

## **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: Appelon-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.