:O1	
A:O2	A: X-O1, Y-O2, S-O1, T-O2, R-	A:X-O2, Z-O1, S-O1, T-O2, R-	A:O2	A:X
A:O3	B: S-O1, T-O2, R-	B: S-O1, T-O2, R-	A:O3	B:X
B:O2/O3			B:O2/O3	

Nota: A y B son dos contextos (restaurantes) diferentes, X, Y, Z, S, T y R son nombres de alimentos; O1 y O2 fueron diarrea y estreñimiento; O3 fue vómito. Los tratamientos relevantes aparecen en negrita.

que veíamos arriba. Así, García-Gutiérrez y Rosas (2003a) diseñaron una situación que permite discriminar la interpretación de reinstauración como un caso especial de renovación de las que interpretan el fenómeno como un nuevo aprendizaje. Observaron que si la reinstauración era debida a un cambio en el contexto causado por los emparejamientos contexto-consecuencia, el cambio contextual debiera de producirse igual independientemente de cuál fuera la consecuencia utilizada. El cambio de contexto debiera ser idéntico cuando presentamos en el contexto la consecuencia emparejada primero con la clave, la interferente o una consecuencia nueva, en todos los casos debería producir una recuperación de la relación original clave-consecuencia.

La Tabla 1 presenta el diseño utilizado en el experimento 4 de García-Gutiérrez y Rosas (2003a). Todos los grupos recibieron el mismo tratamiento en el que la clave X se relacionaba primero con la consecuencia O1 y después con la consecuencia O2 en el contexto A, recibiendo un tratamiento de interferencia. Al mismo tiempo, la clave Y se relacionaba con la consecuencia O2 y no se presentaba durante el tratamiento de interferencia. Inmediatamente después de la interferencia se expuso a los sujetos al contexto emparejado con una consecuencia en ausencia de la clave. Los nombres de los grupos reflejan la consecuencia con la que se emparejó el contexto, O1 en A:O1, O2 en A:O2, y una consecuencia nueva (vómito) en A:O3. El grupo B:O2/O3 fue un grupo de control en el que la mitad de los sujetos recibieron exposición a la consecuencia O2 y la otra mitad a la consecuencia O3 en un contexto distinto al contexto donde recibían el entrenamiento de interferencia, pero igualmente familiar. El resto de los estímulos utilizados se emplearon como distractores y con el fin de mantener la experiencia con consecuencias equivalente entre contextos y grupos. En este caso los contextos utilizados fueron parajes naturales en los que los sujetos evaluados habían ingerido los alimentos mientras estaban de picnic.

La Figura 2 recoge la diferencia entre la probabilidad adscrita a la relación X-O1 y la adscrita a la relación X-O2 durante la prueba final tras el tratamiento de reinstauración. Diferencias negativas implican una actuación de acuerdo a la situación de interferencia (probabilidad adscrita a X-O2 mayor que X-O1), mientras que diferencias positivas impli-

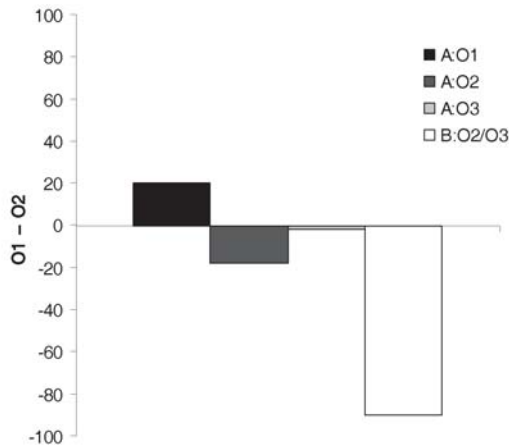


Figura 2. Diferencias medias entre las probabilidades adscritas a la relación X-O1 y X-O2 en la prueba realizada tras el tratamiento de reinstauración en los grupos A:O1, A:O2, A:O3 y B:O2/O3. El tratamiento de reinstauración produjo una atenuación de la interferencia independientemente de la consecuencia utilizada (Basado en García-Gutiérrez y Rosas, 2003a).

can recuperación de la primera relación aprendida (X-O1) en detrimento de la información interferente (X-O2). Tal y como puede observarse en la figura, en el grupo control (B:O2/O3) los sujetos juzgaron que la clave X iba seguida por la consecuencia O2 (diferencias negativas muy altas) mostrando que la presentación de las consecuencias fuera del contexto de prueba tenía poca influencia sobre la interferencia. Sin embargo, la presentación de la consecuencia en el contexto de prueba produjo una recuperación de la relación X-O1 en detrimento de la relación X-O2, y esto ocurrió independientemente de que la consecuencia presentada fuera la primera asociada con X, la segunda, o una consecuencia nueva (para más detalles véase García-Gutiérrez y Rosas, 2003a).

Estos resultados confirman las predicciones de la teoría de la recuperación de la información sugiriendo que la reinstauración es un caso especial de renovación AAB, en el que los emparejamientos contexto-consecuencia cambian las características asociativas del contexto, lo convierten en un contexto nuevo y como resultado se dificulta la recuperación de la información interferente y facilita la recuperación de la información interferida (véase también García-Gutiérrez, Rosas y Nelson, 2005).

Concluyendo este apartado, los resultados encontrados en nuestro laboratorio sugieren que reinstauración, renovación y recuperación espontánea son todos efectos en los que la recuperación de la información interferente se ve dificultada por un cambio en las características del contexto, sean éstas asociativas, físicas o temporales, respectivamente. Así, tendríamos dos factores principales que regulan la recuperación de la información, la interferencia y el cambio de contexto. En el apartado final de este capítulo vamos a analizar a que tipo de información afectan estos factores y qué determina que esa información se vea afectada por el olvido.

¿Qué hace a la información dependiente del contexto?

De acuerdo con lo expuesto, existen informaciones que no son dependientes del contexto e informaciones que sí lo son (véase la Figura 1). Tras una revisión exhaustiva de la literatura, Bouton (1993, 1994) sostiene que hay dos razones por las que una información se convierte en más dependiente del contexto que otra: que sea una información inhibitoria (v. gr., la extinción) o que sea aprendida en segundo lugar. En la situación experimental que venimos presentando en la que un mismo estímulo se relaciona secuencialmente con dos consecuencias distintas, el efecto del cambio de contexto sobre la recuperación de la información se produciría por dos razones. Por una parte, cuando relacionamos la clave con su segunda consecuencia estamos extinguiendo la relación clave-primera consecuencia, la extinción es información inhibitoria y se verá afectada por el cambio de contexto llevando a una reactivación de la asociación clave-primera consecuencia durante la prueba. Por otra parte, la relación clave-segunda consecuencia es lo segundo que el sujeto aprende acerca de la clave, información que de acuerdo con Bouton (1994) también se verá afectada por el cambio de contexto que dificultará su recuperación. Los efectos del cambio de contexto encontrados en nuestro laboratorio coinciden con esta interpretación ya que el efecto observable de recuperación de la información con el cambio contextual resulta de la combinación de una disminución en la recuperación de la relación clave-O2 y un aumento en la recuperación clave-O1 (véase García-Gutiérrez et al., 2005).

Sin embargo, un análisis más detallado de la situación nos lleva a observar que la información inhibitoria en este caso (extinción de la relación clave-O1) es al mismo tiempo información aprendida en segundo lugar. Siguiendo este razonamiento en una serie experimental en la que contrasta la dependencia contextual de la inhibición y la excitación condicionada, Nelson (2002) encuentra que el efecto de cambio de contexto sólo se manifiesta cuando la información, sea excitatoria o inhibitoria, se aprende en segundo lugar, revelando una simetría entre la excitación y la inhibición.

El análisis realizado hasta ahora parece sugerir que la dependencia contextual de la información se debe a una característica de ésta, el ser lo segundo que aprende un sujeto acerca de una misma clave. Esta aproximación, determinar qué tipo de información es dependiente del contexto, aunque útil, resulta una mera redescipción de la realidad. Así, una vez establecido que la información dependiente de contexto es la aprendida en segundo lugar (Nelson, 2002), la pregunta de verdadero interés teórico es *por qué esa información se hace dependiente del contexto*.

Bouton (1993) tiene una respuesta natural a esta cuestión, sosteniendo que la segunda información acerca de una clave se hace dependiente del contexto porque cuando la clave pasa de proporcionar una información consistente a proporcionar una información ambigua los sujetos comienzan a prestar atención al contexto en una búsqueda automática que rompa la ambigüedad en el significado de la clave. Esto explicaría por qué la segunda información adquirida acerca del significado de una clave es dependiente del contexto donde se aprende, mientras que no ocurre lo mismo con la primera información, que se transfiere de forma similar a distintos contextos (v.gr., Paredes-Olay y Rosas, 1999; Rosas y Bouton, 1998; Rosas et al., 2001).

Tabla 2*Extracto del diseño del Experimento 2 de Rosas et al. (2005)*

Adquisición	Interferencia	Prueba
A: X-O1 , D-O2, F- B: Z-O1 , E-O2, F-	A: X-O2 , Y-O1 , F- B: Z-O1 , E-O2, F-	A: X, Y, Z B: X, Y, Z

Nota: A y B son dos contextos (restaurantes) diferentes, X, Y, Z, D, E y F son nombres de alimentos; O1 y O2 son diarrea y estreñimiento. Los tratamientos relevantes aparecen en negrita. Para las claves X e Y el contexto A fue el mismo del entrenamiento, mientras el contexto B fue un contexto distinto; lo contrario es cierto para la clave Z.

Sin embargo, si profundizamos en este razonamiento la especificidad del contexto no dependerá de que la información posea alguna característica especial como sugiere Bouton (1993), sino más bien de alguna característica específica de la situación que conduce a los participantes a prestar atención al contexto. Nuestra interpretación de este razonamiento tiene una predicción específica que permite contrastarlo directamente: si lo que hace a la información específica del contexto es que la ambigüedad de la información lleva a que los participantes comiencen a prestarle atención, una vez que el contexto se convierte en relevante por la aparición de información ambigua, cualquier información que los participantes aprendan en dicho contexto se convertirá en específica del mismo, independientemente de si la información es la primera o la segunda cosa que los participantes aprenden acerca de la clave o de si este aprendizaje es inhibitorio.

Recientemente hemos realizado varias series experimentales para comprobar esta hipótesis en nuestro laboratorio. En la Tabla 2 se presenta el diseño que utilizamos en uno de nuestros experimentos (Rosas, García-Gutiérrez y Callejas-Aguilera, 2005). Los participantes en este experimento recibieron entrenamiento secuencial en las relaciones X-O1 y X-O2 en el contexto A. Mientras recibían el tratamiento X-O2 durante la segunda fase del experimento se les presentó la relación Y-O1. Además, recibieron exposición a la relación Z-O2 en el contexto B durante todo el experimento. El resto de claves y consecuencias se presentaron para igualar la exposición a contextos, consecuencias y las posibles relaciones contexto-consecuencias a lo largo de las distintas fases del experimento. En la prueba final los participantes tenían que juzgar la relación entre cada una de las claves (X, Y y Z) y cada una de las consecuencias (O1 y O2) en los dos contextos donde tuvo lugar el entrenamiento (A y B). De acuerdo con los resultados de la literatura esperábamos encontrar renovación de la relación X-O1 en el contexto B (García-Gutiérrez y Rosas, 2003c). Además, si la especificidad contextual de la información depende de que la ambigüedad lleve a los sujetos a prestar atención al contexto, cambiar el contexto debiera llevar a una disminución en los juicios acerca de la relación Y-O1, incluso aunque dicha información nunca haya sido ambigua. Por último, con la clave Z se trataba de comprobar la generalidad del efecto de cambio contextual. Si dependiera de que la información se adquiriera en el mismo contexto en el que ha ocurrido la interferencia, no

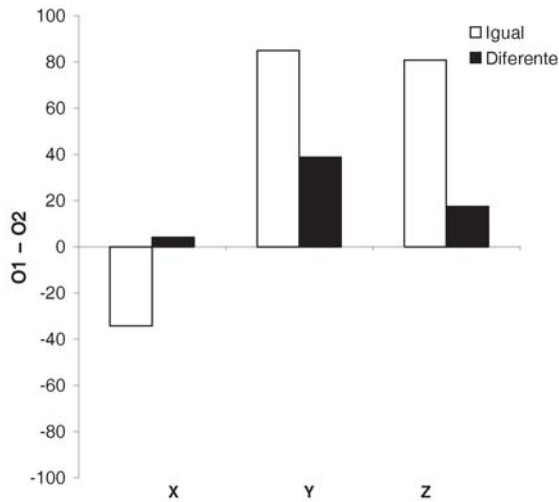


Figura 3. Diferencias medias entre las probabilidades adscritas a la relación de X, Y y Z con las consecuencias O1 y O2 en el contexto donde tuvo lugar el entrenamiento (igual) y en el contexto diferente. X había recibido un tratamiento de interferencia, mientras Y y Z siempre fueron entrenados con la misma consecuencia. Basado en Rosas et al. (2005).

debería encontrarse un efecto de cambio de contexto en esta clave, dado que la relación Z-O1 fue consistente a lo largo de todo el experimento y aprendida en un contexto (B) en el que no se presentó información ambigua. No obstante, si la interferencia llevara a que los participantes prestaran atención a cualquier contexto, deberíamos encontrar un efecto perjudicial del cambio de contexto también en la clave Z.

La Figura 3 presenta los resultados de este experimento. Recoge la diferencia media entre las probabilidades adscritas a O1 y a O2 en las claves X, Y y Z en el contexto donde tuvo lugar el entrenamiento de cada clave (igual) y en otro contexto (diferente). Tal y como puede observarse, cuando la prueba tuvo lugar en el contexto de entrenamiento las diferencias fueron negativas en X, reflejando el tratamiento de interferencia recibido, y positivas en Y y Z reflejando el entrenamiento consistente con la consecuencia O1. La prueba en el contexto diferente llevó a que las diferencias en A fueran más positivas replicando el bien conocido efecto de renovación. Lo más interesante es lo que ocurrió con las claves X y Z, que sólo habían recibido entrenamiento consistente; en ambas claves, el cambio de contexto llevó a que los participantes juzgaran que la relación con O1 era más baja. En otras palabras, lo aprendido acerca de las claves X y Z fue dependiente de contexto en una situación en la que ambas claves habían sido entrenadas sólo con una consecuencia (Rosas et al., 2005).

Una vez los participantes han aprendido que el contexto puede ser importante durante la interferencia, parece que comienzan a codificar toda la información como específica del contexto, independientemente del tipo de información presentada y del contexto donde

se haya adquirido. Así, no parece que exista ningún tipo especial de información que sea específica del contexto, al contrario, parece que exista un tipo especial de diseños que hace que los participantes presten atención a los contextos, con lo que la información se convierte en específica del contexto donde se adquiere.

No obstante, estas inferencias deben ser tomadas con cautela: el efecto del cambio de contexto sobre la relación Z-O1 apareció cuando dicha relación fue evaluada en un contexto donde se presentó información ambigua (A). Por tanto, es posible que la especificidad del contexto dependiera de alguna característica del contexto utilizado, en vez de ser un efecto general debido a que los participantes prestaran atención a cualquier contexto presentado una vez que aprendieron que uno de estos contextos es importante. Con el fin de comprobar la generalidad de la atención prestada a los contextos y contrastar la posibilidad de que los participantes prestaran atención a cualquier contexto una vez aprendida la relevancia de uno de ellos realizamos una serie de experimentos donde se requería que los participantes realizaran dos tareas distintas de manera secuencial, lo que permitiría observar si la ambigüedad de la primera tarea podía afectar a la especificidad del contexto en la segunda tarea, donde no se presentaba información ambigua. Después del entrenamiento en interferencia los participantes parecían codificar toda la información dentro de la tarea como dependiente del contexto. La cuestión que nos planteamos fue si este efecto se limita a la tarea específica, o si por el contrario la interferencia hace que los participantes codifiquen toda la información como específica del contexto, independientemente de la tarea. Nuestros resultados preliminares en esta aproximación sugieren que una vez que los participantes comienzan a prestar atención al contexto en una tarea, esta atención se generaliza a una tarea diferente. No obstante, tenemos que tener en cuenta que las tareas que utilizamos mantienen en común muchos aspectos, como el laboratorio, los experimentadores o los propios ordenadores donde se realizan, por lo que es probable que la especificidad contextual desapareciera si los participantes realizaran una tarea verdaderamente diferente.

Recapitulación

Comenzábamos este capítulo dando nuestra definición operacional de contexto como aquel conjunto de estímulos que forman parte de la tarea pero que no son la tarea misma, que al principio juegan un papel incidental pero que por las propias características de la situación de aprendizaje pueden terminar jugando un papel intencional. De acuerdo con esta interpretación del contexto y su papel en la recuperación de la información aprendida, presentábamos distintos resultados de nuestro laboratorio en los que el contexto jugaba secuencialmente esos papeles, no codificándose como parte de la tarea mientras la información era consistente y pasando a codificarse junto a ella cuando la información se hacía ambigua. La recuperación espontánea, la renovación y de acuerdo con nuestros resultados también la reinstauración muestran que una vez aparece la interferencia, la información interferente se hace dependiente del contexto donde se adquiere.

La parte final del capítulo incide en el tipo de información que es dependiente de contexto, concluyendo que la dependencia contextual no viene determinada por características específicas de la información; que ésta sea inhibitoria o excitatoria, información consistente o información interferente parece irrelevante de acuerdo con nuestros resultados. Todo apunta a que el efecto del cambio de contexto en el aprendizaje y la recuperación de la información es un fenómeno de cambio atencional, siendo el propio diseño de la situación el que conduce a que el sujeto preste atención al contexto donde la información se presenta codificándose junto a ella a partir de entonces y convirtiéndose la recuperación de la información que se aprenda a partir de ese momento en dependiente del contexto.

Llegado este punto conviene realizar una precisión. En la investigación recogida en este capítulo subyace la idea de la búsqueda de los mecanismos de recuperación de la información que nos son comunes a distintas especies animales. De la lectura del resto de los capítulos de este volumen debería quedar claro que este análisis de los mecanismos comunes sólo recoge una parte de los mecanismos de recuperación en aprendizaje causal humano. Así, en la mayoría de estos capítulos se hace incidencia en la necesidad de estudiar el papel que juegan procesos superiores como el razonamiento y el lenguaje en esa recuperación de la información, aspectos que sin duda nos diferencian de otras especies animales. Desde nuestra perspectiva, el análisis completo de los mecanismos de recuperación de la información en aprendizaje causal provendrá de la investigación coordinada de aquellos mecanismos que compartimos con otros animales, de aquellos exclusivos del ser humano y del estudio de la interacción entre ellos.

Nota de los autores

La investigación recogida en estas páginas ha sido financiada por el grupo de investigación HUM642 de la Junta de Andalucía y por el proyecto BSO2002-03398 del Ministerio de Ciencia y Tecnología, España, cofinanciado con fondos FEDER de la comunidad europea. La correspondencia puede dirigirse a Juan M. Rosas, Departamento de Psicología, Universidad de Jaén, 23071 Jaén. E-mail: jmrosas@ujaen.es.

Referencias

- Baddeley, A. D. (1992). Domains of recollection. *Psychological Review*, 89, 708-792.
- Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of pavlovian learning. *Psychological Bulletin*, 114, 80-99.
- Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science*, 3, 49-53.
- Bouton, M. E., Nelson, J. B. y Rosas, J. M. (1999). Stimulus generalization, context change, and forgetting. *Psychological Bulletin*, 125, 171-186.
- Bouton, M. E. y Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning & Behavior*, 22, 317-324.

- Bouton, M. E. y Swartzentruber, D. (1986). Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*, 333-350.
- Brooks, D. C. y Bouton, M. E. (1993). A retrieval cue for extinction attenuates spontaneous recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 77-89.
- Brooks, D. C. y Bouton, M. E. (1994). A retrieval cue for extinction attenuates response recovery (renewal) caused by a return to the conditioning context. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*, 366-379.
- Brooks, D. C. y Bowker, J. L. (2001). Further evidence that conditioned inhibition is not the mechanism of an extinction cue's effect: A reinforced cue prevents spontaneous recovery. *Animal Learning & Behavior*, *29*, 381-387.
- Brooks, D. C., Hale, B., Nelson, J. B. y Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior*, *23*, 383-390.
- Brooks, D. C., Palmatier, M. I., Garcia, E. O. y Johnson, J. L. (1999). An extinction cue reduces spontaneous recovery of a conditioned taste aversion. *Animal Learning & Behavior*, *27*, 77-88.
- García-Gutiérrez, A. y Rosas, J. M. (2003a). Context change as the mechanism of reinstatement in causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*, 292-310.
- García-Gutiérrez, A. y Rosas, J. M. (2003b). Empirical and theoretical implications of additivity between renewal and reinstatement after interference in human causal learning. *Behavioural Processes*, *63*, 21-31.
- García-Gutiérrez, A. y Rosas, J. M. (2003c). Recuperación de la relación clave-consecuencia por el cambio de contexto después de la interferencia en aprendizaje causal. *Psicológica*, *24*, 243-269.
- García-Gutiérrez, A. y Rosas, J. M. (2003d). The role of number of cues on retroactive interference in causal learning. *Psicológica*, *24*, 271-285.
- García-Gutiérrez, A., Rosas, J. M. y Nelson, A. (2005). Extensive interference attenuates reinstatement in human predictive judgments. *International Journal of Comparative Psychology*. *En prensa*.
- Nelson, J. B. (2002). Context specificity of excitation and inhibition in ambiguous stimuli. *Learning and Motivation*, *33*, 284-310.
- Paredes-Olay, C. y Rosas, J. M. (1999). Within-subjects extinction and renewal in predictive judgments. *Psicológica*, *20*, 195-210.
- Pavlov, I. P. (1927). *Conditioned reflexes*. (G. V. Anrep, trans.). London: Oxford University Press.
- Rescorla, R. A. y Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 88-96.
- Robbins, S. J. (1990). Mechanisms underlying spontaneous recovery in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 235-249.
- Romero, M. A., Vila, N. J. y Rosas J. M. (2003). Time and context effects after discrimination reversal in human beings. *Psicológica*, *24*, 169-184.
- Rosas, J. M. y Alonso, G. (1997a). The effect of context change upon long-term memory of CS duration. *Behavioural Processes*, *39*, 69-76.
- Rosas, J. M. y Alonso, G. (1997b). Temporal discrimination and forgetting of CS duration in conditioned suppression. *Learning and Motivation*, *28*, 43-57.
- Rosas, J. M. y Bouton, M. E. (1996). Spontaneous recovery after extinction of a conditioned taste aversion. *Animal Learning & Behavior*, *24*, 341-348.
- Rosas, J. M. y Bouton, M. E. (1997a). Additivity of the effects of retention interval and context change on latent inhibition: Toward resolution of the context forgetting paradox. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 283-294.
- Rosas, J. M. y Bouton, M. E. (1997b). Renewal of conditioned taste aversion upon return to the conditioning context after extinction in another one. *Learning and motivation*, *28*, 216-229.
- Rosas, J. M. y Bouton, M. E. (1998). Context change and retention interval have additive, rather than interactive, effects after taste aversion extinction. *Psychonomic Bulletin & Review*, *5*, 79-83.

- Rosas, J. M., García-Gutiérrez, A. y Callejas-Aguilera, J. E. (2005). Effects of context change upon retrieval of first and second-learned information in human predictive learning. *Psicológica*. En prensa.
- Rosas, J. M., Vila, N. J., Lugo, M., & López, L. (2001). Combined effect of context change and retention interval upon interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 153-164.
- Vila, N. J. y Rosas, J. M. (2001a). Reinstatement of acquisition performance by presentation of the outcome after extinction in causality judgments. *Behavioural Processes*, 56, 147-154.
- Vila, N. J. y Rosas, J. M. (2001b). Renewal and spontaneous recovery after extinction in a causal learning task. *Mexican Journal of Behavior Analysis*, 27, 79-96.
- Westbrook, R. F., Iordanova, M., McNally, G., Richardson, R. y Harris, A. H. (2002). Reinstatement of fear to an extinguished conditioned stimulus: Two roles for context. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 97-110.

Diferencias en la recuperación de la respuesta condicionada después de la interferencia retroactiva en animales y humanos

Javier Vila y Angélica Alvarado

F. E. S. Iztacala, Universidad Nacional Autónoma de México, México

“Sin la palabra no habría historia ni amor; seríamos como el resto de los animales, mera perpetuación y mera sexualidad. Hablar nos une como parejas, como sociedades, como pueblos. Hablamos porque somos, pero somos porque hablamos”

J. Cortázar

La hipótesis de la recuperación de información supone que tiempo y contexto modulan la recuperación de lo aprendido después de la interferencia provocada por un segundo aprendizaje, produciendo o no la expresión del aprendizaje adquirido inicialmente (Bouton, 1993, 1997). Recientemente este supuesto ha sido aplicado con éxito al estudio del aprendizaje humano principalmente en preparaciones de aprendizaje causal y ha generado una gran cantidad de investigación que supone mecanismos similares de recuperación entre humanos y animales (Miller y Escobar, 2002; Pineño y Miller, 2005; Pineño, Vegas y Matute, 2003; Rosas, García-Gutiérrez y Romero, 2003; Rosas, Vila, Lugo y López, 2001), lo que ha tenido como consecuencia una extrapolación de variables metodológicas y procesos observados en estudios con animales al aprendizaje humano, olvidando así su naturaleza única.

El presente capítulo reconsidera algunos de los supuestos más comunes en el análisis teórico de los efectos de recuperación. Para ello se presentan una serie de datos recientes obtenidos en preparaciones de aprendizaje causal que sugieren diferencias básicas en la recuperación de la información entre animales y humanos. Finalmente, se señala la necesidad de la inclusión de procesos controlados como las instrucciones en las *teorías de recuperación*, así como de la creación de una metodología estandarizada para un adecuado estudio del aprendizaje humano.

Antecedentes teóricos

Originalmente Pavlov (1927), al estudiar las condiciones necesarias para la ocurrencia del condicionamiento, manipuló las condiciones originales de la adquisición después del establecimiento de la respuesta condicionada (RC) ya sea presentando el estímulo condicionado (EC) solo o seguido de un estímulo incondicionado (EI) distinto al original, y observó como resultado una disminución de la RC. Posteriormente, el paso del tiempo por sí mismo (recuperación espontánea) o la presentación de un EC novedoso (inhibición externa) producían el restablecimiento de la RC original. Bouton (1993) interpretó estos efectos como ejemplos de interferencia retroactiva entre consecuencias en las que el nuevo aprendizaje durante la segunda fase interfiere con lo aprendido durante el condicionamiento original.

Así los efectos de recuperación como la recuperación espontánea, la inhibición latente y recientemente los efectos de renovación y de reinstauración han sido interpretados como la disminución de la interferencia retroactiva producida por la asociación aprendida en una segunda fase que lleva a la expresión de la primera asociación aprendida durante la primera fase de adquisición (Bouton, 1997). La renovación se define como la recuperación de la RC original producida por un cambio de contexto posterior a la extinción o contracondicionamiento (Bouton y Bolles, 1979), mientras que la reinstauración se define como la recuperación de la RC producida por la presentación del EI solo después de la extinción (Rescorla y Heth, 1975)

El modelo propuesto por Bouton (1993, 1997) para explicar los efectos de recuperación supone que el segundo aprendizaje es dependiente de contexto y del paso del tiempo (como un cambio de contexto novedoso). Así, esta teoría mantiene que al aprender de manera consecutiva dos memorias contradictorias acerca del EC y sus consecuencias (EC-EI vs. EC- no EI), éstas compiten por su expresión en una prueba final de condicionamiento. A su vez, la expresión de estas memorias se facilita por las señales dadas por el contexto físico o temporal. De acuerdo con esto, si el contexto de la segunda fase y el de prueba son iguales se observará un efecto de recencia en el que la RC será similar en la segunda fase y en la prueba final. Pero si el contexto de prueba y el de la segunda fase son diferentes se producirá una recuperación de la memoria de la primera fase y se observará un incremento de la RC inicial. Por tanto la ocurrencia de los efectos de recencia o recuperación dependerá del contexto físico o temporal que acompañe al EC durante la prueba.

Recientemente, Miller y Escobar (2002) han propuesto una modificación al supuesto anterior, en la que a los efectos de recuperación estudiados en situaciones de interferencia entre consecuencias, como la extinción, el contracondicionamiento o la discriminación inversa (Bouton, 2004; Rosas y col., 2001; Rosas y col., 2003; Vila, Alvarado, Jara y Flores, 2003; Vila, Romero y Rosas, 2002), se han agregado los de la interferencia entre claves (Escobar Matute y Miller, 2001; Escobar, Pineño y Matute 2002; Matute y Pineño, 1998, 2000). En la interferencia entre consecuencias, en una primera fase una clave A produce la consecuencia *x*, y en la segunda fase la misma clave produce una consecuencia distinta *y*. En la interferencia entre claves, en una primera fase la clave A produce la consecuencia *x*, en la segunda fase otra clave produce la misma consecuencia *y*. Debido a que en ambos tipos de interferencia se observan los efectos de renovación y de recuperación espontánea, se ha propuesto que la interferencia entre consecuencias y entre claves se basa en un mecanismo común cuyo origen está en que en ambas interferencias existe un elemento común (clave o consecuencia) que es determinante para que la interferencia ocurra y por tanto se observen efectos de recuperación cuando cambia el contexto físico o temporal (Pineño y col, 2003; Pineño y Matute, 2000).

Tanto el modelo de Bouton en su versión mas reciente (2002, 2004) como la propuesta de integración de los dos tipos interferencia en un mecanismo común asociativo de Miller y Escobar (2002) suponen que los mecanismos responsables de los efectos de recuperación y su modulación por el contexto físico y temporal son similares y en algunos casos iguales

entre animales y humanos, lo que ha llevado a una extrapolación de procesos, funciones y estrategias metodológicas de la investigación animal a los estudios con humanos. Así, se ha llegado a suponer que las tareas con humanos son no verbales aun y cuando se reconoce el uso de instrucciones en ellas (Miller y Escobar, 2002, pp. 168-169), que el efecto de las instrucciones de prueba es equiparable en humanos al de los contextos físicos de prueba empleados en ratas (Matute, Vegas y De Marez, 2002; Pineño y col., 2003), que los contextos físicos y la frecuencia con la que se solicita el juicio a los participantes en una tarea causal (frecuencia del juicio o demandas de la tarea) son equivalentes (Vadillo, Vegas y Matute, 2004) y el empleo de etiquetas de texto como análogo a los contextos físicos o temporales empleados en animales (García-Gutiérrez y Rosas, 2003; Matute y Pineño, 1998; Pineño y Matute, 2000; Pineño y Miller, 2004b; Rosas y col., 2001).

Este punto de vista ha sido común y supone que las variables empleadas en los estudios con animales y las variables empleadas en los experimentos con humanos son metodológicamente equivalentes y que por tanto comparten la misma función. El argumento que ha validado esta extrapolación ha sido la observación de resultados similares entre animales y humanos en los efectos de recuperación conocidos. Sin embargo, no deja de ser difícil pensar en cómo una instrucción que hace alusión a un espacio tiempo-concreto (*estás en el país ... en el año...*) es análoga funcionalmente a un contexto o al paso del tiempo, sin pensar en sus propiedades funcionales como comportamiento verbal. Esta generalización de variables ha producido un incremento de la variedad de las tareas empleadas con humanos y se ha preocupado poco por la validez de las mismas, lo que en ocasiones provoca que se obtengan resultados contradictorios al manipular variables similares (Matute, Arce-diano y Miller, 1996; Cobos, Caño, López, Luque y Almaraz, 2000), o dificultades en la replicación de algunos efectos con tareas de aprendizaje causal (Matute y Pineño, 1998).

En general, la aplicación de los modelos asociativos al aprendizaje humano (principalmente causal) ha ignorado la participación de factores verbales o concientes (procesos controlados o dirigidos) debido a que estos modelos en su planteamiento original nos dicen poco o nada del papel de los procesos controlados sobre el aprendizaje asociativo humano (De Houwer, Vandorpe y Beckers, 2005), lo que quizá ha producido que los investigadores hagan fácilmente extrapolaciones de la investigación animal a la humana tanto de procesos como de estrategias metodológicas.

Sin embargo, recientemente algunos investigadores han señalado la necesidad de considerar las variables típicamente humanas como las instrucciones y su contenido verbal en el estudio de los modelos de aprendizaje y los efectos de recuperación (De Houwer y col., 2005; De Houwer, Beckers y Vandorpe, en prensa; Lovibond, 2003, 2004; Vila y col., 2003). La evidencia experimental que demuestra la participación de estas variables en el aprendizaje humano no es nueva (véanse las revisiones de Baron y Galitzio, 1983; Brewer, 1974; Davey, 1987) y existen en trabajos recientes que demuestran su participación durante la adquisición del aprendizaje (De Houwer y col., 2005; Lovibond, 2004). Sin embargo, al presente existen pocos trabajos que sugieran la participación de procesos controlados en los efectos de recuperación con humanos (Alvarado, Jara, Vila y Rosas, 2005; Vila y col.,

2003). Se hace necesaria así una revisión de algunos estudios que sugieren el papel que tienen los procesos controlados en algunos de los efectos de recuperación de la información.

Recuperación espontánea en seres humanos

Uno de los fenómenos del aprendizaje asociativo más conocidos es la recuperación espontánea, en la cual la introducción de un periodo de tiempo (intervalo de retención, IR) después de la extinción de un EC provoca que la RC se observe nuevamente (Pavlov, 1927). Este efecto ha sido demostrado con éxito en aprendizaje causal (Rosas y col., 2001; Vila y Rosas, 2001). Al presente, las explicaciones propuestas para este fenómeno no son completamente claras y han sido derivadas en gran medida de los estudios con animales (Alvarado y col., 2005; Pineño y Miller, 2005; Rescorla, 2004) y suponen por un lado que, posteriormente a la extinción, la memoria EC-no EI es relativamente más reciente en relación a la primera memoria EC-EI, y por tanto en primera instancia se observará un efecto de recencia de la última memoria aprendida que con el paso del tiempo decrece hasta que el aprendizaje original sea más fuerte y su expresión sea mayor (primacía) generando así la recuperación espontánea como un “cambio recencia-primacía” (Knoedler, Hellwig y Neath, 1999; Pineño y Miller, 2005), y que dependería tanto de los efectos de orden del tipo de ensayos presentados como del valor del IR. Un segundo punto de vista también sostiene que la recencia se desvanece con el paso del tiempo, pero que el aprendizaje de ambas memorias (EC-EI y EC no EI) se mantiene intacto y su expresión depende de la similitud relativa de las condiciones de entrenamiento y prueba, lo que produciría que se recuerden ambas memorias bajo las condiciones adecuadas (integración de memorias) y que su expresión sea acorde a la historia asociativa total del EC (Matute y col., 2002; Vadillo y col., 2004)

Datos recientes de nuestro laboratorio empleando una preparación de aprendizaje causal con humanos demostraron que la recuperación espontánea posterior a la interferencia retroactiva entre consecuencias (A_x / A_y) no incrementa como una función del valor del IR (Alvarado y col., 2005). Cinco grupos ($n = 10$) recibieron un entrenamiento en el que un medicamento ficticio producía de manera sucesiva dos síntomas (c_1 y c_2) durante 20 ensayos cada uno; cada cinco ensayos se les preguntó a los participantes en qué medida creían que el medicamento A producía los síntomas c_1 y c_2 . Posteriormente cada grupo fue asignado a un valor distinto del IR (0, 0.75, 1, 24 y 48 hrs.), finalmente todos los grupos recibieron una prueba en la que se les pidieron sus juicios causales acerca de la relación del medicamento A con c_1 y c_2 . Los juicios emitidos por los participantes para ambas consecuencias fueron del 50% (juicio proporcional, $c_1/c_1+c_2 = 0.5$) en todos los valores del intervalo mayores a 0 hrs. Así la recuperación observada no fue una función del valor del IR como ocurre en animales, sino que fue constante a lo largo de los cuatro valores del IR. Para el IR igual a 0 se observó un efecto de recencia en el que los juicios para c_1 fueron de 0% y para c_2 de 100% (juicio proporcional, $c_1/c_1+c_2 = 0$) tal y como se esperaba. En la parte izquierda de la Figura 1 se presentan los datos de la proporción de

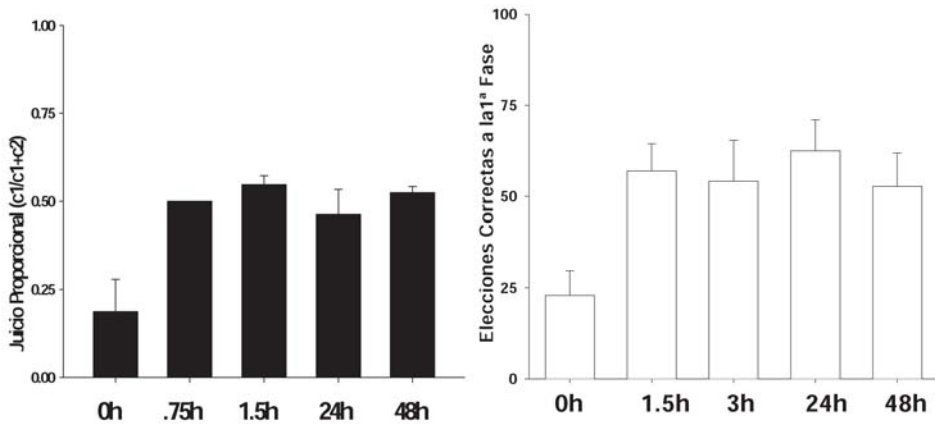


Figura 1. En el panel izquierdo se presenta la recuperación de cinco grupos en una tarea causal de interferencia retroactiva, como juicio proporcional $C1/C1+C2$ del último juicio de cada consecuencia, a cada grupo se le asignó un valor de IR (adaptado de Alvarado y col., 2005). El panel derecho muestra la recuperación espontánea como el porcentaje de elecciones correctas de la fase uno, de cinco grupos en una tarea conductual de discriminación de relación forma-color para valores de IR diferentes en cada grupo.

juicio $c1/c1+c2$ para cada valor del intervalo. Por tanto la recuperación espontánea del aprendizaje causal no depende del valor del IR.

Por otro lado en nuestro laboratorio hemos realizado una serie de experimentos con una tarea conductual de discriminación de relaciones (forma y color) presentada individualmente en un monitor de ordenador en la que cincuenta participantes aprendían durante 24 ensayos a elegir uno de dos posibles estímulos en base al color de un tercer estímulo central (mismo color, diferente forma) en una primera fase (si hacían la elección correcta se les reforzaba con la palabra *acierto* en el monitor). En una segunda fase esta relación se invirtió y los participantes tenían ahora durante 24 ensayos que elegir uno de dos estímulos en base a la forma del estímulo central (misma forma, diferente color). Finalmente se presentó una prueba de 6 ensayos sin reforzamiento después de cinco valores del IR (0, 1'5, 3, 24, y 48 hrs.), uno para cada grupo de 10 participantes cada uno. Todos los grupos aprendieron a elegir correctamente el estímulo adecuado durante las dos primeras fases. Sin embargo durante la prueba sólo el grupo con un IR de 0 hrs. tuvo un número menor de elecciones correctas. No existieron diferencias entre los otros cuatro grupos con los demás valores del IR, estos datos se presentan en la parte derecha de la Figura 1. Los datos demuestran que la recuperación espontánea en humanos en una tarea conductual no es una función del IR.

En conjunto, estos datos contradicen una interpretación de la recuperación espontánea en humanos basada en el supuesto del “cambio recencia-primacia” (Knoedler, y col., 1999; Pineño y Miller, 2005; Stout, Amundson y Miller, en prensa), ya que a partir del valor mínimo del IR los participantes dieron su juicio causal en relación a la contingencia total del EC durante ambas fases, de manera similar a cuando se pide un solo juicio causal

al final del entrenamiento. Esto nos permite suponer que los participantes recuerdan ambas fases después de un IR corto y que esta memoria perdura al menos 48 hrs. (Matute y col., 2002). Recientemente, Pineño y Miller (2005) han supuesto que en la recuperación espontánea las claves producidas en el regreso a la situación experimental con IR largos estarían potenciando el recuerdo de la primera asociación sobre la segunda, lo que produciría el cambio de recencia a primacía tanto en animales como en humanos.

Sin embargo en el tercer experimento de Alvarado y col. (2005) se alteró la contingencia de la clave en las dos fases de interferencia aumentando el número de ensayos Ax a 30 y reduciendo el número de ensayos Ay a 10 (contingencia Ax = 0'75; contingencia Ay = 0'25). Para la mitad de los participantes los ensayos se presentaron en orden, primero 30 ensayos Ax y después 10 ensayos Ay, para la otra mitad los ensayos fueron entremezclados. Se presentó un IR de 48 hrs. para la mitad de los participantes y de 0 hrs. para la otra mitad, dando lugar así a cuatro grupos. Los juicios obtenidos en la prueba final para los dos grupos con un IR de 48 hrs. y para el grupo entremezclado de 0 hrs. igualaron a la contingencia programada para cada consecuencia. De este modo un IR de 48 hrs. sólo tiene efectos de primacía cuando los ensayos se presentan en el orden Ax /Ay, pero no cuando se presentan de forma entremezclada.

Estos datos demuestran que después de un IR los humanos recuerdan ambas fases y por tanto mantienen ambas memorias como un solo episodio, ya que al emitir un juicio lo hacen tomando en cuenta la contingencia total programada para el EC (Matute y col., 2002; Vadillo y col., 2004), quizá simplemente porque los humanos al regresar a la situación experimental recuerdan la situación de una manera global y no como dos memorias distintas. Esto último tiene sentido sólo para quien programa el experimento, dado que para los participantes ambas fases ocurren en la misma situación espacio-temporal. Así, la aproximación sugerida por Maldonado y col. en este mismo volumen parece más acertada al sugerir que la recencia se pierde al vaciarse el almacén de memoria de trabajo, recuperándose con el paso del tiempo lo almacenado en la memoria de referencia.

Por otro lado no se debe de olvidar la capacidad de descripción verbal de las contingencias que tenemos los humanos y su posible papel en las inferencias causales (Baron y Galitzio, 1983; De Houwer y col., 2005), lo que sugiere la posibilidad de que al presentarse un IR mayor a cero los participantes se decidan a cambiar la estrategia seguida hasta ese momento en función de la descripción verbal de la situación. La recuperación de información después de un IR en humanos es por tanto un efecto complejo que no puede ser comprendido sólo a partir de modelos derivados de la investigación animal (Bouton, 1993; Pineño y Miller, 2005).

Interferencia entre claves en humanos

Al presente los efectos de recuperación son considerados como equivalentes independientemente de si han sido producidos por una interferencia entre claves Ax/Bx (Matute y Pineño, 1998; Pineño y Matute, 2000; Pineño y col., 2003) o por una interferencia entre consecuencias Ax/Ay (Rosas y col., 2001; Vila y col., 2003). De hecho en la literatura

existen intentos de analizar ambos tipos de interferencia en los mismos términos (Underwood, 1966; Miller y Escobar, 2002) y es fácil suponer un mecanismo único a partir de que en ambas interferencias existe siempre un elemento común, ya sea este una clave o una consecuencia (Pineño y Matute, 2000).

La evidencia experimental existente de la interferencia entre claves con humanos procede de tareas conductuales de castigo discriminado (Matute y Pineño, 1998; Pineño y col., 2003). En general, en estas tareas el responder ante un estímulo A produce refuerzo positivo –ganar puntos– y responder a un segundo estímulo B produce castigo –perder puntos– (Dinsmoor, 1952). Un problema con estas tareas es que no distinguen si la disminución de la respuesta dada ante el estímulo B disminuye sólo por los efectos de la extinción (no se ganan puntos) o por los efectos del castigo (se pierden puntos). De esta forma, estas tareas serían equivalentes a un programa múltiple en el que se entrena una discriminación entre varios estímulos sucesivos asociados a diferentes contingencias de refuerzo que requiere de una sola respuesta (Catania, 1968). Sin embargo, en la literatura se les considera equivalentes a la interferencia entre consecuencias observada en tareas causales (Rosas y col., 2001), y actualmente son consideradas como tareas válidas en el estudio de la interferencia y sus efectos de recuperación (Miller y Escobar, 2002; Pineño y col., 2003). No obstante, al presente la única demostración en aprendizaje causal de la interferencia entre claves proviene del trabajo de Matute y Pineño (1998), en el que para poder replicar la interferencia entre claves se tiene que tomar un juicio predictivo en lugar de un juicio causal; o se tiene que instruir a los participantes para que emitan sus juicios causales en un supuesto contexto temporal determinado. Así, cabe suponer que una posibilidad para poder observar interferencia entre claves en aprendizaje causal podría depender de un análisis de la tarea en términos de los procesos dirigidos presentes en la situación. En las tareas conductuales, cuando está presente un estímulo los participantes deben inferir que el segundo estímulo no lo está. Esto es, cuando está presente el estímulo A responder produce puntos, entonces el estímulo B no está presente y viceversa. Esta regla de incompatibilidad de presencia entre las claves (si está A no está B) es al parecer una variable necesaria para la observación de interferencia.

Datos preliminares obtenidos en nuestro laboratorio han demostrado la interferencia entre claves en una tarea de aprendizaje causal cuando se da a los participantes una instrucción explícita acerca del estado de ambas claves. Se empleó una tarea causal de interferencia entre claves en la que cuatro grupos de participantes aprendían durante la tarea que el derrame de dos sustancias químicas A y B (Alfa y Beta) provocaba el encendido de la luz de un indicador (rojo o verde) en el panel de seguridad de una planta química. La tarea se presentó en pantallas en las que en el grupo implícito aparecía un enunciado que le decía al participante que una sola sustancia se había derramado, mientras que para el grupo explícito se indicó el estado de las dos sustancias con una leyenda debajo del nombre de la sustancia que decía “SÍ se derramó” o “NO se derramó” según el diseño experimental. De esta manera, en el grupo implícito se presentó información sólo de la presencia de una sustancia cada vez, mientras que en el grupo explícito se le indicó al participante la presen-

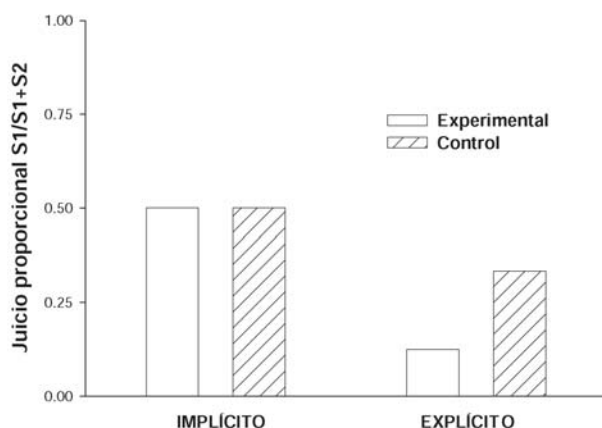


Figura 2. Se presenta el juicio proporcional ($S1/S1+S2$) del último juicio para cada clave en los grupos experimental y control de interferencia entre claves en función del entrenamiento explícito o implícito.

cia de una de las sustancias y la ausencia de la segunda sustancia. A continuación se hizo la siguiente pregunta “¿Qué indicador se encenderá?”, debajo de la pregunta aparecía un rectángulo que contenía dos círculos de colores (rojo y verde) con un número cada uno de los círculos, 1 para el círculo verde y 3 para el círculo rojo. Se contrabalanceó la posición de los indicadores de colores y de las claves. La pantalla de retroalimentación fue un enunciado que le decía al participante que el derrame de la sustancia precedió al encendido de indicador rojo o verde según el diseño experimental. Los juicios causales se solicitaron cada 10 ensayos, en todos los grupos se presentó una pantalla en la que se preguntó a cada participante en qué medida consideraba que el derrame de una de las dos sustancias provocaba el encendido de ambos indicadores, marcando con una “X” el recuadro de una escala con 10 recuadros que se presentó del lado derecho del círculo con el color rojo y otra escala para el círculo verde. La tarea constó de dos fases con 20 ensayos cada una. En el grupo explícito se llevó a cabo el mismo entrenamiento que en el grupo anterior, sólo que siempre se informó al participante del estado de las dos sustancias. En los grupos control (implícito y explícito) se entrenó la relación Ax, durante la primera fase y la relación By, durante la segunda fase.

Los resultados se presentan en la Figura 2 y muestran que la interferencia entre claves sólo ocurrió en el grupo experimental explícito en el que los juicios para cada sustancia fueron diferentes. El juicio proporcional calculado como clave uno entre el total de ambos juicios ($S1/S1+S2$) fue de 0’33 en el grupo experimental y de 0’125 para el grupo control. En el grupo control explícito y los grupos implícito control y experimental los juicios a cada clave no difieren y su proporción de juicios $S1/S1+S2$ fue de 0’5 indicando valoraciones iguales para cada clave y por tanto una ausencia de interferencia ya que ambas claves se valoran por igual. Este hallazgo demuestra que la información presentada en el grupo

explícito acerca del estado de ambas sustancias (una presente y la otra ausente) es un factor determinante para la ocurrencia de la interferencia entre claves en una tarea de juicios causales.

Un segundo experimento replicó el estudio anterior pero empleando una tarea de interferencia entre consecuencias (Ax/Ay), la tarea fue la misma que en el estudio anterior, sólo que ahora la relación que se presentó durante la primera fase fue Ax , donde el derrame de una sustancia provocó el encendido de un indicador durante la primera fase (verde), mientras que esa misma sustancia en la fase dos provocó el encendido de un segundo indicador (rojo), (relación Ay), contrabalanceando las consecuencias. De la misma manera que en el experimento anterior se entrenó a un grupo con información implícita y a otro con información explícita, en la que se informaba a los participantes sobre el estado de cada indicador. Ambos grupos de interferencia entre consecuencias, implícito y explícito, dieron juicios de cero para la relación de la clave A con la consecuencia uno y de 100 para la relación de A con la consecuencia dos (juicio proporcional, $C1 / C1 + C2 = 1$), lo que demuestra que en aprendizaje causal la interferencia entre consecuencias es un efecto robusto que no depende de instrucciones adicionales para su ocurrencia y que es de mayor magnitud que la interferencia entre claves.

Estos hallazgos sugieren que la interferencia entre consecuencias ocurre fácilmente en varias preparaciones (Bouton, 1993; Rosas y col., 2001; Underwood, 1966; Vila y col., 2002) y que se produce de manera distinta a la interferencia entre claves en tareas de aprendizaje causal, ya que no requiere de una instrucción que informe a los participantes del estado de ausencia de la segunda consecuencia. Esto sugiere una mayor participación de procesos dirigidos en la interferencia entre claves. Un análisis de las tareas conductuales en función del tipo de inferencia empleada por los participantes nos indica que para poder responder ante un estímulo y ganar puntos es necesario el supuesto o inferencia de que el segundo estímulo no está operativo (para un punto de vista alternativo véase Luque, Cobos y López, en este volumen). Esta inferencia no existe en la interferencia entre consecuencias, en la que el significado de una clave que produce una primera consecuencia se contradice en una segunda fase produciendo ambigüedad tal y como ocurre en un discriminación inversa (Vila y col., 2002), por lo que se hace necesario el estudio de las diferencias entre los dos tipos de interferencia en humanos antes de poder hacer generalizaciones válidas entre ellas.

El efecto de las instrucciones en la recuperación

Es bien conocido que las instrucciones dadas a los participantes pueden afectar el aprendizaje. Por ejemplo, el solo hecho de informar a un participante de que el EC ya no se verá seguido por el EI produce la extinción de la RC (Colgan, 1970; Hugdahl y Öhman, 1977). Las instrucciones han demostrado también tener propiedades similares a las de un estímulo discriminativo en una tarea operante en humanos (Okouchi, 1999), o de funcionar como un contexto físico en la renovación (Romero, Vila y Rosas, 2005). Recientemente, Lovibond (2004) ha sugerido que el desarrollo de una RC en humanos depende en gran

medida del lenguaje, ya que puede aumentar o disminuir por el efecto de instrucciones que describan la relación EC-EI. Sin embargo, aun y cuando en varios estudios de aprendizaje causal se manipulan variables instruccionales, apenas se menciona el papel que las instrucciones tienen sobre el aprendizaje humano (Cobos y col., 2000; Matute y col., 1996; Matute y col., 2002), lo que quizá sea resultado del interés en explicar el aprendizaje humano sólo a partir de modelos derivados de la investigación animal.

En el estudio del aprendizaje causal el efecto del modo de respuesta o frecuencia del juicio se produce cuando al solicitar un solo juicio al final de dos fases de entrenamiento se produce una integración de ambas, mientras que cuando se solicita un juicio en cada ensayo durante ambas fases se produce un efecto de recencia en el que sólo se manifiesta la información de la segunda fase (Catena, Maldonado y Candido, 1998). Recientemente se han demostrado los efectos de las instrucciones sobre la frecuencia del juicio (Matute y col., 2002). En este estudio, cuando a participantes entrenados con un modo de respuesta ensayo a ensayo en dos fases (adquisición y extinción) se les presentó una pantalla antes de la prueba final en la que se les instruyó a responder en función de toda la información presentada, los juicios emitidos fueron de integración de ambas fases en comparación a un grupo que dio un juicio de recencia cuando no se presentan instrucciones. Empleando una lógica similar, Vila, Alvarado, Jara, Flores y Flores (2002) realizaron un experimento en el que a cuatro grupos de 11 participantes se les entrenó en una tarea causal de interferencia entre consecuencias de 40 ensayos en la que en una primera fase un medicamento producía un síntoma *x* y en una segunda fase un síntoma *y*. Cada cinco ensayos se solicitaban los juicios causales a los participantes acerca de en qué medida creían que el medicamento producía ambos síntomas. A dos grupos se les pidió su último juicio inmediatamente (0 h) y a dos grupos a las 48 h. A un grupo de 48 y uno de 0 hrs. antes de la prueba final se les presentó una pantalla con instrucciones en la que se les indicaba a los participantes del grupo 0 hrs. que considerasen toda la información presentada, y a los participantes del grupo 48 hrs. que sólo considerasen la última información presentada. Los juicios de los cuatro grupos se presentan en la Figura 3, en la que se observa una interacción de las instrucciones con el intervalo de retención. Esto es, las instrucciones llevan a que el efecto de recuperación producido por un IR de 48 hrs. se invierta, ya que el grupo 0 hrs. con instrucciones da juicios de 50 % para cada consecuencia ($C1/C1+C2 = 0.5$), de manera similar a los juicios emitidos por grupo 48 hrs. sin instrucciones, y por otro lado el grupo de 48 hrs. con instrucciones responde con juicios de 0 para la consecuencia uno y de 100 % para la segunda consecuencia ($C1/C1+C2 = 0$), de la misma forma que lo hace el grupo de 0 hrs. sin instrucciones.

De esta manera, estos datos demuestran que las instrucciones son un modulador de la expresión de la información aprendida en ambas fases. Sin embargo, estos datos pueden interpretarse fácilmente argumentando que las instrucciones al igual que los contextos físicos y temporales pueden funcionar de la misma manera, tal y como se observa en los estudios con animales (Matute y col., 2002; Vadillo y col., 2004), aun y cuando sólo ha sido demostrada una interacción aditiva del contexto y tiempo (Rosas y col., 2001; Vila y

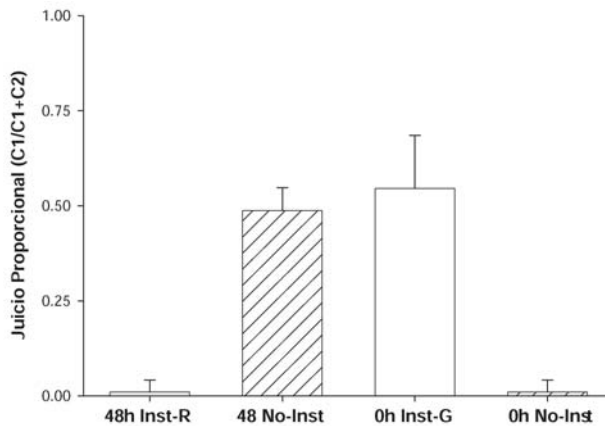


Figura 3. Se presenta el juicio proporcional ($C1/C1+C2$) del último juicio para cada consecuencia en función del IR (0 y 48 h) y su interacción con las instrucciones de recencia (Inst-R), las instrucciones globales (Inst-G) y sin instrucciones (No-Inst).

col., 2002) que no es equivalente a la interacción del IR con las instrucciones demostrada en el experimento anterior.

Por esta razón se hace necesario demostrar que las instrucciones pueden modular el comportamiento de los participantes durante toda la tarea y no sólo durante la prueba final. Así, Alvarado, Vila y Flores (2004) realizaron un estudio en el que a seis grupos se les presentó la tarea previa en la que una medicina era asociada a dos síntomas diferentes en dos fases. A tres grupos de 10 participantes se les requirió un solo juicio causal de la relación entre la medicina y ambos síntomas al final de la tarea (modo global), a otros tres grupos se les solicitó su juicio causal cada cinco ensayos (modo cada 5 ensayos). A dos grupos, uno global y otro cada cinco ensayos se les presentó una pantalla con instrucciones antes de comenzar la tarea, y a los dos grupos restantes se les presentó una pantalla con instrucciones antes de la prueba final. La instrucción para el caso de los grupos globales fue: “*Ten cuidado porque los primeros expedientes son falsos y tu última valoración puede ser errónea*” y en los grupos del modo cada cinco ensayos fue: “*Ten cuidado porque los últimos expedientes son falsos y tu última valoración puede ser errónea*”. Los resultados obtenidos muestran un efecto general de las instrucciones en todos los grupos y de si estas se presentan al final o al principio de la tarea. Los datos para ambos modos de respuesta se presentan en los dos paneles de la Figura 4.

Las instrucciones producen recencia en los grupos globales y un efecto de primacía en los grupos del modo cada cinco ensayos. Sin embargo, este efecto de las instrucciones es mayor cuanto más próximas sean las instrucciones en relación a la prueba final. Así, si las instrucciones se presentan al inicio de la tarea pueden modular toda la tarea. Por tanto las instrucciones no pueden ser consideradas sólo como demandas de la prueba como se ha sugerido en algunos estudios (Matute y col., 2002), ya que sus efectos son más similares a

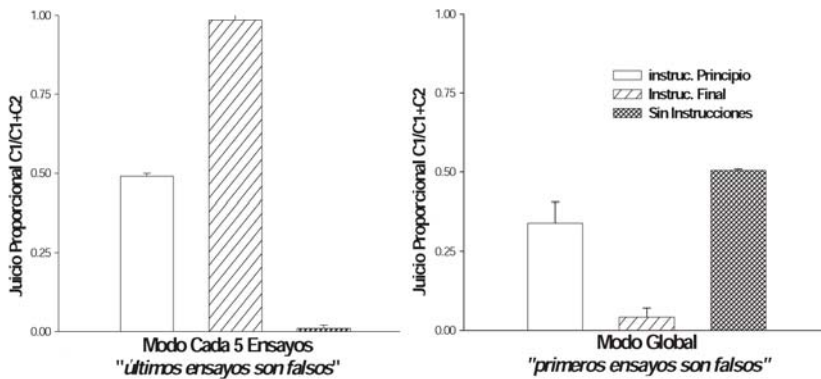


Figura 4. El panel izquierdo presenta los juicios proporcionales ($C1/C1+C2$) del último juicio para cada consecuencia de grupos con la instrucción “*Los últimos ensayos son falsos*” presentada al inicio de la tarea, al final y sin instrucciones, entrenados con un modo de respuesta cada 5 ensayos. En el panel derecho se presentan los juicios proporcionales de tres grupos entrenados con un modo global de respuesta con la instrucción “*Los primeros ensayos son falsos*” presentada al inicio de la tarea, al final y sin instrucciones.

los efectos de estímulos moduladores de otras asociaciones (*occasion setters*, Holland, 1992).

Un estudio de Lobivond (2003) sugiere que las instrucciones además pueden sustituir a las contingencias. Observó un efecto de reevaluación retrospectiva tanto cuando presentó los ensayos a los participantes o cuando el ensayo se describía verbalmente a través de una instrucción, lo que demuestra la posibilidad de observar efectos de aprendizaje casual con ensayos instruidos. Un estudio inédito de nuestro laboratorio intentó observar interferencia presentando en cada una de dos fases de interferencia ensayos instruidos en sustitución de los ensayos experimentados. A tres grupos de diez participantes se les dio un entrenamiento de dos fases en las que un medicamento producía un primer síntoma durante 20 ensayos y un segundo síntoma en los 20 ensayos finales. Se pidió a los participantes un juicio causal acerca de la relación del medicamento con cada uno de los síntomas. A un grupo no se le dio ninguna instrucción, a los dos grupos restantes se les presentó la instrucción “*La medicina produjo en los últimos cinco ensayos el síntoma...*” antes de preguntar por su juicio. Esta instrucción se presentó en cuatro ocasiones, a uno de los grupos se le presentaron los ensayos de instrucción en la fase uno y al otro en la fase dos. Los resultados no mostraron diferencias en los juicios proporcionales ($C1/C1+C2 = 0$) de los grupos sin instrucciones y con instrucciones en las fases uno y dos. Los tres grupos emitieron juicios de cero a la primera consecuencia y de 100 para la segunda consecuencia. Estos datos son coherentes con los de Lovibond (2003) y demuestran la posibilidad de observar interferencia entre consecuencias con ensayos instruidos de manera similar a cuando los ensayos son experimentados. El hecho de que las instrucciones por sí mismas produzcan los efectos de los ensayos reales nos sugiere una función de referencia en las instrucciones que permitiría a los participantes por medio del lenguaje tener acceso a estímulos o contingencias ausentes.

Es por ello que los efectos de las instrucciones en las tareas de aprendizaje causal y sus efectos de interferencia y recuperación debe reconsiderarse.

Conclusión

Se han presentado aquí una serie de efectos de recuperación de la interferencia retroactiva en humanos que son cualitativamente distintos a su contraparte con animales. Los modelos actuales que explican la interferencia en el aprendizaje asociativo (Bouton, 1993, 1997; Miller y Escobar, 2002) y aquellos que han intentado explicar algunos de los efectos de recuperación (Pineño y Miller, 2005; Rescorla, 2004) se han caracterizado por la explicación de datos animales y humanos a partir de principios similares. Lo que ha llevado a los investigadores a una extrapolación de variables metodológicas y conceptos propios de la investigación animal a los estudios con humanos y que ha tenido como resultado que variables típicamente humanas, como las instrucciones o el papel del lenguaje en las tareas estudiadas haya sido minimizado. Sin embargo, recientemente se ha comenzado a poner atención en el papel de los procesos dirigidos en el aprendizaje asociativo humano (De Houwer y col., 2005, en prensa; Lovibond, 2004). El presente capítulo es en esta misma medida una propuesta más para considerar los procesos dirigidos o controlados en la recuperación de información posterior a la interferencia retroactiva.

Una pregunta interesante es si el aprendizaje asociativo humano depende sólo de procesos controlados, o de procesos automáticos similares a los encontrados con animales. Por ahora debemos de considerar que ninguna de las dos aproximaciones puede por sí sola explicar satisfactoriamente los efectos de recuperación en humanos, tales como la recuperación espontánea o la renovación. Sin embargo, a manera de propuesta, consideramos que el estudio detallado de las interacciones entre variables humanas como las instrucciones o el modo de respuesta y las variables estudiadas en animales como el cambio de contexto o el IR podría explicarnos la medida en la que el aprendizaje animal participa en los mecanismos de recuperación estudiados con humanos. Los trabajos resumidos en este capítulo pueden ser considerados un inicio en esta dirección. Así el estudio de las interacciones entre las instrucciones y el IR o de las instrucciones y el modo de respuesta son algunos de los puntos que nos indicarían en qué medida se puede generalizar entre ambas especies. Desgraciadamente al presente el estudio de este tipo de interacciones en humanos ha sido mas bien escaso (sin embargo véase Alvarado y col., 2005; Rosas y col., 2001; Vadillo y col., 2004; Vila y col, 2002).

Un problema en el estudio del aprendizaje asociativo humano ha sido el diseño de tareas adecuadas para replicar efectos de recuperación observados en animales, pero que sean además herramientas adecuadas para el estudio estandarizado del aprendizaje humano. Actualmente existen en la literatura un gran número de tareas para el estudio del aprendizaje humano sin que se cuente con una comparación adecuada de su validez y utilidad práctica. Por ejemplo, las tareas conductuales se han tomado como equivalentes de las de aprendizaje causal sin estudiar a fondo las variables responsables de cada una de ellas. Se han mezclado así de manera indistinta tareas instrumentales con tareas pavlovias-

nas olvidando el efecto que tiene en los humanos el refuerzo o el castigo de conductas. Es necesaria una estandarización y validación de las tareas empleadas en los estudios con humanos, del mismo modo que se tiene en los estudios animales. Debemos recordar que las tareas para el estudio del aprendizaje humano no deben ser una cuestión de creatividad sino de validez. Resulta paradójico que en las tareas con humanos se hayan generalizado los efectos y explicaciones derivados de los estudios con animales, pero no así el riguroso control y estandarización que caracteriza a estos estudios.

En la explicación del efecto de la recuperación espontánea producido por un IR posterior a la interferencia en humanos se han planteado dos aproximaciones; una de ellas como una extensión de los estudios con animales (Pineño y Miller, 2005) y la segunda como la aplicación de principios derivados de resultados observados con humanos (Alvarado y col., 2005). La evidencia experimental ha favorecido esta última al demostrar por un lado que la recuperación espontánea en humanos no es una función del valor del IR, en la que se aplique un cambio de recencia a primacía, y por otro lado que esta recuperación podría ser debida a una integración de fases en una sola memoria de referencia que se ajusta a la contingencia total experimentada durante todo el entrenamiento.

No podría existir un estudio del aprendizaje humano adecuado sin considerar la participación del lenguaje o al menos de las variables verbales que en él participan. En un trabajo anterior (Vila y col., 2003) ya se señalaba que el lenguaje no se define por sus dimensiones físicas ni por su correspondencia con las dimensiones físicas de los estímulos, sino por sus efectos en la comunidad que lo utiliza y que es esta característica precisamente la que dificulta su conceptualización y estudio como variable experimental. Así, considerando la función de referencia del lenguaje, que permite que los humanos respondan ante estímulos ausentes como si estuviesen presentes físicamente sólo por escuchar la palabra que los nombra, se debe pensar que esta función permite a los humanos responder ante adverbios de tiempo y lugar como “*aquí, allá, entonces o ahora*” como si fuesen contextos físicos o el paso del tiempo, sin que realmente lo sean, sino sólo de una forma referida o metafórica en la que podemos comportarnos *como si* ese cambio realmente hubiese ocurrido. Otra función del lenguaje que es importante mencionar es la capacidad de nombramiento de estímulos que tenemos los humanos que puede generar respuestas diferentes a las consideradas en la situación y que ocasiona nuevos estímulos, discriminaciones y memorias de los arreglos presentados (Horne y Lowe, 1996). Hasta el momento ninguna de estas propiedades tan comunes en los humanos ha sido demostrada aún con animales. Las implicaciones del estudio de estas propiedades en el razonamiento inferencial de los participantes en un estudio de interferencia retroactiva y su posterior recuperación en una situación de aprendizaje causal son grandes y aún no se han estudiado.

La demostración de algunas diferencias entre la recuperación de la RC en humanos y animales se plantea como un inicio en el estudio de los procesos controlados en los efectos de recuperación con humanos, lo que permitiría entender cuáles son los mecanismos que determinan las similitudes y diferencias de esta situación entre humanos y animales. Si bien el planteamiento de tratar de entender el comportamiento humano a partir de los estudios

con animales parece ser una estrategia adecuada, no debemos olvidarnos de sus limitaciones y dejar de mirar aquellos factores que hacen que los humanos tengamos características únicas.

Nota de los autores

Este trabajo ha sido realizado en el marco y con la ayuda de los proyectos de CONACyT 40345 de México y de la DGAPA de la UNAM IN302605. La correspondencia relacionada con el presente trabajo puede ser enviada a N. Javier Vila-Carranza, Facultad de Estudios Superiores, UNAM, División de Investigación y Posgrado, A.P. 314, 54090, Edo. de México, MEXICO.

Referencias

- Alvarado, A., y Vila, N. J. (2004). Efectos de las instrucciones en una tarea causal de interferencia con dos consecuencias. *XVI Congreso de la Sociedad Española de Psicología Comparada*. Universidad de Oviedo.
- Alvarado, A., Jara, E., Vila, J. y Rosas, J. M. (2005). Time and order effects on causal learning. *Manuscrito remitido para su publicación*.
- Baron, A. y Galizio, M. (1983). Instructional human behavior. *The Psychological Record*, 33, 495-520.
- Bouton, M. E y Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*. 10, 445-466.
- Bouton, M. E. (1993). Context, time and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80-99.
- Bouton, M. E. (1997). Signals for whether versus when an event will occur. En Bouton, M. E., Fanselow, M. S. (Eds.) *Learning, Motivation and Cognition. The Functional Behaviorism of Robert C. Bolles*. Washington: American Psychological association.
- Bouton, M. E. (2002). Context, ambiguity and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, 52, 976-986.
- Bouton, M. E. (2004). Context and Behavioral Processes in Extinction. *Learning and Memory*, 11, 485-494.
- Brewer, W. F. (1974). There is no convincing evidence of conditioning in adult humans En W. B. Weirner y D. S. Palermo (Eds). *Cognition and the symbolic processes*. Hillsdale, N. J: Erlbaum.
- Catania, C. (1968). *Contemporary Research in Operant Behavior*. Glenview: Scott Foresman.
- Catena, A., Maldonado, A. y Cándido, A. (1998). The effect of the frequency of judgment and type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 2, 481-495.
- Cobos, P. L., Caño, A., López, F. J., Luque, J. L. y Almaraz, J. (2000). Does the type judgment required modulate cue competition? *Quarterly Journal of Experimental Psychology*, 53B, 193-207.
- Colgan, D. M. (1970). Effect of instructions on the skin conductance response. *Journal of Experimental Psychology*, 86, 108-112.
- Davey, G. C. L. (1987). An integration of human and animal models of pavlovian conditioning: Associations, cognitions and attributions. En G. C. L. Davey (Ed). *Cognitive processes and Pavlovian conditioning in humans*. Chichester: Wiley.
- De Houwer, J., Vandorpe, S. y Beckers, T. (2005). On the role of controlled cognitive processes in human associative learning. En A. J. Wills (Ed.). *New directions in human associative learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc
- De Houwer, J. Beckers, T. y Vandorpe, S. (*en prensa*). Evidence for the role of higher-order reasoning in cue competition and other learning phenomena. *Learning & Behavior*.
- Dinsmoor, J. A. (1952). A discrimination based on punishment. *Quarterly Journal of Experimental Psychology*, 4, 27-45.

- Escobar, M., Matute, H. y Miller, R. R. (2001). Cues Trained apart compete for behavioral control in rats: Convergence with the associative interference literature. *Journal of Experimental Psychology: General*, 130, 97-115.
- Escobar, M., Pineño, O. y Matute, H. (2002). A comparison between elemental and compound training of cues in retrospective reevaluation. *Animal Learning & Behavior*, 30, 228-238.
- García-Gutiérrez, A. y Rosas, J. M. (2003). Context change as the mechanism of reinstatement in a causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 29, 292-310.
- Holland, P. C. (1992). Occasion setting in Pavlovian conditioning. *The Psychology of Learning and Motivation*, 28, 69-125.
- Horne, P. J. y Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185-241.
- Hugdahl, K. y Öhman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 608-618.
- Knoedler, A. J., Hellwig, K. A. y Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 474- 487.
- Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective reevaluation induced by experience and instruction. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 97-106.
- Lovibond, P., F. (2004). Cognitive Processes in Extinction. *Learning and Memory*, 11, 495-500.
- Matute, H., Arcediano, F. y Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22, 182-196.
- Matute, H. y Pineño, O. (1998). Cue competition in the absence of compound training. Its relation to paradigms of interference between outcomes. En D. L. Medin (Ed.). *The Psychology of Learning and Motivation*, Vol. 38. San Diego, CA: Academic Press.
- Matute, H., Vegas, S. y De Marez (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition*. Vol. 28, 4, 714-725.
- Miller, R. R., y Escobar, M. (2002). Associative interference between cues and between outcomes presented together and presented apart: An integration. *Behavioral Processes*, 57, 163-185.
- Okouchi, H. (1999). Instructions as discriminative stimuli. *Journal of Experimental Analysis of Behavior*, 72, 205-214.
- Pavlov, I. (1927). *Conditioned Reflexes*. London: Oxford University Press.
- Pineño, O. y Miller, R., R. (2004). Signaling a change in cue-outcome relations in human associative learning. *Learning and Behavior*, 32, 360-375.
- Pineño, O. y Miller, R. R. (2005) Primacy and recency effects in extinction and latent inhibition: A selective review with implications for models of learning. *Behavioural Processes*, 69, 223-235.
- Pineño, O., Vegas, S. y Matute, H. (2003). Factores que median la expresión del aprendizaje asociativo humano. En J. Vila, J. Nieto y J. M. Rosas (Eds.), *Investigación contemporánea en aprendizaje asociativo: estudios en España y México*. Jaén: del lunar-UNAM.
- Pineño, O. y Matute, H. (2000). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology*, 13, 16-33.
- Rescorla, R. (2004). Spontaneous Recovery. *Learning and Memory*, 11, 501-509.
- Rescorla, R. y Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavioral Processes*, 1, 88-96.
- Romero, M. A., Vila, J. y Rosas, J. M. (2005). Instructions inadequate to the task control response recovery after discrimination reversal in human beings. *Manuscrito remitido para su publicación*.
- Rosas, J. M., García-Gutiérrez, A. y Romero, M. A. (2003). Contexto y Tiempo en la recuperación de la información. En J. Vila, J. Nieto y J. M. Rosas (Eds.), *Investigación contemporánea en aprendizaje asociativo: estudios en España y México*. Jaén: del lunar-UNAM.
- Rosas, J. M., Vila, J., Lugo, M. y López, L. (2001). Combined effect of context change and retention interval upon proactive interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 2, 153-164.

- Stout, S., Amundson, J. C. y Miller, R. R. (*en prensa*). Trial order and retention interval in human predictive judgments. *Memory and Cognition*.
- Underwood, B. J. (1996). *Experimental Psychology (2nd ed.)*. New York: Appleton-Century-Crofts.
- Vadillo, M. A., Vegas, S. y Matute, H. (2004). The frequency of judgment as a context-like determinant of predictive judgments. *Memory and Cognition*.
- Vila, N. J., Alvarado, A., Jara, E., Flores, J. y Flores, S. (2002). Efectos de las instrucciones de prueba y del intervalo de retención en la recencia e integración de fases en una tarea de aprendizaje causal con interferencia. *XIV Congreso de la Sociedad Española de Psicología Comparada*. Universidad de Sevilla.
- Vila, J. y Rosas, J. M. (2001). Renewal and spontaneous recovery after extinction in a causal-learning task. *Mexican Journal of Behavior Analysis*, 27, 79-96.
- Vila, J., Romero, M. A. y Rosas, J. M. (2002). Retroactive interference after discrimination reversal decreases following temporal and physical context changes in human subjects. *Behavioural Processes*, 59, 47-54.
- Vila, J., Alvarado, A., Jara, E. y Flores, J. (2003). La extinción experimental en aprendizaje causal. En J. Vila, J. Nieto y J. M. Rosas (Eds.), *Investigación contemporánea en aprendizaje asociativo: estudios en España y México*. Jaén: del lunar-UNAM.

Índice

Presentación	5
Procesamiento de la información en el aprendizaje causal: ¿qué se aprende? <i>A. Maldonado, A. Herrera, A. Catena, A. Cándido y J. C. Perales.</i>	7
Procesos de razonamiento causal en la interferencia entre señales. <i>D. Luque, P. L. Cobos y F. J. López.</i>	18
Nuevas perspectivas teóricas para el estudio de la extinción en juicios de causalidad. <i>M. A. Vadillo y H. Matute.</i>	31
Contexto y recuperación de la información: ¿qué hace que la recuperación de la información sea dependiente del contexto? <i>J. M. Rosas, A. García-Gutiérrez, M. J. F. Abad y J. E. Callejas-Aguilera</i>	47
Diferencias en la recuperación de la respuesta condicionada después de la interferencia retroactiva en animales y humanos <i>J. Vila y A. Alvarado</i>	62



Universitas

- Michael Domjan. *Bases del aprendizaje y el condicionamiento.*
- Juan M. Rosas y José E. Callejas. *Aprende condicionando.*
- Andalucía Acoge. *La inserción sociolaboral de inmigrantes*
- Edward C. Tolman. *Una nueva fórmula para el conductismo y otros escritos.*
- Angel Cagigas. *Guía de viaje por una historia de la psicología.*
- Juan M. Rosas. (Ed.). *Teorías asociativas del aprendizaje.*
- J. Vila, J. Nieto y J. M. Rosas. (Eds.). *Investigación contemporánea en aprendizaje asociativo.*
- José María Colmenero. *Percepción visual y auditiva.*
- J. M. Augusto Landa, E. López Zafra, R. A. Martínez de Antoñana Ugarte. *Introducción a la Psicología Social.*
- Francisco Balbuena Rivera. *Psicoterapias cognitivo-conductuales. Desarrollo histórico y estado actual.*
- Esther López Zafra y Pilar Berrios Martos (Dir.). *Violencia en las aulas.*
- Juan M. Rosas, Ana García y José E. Callejas. *Fundamentos del aprendizaje humano.*
- José María Colmenero. *La memoria.*
- Francisco Pérez Fernández. *Imbéciles morales.*
- Javier Vila y Juan M. Rosas. *Aprendizaje causal y recuperación de la información.*

El sillón de orejas

- Georg Groddeck. *Escritos.*
- Angel Cagigas. *Georg Groddeck: el soñador de mundos.*
- Georg Groddeck. *La vista, el mundo del ojo y ver sin ojos.*
- Karl Abraham. *Sueño y mito.*
- Georg Groddeck. *Las tripas.*
- Angel Cagigas. *Genio y figura. Georg Groddeck en imágenes y textos.*
- Sandor Ferenczi y Georg Groddeck. *Correspondencia (1921-1933).*
- Dezso Kosztlányi. *Cuentos psicoanalíticos.*

Heterohistorias

- William James. *La inmortalidad humana.*
- Max Nordau. *Fin de siglo*
- Jean Martin Charcot y Paul Richer. *Los endemoniados en el arte.*
- Wilhelm Wundt. *Hipnotismo y sugestión.*
- Gustav Theodor Fechner. *Anatomía comparada de los ángeles.*
- Jean Martin Charcot, Valentin Magnan y Alfred Binet. *Perversiones.*
- Jean Martin Charcot y Paul Richer. *Los deformes y los enfermos en el arte.*
- Jean Martin Charcot. *Histeria. Lecciones del martes.*
- Angel Cagigas. *La histeria de Charcot.*
- Paul Bourget. *Taine.*
- William James. *Investigación psíquica.*

Del nagual

- Reny Poch. *Poemigas.*
- Reny Poch. *Columnata.*