

Aprendizaje de relaciones de contingencia y causalidad: Una aproximación a las tendencias teóricas actuales

José C. Perales*, Andrés Catena*, Manuel M. Ramos** y Antonio Maldonado*

*Universidad de Granada; **Universidad de Jaén

En las dos últimas décadas el estudio de la causalidad desde la perspectiva de la Psicología del Aprendizaje ha experimentado un auge notable. Sin embargo, el crecimiento del conjunto de datos experimentales no ha ido acompañado de un desarrollo teórico que permitiese integrarlos. Cualquier marco teórico que pretenda ser explicativo debe dar cuenta de estos datos, así como su relación con los tipos de tarea utilizados, a menudo de forma indiferenciada. Esta revisión pretende ofrecer una visión organizada del corpus experimental desarrollado dentro del campo del aprendizaje causal humano y esbozar algunas líneas posibles para la integración teórica. El análisis de las similitudes e incompatibilidades entre distintas aproximaciones (hasta la fecha necesariamente parciales) podrá llevar en el futuro a modelos de organización serial que abarquen las distintas etapas del procesamiento en este tipo de aprendizaje.

Palabras clave: Aprendizaje de causalidad, contingencia, covariación, modelos asociativos, modelos de reglas.

1. Introducción.

Se denomina *aprendizaje causal* al proceso mediante el que los organismos son capaces de captar las relaciones entre acontecimientos del medio que les rodea y así adaptar su comportamiento de forma apropiada. Según Kareev (1995), la adquisición de información sobre la textura causal del medio es una capacidad psicológica fundamental porque permite al organismo predecir acontecimientos futuros basándose en la información actual, pero también manipular el medio de manera efectiva para provocar consecuencias deseables o para evitar las negativas.

* Correspondencia a: Andrés Catena / José C. Perales. Depto. de Psicología Experimental y Fisiología del Comportamiento. Universidad de Granada. Facultad de Psicología. Campus de Cartuja s/n. 18071 Granada. E-mails: acatena@platon.ugr.es y jperales@platon.ugr.es. Fax: 958-246339

En general, el concepto de relación de causalidad es claramente diferente del de *contingencia*. La diferencia entre ellos es un tema central para la filosofía de la causalidad. Para la ciencia psicológica, la cuestión más importante estriba en determinar hasta qué punto aprender que dos fenómenos ocurren conjuntamente (relación de contingencia) y aprender que uno produce el otro (relación de causalidad) son habilidades que se basan en procesos cognitivos comunes. En este sentido, parece claro que si dos eventos covarían ello no implica que el primero (A) sea causa del segundo (B). La covariación entre los acontecimientos A y B puede deberse a que, efectivamente, el antecedente A sea la causa del consecuente B. Sin embargo, aunque un descenso en el indicador de un barómetro casero suele acompañarse de la aparición de precipitaciones, pocos interpretarían que el indicador del barómetro tenga algún tipo de influencia (causal) misteriosa sobre los fenómenos atmosféricos. Más bien, la interpretación general sería que existe un tercer acontecimiento (una disminución de la presión atmosférica) que es simultáneamente la causa de la lluvia y de la bajada del barómetro, llevando a que la bajada del barómetro permita predecir la lluvia. Estas relaciones de predicción pueden establecerse en otros casos debido a que realmente existe un nexo causal entre A y B, aunque la consecuencia se manifieste en primer lugar. Por ejemplo, la aparición de un síntoma (consecuencia) es previa a la constatación de la enfermedad (la causa) que lo provoca.

Obviamente, las personas son capaces de verbalizar la diferencia entre una relación de causalidad y una de mera concurrencia o predicción (un acontecimiento predice la presencia de otro, aún cuando no sea su causa). Esa posibilidad no existe en los estudios con animales no humanos. La disociación entre ambos aprendizajes es discutible en animales, sobre todo porque la comprobación ha de ser indirecta (Miller y Matute, 1996; Premack, 1993). La importancia de esta discusión radica en la posibilidad de extrapolar modelos teóricos surgidos en el campo del aprendizaje animal al aprendizaje causal humano (Dickinson y Shanks, 1996; Shanks, 1993; Wasserman, 1993).

En este trabajo perseguimos dos objetivos. En primer lugar, estamos convencidos de que el progreso en un campo de investigación está relacionado de forma directa con su sistemática. En Psicología la sistemática depende básicamente del análisis de las tareas. En otras palabras, el tipo de tarea que se emplee puede determinar qué variables van a producir efecto y, por ende, qué clase de explicaciones teóricas pueden realizarse. Por consiguiente, presentaremos una revisión de la literatura no tanto basada en fenómenos, cuanto basada en tareas y variables que las afectan. Consideramos que esta perspectiva facilitará la comprensión del campo en su estado actual y también en su desarrollo pasado y futuro. En segundo lugar, presentaremos las aproximaciones teóricas más influyentes en la explicación del aprendizaje de relaciones de contingencia y causalidad, y concluiremos argumentando que para soslayar las dificultades de estos modelos probablemente sea necesario postular un modelo que organice de forma serial las etapas de procesamiento.

2. Tareas de aprendizaje causal.

Todas las tareas utilizadas en el estudio del aprendizaje de relaciones causales siguen un mismo esquema básico. El objetivo es registrar la intensidad de la relación percibida por la persona entre dos variables a las que llamaremos, genéricamente, variable de entrada (*input*) y variable de salida (*output*). En el caso más simple, se emplean variables dicotómicas asimétricas, cuyos valores vienen definidos por la presencia o ausencia de un determinado estímulo o evento ambiental (ver Trolier y Hamilton, 1986, para una comparación de los juicios de contingencia con variables continuas y dicotómicas). Normalmente, a la variable que se presenta en primer lugar se le denomina **clave** (C) mientras que a la variable que se presenta en segundo lugar se le denomina **resultado** (R). La definición de un determinado evento ambiental como clave o resultado depende exclusivamente de la posición temporal del mismo, no del papel causal que ese estímulo o evento desempeña. Por tanto, cabe distinguir dos tipos de tareas: predictivas y diagnósticas (Morris y Larrick, 1995; Waldmann y Holyoak, 1992). En las **tareas predictivas** la clave es la causa potencial, mientras que el resultado es la consecuencia. Por ejemplo, una tarea en la que la persona debe estimar la relación entre la ingestión de un determinado producto y la aparición de una enfermedad sería una tarea predictiva. Por el contrario en las **tareas diagnósticas** la clave es la consecuencia mientras que el resultado es la causa. Por ejemplo, cuando se debe estimar en qué medida un determinado síntoma (el efecto, la clave) permite anticipar la presencia o ausencia de una enfermedad (la causa).

Así pues, en las tareas más sencillas en las que sólo se presentan una clave y un resultado, pueden definirse cuatro tipos de ensayos que resultan de la combinación de los valores posibles de las variables de entrada y salida. Según la notación más comúnmente utilizada (por ejemplo Allan, 1993; Arkes y Harkness, 1983; Chapman y Robbins, 1990; Levin, Wasserman y Kao, 1993), los ensayos en los que tanto la clave como el resultado están presentes se denominan ensayos **tipo a**; los ensayos en los que la clave está presente pero no el resultado se denominan **tipo b**; en los ensayos **tipo c** no se presenta la clave, pero sí el resultado; por último, en los **tipo d** no se presentan ni la clave ni el resultado.

Tabla 1. Tipos de ensayo posibles en una tarea estándar de juicios de contingencia.

	Resultado Presente	Resultado ausente
Clave presente	a	b
Clave ausente	c	d

$$P(R/C) = a / (a+b)$$

$$P(R/\text{no } C) = c / (c+d)$$

Es importante señalar de antemano que, con independencia de la respuesta pedida a las personas, suele asumirse que ésta es un índice del funcionamiento del mecanismo de aprendizaje encargado de computar la relación existente entre la variable de entrada y la variable de salida. Esto implica que prácticamente todos los modelos asumen una arquitectura de procesos muy simple, en la que entre la entrada y la salida hay un solo proceso, esto es, son modelos de **mecanismo único**. El corolario es evidente: las variaciones que se registren en la salida del sistema (respuesta) se deben única y exclusivamente a los cambios en ese mecanismo de aprendizaje. En otras palabras, las variaciones en la respuesta pueden interpretarse directamente como variaciones del mecanismo que computa.

Esta tarea básica permite la manipulación de un número relativamente restringido de variables. En efecto, principalmente es posible variar el tipo de respuesta que se pide a los sujetos, la relación definida entre la clave y el resultado y la complejidad de la tarea.

2.1. Efecto del tipo de respuesta.

En las **tareas de juicios de causalidad / contingencia**, la persona debe emitir un juicio numérico sobre la relación existente entre las dos variables. Si el juicio se refiere simplemente a la relación de covariación o concurrencia entre ambas, se trata de **tareas de juicios de contingencia**. Pero si se pide a la persona que estime la capacidad de una determinada causa potencial para provocar o prevenir una determinada consecuencia se trata entonces de una **tarea de juicios de causalidad**. Prácticamente en todos los casos el juicio se recoge mediante una escala numérica graduada, que puede ser unipolar (con valor cero en un extremo y máximamente positivo o negativo en otro) o bipolar (con valor máximamente positivo en un extremo y máximamente negativo en otro). Las escalas bipolares tienen la ventaja de permitir estudiar en una misma tarea tanto la causalidad generativa (relación causal directa) como la causalidad preventiva (relación causal inversa), pero resultan menos intuitivas y más difíciles de utilizar.

Las tareas también pueden clasificarse según la naturaleza de la clave utilizada. La clave puede ser un evento ambiental bajo el control del experimentador o bien puede ser una respuesta del sujeto. En el primer caso tendríamos una situación análoga a un procedimiento de condicionamiento clásico, mientras que en el segundo tendremos un análogo de condicionamiento operante. No se debe olvidar, en cualquier caso, que en las tareas de juicios de causalidad/contingencia la variable comportamental que se registra es por definición el juicio verbal-numérico de la persona sobre la relación estimada entre clave y resultado¹.

¹ Normalmente la persona debe emitir sus juicios cada cierto número de ensayos, lo cual permite representar curvas de aprendizaje similares a las observadas en los estudios de condicionamiento animal. Cuando la presentación se hace en tablas resumen, toda la información relevante sobre la concurrencia de clave y resultado se presenta de forma simultánea en una tabla de doble entrada en la que una entrada representa la

La estructura de las **tareas predictivas** suele ser similar a la de juicios de causalidad/contingencia. En el formato habitual, se presentan una serie de ensayos en que el experimentador controla la covariación entre clave y resultado. La respuesta del sujeto se intercala entre los dos eventos: normalmente se le pide que prediga (y/o indique la probabilidad de) el resultado (respuesta dicotómica). Estas tareas son conceptual y normativamente distintas de las primeras. Las tareas predictivas no permiten discriminar el aprendizaje causal del simple aprendizaje predictivo (si se asume que se trata de tipos distintos de aprendizaje). Además, aunque hay discrepancias sobre cuál es el referente normativo de los juicios causales (Anderson y Sheu, 1995; Cheng y Novick, 1992, 1995; Shanks, 1996), se acepta de forma general que éste depende tanto de la probabilidad de aparición del resultado en presencia de la clave $-P(R/C)$ - como de la probabilidad de aparición del resultado en ausencia de la clave $-P(R/noC)$ -, mientras que el referente normativo de los juicios predictivos es únicamente la probabilidad de aparición del resultado en presencia de la clave ($P[R/C]$) para los ensayos en los que se presenta la clave, o la probabilidad de aparición del resultado en ausencia de la clave ($P[R/no C]$) para los ensayos en los que no se presenta la clave. A pesar de estas notables diferencias, muchos autores interpretan las respuestas de la misma manera y las utilizan indistintamente. La idea subyacente es que, aún siendo conceptualmente distintas, ambas respuestas son indicadores de un mismo proceso psicológico, común al aprendizaje causal y al aprendizaje predictivo. De hecho, un buen número de manipulaciones experimentales afectan a los dos tipos de respuesta de la misma manera. Por ello, son pocos los modelos teóricos que hacen una distinción clara entre los juicios de causalidad/contingencia y los juicios predictivos/probabilísticos, aún cuando estos datos indican que podría ser necesario considerar los dos tipos de tarea de forma diferenciada. Por ejemplo, Wasserman, Elek, Chatlosh y Baker (1993) demostraron que los juicios emitidos por los participantes en una tarea de juicios de contingencia eran relativamente independientes de sus estimaciones de las probabilidades $P(R/C)$ y $P(R/no C)$. Además, Mandel y Lehman (1998) han descrito sesgos de estimación que afectan significativamente a los juicios de causalidad y contingencia, pero no a las estimaciones de probabilidad. Por último, Waldmann y Holyoak (1997) resaltan la importancia de poner a prueba las teorías de aprendizaje causal con tareas que hagan referencia explícita a la naturaleza causal de la relación entre la clave y el resultado.

Se ha propuesto que las respuestas emitidas en tareas de **condicionamiento instrumental** podrían ser una alternativa a las respuestas de estimación o de juicio. Mientras la aparición de una respuesta clásicamente condicionada sólo requiere conocimiento acerca de la concurrencia de dos

presencia/ausencia de la clave y la otra la presencia/ausencia del resultado. En las casillas interiores se representa la frecuencia de presentación de cada uno de los tipos de ensayos (para una discusión sobre las implicaciones del uso de presentación en tablas o serial ver Shanks, 1991).

estímulos, instaurar una respuesta instrumental requiere que se capte el nexo causal que existe entre una respuesta emitida por el sujeto y la consecuencia que se deriva de ésta. Según Shanks (1993) las principales manipulaciones experimentales que afectan a la conducta operante en animales y humanos tienen un efecto similar sobre los juicios causales. Así ocurriría con condiciones como la contigüidad, la contingencia, el número de ensayos de aprendizaje, el valor del resultado o la interacción entre claves. Sin embargo, es difícil establecer una valoración porque los estudios de aprendizaje causal que utilizan la tasa de respuesta operante como variable dependiente en los últimos años han sido muy escasos (Chatlosh, Neunaber y Wasserman, 1985; ver también Wasserman, 1993).

2.2. Efecto de la fuerza de la relación entre clave y resultado: Ajuste de los juicios a ΔP .

Aunque se han propuesto distintos indicadores de contingencia, el más popular en el campo del aprendizaje animal y humano es la regla P , que mide la dependencia unidireccional del resultado respecto de una clave. P se define como la probabilidad de ocurrencia del resultado en presencia de una clave menos la probabilidad de ocurrencia de ese resultado en ausencia de la clave.

$$(1) P = P(R/C) - P(R/\text{no}C)$$

donde $P(R/C)$ y $P(R/\text{no}C)$ se calculan a partir de las frecuencias de cada uno de los tipos de ensayos (a, b, c y d) que resultan de cruzar los valores de las variables de entrada y salida (ver tabla 1), de tal manera que

$$P = a/(a+b) - c/(c+d)$$

Tomando como base los resultados de dos estudios realizados en los años 60 (Jenkins y Ward, 1965; Smedslund, 1963) se creyó inicialmente que, en el mejor de los casos, los juicios de contingencia sólo están marginalmente influidos por la contingencia programada entre la clave y el resultado. Sin embargo, actualmente está bien establecido que los juicios de contingencia/causalidad humanos son, si se mantienen constantes otros factores, altamente sensibles a la contingencia establecida por el experimentador. Este hecho fue inicialmente corroborado por Alloy y Abramson (1979) en un estudio en el que las personas debían estimar la relación entre una respuesta discreta y el encendido de una luz. Sus resultados mostraron que, aunque aparecían ciertos sesgos sistemáticos dependientes del incentivo y del estado de ánimo, la manipulación de la contingencia en un rango de 0.00 a 0.75 afectó significativamente a las estimaciones realizadas por las personas. En un experimento de Chatlosh et. al. (1985) en el que los participantes debían juzgar en qué medida la respuesta de pulsar una tecla

telegráfica era capaz de producir o prevenir un *flash* luminoso, el 98'11% de la varianza de los juicios venía explicada por la manipulación de la contingencia a nueve niveles. En un estudio posterior de similares características, Wasserman (1990b) encontró que la manipulación de la contingencia entre una respuesta de pulsación (*tapping*) y el encendido de una luz afectaba tanto a la tasa de respuesta como a las estimaciones de la contingencia respuesta-resultado. Este último efecto también se pone de manifiesto en situaciones en las que la clave es un estímulo ambiental y no una respuesta (para un ejemplo reciente ver Lober y Shanks, en prensa).

La sensibilidad de los juicios a la manipulación de la contingencia medida mediante P se ha interpretado (1) como una prueba de la adecuación racional o normatividad del aprendizaje humano (Shanks, 1995, 1996) y (2) como evidencia en favor de que P es efectivamente el algoritmo utilizado por las personas en el cómputo de la contingencia a través de una serie de ensayos. La primera interpretación de P se denomina *normativa, computacional o funcional*, e implica, por un lado que P es la norma racional contra la que contrastar el comportamiento de los sujetos y, por otro, que la respuesta (instrumental o de juicio) tiende a ajustarse asintóticamente a dicha norma. La segunda interpretación se conoce como *algorítmica*, y supone de facto un modelo teórico de funcionamiento mental.

La regla P como modelo algorítmico se inserta dentro de los llamados **modelos de reglas** o **modelos estadísticos**. Estos modelos suponen que las personas, y probablemente también los animales, se comportan como *científicos intuitivos* (Kelley, 1973) y computan la relación entre clave y resultado a partir de las frecuencias relativas de los cuatro tipos de ensayos (a, b c ó d) de la tabla de contingencia (ver tabla 1).

Sin embargo, la regla P hace predicciones cuantitativas precisas sobre la magnitud de los juicios que en ciertas ocasiones no se ajustan a los datos experimentales recogidos en la literatura. Ciertas desviaciones pueden explicarse por el hecho de que la magnitud de los juicios de causalidad/contingencia sean una función no lineal del resultado del proceso o procesos psicológicos subyacentes. Ciertamente, se sabe que la función lineal no es la que mejor describe la relación psicofísica entre la contingencia P y el juicio (Ortega, 1991). De esta forma podría darse cuenta de las infra o sobre-estimaciones situacionalmente regulares y no interactivas que son frecuentes en la estimación de la contingencia.

Existen sin embargo desviaciones bien documentadas que no pueden explicarse basándose en la relación psicofísica contingencia-juicio y que, por tanto, suponen un verdadero problema para la regla P como modelo explicativo del aprendizaje. Un fenómeno ampliamente corroborado es el **sesgo de densidad**, según el cual, para un nivel fijo de contingencia, las estimaciones de las personas son una función directa de la probabilidad de ocurrencia del resultado - $P(R)$ -. (Allan y Jenkins, 1980, 1983; Alloy y Abramson, 1979; Baker, Berbrier y Vallee-Tourangeau, 1989; Chatlosh et al, 1985; Dickinson, Shanks y Evenden, 1984; Shanks, 1985a; Shanks y Dickinson, 1991). En un intento de resolver este problema, se han postulado

diversas modificaciones de la regla P . Wasserman, Elek, Chatlosh y Baker (1993) intentaron comprobar si las desviaciones de la regla P estaban fundamentadas en una estimación incorrecta de las probabilidades necesarias para su cómputo - $P(R/C)$ y $P(R/\text{no } C)$ -. Sin embargo, encontraron que la correlación con los juicios subjetivos para P calculada sobre la base de las estimaciones de probabilidad individuales era incluso peor que para P calculada sobre las contingencias objetivas.

Otras modificaciones de la regla P , por ejemplo P ponderada, (Heit, 1998) se basan en que no concedemos el mismo peso a cada uno de los cuatro tipos de ensayo (Kao y Wassermann, 1993; Levin, Wasserman y Kao, 1993; Maldonado, Catena, Cándido y García, 1999). Este dato se ha obtenido tanto analizando juicios sobre la importancia de cada una de las casillas de la tabla de contingencia, como analizando los juicios de contingencia directos en problemas de estimación de contingencia que difieren únicamente en una casilla. La estrategia de investigación ha sido asignar pesos a los cuatro tipos de ensayo en función de la importancia concedida a cada uno de ellos.

Otros autores han propuesto modelos de reglas diferentes a P , que no son normativamente correctos. La regla D , por ejemplo, se define como la diferencia entre las diagonales de la tabla de contingencia. Formalmente:

$$(2) \quad D = [a+d] - [c+b] / N$$

donde a , b , c y d son las frecuencias relativas de cada uno de los tipos de ensayo de la tabla de contingencia y N es el total de ensayos ($N = a + b + c + d$). Así por ejemplo, para variables dicotómicas asimétricas D se define como la diferencia entre la proporción de ensayos positivos -que tienden a confirmar la relación directa- y negativos -que tienden a confirmar la relación inversa - (Allan y Jenkins, 1983).

Aplicando la misma lógica de asignación de pesos a los tipos de ensayo que se utilizan para calcular P ponderada, se puede obtener también una versión de D semiponderada.

$$(3) \quad D = (w_a a - w_b b - w_c c + w_d d) / N$$

donde a , b , c y d son las frecuencias absolutas de cada uno de los tipos de ensayo, w_a , w_b , w_c , w_d son los pesos asignados a cada uno de ellos y N el número total de ensayos presentados a la persona (Catena, Maldonado y Cándido, 1998; Kao y Wasserman, 1993; Maldonado et al, 1999).

Otros estadísticos propuestos como reglas de ajuste para los juicios de contingencia/causalidad en la línea normativa de P , son los coeficientes de correlación Φ , r , ó la pendiente de la recta de regresión de la variable de salida (presencia/ausencia del resultado) sobre la variable de entrada (presencia/ausencia de la clave), que estima la magnitud del cambio producido

en la variable de salida por un cambio en la de entrada¹ (Ortega, 1991; Shanks, 1996).

Prácticamente todas las reglas propuestas presentan altas correlaciones entre sí y con los juicios emitidos por las personas (Mckenzie, 1995). Conocidos los problemas que presentaba P como regla descriptiva de los juicios de contingencia humanos, la estrategia de investigación dominante durante la década de los 80 fue buscar la regla que mejor se ajustase a dichos juicios. En la actualidad tiende a aceptarse que las reglas que mejor se ajustan a los juicios de contingencia humanos en niveles de aprendizaje asintóticos y con una sola variable de entrada asimétrica son P (Shanks, 1996; Wasserman, 1990b; Wasserman et. al, 1983) y D (Allan y Jenkins, 1983), o sus versiones ponderadas (Maldonado et al, 1999).

Hay, sin embargo, otro conjunto de efectos que resulta más problemático para los modelos de reglas. Según todos los modelos teóricos citados hasta ahora, la contingencia, sea cual sea la regla que se aplique y si no se tienen en cuenta supuestos adicionales, se calcula a partir una serie de ensayos completa. Por tanto, si se mantiene constante la frecuencia absoluta de cada uno de los tipos de ensayos en la serie, la magnitud del juicio final debería ser siempre la misma, independientemente del orden en que se presenten tales ensayos. Los *efectos de orden de los ensayos* se caracterizan por la constatación de una influencia significativa del orden en que se presentan los ensayos de una serie sobre la magnitud del juicio finalmente emitido. Yates y Curley (1986) realizaron un estudio en el que los participantes debían estimar la relación entre el color de una planta y su lugar de origen. Las respuestas de los sujetos mostraron una mayor influencia de la información presentada en el primer bloque, pero sólo cuando las personas no eran informadas de que al final de la serie deberían emitir un juicio sobre la información presentada. Un efecto similar ha sido descrito por De Jong, Merckelbach y Arntz (1990). Por el contrario, Wasserman, Kao, van Hamme, Masayoshi & Young (1996) no encontraron efecto alguno ni de primacía ni de recencia en un experimento en el que había que estimar en qué medida un fertilizante facilitaba o no la consecuencia de que una planta floreciera. Utilizando claves compuestas en lugar de claves simples, esto es, en tareas en las que se presenta más de una variable de entrada, se ha podido observar que la manipulación del orden de los ensayos afecta también significativamente a la magnitud del juicio final (Chapman, 1991; López, Shanks, Almaraz y Fernández, 1998).

¹ Todas las reglas citadas hasta el momento se denominan *reglas complejas* porque combinan de una u otra manera todas los tipos de información de la tabla de contingencia. Existen, no obstante una serie de reglas, denominadas *estrategias simples*, que se computan selectivamente a partir de las frecuencias relativas de algunos de los tipos de ensayos que se pueden presentar. Por ejemplo, en la regla F_{a-c} se computa la diferencia entre la proporción de ensayos tipo a y la proporción de ensayos tipo c. En F_{a-b} , similar a la anterior, se calcula la diferencia entre la proporción de ensayos tipo a y la proporción de ensayos tipo b.

2.3. Efecto de la complejidad de la tarea: Interacción de claves.

La complejidad de la tarea puede incrementarse o disminuirse de diversas formas. Es posible, por ejemplo, manipular el grado de experiencia que el sujeto tiene con la tarea, o la cantidad de información que se tiene que integrar entre dos respuestas consecutivas. Sin embargo, las manipulaciones más frecuentes y de las cuales se derivan mayores implicaciones teóricas son las realizadas sobre el número de claves o resultados. Especialmente interesantes son las tareas denominadas de **interacción de claves**, en las cuales hay que relacionar más de una clave con un resultado común. Dichas tareas permiten diseñar experimentos de **interacción de claves** análogos a los que aparecen en la literatura sobre condicionamiento animal, como los diseños de bloqueo, validez relativa, ensombrecimiento o inhibición condicionada. Los efectos de interacción entre claves han supuesto uno de los impulsores de cambio más importantes en las teorías del aprendizaje animal y han adquirido también una importancia notable en el campo del aprendizaje causal humano, como fenómenos que permiten discriminar entre modelos teóricos. Dentro de estos fenómenos se incluyen la **inhibición condicionada**, la **validez relativa**, el **bloqueo** y el efecto de **señalización**.

En un diseño estándar de inhibición condicionada (Pavlov, 1927; Rescorla, 1969) un grupo de animales es expuesto a una relación consistente entre un EC (A) y un EI (+). Sin embargo, cuando ese EC se presenta en un compuesto con un segundo EC (B) la consecuencia es la no aparición del EI (-). El grupo de control es expuesto únicamente al segundo EC (B) seguido de la ausencia del EI.

El resultado habitual de estos experimentos de condicionamiento clásico es que la clave B se convierte en un predictor negativo del EI en el grupo de inhibición condicionada, pero no en el grupo de control. Trasladado al terreno del aprendizaje causal, el efecto de inhibición condicionada supondría que para establecer una clave como causa preventiva o como un predictor negativo del resultado es necesario que esta vaya acompañada de una causa generativa o un predictor positivo. Esto es, los juicios a la clave negativamente relacionada con el resultado (B) deben ser menores en la condición de inhibición condicionada que en la condición de control.

Chapman y Robbins (1990, exp. 2) llevaron a cabo un experimento en el que había que establecer relaciones entre el cambio de agentes bursátiles (llamados genéricamente P, X, I y N) como predictores de un criterio consistente en los incrementos en el índice general de la bolsa. La tarea constaba de dos fases. En la primera, el compuesto XP iba seguido consistentemente de una elevación en el precio general del mercado, mientras que X iba seguido de la ausencia dicho resultado. En esta primera fase, P se establecía como un predictor positivo del resultado. En una segunda fase, los dos tipos de ensayos presentados en la primera continuaron apareciendo. Además, se añadieron dos nuevos tipos de ensayo. En uno de ellos un compuesto XPI iba seguido de la ausencia del resultado. Un último compuesto NX también iba seguido de la ausencia del resultado. Es de esperar que, al ir I acompañado por un predictor positivo (en el compuesto

XPI) esta clave se convierta en un predictor negativo. La clave N, por el contrario, va acompañada por una clave no predictiva (en el compuesto NX). El *producto* N y el *producto* I, aún manteniendo la misma contingencia objetiva (P) con el resultado (la elevación en el precio general del mercado) fueron evaluados de forma muy distinta por las personas. En concreto, los juicios sobre la relación entre la clave I y el resultado fueron significativamente menores (más negativos) que los juicios sobre la relación entre la clave N y el resultado.

Tabla 2. Diseño de inhibición condicionada intra-sujetos (Chapman y Robbins, 1990. Exp 2)

Fase 1		Fase 2				Test
XP+	X-	XP+	X-	XPI-	XN-	N > I

En este experimento, los juicios para la clave N también fueron negativos. La interacción entre claves se demuestra porque el simple hecho de ir acompañada de una clave predictiva lleva a que la clave objetivo (I) reciba estimaciones aún más negativas que la clave de control (N). Sin embargo, en un experimento realizado en nuestro laboratorio (Perales, Catena y Maldonado, remitido) no se consiguió replicar este efecto (ver también Williams et al.). En nuestro trabajo, los juicios fueron sensibles a una contingencia normativa negativa (-0.5), independientemente de que la clave objetivo fuese acompañada o no de otra altamente predictiva.

La validez relativa también es un fenómeno que ha sido descrito previamente en la literatura de condicionamiento animal (Wagner, Logan, Haberlandt y Price, 1968; Wasserman, 1974; ver Wagner, 1969c para una revisión). Ya en el contexto de juicios, Wasserman (1990a) realizó un experimento en el que las personas debían estimar la relación existente entre tres comidas (genéricamente A, B y X) y una reacción alérgica. Las claves (comidas) eran presentadas siempre en forma de compuesto (AX y BX) y las frecuencias de aparición/no aparición del resultado de reacción alérgica en presencia y en ausencia de cada uno de los compuestos se presentaba en forma de tablas de contingencia 2x2. A cada una de las personas se le presentaba una serie de tablas tras cada una de las cuales debían emitir un juicio independiente sobre la efectividad causal de A, de B y de X. En todas las tablas la contingencia para la clave X se mantuvo constante, manipulándose a cinco niveles la diferencia entre la contingencia para el compuesto AX y la contingencia para el compuesto BX (0, 0.25, 0.50, 0.75 y 1.00). Dicha manipulación permite ir incrementando la capacidad predictiva de la clave A (y disminuyendo simultáneamente la de la clave B) sin afectar a la contingencia entre X y la reacción alérgica. Los resultados muestran que las estimaciones de efectividad causal para la clave X son tanto menores cuanto mayor la

capacidad predictiva de A. Esta dependencia de las estimaciones sobre la relación clave-consecuencia de la capacidad predictiva de otras claves presentes simultáneamente en el medio es lo que se conoce como efecto de **validez relativa**.

Otro procedimiento que muestra interacción es el de bloqueo, descrito también en su origen en el campo del aprendizaje animal (Kamin, 1968). El procedimiento estándar consta de dos fases. En la primera, el grupo experimental es expuesto a una relación contingente entre un estímulo (A) y un EI (+). Una vez establecido este aprendizaje el sujeto es expuesto en una segunda fase a una relación contingente de un compuesto formado por el estímulo presentado en primera fase y un nuevo estímulo (B) con el EI. Junto a este grupo experimental pueden utilizarse diversos controles. Por ejemplo, uno recibiría en primera fase el mismo número de presentaciones del estímulo A y el EI, pero sin que haya correlación alguna entre ellos ($P=0$). En la segunda fase los sujetos recibirían exactamente el mismo tratamiento que los del grupo experimental¹.

El resultado habitual es que la exposición a la relación contingente entre el estímulo A y el EI impide o disminuye (bloquea) el aprendizaje sobre el estímulo añadido en la segunda fase, lo cual se demuestra porque la RC eliciteda por B es significativamente menos intensa en el grupo experimental que en el de control.

La réplica del efecto de bloqueo en humanos utilizando este paradigma de condicionamiento clásico ha resultado más problemática. Por ejemplo, Lovibond, Siddle y Bond (1988) fracasaron al intentar demostrar la aparición de bloqueo en una tarea de condicionamiento electrodermal.

En tareas de juicios de contingencia/causalidad el efecto de bloqueo se constata cuando la preexposición a una relación contingente entre una clave (C1) y un resultado (R) bloquea en una segunda fase el aprendizaje sobre una segunda clave (C2) que se presenta en compuesto con la primera. El bloqueo se demostraría por una disminución de la magnitud de los juicios sobre C2 en relación con los de un grupo adecuado de control. Shanks (1985b, exp.1) realizó un experimento en el que demostraba que la observación de la relación entre una causa potencial (clave 1) y una consecuencia (resultado) de hecho bloqueaba o reducía el aprendizaje sobre una segunda causa potencial (clave 2) presentada posteriormente en compuesto con la primera. Sorprendentemente, en un grupo en el que el orden de presentación de las fases se invertía (AB+ A+) también se observaba una disminución en los juicios de efectividad finales para B. En este caso el aprendizaje en la segunda fase de una relación contingente entre una clave (A) y un resultado interfiere a posteriori con un aprendizaje previamente establecido sobre una clave B que se había presentado en compuesto con la clave A. Este efecto se conoce como

¹ Este grupo de control, aún siendo el que mejor controla todos los factores que podrían explicar las diferencias entre los grupos y, por tanto, el más adecuado, no es el control utilizado originalmente por Kamin, sino que fue introducido posteriormente por Rescorla (1971).

bloqueo hacia atrás y, por contraposición, al bloqueo estándar se le ha denominado bloqueo hacia adelante. En este experimento los grupos de control utilizados fueron dos grupos de *ensombrecimiento*, que reciben solamente la fase de aprendizaje con el compuesto. Cada grupo de bloqueo es contrastado con un grupo de ensombrecimiento en el que la fase de aprendizaje con el estímulo simple (A) es sustituida por un intervalo de tiempo en blanco.

Tabla 3. Diseño de bloqueo entre grupos (Shanks, 1985).

Grupo	Fase 1	Fase 2	Test
Bloqueo estándar	A -> R	AB -> R	B?
Bloq. hacia atrás	AB -> R	A -> R	B?
Ensombrecimiento 1	-	AB -> R	B?
Ensombrecimiento 2	AB -> R	-	B?

Chapman y Robbins (1990, exp. 1) demostraron un efecto de la misma naturaleza con un diseño notablemente diferente. Utilizaron la tarea antes descrita en la que las personas debían estimar la relación entre la subida del precio de diversos *productos* (genéricamente P, N, B y C) y una subida en el precio general del mercado. La tarea constaba de dos fases. En la primera, siempre que P subía su precio, también lo hacía el precio global, mientras que la subida de N no iba seguida de dicho resultado. En la segunda fase, cuando los *productos* P y B subían (compuesto PB) también subía el precio global, al igual que ocurría cuando los *productos* N y C subían su precio simultáneamente (compuesto NC). En otras palabras, en la segunda fase la clave B iba acompañada de una clave (P) que previamente había sido establecida como una clave predictiva. Por el contrario, la clave C va acompañada de una clave que ha sido previamente establecida como no predictiva (N). En este caso las condiciones de bloqueo y control se presentan intra-sujetos. Así, la condición de control es equivalente a una condición de *supercondicionamiento*, en la que en la primera fase de aprendizaje la contingencia programada entre clave y resultado (en este caso entre la subida de N y la subida global del precio del mercado) no es nula sino negativa.

El efecto de interacción se constata porque los juicios respecto de la clave bloqueada (B) son de una magnitud significativamente menor que los juicios relativos a la clave no bloqueada (C). En todos los casos citados los sujetos han experimentado los dos estímulos que interaccionan en algún momento de la tarea. Sin embargo, Matute y Piñeno (1998) han encontrado interacción entre estímulos entrenados individualmente. Simplemente, el hecho

de que en una primera fase el sujeto reciba A+, deteriora la adquisición de la relación B+. Si tenemos en cuenta que, además, la obtención de la interacción entre claves puede depender de las instrucciones (Matute, Arcediano y Miller, 1996) o de si la forma de manipulación permite al sujeto realizar una comparación entre las distintas claves (Ramos y Catena, remitido), entonces quizá quepa poner en tela de juicio la generalidad e importancia de este fenómeno en humanos.

Tabla 4. Diseño de bloqueo intra-sujetos (Chapman y Robbins, 1990. Exp 1).

Fase 1		Fase 2		Test
P+	N-	PB+	NC+	B < C

Price y Yates (1993) han encontrado un efecto de interacción muy similar al bloqueo utilizando un diseño de una sola fase, denominado *bloqueo simultáneo*. En este diseño, los ensayos pertenecientes a la primera y segunda fase del diseño estándar se entremezclan dentro de una sola fase. Al igual que con los otros procedimientos, la clave bloqueada recibe juicios de menor magnitud que la clave no bloqueada. Este diseño es conceptualmente análogo al efecto de validez relativa, pero se diferencia de Wasserman (1990b) por la forma de presentación de los estímulos. En este caso la información se presentaba ensayo-a-ensayo y no resumida en tablas.

Por último, en un diseño de **señalización** (Shanks, 1989) la persona debe estimar la relación entre una clave y un resultado para los que se programa una contingencia nula. En el grupo 'señalado', las apariciones no contingentes del resultado, esto es, aquellas en las que éste no va precedido de la clave, van acompañadas de una clave alternativa de cuya aparición no se había informado a las personas. En esta condición, los juicios sobre la relación entre la clave objetivo y el resultado son significativamente mayores que en un grupo control no señalado.

Este conjunto de efectos se caracteriza porque la inclusión de una segunda clave produce una desviación de los juicios de causalidad/contingencia respecto de la contingencia programada, aún cuando esta se mantiene constante a través de las distintas condiciones experimentales. Ninguno de los modelos que hemos considerado hasta ahora pueden dar cuenta de este conjunto de datos. Los modelos tradicionales de reglas se consideran hoy funciones que describen más o menos adecuadamente la ejecución en las tareas de juicios de causalidad/contingencia. Los modelos de reglas actuales son la evolución lógica de los modelos propuestos durante los años 70 y 80, y se han refinado y complicado para dar cuenta de los datos provenientes de los experimentos de interacción entre claves. Se siguen manteniendo, sin embargo, los presupuestos básicos. A saber, que los sujetos operan con los distintos tipos de ensayos de la tabla de contingencia como si

fueran estadísticos intuitivos, aplicando una regla matemática (no necesariamente de forma consciente) que les permita computar la relación entre claves y resultados.

Un segundo grupo de teorías surge de la similitud formal entre algunos de los efectos descritos en la literatura de aprendizaje animal y los descritos en tareas de aprendizaje causal. Este paralelismo ha llevado a un buen número de autores a defender la idea de que los modelos asociativos nacidos en el estudio del condicionamiento animal permiten explicar adecuadamente la ejecución de los seres humanos en tareas de juicios de causalidad/contingencia. Tanto los seres humanos como el resto de los animales necesitan captar las relaciones causales presentes en el medio. En este sentido, los problemas a los que unos y otros deben enfrentarse son similares, de ahí que los mecanismos cognitivos sean esencialmente similares.

La dialéctica entre los modelos asociativos y los modelos de reglas ha centrado la mayor parte del trabajo teórico y experimental desarrollado en el campo del aprendizaje causal en las últimas dos décadas. A continuación revisaremos las aportaciones más recientes y los problemas que se plantean.

3. Modelos actuales de aprendizaje causal.

Las teorías propuestas hasta la fecha intentan integrar el conjunto de datos experimentales enumerados en el apartado anterior. Ya hemos visto que algunos de estos modelos parten de la idea de que el sujeto aplica algún tipo de regla o estadístico de forma intuitiva, a partir de la frecuencia de los distintos tipos de ensayos presentados durante la tarea. Estos modelos se han denominado **modelos basados en reglas**. Una segunda aproximación, no necesariamente incompatible con la anterior, mantiene que el aprendizaje de las relaciones entre causas y consecuencias en ambientes naturales se produce automáticamente, de forma similar a como se establece el aprendizaje por condicionamiento. Partiendo de esta idea se han sucedido varios intentos de aplicar los modelos provenientes del aprendizaje animal al aprendizaje causal. A continuación repasaremos las aportaciones más recientes provenientes de una y otra línea.

3.1. ΔP condicional y la Teoría de Potencia Causal

Tradicionalmente se ha sostenido que P es el referente normativo adecuado frente al cual contrastar la ejecución de los sujetos en tareas de juicios de causalidad/contingencia. Esto se postula igualmente en situaciones en las que sólo existe una clave predictiva y un resultado, o una causa potencial y una consecuencia. Sin embargo, todos sabemos por experiencia propia que la textura causal del medio es compleja, y que es muy raro el caso en el que una clave o una posible causa aparece de manera aislada. Más bien, en ciertas ocasiones es necesaria la presencia simultánea de dos o más claves para que se produzca el resultado, o en otras, una determinada clave predice la presencia de un resultado si y sólo si una segunda clave está presente. En muchas de tales situaciones complejas, ajustar nuestro comportamiento al

estándar P no es lo más *racional*, en tanto que P no es la norma que mejor refleja las relaciones causales o predictivas que realmente se dan en el ambiente. El **modelo de contrastes probabilísticos** (Cheng y Novick, 1990; 1992) mantiene que las personas computan P para una clave potencialmente predictiva o una posible causa en un contexto (llamado *focal set*) en el que el resto de las claves o causas alternativas se mantienen constantes. Pongamos un ejemplo: ¿qué ocurriría en el grupo experimental del diseño de señalamiento que antes hemos comentado?. Computada sobre el total de los ensayos presentados, la contingencia P para la clave objetivo es igual a cero. Concretando, tanto $P(R/C)$ como $P(R/\text{no } C)$ son iguales a 0.5. Sin embargo, según el modelo de contrastes probabilísticos (MCP en adelante), P no se calcula sobre el total de ensayos, ya que en algunos de ellos (los ensayos señalados) aparece una clave alternativa mientras que en otros no (los ensayos no señalados). En este caso el sujeto tiene dos alternativas: computar P únicamente sobre los ensayos no señalados, en los que la clave alternativa está permanentemente ausente o bien sobre los ensayos señalados, en los que la clave alternativa está siempre presente. En el primer caso hay que eliminar del cómputo todos los ensayos tipo c (en los que no aparece la clave pero sí el resultado) y, por tanto, el cómputo quedaría de la siguiente manera:

$$(4) \quad P = P(R/C) - P(R/\text{no } C) = 0.5 - 0.0 = \mathbf{0.5}$$

En el segundo caso P no sería computable. Por tanto, según la teoría MCP los juicios se ajustarán al valor de P calculado en el único contexto en el que las claves alternativas se mantienen constantes; esto es, a través de los ensayos no señalados. Bajo estos presupuestos, P calculada sobre un *focal set*, también llamada ΔP **condicional** permite dar cuenta de la mayoría de los efectos de interacción entre claves antes descritos.

Cheng y Novick (1992) hacen hincapié en que es P condicional el referente normativo adecuado frente al que contrastar el comportamiento de los sujetos. Siendo X y Z dos claves predictivas o dos causas potenciales y R el resultado, pueden definirse distintos casos. Imaginemos que X y Z correlacionan perfectamente, esto es, que en todos aquellos ensayos en los que X aparece, también lo hace Z , y viceversa cuando está ausente. En estos casos no puede definirse un *focal set* para ninguna de las dos claves potenciales. Si Z se mantiene siempre presente e intentamos calcular P para X en este contexto, $P(R/\text{no } X)$ no es computable, ya que no hay ningún ensayo en el que X no aparezca. Si, por el contrario, mantenemos Z ausente, $P(R/X)$ no es computable, ya que no hay ningún ensayo en el que X aparezca. Estos casos se denominan casos de indeterminación causal y la teoría predice que los sujetos prefieren no emitir ningún juicio si se les permite no hacerlo. Se ha constatado asimismo que si damos a las personas la oportunidad de estimar en qué medida la información de la que dispone en una tarea de juicios de contingencia es adecuada para realizar un juicio causal, dichas estimaciones se reducen notablemente si las claves estaban perfectamente correlacionadas

(Perales y Catena, en preparación). En caso de que se les fuerce a emitir un juicio de causalidad/ contingencia, el valor de éstos tiende a ser medio/bajo, como ocurre en la mayoría de los diseños de bloqueo ante los predictores añadidos.

En aquellos casos en los que P para X puede calcularse manteniendo constante la presencia de Z , pero no para su ausencia (como en el caso del diseño del señalamiento o de bloqueo), los juicios deberían ajustarse al único valor de P calculable. Nos quedaría:

$$(5) \quad P_x = P(R/X \cdot Z) - P(R/\text{no}X \cdot Z)$$

Por último, existe otro tipo de casos en los que P_x puede calcularse manteniendo constante tanto la presencia como la ausencia de Z . Por ejemplo,

$$(6) \quad P(R/X \cdot Z) - P(R/\text{no}X \cdot Z) > 0 ; P(R/X \cdot \text{no}Z) - P(R/\text{no}X \cdot \text{no}Z) = 0$$

Cuando la contingencia para X es positiva en presencia de Z , pero resulta nula en ausencia de Z , se dice que Z es una condición «posibilitante» (*enabling condition*). Por ejemplo, la presencia de oxígeno (Z) es una condición necesaria para que una chispa (X) produzca fuego (Y). Hay que recalcar que las condiciones posibilitadoras $\neg Z$ pueden interpretarse como causas verdaderas si lo que se mantiene constante es la presencia de X , y viceversa. Por ejemplo, encender una chispa (Z) es una condición necesaria para que la entrada de oxígeno en una bureta (X) produzca fuego (R).

Según Cheng y Novick (1990; 1992; ver también Cheng, 1997) su modelo teórico no es algorítmico, sino computacional, en términos de la metateoría de Marr (1982). Esto quiere decir que en la asíntota, el comportamiento de los sujetos se ajustará a la norma racional definida por el MCP, pero en ningún momento se hace alusión a cuál sería el algoritmo mediante el que se hace realmente el cómputo de la contingencia. En este sentido, el modelo de contrastes no es incompatible con el otro tipo de modelos explicativos de los juicios. Como más adelante veremos, MCP y algunos modelos asociativos hacen predicciones similares en la asíntota en una variedad de condiciones (Shanks, 1996). Esto ha llevado a algunos autores a distinguir entre una teoría MCP *computacional* y una teoría MCP *estadística* (Lober y Shanks, remitido).

Además de los estudios de interacción entre claves, existe evidencia directa de que, al menos en determinadas circunstancias, los juicios de causalidad humanos se ajustan a este tipo de cómputos condicionales, es decir para una clave relativo al resto. En una tarea ensayo-a-ensayo en la que las personas debían estimar la influencia de dos tipos de fertilizantes sobre el florecimiento de una planta, Spellman (1996) demostró que los participantes ajustaban sus juicios causales sobre cada uno de los fertilizantes a la presencia o ausencia constante del otro fertilizante. En tareas de juicios de contingencia,

la evidencia resulta más escasa y controvertida. Así, Ramos y Catena (remitido) han demostrado que bajo ciertas circunstancias los juicios de contingencia son mejor predichos por P incondicional, es decir considerando cada clave de manera independiente.

La teoría MCP presenta además una serie de problemas. En las ocasiones en las que existen varios *focal set* viables, no es posible definir a priori cuál de ellos utilizará el sujeto para realizar el cómputo de la contingencia. En estos casos, la corroboración de la teoría puede hacerse únicamente a posteriori. Por otra parte, la teoría MCP pretende ser un modelo de aprendizaje causal propiamente dicho. Sin embargo, tal y como está planteada, puede aplicarse tanto al aprendizaje causal como al aprendizaje predictivo. En otras palabras, la teoría de Cheng y Novick no hace explícitos (a) los principios de elección de un determinado focal set frente a otros y (b) las condiciones bajo las cuales la contingencia condicionada es utilizada por el sujeto como un verdadero índice de influencia causal.

Una variante del modelo de contrastes, la **teoría de potencia causal** (*Power PC Theory*, Cheng, 1997) pretende solventar estos problemas. En primer lugar, establece una distinción clara entre *contingencia percibida*, que sería la contingencia condicional estimada por el sujeto entre una clave y un resultado y la *potencia causal percibida (causal power)*, que sería la estimación que el sujeto realiza de la influencia causal unidireccional que la clave (causa potencial) ejerce sobre el resultado. La teoría de Potencia Causal retoma, por tanto, una cuestión heredera del estudio filosófico de la causalidad, esto es si la diferencia entre el aprendizaje de las relaciones predictivas o de contingencia y el aprendizaje de relaciones causales es una distinción meramente conceptual o es una distinción con contenido psicológico.

Para la teoría de Potencia causal, siendo X y Z dos causas potenciales alternativas, y R el resultado o consecuencia, la relación entre potencia causal para X (una causa potencialmente generativa) y su P condicional viene dada por la siguiente ecuación:

$$(7) P_x = (P_x - [P(z/x) - P(z/no x)] * P_z) / (1 - P(z/x) * P_z)$$

donde P_x es la potencia causal para la clave x , P_x es la contingencia entre x y el resultado, $[P(z/x) - P(z/no x)]$ es la contingencia entre las dos claves, P_z es la potencia causal para z y $P(z/x)$ es la probabilidad de aparición de la clave z dada la aparición de la clave x (para una justificación racional de esta ecuación ver Cheng, 1997, p. 374).

P_x y P_z son parámetros internos, desconocidos para el experimentador. Si tanto z como x son claves nuevas, esto es, si P_x y P_z no tienen ningún valor previo para el sujeto, la potencia causal para x no es calculable a partir de P_x , ya que P_z es desconocido. La única excepción a esta afirmación se produce cuando las dos claves no covarían entre sí. En ese caso, la ecuación (7) puede reescribirse como sigue:

$$(8) P_x = P_x / (1 - P(R/\text{no } x))$$

esto es, P es un estimador conservador de la potencia causal; tanto más conservador cuanto mayor es la probabilidad de ocurrencia del resultado en ausencia de la clave objetivo (x).

Para causas potencialmente preventivas, y manteniendo la condición de que z e x no covaríen, puede deducirse una ecuación similar:

$$(9) P_x = - P_x / (P(R/\text{no } x))$$

En este caso, P_x se refiere a la potencia causal preventiva de la clave x .

Uno de los objetivos de esta teoría es solucionar el problema de la elección del *focal set*. Para causas generativas se elegirán aquellos focal sets en los que las causas alternativas se mantengan ausentes de forma constante (y así que la covariación entre claves sea nula) y $P(R/\text{no } x)$ se aleje de uno. En esas condiciones P_x es calculable a partir de P_x . Por el contrario, para causas preventivas se tenderá a elegir un focal set en el que una causa alternativa esté continuamente presente y $P(R/\text{no } x)$ se aleje de cero. En esas condiciones, P_x es un estimador más o menos conservador de P_x .

Como consecuencia, la teoría de potencia causal mantiene que los juicios de causalidad son un índice de P_x , esto es, de la potencia causal de la clave objetivo. Ello quiere decir que en tareas de juicios de causalidad, los juicios se verán influidos no sólo por la contingencia P_x condicional, sino también por la tasa base de ocurrencia del resultado o consecuencia ($P(R/\text{no } x)$). Dicha predicción ha sido contrastada y confirmada en varias ocasiones (i.e. Buehner y Cheng, 1997), aunque la interpretación del resultado no es siempre coincidente con la teoría de potencia causal (Lober y Shanks, en prensa; Vallée-Tourengeau et al, 1997).

Un problema grave al que se enfrenta esta teoría se refiere a las condiciones de aplicación de las ecuaciones para causas preventivas o generativas. El sujeto debe tener cierto conocimiento a priori que le permita (a) discriminar si la causa es potencialmente generativa o preventiva y (b) utilizar el algoritmo adecuado para el tipo de causa al que se enfrenta. Se deduce de aquí que cuando el sujeto se enfrenta a claves causales totalmente nuevas no puede saber a priori qué algoritmo aplicar. El problema se agrava si consideramos que dicho conocimiento también es necesario para elegir el *focal set* de referencia adecuado.

El mayor mérito de este modelo es el de distinguir explícitamente entre estimaciones de causalidad y estimaciones de contingencia. Esta distinción, como veremos, es ignorada (aunque no explícitamente negada) en las teorías de carácter asociativo. Una segunda aportación importante, e implícitamente asumida por la práctica totalidad de las aproximaciones teóricas, es que el cómputo de la contingencia es al menos una condición previa necesaria para la estimación de la causalidad. Los mecanismos que realizan dicho cómputo, ya

sea en tareas de aprendizaje causal, ya en tareas de aprendizaje meramente predictivo, son probablemente comunes.

3.2. Modelos asociativos

Los modelos asociativos suponen una aproximación teórica totalmente distinta al estudio del aprendizaje causal. Todos estos modelos comparten una serie de principios comunes:

Los mecanismos cognitivos que rigen el aprendizaje predictivo en humanos y animales son de la misma naturaleza.

Los mecanismos que subyacen al aprendizaje predictivo y al aprendizaje causal son, al menos parcialmente, comunes.

El aprendizaje se produce por la acumulación de fuerza asociativa, el fortalecimiento del lazo que une los nodos de la clave y el del resultado. La acumulación de fuerza asociativa se produce por la activación conjunta de las representaciones mentales de la clave (*input*) y el resultado (*output*).

La magnitud de la respuesta aprendida es una función de la activación del nodo o representación mental del resultado, y ésta, a su vez es una función de la activación del nodo o representación mental de la clave ponderada por la fuerza asociativa del lazo que los une.

La acumulación de fuerza asociativa sigue un patrón negativamente acelerado que termina por llegar a un valor asintótico, representando el aprendizaje máximo que se puede alcanzar para una clave y un resultado.

Los modelos asociativos nacieron dentro del campo del aprendizaje animal. El más conocido de los modelos asociativos es el formulado por Rescorla y Wagner (1972; aunque ver también Mackintosh, 1975; Pearce y Hall, 1980). El principio que subyace a este modelo es que la acumulación de fuerza asociativa en cada ensayo es una función de la discrepancia entre el valor predictivo de la clave (*input*), representado por la activación que ésta comunica al nodo de salida (*output*), y el valor de activación real del *output*. Esa discrepancia tiende a reducirse conforme avanza el aprendizaje, lo cual predice curvas de aprendizaje negativamente aceleradas. Además, se supone que la cantidad de fuerza asociativa que puede acumularse en los lazos que unen las representaciones de las diversas claves (*input*) que ocurren simultáneamente en el mismo contexto y la representación del resultado (*output*) es limitada, de lo cual se deriva que se establecerá un proceso de competición entre las distintas claves por la fuerza asociativa disponible.

El modelo de Rescorla y Wagner (R-W en adelante) se expresa matemáticamente como sigue:

$$(10) \quad V_a = (V_x - V_x)$$

donde ΔV_a es el incremento de fuerza asociativa en un ensayo entre una clave (a) y un resultado, α y β son parámetros de actualización que dependen, respectivamente, de la saliencia de la clave y el resultado, λ es el máximo de fuerza asociativa que la representación del resultado puede soportar (cuyo valor es 0 en los ensayos en los que el resultado no aparece) y V_x es la fuerza asociativa acumulada por todas las claves simultáneamente presentes en el ambiente. Esta regla es matemáticamente equivalente a la **regla LMS** (Least Mean Squares) o **regla delta**, utilizada como regla de actualización de pesos en muchos modelos conexionistas.

Este modelo se ideó expresamente para dar cuenta de los efectos de interacción entre claves en condicionamiento animal y, por tanto, explicar de forma adecuada los análogos de tales efectos en aprendizaje causal. El cómputo de la contingencia se supone un proceso totalmente dirigido por los datos. No se considera la posibilidad de que existan influencias *top-down* en el cómputo de la contingencia. De manera no explícita se asume que cualquier influencia de arriba abajo que pudiera darse no se produciría sobre el mecanismo de cómputo propiamente dicho, sino sobre la emisión de la respuesta.

Sin embargo, los modelos asociativos no están libres de problemas. El más obvio es que sólo pueden aplicarse a situaciones de aprendizaje ensayo-a-ensayo. En las tareas en las que la información sobre la contingencia clave-resultado se presenta en forma resumida (i.e. tablas) es necesario postular un mecanismo distinto, no asociativo. Este problema ha dado lugar a los llamados **modelos duales**, que asumen la puesta en marcha de distintos mecanismos de procesamiento en función de cuál sea el tipo de información sobre la que se trabaja (Shanks, 1991).

También hay efectos que suponen un problema para modelos como R-W, o al menos para su formulación original. Por ejemplo, Wassermann y sus colaboradores (Kao y Wasserman, 1993; Levin, Wasserman y Kao, 1993) han descubierto que las personas no conceden la misma importancia a los distintos tipos de ensayo de la tabla de contingencia. Esto hecho se ha comprobado (a) analizando los juicios subjetivos de los sujetos sobre la importancia relativa concedida a cada tipo de ensayo y (b) analizando los juicios de contingencia directos en distintas tareas que se diferenciaban en un solo ensayo. Ambos tipos de evidencia convergen en la idea de que se concede un mayor peso a los ensayos del tipo **a** que a los de tipo **b** y **c**, y más a éstos a los del tipo **d**. La ponderación diferencial de los ensayos forzó la modificación de los modelos de reglas y también ha dado lugar a una versión modificada del modelo R-W (Markman, 1989; Tassoni, 1995), en la que los parámetros α y β pueden tomar distintos valores (positivos o negativos) en función de que la clave aparezca o no. Esta modificación permite, además, postular que se producirá aprendizaje sobre una clave incluso cuando esa clave no se presenta, posibilidad que no contempla la versión original de R-W. Sin embargo, Maldonado et al (1999) han demostrado que los pesos de los ensayos pueden también cambiar a lo largo de la tarea de aprendizaje, efecto

que no es predicho por ninguno de los modelos asociativos planteados hasta la fecha.

Los modelos asociativos también tienen problemas para dar cuenta de los llamados **efectos del modo de respuesta**, bien conocidos en la literatura sobre actualización de creencias y formación de impresiones (para una revisión ver Hastie y Pennington 1995; Hogarth y Einhorn, 1992). En el tipo de tareas que a nosotros nos ocupa, el modo de respuesta se refiere básicamente a la frecuencia con la que el sujeto debe emitir sus juicios de causalidad o contingencia. Los modos de respuesta más utilizados son el tipo *global*, cuando se pide una única estimación al finalizar una serie de ensayos y el tipo *ensayo-a-ensayo* se debe realizar una estimación tras cada pieza de información. Ninguno de los modelos que hemos analizado hasta aquí prevé un efecto principal del modo de respuesta sobre la magnitud de los juicios emitidos (Arkes y Harkness, 1983), aunque a posteriori los modelos asociativos son suficientemente versátiles como para acomodar este tipo de efecto. Se podría argumentar, por ejemplo, que el modo de respuesta afecta a los valores de los parámetros de actualización de los pesos (α y β). Más problemático resulta un efecto descrito recientemente y que se ha llamado **efecto de la frecuencia del juicio** (Catena et al, 1998). Catena y cols (1998, exp. 1) realizaron un experimento en el que se manipuló la frecuencia con la que los sujetos debían emitir sus juicios, a dos niveles: alta frecuencia (después de cada ensayo) y baja frecuencia (cada 8 ensayos). Cada serie de ensayos estaba compuesta de 4 bloques de 8 ensayos, de tal manera que el número total de ensayos fue de 32. La contingencia era igual en cada uno de los bloques e igual a la contingencia global de la tarea (0.5). En cada serie de 32 ensayos se manipuló el tipo de ensayo final de cada uno de los bloques, de tal manera que un bloque finalizaba en un ensayo tipo a, otro en un ensayo tipo b, otro en un ensayo tipo c y otro en un ensayo tipo d. La secuencia de los tipos de ensayo al final de bloque se contrabalanceó a través de los sujetos. El efecto más interesante fue la interacción entre la frecuencia del juicio y el tipo de ensayo final de bloque. En el grupo que emitía sus juicios ensayo a ensayo, la magnitud de los juicios era dependiente del tipo de ensayo final de bloque. Más concretamente, la magnitud de los juicios emitidos tras un ensayo tipo a era algo mayor que la de los emitidos tras un tipo d y ésta algo mayor que la de los emitidos tras un ensayo tipo b o c. Tal dependencia no aparecía en absoluto en los sujetos que emitían sus juicios cada 8 ensayos (grupo de baja frecuencia). Este patrón de resultados tiende a atenuarse con frecuencias de juicio intermedias, hasta desaparecer totalmente con frecuencias de juicio menores a un juicio cada cuatro ensayos (exp. 2). Como veremos más adelante, sólo modelos más complejos que los vistos hasta ahora (tanto asociativos como estadísticos) pueden explicar este efecto.

Un último conjunto de datos que resulta problemático para la formulación original de los modelos asociativos proviene, sorprendentemente, de los estudios sobre interacción entre claves. Shanks (1985b) demostró que, si en un diseño de bloqueo estándar se invierten las fases del grupo experimental, el efecto de bloqueo también aparece (bloqueo hacia atrás).

En un grupo de bloqueo hacia atrás los sujetos son expuestos en una primera fase a la relación contingente entre un compuesto estimular (AB) y un resultado (R). En una segunda fase el sujeto es expuesto a una relación contingente entre uno de los estímulos del compuesto (B) y el mismo resultado (R). Sorprendentemente, esta segunda fase (en la que no se presenta el estímulo A) interfiere con lo aprendido sobre A en primera fase, produciendo una reducción de los juicios sobre esta clave. El bloqueo hacia atrás pertenece a un conjunto de efectos que se conoce como *efectos de reevaluación retrospectiva*, y que se caracteriza por la modificación de lo aprendido sobre una clave en virtud de una experiencia posterior de aprendizaje en la que no se presenta esa clave (Dickinson y Burke, 1996).

La estrategia que se ha seguido para adaptar el modelo R-W a este tipo de efectos ha sido similar a la seguida para dar cuenta de la ponderación diferencial de los ensayos. Básicamente, se permite que la regla produzca aprendizaje sobre una clave en una situación en la que esa clave no aparece. La condición fundamental que debe ocurrir para que eso suceda es que dicha clave sea esperada por el sujeto en ese contexto (Tassoni, 1995; Wasserman, 1994). Según esta variante asociativa, en la primera fase del bloqueo hacia atrás no sólo se asocian ambas claves con el resultado, sino que también se asocian ambas claves entre sí. En virtud de la asociación intra-compuesto entre claves, el sujeto espera en la segunda fase la aparición de la clave. La disonancia producida entre la expectativa de aparición de la clave y su no aparición real permite aprender sobre esa clave. El tipo de aprendizaje que se produce es inhibitorio, ya que para este caso se supone que α adopta un valor negativo. Este aprendizaje inhibitorio sería por tanto la causa de la interferencia que se produce en los experimentos de reevaluación retrospectiva.

3.3. Teoría del modelo causal.

Para el modelo R-W, y los modelos asociativos en general, las representaciones o nodos de input corresponden siempre a las claves, esto es, los estímulos que temporalmente aparecen en primer lugar en la tarea, mientras que los nodos de output representan a los resultados. Según este punto de vista, el papel causal de las claves y el de los resultados es indiferente. Tanto en el caso de que las claves sean causas potenciales y los resultados las consecuencias, como en el caso contrario -en el que las claves son las consecuencias aunque temporalmente sean anteriores- la dirección del lazo asociativo va siempre de clave a resultado y no al revés.

La teoría del modelo causal (Waldmann, 1996; Waldmann y Holyoak, 1992, 1997) mantiene que los juicios causales se ajustan a P condicional, calculada sobre las causas potenciales y las consecuencias, independientemente de la ordenación temporal de las mismas. En otras palabras, siendo X y Z dos causas potenciales y R una consecuencia, P condicional para X se calcularía de la siguiente manera en caso de que se estableciesen como focal set los ensayos en los que Z está presente:

$$(11) P_x = P(R/X \cdot Z) - P(R/\text{no}X \cdot Z)$$

o bien se calcularía de la siguiente manera en caso de que se estableciesen como focal set los ensayos en los que Z está ausente.

$$(12) P_x = P(R/X \cdot \text{no}Z) - P(R/\text{no}X \cdot \text{no}Z)$$

Recordemos que P condicionada predice efectos de interacción entre dos claves por la necesidad de condicionar una de ellas a la presencia o la ausencia de la otra. La teoría del modelo causal predice lo mismo en el caso de que haya dos causas potenciales y una consecuencia (**tareas predictivas**). Sin embargo, si las dos claves cumplen el papel de consecuencias y el resultado cumple el papel de posible causa común para ellas (en las llamadas **tareas diagnósticas**), la teoría predice la ausencia de interacción entre las claves. Dicho de otro modo, las causas compiten entre sí, pero no las consecuencias. Esta predicción se deriva del principio de la teoría que mantiene que las variables de entrada para el cómputo de la contingencia son siempre las causas, independientemente de la posición temporal que ocupen en la tarea.

Waldmann y Holyoak (1997) hacen hincapié en que la teoría del modelo causal es *explícitamente* una teoría de aprendizaje causal. Esto delimita de manera muy estricta el ámbito de tareas en las que la teoría puede ponerse a prueba. Concretamente, en la tarea debe quedar clara la naturaleza causal de la relación existente entre la clave o claves y el resultado o resultados. Además, el juicio registrado debe ser un juicio causal. Bajo estas condiciones se ha encontrado que, como predice la teoría, las causas compiten entre sí mientras que las consecuencias no lo hacen (Waldmann, 1996; Waldmann y Holyoak, 1992; aunque ver Shanks y López, 1997 y Waldmann y Holyoak, 1998).

La teoría del modelo causal se inserta en la tradición filosófica Kantiana en la medida en que supone que las personas contamos con alguna clase de conocimiento previo que permite discriminar qué clase de acontecimientos pueden ser causas y qué clase de acontecimientos pueden desempeñar el papel de consecuencias. Existen distintas aproximaciones sobre cuál podría ser la naturaleza de dicho conocimiento previo (ver Cheng y Lien, 1996). Sin embargo, la cuestión teórica de más importancia actualmente a este respecto es si el conocimiento previo influye de arriba abajo (top-down) sobre el cómputo mismo de las relaciones causales, tal y como mantiene la teoría del modelo causal, o si el conocimiento previo está implementado simplemente como un conjunto de expectativas previas en forma de valores iniciales más o menos altos para los pesos asignados a los lazos en una red asociativa.

4. Conclusiones: Perspectivas de integración presentes y futuras.

El estudio experimental del aprendizaje causal tal y como lo conocemos abarca ya tres décadas. El corpus de datos recogido a lo largo de todos estos años tiene ya un volumen considerable. Cualquier teoría que pretenda ser explicativa debe dar cuenta de los efectos derivados de la manipulación de la contingencia, las distintas desviaciones de la regla P que se han observado, los efectos de orden de los ensayos y de la interacción entre claves, los efectos debidos al modo de respuesta y la frecuencia del juicio, la reevaluación retrospectiva y los efectos de la manipulación de la estructura causal de la tarea, entre otros. Ninguna de las teorías propuestas explica todos y cada uno de los datos recogidos en la literatura, de ahí la necesidad de hacer una integración razonable de las distintas aportaciones teóricas que han ido surgiendo a lo largo de los años.

Una primera estrategia que ha dado buenos resultados ha sido considerar diferentes teorías a distintos niveles de explicación psicológica. Puede demostrarse matemáticamente que en tareas de juicios de contingencia/causalidad, asumiendo igualdad de parámetros de tasa, cuando se presenta una sola clave y un solo resultado, las predicciones de la regla LMS y P coinciden asintóticamente (Chapman y Robbins, 1990). Es más, cuando existe más de una clave y los distintos compuestos son anidables (cuando cada compuesto de un grado inferior es un subconjunto de un compuesto de grado superior, p.e. A, AB y ABC), las predicciones asintóticas de la regla LMS y P condicional también coinciden (Shanks, Holyoak y Medin, 1996). Esto ha llevado a distintos autores a proponer que simplemente se trata de teorías no incompatibles situadas en distintos niveles de organización. Cheng y Novick (1995) plantean la teoría del contraste probabilístico como una teoría situada a nivel computacional en los términos de D. Marr (1982). Cheng (1997) también plantea la teoría de potencia causal como una teoría situada en el nivel computacional. Por el contrario, los modelos de reglas y asociativos son modelos algorítmicos, esto es, teorizan sobre los procesos cognitivos que el sujeto efectivamente realiza. Shanks (1996) por ejemplo, considera que el equivalente computacional de los modelos basados en la regla LMS es la teoría del contraste probabilístico. Igualmente el algoritmo de cómputo de la contingencia en la teoría de potencia causal podría ser de naturaleza tanto asociativa como estadística.

Una segunda posibilidad de integración proviene de la constatación de que necesariamente deben existir varios tipos de mecanismos para el cómputo de las relaciones entre claves y resultados. Los modelos asociativos no son aplicables cuando la presentación de la información se hace en formato de tablas. En este contexto sólo son aplicables los modelos de reglas. De forma similar, cuando en tareas con presentación serial se produce experimentalmente una sobrecarga de la memoria de trabajo (por ejemplo mediante una tarea simultánea, o presentando un gran número de claves y resultados) no es lógico pensar que se adopte una estrategia de tipo estadístico. Sin embargo ni en las tareas en las que la información se presenta resumida, ni en las tareas seriales en las que se sobrecarga la memoria de

trabajo los juicios de contingencia sufren deterioros considerables en su ajuste. En otras palabras, tanto un mecanismo asociativo como uno estadístico son necesarios. La cuestión a resolver, por tanto, es determinar cuál es el mecanismo aplicado por defecto en situaciones en las que cualquiera de los dos podría aplicarse. Los modelos futuros deberán complicarse para establecer cuáles son los distintos mecanismos implicados en el aprendizaje causal y qué tipo de relaciones se establecen entre ellos.

El modelo de revisión de creencias de Catena y colaboradores (Catena et al, 1998; Maldonado et al, 1999; Perales, Catena y Maldonado, en preparación) planteado para explicar efectos relacionados con el modo de respuesta y de la frecuencia del juicio adopta esta perspectiva integradora. Según este modelo, el efecto de la frecuencia del juicio se produce porque las personas no utilizan directamente toda la información presentada durante la tarea experimental para realizar la estimación de contingencia, sino que integran la información presentada desde el último juicio emitido con el valor de ese último juicio. El modelo se expresa matemáticamente como sigue:

$$(13) J_i = J_{i-1} + \beta(EN - J_{i-1})$$

Donde J_i representa el juicio actual, J_{i-1} representa el juicio anterior emitido por el sujeto y β es un parámetro de actualización, cuyo valor depende de variables de tipo estratégico. EN es la parte nuclear del modelo, y representa una función que acumula evidencia sobre la relación entre la clave y la consecuencia entre dos juicios consecutivos. En la formulación actual del modelo esta función es semejante a una delta D ponderada (ver ecuación 3). Sin embargo, podría tratarse también de un mecanismo asociativo, tipo regla LMS. Por consiguiente, el modelo es independiente del tipo de mecanismo responsable de la acumulación de evidencia (el mecanismo de cómputo de la contingencia). De hecho ambos mecanismos (asociativo y estadístico) podrían funcionar en paralelo, siendo factores extra tarea, como la comprensión que el sujeto alcance de la tarea o sus expectativas, quienes determinen cuál de ellos va a ser utilizado en una situación determinada (Ramos y Catena, remitido). La única restricción que el mecanismo de actualización de juicios (el *mecanismo de integración de información*, expresado en la ecuación 12) impone sobre el mecanismo de cálculo de la relación clave-consecuencia es que éste debe reiniciarse siempre que se emite un nuevo juicio. La reiniciación del mecanismo de cálculo conllevaría la desaparición de las curvas estándar de aprendizaje. Esa predicción ha sido contrastada con éxito en nuestro laboratorio (Perales, Catena y Maldonado, en preparación).

La integración de distintas aportaciones teóricas debe hacerse en cualquier caso de forma cuidadosa, ya que tal y como ocurre en el modelo de Catena y cols., mantener determinados postulados a un nivel puede conllevar restricciones a otro nivel. En el futuro, el análisis de las coincidencias e incompatibilidades entre distintas aproximaciones teóricas podrá llevar a modelos más completos que los planteados hasta la fecha.

ABSTRACT

Learning of Contingency/Causality relationships: An approach to current theoretical trends. The study of causality learning has greatly increased along the two last decades. However, the rise of experimental data has not been accompanied by a parallel theoretical development that allows its integration. Any theoretical approach that claims to be explanatory should be able to explain these data, as well as its relationship with the types of task. This review presents an organized account of the experimental corpus of human causality learning, as well as its possible theoretical explanations. The analysis of the similarities and incompatibilities among different theoretical approaches led to consider models of serial organization that make explicit the different processing stages implied in this learning.

Key Words: Causality learning, contingency, covariation, associative models, rule models.

REFERENCIAS

- Allan, L. G. & Jenkins, H. M. (1980). The judgment of contingency and the nature of response alternatives. *Canadian Journal of Psychology*, 34, 1-11.
- Alloy, L. B. & Abramson, L. Y. (1979). Judgements of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology : General*, 108, 441-485.
- Anderson, J. R. & Sheu, C. (1995). Causal inferences as perceptual judgements. *Memory and Cognition*, 23, 510-524.
- Arkes, H. R. & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology : General*, 112, 117-135.
- Baker, A.G., Berbrier, M. W., Vallée-Tourangeau, F. (1989). Judgments of a 2 x 2 contingency table : Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology*, 41B, 65-97.
- Buehner, M. J. & Cheng, P. W. (1997). Causal induction : The Power-PC theory vs. Rescorla-Wagner theory. In Roitblat, H. L. & Meyer, J. A. (Eds.), *Proceedings of the nineteenth annual conference of the Cognitive Science Society*. Erlbaum. Mahwah, New Jersey.
- Catena, A., Maldonado, A. & 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.
- Chapman, G. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 414-432.
- Chapman, G. B. & Robbins, S. J. (1990). Cue interaction in human contingency judgement. *Memory and Cognition*, 18, 537-545.
- Chatlosh, D.L., Neunaber, D. J. & Wasserman, E. A. (1985). Response-outcome contingency : Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, 16, 1-34.

- Cheng, P. (1997). From covariation to causation : A causal power theory. *Psychological Review*, *104*, 367-405.
- Cheng, P. & Lien, L. (1996). The role of coherence in differentiating genuine from spurious causes. In Sperber, D., Premack, D. & Premack, A. (eds.) *Causal Cognition: A multidisciplinary debate*. Clarendon Press. Oxford.
- Cheng, P. & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545-567.
- Cheng, P. & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365-382.
- De Jong, P. J., Merckelbach, H. & Arntz, A. (1990). Illusory correlation, on-line probability estimates and electrodermal responding in a quasi-conditioning paradigm. *Biological Psychology*, *31*, 201-212.
- Dickinson, A. and Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *49B*, 60-80.
- Dickinson, A., Shanks, D. R. & Evenden, J. (1984). Judgment of the act-outcome contingency : The role of selective attribution. *Quarterly Journal of Experimental Psychology*, *36A*, 29-50.
- Dickinson, A. & Shanks, D. (1996). Instrumental action and causal representation. In Sperber, D., Premack, D. & Premack, A. (eds.) *Causal Cognition: A multidisciplinary debate*. Clarendon Press. Oxford.
- Jenkins, H. M. & Ward, W. C. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs*, *79*.
- Hastie, R. & Pennington, N. (1995). Cognitive approaches to judgment and decision making. *The Psychology of Learning and Motivation*, *32*, 1-31.
- Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology : Learning, Memory and Cognition*, *24*, 712-731.
- Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating : The belief adjustment model. *Cognitive Psychology*, *24*, 1-55.
- Kao, S. F. & Wasserman, E. A. (1993). Assessment of an 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*, 1363-1386.
- Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review*, *102*, 490-502.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *28*, 107-128.
- Levin, I. P., Wasserman, E. A. & Kao, S. F. (1993). Multiple methods of examining biased information use in contingency judgements. *Organizational Behavior and Human Decision Processes*, *55*, 228-250.
- Lober, K. & Shanks, D. R. Experimental falsification of Cheng's (1997) Power PC Theory of causal induction. Remitido.
- López, F., Shanks, D. R., Almaraz, J. & Fernández, P. (1998). Effects of trial order on contingency judgments : A comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology : Learning, Memory and Cognition*, *24*, 672-694.

- Lovibond, P. F., Siddle, A. T. & Bond, N. (1988). Insensitivity to stimulus validity in human pavlovian conditioning. *The Quarterly Journal of Experimental Psychology*, 40B, 377-410.
- Mackintosh, N. J. (1975). A theory of attention : Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-289.
- Maldonado, A., Catena, A., Cándido, A. & García, I. (1999). Assymmetrical effects of previous noncontingency on human covariation learning. *Animal Learning & Behavior*, 27, 168-180.
- Mandel, D. R. & Lehman, D. R. (1998). Integration of contingency information in judgements of cause, covariation and probability. *Journal of Experimental Psychology : General*, 127, 269-285.
- Markman, A. B. (1989). LMS rules and the inverse base-rate effect : Comments on Gluck and Bower (1988). *Journal of Experimental Psychology : General*, 118, 417-421.
- Marr, D. (1982). *Vision : A computational investigation into the human representation and processing of visual information*. Freeman. San Francisco.
- Matute, H., Arcediano, F. y Miller, R.R. (1996) Test question modulates cue competition between causes and effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 22, 182-196
- Matute, H. y Piñeno, O. (1998). Stimulus competition in the absence of compound conditioning. *Animal Learning and Behavior*, 26, 3-14.
- Miller, R. & Matute, H. (1996). Animal analogues of causal judgement. *The Psychology of Learning and Motivation*, 34, 133-166.
- Ortega, A. R. (1991). *Contingencia y Juicios de Covariación en humanos*. Tesis Doctoral. Universidad de Granada.
- Pearce, J. M. & Hall, G. (1980). A model for pavlovian conditioning Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Perales, J.C., Catena, A. & Maldonado, A. The frequency of judgment effect: The role of contingency and trial number. Remitido.
- Pavlov, I.P. (1927). *Conditional Reflexes*. Oxford University Press. Oxford.
- Premack, D. (1993). Prolegomenon to evolution of cognition. In Poggio, T. & Glaser, D. (eds.) *Exploring brain functions : Models in Neuroscience*.
- Price, P.C. & Yates, F. J. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 21, 1639-1655.
- Ramos, M. M. & Catena, A. Adaptive learning : Evidence in favour of an unconditional computation of causal relationship. Remitido.
- Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, 72, 77-94.
- Rescorla, R.A. (1971). Variations in the effectiveness of reinforcement and non reinforcement following prior inhibitory conditioning. *Learning and Motivation*, 2, 113-123.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of pavlovian conditioning : variations on the effectiveness of reinforcement and non-reinforcement. In Black, A. H. & Prokasy, W. F. (Eds.), *Classical Conditioning II : Current Theory and Research*. Appleton-Century Crofts. New York.
- Shanks, D. R. (1985a). Continuous monitoring of human contingency judgments across trials. *Memory and Cognition*, 13, 158-167.

- Shanks, D. R. (1985b). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, 37B, 1-21.
- Shanks, D. R. (1989). Selectional processes in causality judgment. *Memory and Cognition*, 17, 27-34.
- Shanks, D. R. (1991). On similarities between causal judgments in experienced and described situations. *Psychological Science*, 2, 341-350.
- Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology*, 84, 319-354.
- Shanks, D. R. (1995). Is human learning rational?. *Quarterly Journal of Experimental Psychology*, 48A, 257-279.
- Shanks, D. R. (1996). *The Psychology of Associative Learning*. Cambridge University Press.
- Shanks, D. R. & Dickinson, A. (1991). Instrumental learning and judgement under variations in action-outcome contingency and contiguity. *Memory and Cognition*, 19, 253-360.
- Shanks, D. R., Holyoak, K. J. & Medin, D. L. (Eds.).(1996). *The Psychology of Learning and Motivation (vol. 34) : Causal Learning*. Academic Press. San Diego.
- Shanks, D. R. & López, F. (1996). Causal order does not affect cue selection in human associative learning. *Memory and Cognition*, 24, 511-522.
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-173.
- Spellmann, B. A. (1996). Acting as intuitive scientists : Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, 7, 337-342.
- Tassoni, C. J. (1995). The LMS network with information coding : A model of cue learning. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 21, 193-204.
- Valleé-Tourangeau, F., Murphy, R. A. & Drew, S. (1997). Causal Judgments that violate the predictions of the power PC theory of causal induction. In Shafto, M.G. & Langley, P. (Eds.), *Proceedings of the nineteenth annual conference of the Cognitive Science Society*. Erlbaum. Mahwah, New Jersey.
- Troler, T. K. & Hamilton, D. L. (1986). Variables influencing judgements of correlational relations. *Journal of Personality and Social Psychology*, 50, 879-888.
- Wagner, A. R., Logan, F. A., Haberlandt, K. & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, 76, 171-180.
- Waldmann, M. R. (1996). Knowledge based causal induction. In D. R. Shanks, K.J. Holyoak, & D. L. Medin, (Eds.).(1996). *The Psychology of Learning and Motivation (vol. 34) : Causal Learning*, pp. 47-88. Academic Press. San Diego.
- Waldmann, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models : Asymmetries in cue competition. *Journal of Experimental Psychology : General*, 2, 222-236.
- Waldmann, M. R. & Holyoak, K. J. (1997). Determining whether causal order affects cue selection in human contingency learning : Comments on Shanks and Lopez (1996). *Memory and Cognition*, 25, 125-134.
- Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, 108, 284-297.
- Wasserman, E. A. (1990a). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science*, 1, 298-302.

- Wasserman, E. A. (1990b). Detecting response-outcome relations: Towards an understanding of the causal texture of the environment. In Bower, G. H. (Ed.) *The Psychology of Learning and Motivation*, 26, 27-82.
- Wasserman, E. (1993). Comparative cognition: Toward a general understanding of cognition un behavior. *Psychological Science*, 4, 156-161.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L. & Baker, A. G. (1993). Rating causal relations : The role of probability in judgements of response-outcome contingency. *Journal of Experimental Psychology : Learning, Memory & Cognition*, 19, 174-188.
- Wasserman, E.A., Kao, S.F., van Hamme, L.J., Masayoski, K. & Young, M.E. 1996) Causation and Association. In Shanks, D. R., Holyoak, K. J. & Medin, D. L. (Eds.).(1996). *The Psychology of Learning and Motivation (vol. 34) : Causal Learning*, 208-264.
- Williams , D.A., Sagness, K.B. & McPhee, J.,E. (1994) Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology : Learning, Memory & Cognition*, 20, 694-709.
- Yates, J. F. & Curley, S. P. (1986). Contingency judgments: Primacy effects and attention decrement. *Acta Psychologica*, 62, 293-302.

Revisión aceptada (5/10/99)