

INTRODUCCIÓN AL APRENDIZAJE CAUSAL

MIGUEL A. VADILLO Y HELENA MATUTE
UNIVERSIDAD DE DEUSTO

INTRODUCCIÓN

No es extraño oír a los estudiantes de psicología del aprendizaje cuestionar la utilidad de esta disciplina científica en la que casi todos los experimentos se realizan con animales y se ponen a prueba predicciones e hipótesis extrañas que aparentemente poco o nada tienen que ver con la vida cotidiana de las personas. ¿Para qué nos interesan a nosotros, al fin y al cabo, conceptos tan abstractos como los de recuperación espontánea, competición de claves, o aprendizaje perceptivo? ¿Nos dicen esos términos algo sobre la forma en que las personas pensamos y actuamos? No sólo los estudiantes, sino también los psicólogos profesionales de diversas áreas siguen considerando la psicología del aprendizaje como un mero resquicio de aquella presunta Edad Media de la psicología conocida como conductismo que, si bien nos da algunas ideas útiles para la práctica clínica, poco o nada nos dice sobre los procesos mentales característicamente humanos.

Nada más lejos de la realidad. En pleno siglo XXI la psicología del aprendizaje sigue aportándonos conceptos y teorías extremadamente útiles para entender algunos de los procesos mentales más importantes para los seres humanos. Los modelos asociativos que un día surgieron para explicar unos pocos fenómenos de condicionamiento animal nos permiten hoy explicar aspectos tan vitales y complejos de la vida mental humana como son el aprendizaje de conceptos, el aprendizaje verbal y lector, o la inducción causal. Será precisamente este último proceso mental el que abordaremos en la presente sección de este libro. Echando un breve vistazo a la historia reciente de los estudios sobre aprendizaje causal, veremos cómo los conceptos utilizados en estas investigaciones han sido explícita o implícitamente semejantes a los utilizados en los estudios sobre condicionamiento animal.

@becedario/p.301

PRIMEROS ESTUDIOS MODERNOS

Cualquier intento de distinguir un antes y un después en los estudios sobre razonamiento causal es necesariamente arbitrario y criticable. Los filósofos se han preocupado por las raíces de nuestro conocimiento causal desde el origen de los tiempos. Mucho más desde que David Hume (1739/1964) resaltara en su obra la importancia de la inducción causal en la construcción de nuestro conocimiento del mundo y señalara, además, las graves limitaciones del ser humano en este dominio. Por razones obvias, nuestra historia de la psicología del aprendizaje causal no puede remontarse tan atrás en el tiempo.

Puestos a elegir un punto por el que empezar no sería mala opción comenzar por describir algunos experimentos realizados en la década de los 60 del pasado siglo XX. Durante las décadas inmediatamente anteriores se había puesto de moda una interesante línea de investigación que consistía en comprobar hasta qué punto la actuación de las personas en diversas tareas se adaptaba a lo que podía considerarse un comportamiento racional. Por ejemplo, se estudiaba si la forma en que se toman diversos tipos de decisiones se ajusta a lo que cabría definir como óptimo desde la teoría económica o desde la teoría de juegos (Edwards, 1954). También surgió cierto interés por saber si las personas utilizaban y entendían los conceptos estadísticos y matemáticos de la forma adecuada (Peterson y Beach, 1967). Dentro de este contexto intelectual, algunos autores empezaron a preocuparse por estudiar hasta qué punto se utilizaban correctamente los conceptos de correlación y contingencia, iniciando así una tradición que llevaría a los estudios modernos sobre aprendizaje causal.

De entre los estudios de este tipo cabe destacar dos, realizados por Herbert Jenkins y William Ward en 1965 (Jenkins y Ward, 1965; Ward y Jenkins, 1965). En uno de estos experimentos (Jenkins y Ward, 1965, Experimento 1), se decía a un grupo de participantes que tenían que obtener puntos utilizando una máquina que disponía de dos botones. En cada ensayo, los participantes podían pulsar cualquiera de los dos botones y a continuación veían si habían ganado un punto en ese ensayo o no. La tarea de los participantes era detectar si el hecho de presionar un botón u otro les permitía influir en la probabilidad de recibir un punto. Si los puntos se adjudicaban con la misma probabilidad fuera cual fuera el botón elegido por el participante, se esperaba que juzgaran que no tenían control sobre la obtención del premio. Si, por el contrario, se ganaba el punto con mayor probabilidad presionando uno de los botones que presionando el otro, se esperaba que los participantes juzgaran haber tenido cierto control. En concreto, Jenkins y Ward proponían que lo más racional en esta situación sería que los participantes dieran sus juicios de control de acuerdo con un índice estadístico de contingencia, denominado ΔP , que se define como la diferencia entre la probabilidad del resultado (ganar el punto) si se da una respuesta y la probabilidad de ese mismo resultado si no se da esa respuesta. Los resultados mostraban que los juicios de control de los participantes apenas guardaban relación con esta medida de

contingencia, razón por la cual los autores concluyeron que las personas no entrenadas no tienen un concepto de correlación o contingencia adecuado. Dicho de otra forma, no parecía que los seres humanos fueran especialmente hábiles a la hora de juzgar la intensidad de una relación causal, en este caso entre la pulsación de un botón y la obtención de un punto. Los demás experimentos realizados en aquel estudio, así como otros experimentos similares realizados en la misma época arrojaban resultados parecidos (Smedslund, 1963, Ward y Jenkins, 1965).

Veremos después que investigaciones posteriores cuestionaron esta interpretación de los experimentos de los 60 que concluía que los seres humanos eran poco precisos a la hora de aprender relaciones causales. Por ahora, nos interesan especialmente dos características de estos estudios pioneros realizados en los años 60. En primer lugar, estos experimentos fueron los primeros en los que se usó un tipo de procedimiento que, con pequeñas modificaciones, se sigue utilizando en la actualidad. En el experimento de Jenkins y Ward, los participantes veían en cada ensayo si, tras haber presionado un botón u otro, aparecía el premio o no. Tras haber recibido unos cuantos ensayos de este tipo, los participantes debían dar su opinión sobre el grado de relación entre presionar un botón y obtener el punto. Los procedimientos empleados actualmente son muy similares. En cada ensayo se les dice a los participantes si está presente o no un evento que podría ser la causa (al que generalmente se denomina clave) y si está o no presente un segundo evento que podría ser el efecto (al que generalmente se denomina resultado). Tras ver unos cuantos ensayos en los que pueden coincidir o no clave y resultado se pide a los participantes que juzguen si hay o no una relación causal entre clave y resultado. Como puede verse, el procedimiento apenas ha cambiado desde los experimentos de Jenkins y Ward. (Véase Matute, 2004, para una descripción detallada de algunas de las tareas experimentales más utilizadas en la actualidad.)

El segundo motivo por el que estos estudios nos interesan tiene un carácter más teórico que metodológico. Decíamos más arriba que estos autores consideraban que los juicios de los participantes debían ajustarse al índice estadístico ΔP . Curiosamente casi al mismo tiempo, Robert Rescorla, entonces un joven investigador del condicionamiento clásico, proponía que ese mismo índice estadístico determinaba si los emparejamientos de un estímulo condicionado (EC) con un estímulo incondicionado (EI) producían condicionamiento o no. Según mostraban los experimentos de Rescorla (1966, 1968; véase también Rescorla, 1988, para una interpretación más actual), para que hubiera un buen condicionamiento no bastaba con que un EC se emparejara con un EI. Lo realmente importante era que la aparición del EI fuera contingente con la aparición del EC, es decir, que la probabilidad del EI justo tras el EC fuera mayor que la probabilidad del EI en ausencia del EC. ¡De modo que las ratas que participaban en los experimentos de condicionamiento clásico aprendían el grado de relación entre un EC y un EI, siendo sensibles a la misma información que Jenkins y Ward

consideraban relevante para detectar el grado de relación entre una causa y un efecto!

Hasta donde nuestro conocimiento llega, nadie en aquel momento se percató de las similitudes que existían entre el marco teórico que se utilizaba en los estudios de inducción causal en seres humanos y el marco teórico que comenzaba a surgir en el ámbito del condicionamiento animal. Sin embargo, vista desde la perspectiva que dan 40 años de historia, esta coincidencia parecía anunciar el futuro desarrollo conjunto de ambas áreas de investigación.

EL EFECTO DE LA CONTINGENCIA

Tras esos pocos estudios realizados a principios de los 60, el interés por los procesos de inducción causal y de detección de correlaciones decayó considerablemente. Sin embargo, a finales de los años 70 se reaviva el interés por estos temas con la publicación de un artículo de Alloy y Abramson acerca de la influencia de la depresión sobre la capacidad para detectar relaciones de causalidad (Alloy y Abramson, 1979). En este estudio, los participantes tenían varias oportunidades para realizar una respuesta y ver si después aparecía un determinado resultado. En algunas condiciones experimentales los participantes tenían cierto grado de control sobre este resultado (la probabilidad del mismo era diferente dependiendo de que se diera o no la respuesta) y en otras condiciones los participantes no podían influir en absoluto sobre la aparición del resultado (la probabilidad del resultado era la misma independientemente de que el participante ejecutara la respuesta o no). El hallazgo más importante de Alloy y Abramson fue que los deprimidos eran notablemente más precisos que los no deprimidos en sus juicios sobre el grado de control. Aunque tanto deprimidos como no deprimidos detectaron con cierta precisión el grado de relación entre respuesta y resultado cuando realmente tenían control, sólo los deprimidos se percataron de que no tenían control cuando no había relación alguna entre respuesta y resultado. Además, en el caso de los no deprimidos los juicios de control también se veían influidos por variables aparentemente irrelevantes, como el valor positivo o negativo del resultado, que no tenían efecto alguno sobre los juicios de los participantes deprimidos. Tanto la "ilusión de control" de los no deprimidos como la aparente mayor precisión de los deprimidos a la hora de detectar la ausencia de control (realismo depresivo) se han observado en situaciones muy diversas y son objeto de un renovado interés en la actualidad (Allan, Siegel, y Hannah, 2007; Msetfi, Murphy, Simpson, y Kornbrot, 2005). Ambos conceptos están además estrechamente relacionados con el estudio de la superstición (tendencia a repetir conductas bajo el falso supuesto de que provocan resultados deseables; véase Matute, 1996) y la indefensión aprendida (percepción de falta de control en una situación que puede llevar a renunciar a controlar eventos en otras situaciones;

véase Seligman, 1975; véase también el capítulo de Maldonado, Catena, Cándido, Perales, y Herrera en el presente volumen).

Uno de los aspectos que pueden destacarse del estudio de Alloy y Abramson (1979) es que, a pesar de que los juicios de los no deprimidos tenían algunos sesgos, estos parecían estar limitados sólo a algunas situaciones en las que no existía contingencia real entre respuesta y resultado. En todos los demás casos, sus juicios de control eran igual de precisos que los de los deprimidos. A menos que se tratara de una situación incontrolable, tanto deprimidos como no deprimidos daban juicios similares y aparentemente sensibles al índice estadístico que autores anteriores habían considerado ideal para valorar relaciones de causalidad: el índice de contingencia ΔP . Este resultado supone una incongruencia con los resultados de Jenkins y Ward anteriormente mencionados. Recordemos que estos autores habían concluido que los juicios de control de los participantes apenas parecían guardar relación con ΔP , mientras que Alloy y Abramson observaban ahora que los juicios de los participantes eran bastante próximos a este índice, salvo en las condiciones que daban lugar a la falsa percepción de control, llamada ilusión de control. Estos nuevos resultados parecían sugerir que los juicios de la causalidad no eran tan irracionales como los estudios previos parecían sugerir. Además, demostraban que el índice ΔP , que como ya se ha mencionado, parecía describir con bastante precisión el curso del aprendizaje en los experimentos de condicionamiento clásico con animales (Rescorla, 1966, 1968), era igualmente válido en el campo del aprendizaje causal humano. En efecto, Alloy y Abramson propusieron que los mecanismos responsables del aprendizaje causal bien podrían ser semejantes a los que gobiernan el condicionamiento clásico. Aunque no desarrollaron esta idea más a fondo, se trataba de una afirmación llena de interesantes implicaciones que serían desarrolladas a lo largo de las décadas posteriores y que proporcionaban un nuevo marco para la comprensión de la inducción causal.

Alloy y Abramson no fueron los únicos que estudiaron en aquella época el grado de correspondencia entre el aprendizaje causal, tal y como se manifestaba en los juicios de control de los participantes, y el índice estadístico ΔP . Por un lado, Herbert Jenkins, que anteriormente había publicado sus estudios pioneros con Ward, retomaba el estudio del aprendizaje causal con la colaboración de Lorraine Allan (Allan y Jenkins, 1980, 1983). Por otro lado, Edward Wasserman, investigador anteriormente centrado en el estudio del condicionamiento clásico, comenzaba a trabajar en el aprendizaje causal acompañado de un grupo de investigadores con el que realizaría múltiples experimentos durante los años 80 y 90 (Chatlosh, Neunaber, y Wasserman, 1985, Wasserman, Chatlosh, y Neunaber, 1983). Todos estos autores detectaron una serie de factores que podían explicar la aparente falta de correspondencia entre los juicios de control y ΔP en los estudios anteriores, y ayudaron a pulir los métodos de investigación para obtener juicios de control más "racionales".

MODELOS ASOCIATIVOS DE APRENDIZAJE CAUSAL

Los experimentos realizados hasta mediados de los 80 proporcionaron algunos datos interesantes sobre el aprendizaje causal, pero revelaron relativamente poco acerca de los mecanismos psicológicos que podían subyacer a este proceso. Los resultados comentados hasta ahora indicaban que los juicios sobre la intensidad de una relación causal coincidían en muchas situaciones con la contingencia entre causa y resultado medida con el índice estadístico ΔP . ¿Pero quiere esto decir que los participantes elaboraban sus juicios computando mentalmente este índice? Es decir, ¿comparan los participantes conscientemente la probabilidad del efecto en presencia de la supuesta causa con la probabilidad del efecto en ausencia de dicha causa? Muchos de los estudios revisados hasta ahora parecían dar por sentado que un proceso mental de cómputo de estas características estaba teniendo lugar en la cabeza de los participantes, pero no proporcionaban pruebas al respecto.

Fueron David Shanks y Anthony Dickinson los encargados de desafiar esta idea en los años 80 (Dickinson, Shanks, y Evenden, 1984; Shanks, 1985a; Shanks y Dickinson, 1987). Estos autores proporcionaron una interpretación del aprendizaje causal sin asumir que los participantes elaboraban sus juicios por medio de razonamientos estadísticos u otro tipo de procesos cognitivos de nivel superior. En el ámbito de la investigación sobre condicionamiento clásico con animales el concepto de contingencia estaba siendo sustituido por nuevos modelos que explicaban el aprendizaje como resultado de procesos muy automáticos que se limitaban a asociar las representaciones mentales de los estímulos que se emparejaban en el curso del condicionamiento (para una revisión de la crisis del concepto de contingencia, véase Papini y Bitterman, 1990). En opinión de Shanks y Dickinson, los mismos modelos asociativos que se utilizaban para explicar diversos fenómenos de condicionamiento clásico en animales podían ser extrapolados para entender el aprendizaje causal humano. El modelo de condicionamiento clásico más popular del momento, propuesto por Rescorla y Wagner en 1972, asumía que al emparejarse un EC con un EI se asociaban las representaciones mentales de ambos estímulos de tal modo que, en el futuro, al activarse la representación del EC automáticamente se activaría también la representación del EI (véase el capítulo de Luque, López, y Cobos en el presente volumen para una descripción pormenorizada del modelo). De la misma forma, Shanks y Dickinson proponían que, cuando una causa potencial y un efecto se emparejan repetidas veces, se desarrolla una asociación entre las representaciones mentales de ambos eventos, de tal forma que cuando se active la causa en el futuro se activará también la representación del efecto. Lo más interesante de esta explicación es que asume que el aprendizaje de relaciones causales, al igual que el condicionamiento, tiene lugar de manera automática, sin necesidad de que intervengan procesos cognitivos superiores.

Estos modelos asociativos también permitían explicar por qué los juicios de causalidad se veían afectados por la contingencia causa-efecto. Según el modelo de Rescorla y Wagner (1972), los emparejamientos de un EC y un EI sólo conducen al fortalecimiento de la asociación entre ambos si el organismo aún no sabe predecir el EI con la ayuda de los estímulos disponibles. En otras palabras, lo que se aprende sobre la relación entre un EC y un EI depende de hasta qué punto ese EC ayuda a predecir mejor el EI. En el caso del aprendizaje causal humano, cuando hay una contingencia positiva entre la causa y el efecto, la probabilidad del efecto es mayor en presencia de la causa que en su ausencia. En esta situación, la causa ayuda a predecir mejor el efecto. Según el modelo de Rescorla y Wagner, esto dará lugar al fortalecimiento de la asociación excitatoria causa-efecto, con lo que los juicios de los participantes sobre la relación entre ambos eventos aumentarán.

En una interesante serie de experimentos, Shanks y Dickinson, junto con sus colaboradores, trataron de mostrar (1) algunos paralelismos empíricos entre los fenómenos de condicionamiento y los de aprendizaje causal y (2) algunas pruebas de la adecuación de modelos de condicionamiento concretos, como el de Rescorla y Wagner, para explicar el aprendizaje causal humano. La principal aportación de uno de sus primeros estudios (Dickinson y colaboradores, 1984) fue replicar en el paradigma del aprendizaje causal humano uno de los fenómenos de condicionamiento más ampliamente estudiados: el efecto de bloqueo (véase el capítulo de Pineño, Castro, y Urcelay en el presente volumen). Desde finales de los años 60 se sabía que la fuerza de la respuesta condicionada a un EC se reducía si ese EC se había emparejado con el EI en presencia de otro estímulo que ya permitía predecir el EI (Kamin, 1968). Es decir, si tras una serie de emparejamientos de un EC (A) con el EI ($A \rightarrow EI$) tratamos de condicionar un segundo EC (B) junto con el EC anteriormente entrenado ($AB \rightarrow EI$), el resultado normal suele ser que se observa poca respuesta condicionada a B. Dickinson y cols. (1984) replicaron este efecto, conocido como bloqueo, en un estudio de aprendizaje causal (véase también Shanks, 1985b). Para ello, utilizaron una tarea similar a un videojuego en la que los participantes debían estudiar hasta qué punto disparar un proyectil era efectivo a la hora de lograr la destrucción de un tanque. En la primera fase del experimento, los participantes veían que cuando el tanque entraba en un campo de minas (causa potencial A) el tanque explotaba con una probabilidad alta ($A \rightarrow \text{explosión}$). En una segunda fase, los participantes podían disparar un proyectil al tanque (causa potencial B) mientras pasaba por ese mismo campo de minas. En estos ensayos el tanque también explotaba con la misma probabilidad ($AB \rightarrow \text{explosión}$). Como cabría esperar, los participantes apenas juzgaron que el proyectil que habían disparado fuera eficaz para destruir el tanque. Las explosiones del tanque que tenían lugar al disparar el proyectil podían en realidad atribuirse al campo de minas. En otras palabras, el conocimiento previo de la eficacia de las minas para provocar la explosión había bloqueado el aprendizaje causal sobre los proyectiles. La conclusión que Dickinson y sus colaboradores sacaron de estos resultados es que los modelos de condicionamiento que se habían desarrollado para explicar el bloqueo en animales

deberían ser considerados como explicaciones plausibles del aprendizaje causal humano.

Experimentos publicados por Shanks pocos años después (Shanks, 1985a, 1987) también apuntaban en la misma dirección. En estos dos estudios, Shanks demostró que los juicios de causalidad muestran una curva de aprendizaje similar a la que se observa en el condicionamiento: los juicios empiezan siendo bajos y aumentan a medida que los participantes van adquiriendo más información. Por ejemplo, si sólo hemos tenido la oportunidad de disparar un proyectil al tanque y hemos visto que efectivamente el tanque se destruye, hasta ahí la contingencia entre disparar el proyectil y la explosión es perfecta. Sin embargo, contando sólo con esa evidencia pocos de nosotros nos mostraríamos absolutamente seguros de que el proyectil es completamente efectivo. Sólo nos convenceremos de que el proyectil es una causa efectiva de la explosión después de ver unos cuantos emparejamientos adicionales del proyectil y la explosión, y también después de comprobar que el tanque no explota cuando no se dispara el proyectil. Este incremento gradual de los juicios de causalidad a medida que se van recibiendo nuevas piezas de información es muy similar a lo que sucede en el condicionamiento a medida que se van proporcionando nuevos emparejamientos EC-EI. Cuando un EC y un EI se han emparejado unas pocas veces apenas se observa respuesta condicionada tras la presentación del EC, aunque la contingencia entre ambos estímulos sea perfecta. Sin embargo, a medida que se proporcionan más y más emparejamientos del EC con el EI la respuesta condicionada ante el EC va aumentando progresivamente. De nuevo, estos resultados son fáciles de interpretar desde los modelos asociativos de condicionamiento. Sin embargo, no es tan fácil explicarlos si uno asume que los participantes están calculando la contingencia explícitamente con procesos de razonamiento matemático. Si este fuera el caso, no importaría demasiado cuántos ensayos o emparejamientos causa-efecto hubieran visto: con un único emparejamiento de la causa y el efecto la contingencia ya es perfecta. En principio, no harían falta ensayos adicionales para descubrir que existe una contingencia positiva entre causa y efecto.

Un tercer tipo de estudios realizados por Shanks y Dickinson (1987) trataba de estudiar la influencia que tenía interpolar un intervalo de tiempo entre causa y efecto. Los experimentos de condicionamiento mostraban que el aprendizaje es más fácil cuando hay sólo una breve demora entre el EC y el EI (en el caso del condicionamiento clásico) o entre la respuesta operante y el reforzador (en el condicionamiento instrumental). A medida que el intervalo temporal entre ambos se hace más largo, se va haciendo más difícil observar un buen condicionamiento (véase el capítulo de Escobar y Arcediano en el presente volumen). También en el caso del aprendizaje causal es fácil detectar una relación de causa-efecto cuando el efecto sigue inmediatamente a la causa. Sin embargo, a medida que se reduce la contigüidad temporal entre causa y efecto la relación se va haciendo progresivamente más difícil de percibir. En efecto, Shanks y Dickinson (1987; véase también Shanks y Dickinson, 1991) comprobaron que los juicios sobre el

grado de relación causal entre una acción (p. ej., presionar una tecla en el ordenador) y un evento que le seguía a continuación (p. ej., la aparición de unas figuras geométricas en la pantalla) se reducían a medida que se incrementaba el lapso temporal que transcurría entre ambos eventos.

Todos estos resultados y otros semejantes prestaban apoyo a la hipótesis de que los mecanismos responsables del condicionamiento en animales no humanos podrían ser similares a los que subyacen al aprendizaje causal en humanos, lo que implicaba que las teorías asociativas propuestas para explicar los fenómenos de condicionamiento podían y debían extenderse al estudio del aprendizaje causal. No debemos olvidar que el auge del conexionismo en la década de los 80 reforzaba este interés por el asociacionismo en ámbitos muy diversos. Como veremos en las siguientes secciones, buena parte de la investigación realizada desde entonces se ha llevado a cabo con la intención expresa de contrastar la validez de estas explicaciones asociativas con teorías alternativas que surgieron en años posteriores. Además, estos modelos asociativos se han desarrollado y han sufrido múltiples revisiones para acomodarse a ciertos resultados empíricos que plantearon en su momento dificultades a la hora de ser explicados desde los modelos asociativos originales. Como veremos más adelante, apenas puede entenderse la investigación sobre aprendizaje causal realizada en las dos últimas décadas sin conocer con cierta profundidad estos modelos asociativos y sus variantes más perfeccionadas.

CONTRASTES PROBABILÍSTICOS Y POWER PC

Los 90 suponen, en cierta medida, una reacción a esta visión del aprendizaje causal como algo sencillo, automático e independiente de procesos cognitivos de orden superior. Frente a esta concepción, relativamente *Humeana*, de la inducción de relaciones causales, surgen una serie de perspectivas teóricas a las que cabría denominar *Kantianas*. En lugar de percibir el aprendizaje causal como un proceso pasivo en el cual la mente se limita a asociar eventos que tienden a aparecer juntos, las nuevas teorías van a destacar que en muchas situaciones la propia mente impone sobre los eventos cierta estructura y los analiza de forma activa siguiendo unas pautas que garantizan la "racionalidad" del proceso de inducción causal. Dos son los principales responsables de esta tendencia: Patricia Cheng y Michael Waldmann. Veamos sus aportaciones por separado.

Como ya hemos comentado antes, los primeros experimentos de aprendizaje causal estuvieron guiados por la hipótesis de que los seres humanos detectamos relaciones de causalidad aplicando reglas estadísticas como, por ejemplo, comparar la probabilidad del efecto en presencia y en ausencia de la causa potencial (es decir, computando ΔP). Cheng retoma esta idea pero propone además que estos cálculos estadísticos no se realizan de cualquier manera, sino que incluyen o excluyen determinadas piezas de información. En concreto Cheng

propone que, al realizar estos cálculos, los participantes tienen en cuenta el principio de control que utilizan los científicos siempre que realizan un experimento para comprobar si existe una relación de causalidad (Cheng y Novick, 1990, 1992).

Imagine que desea saber si una sustancia que acaba de descubrirse en un laboratorio estimula la inteligencia. Para comprobar esta hipótesis, lo mejor es inyectar la sustancia a un grupo de personas y ver si su inteligencia se desarrolla más que la de un grupo de control al que no se inyecta esa sustancia. En principio, si la inteligencia aumenta más en los que han sido tratados con la sustancia que en los que no han sido tratados (es decir, si la probabilidad o intensidad del efecto es mayor cuando está presente la causa potencial que cuando está ausente), cabe concluir que la sustancia realmente afecta a la inteligencia. Pero imagine ahora que descubrimos que las personas a las que hemos inoculado esta sustancia también han pasado más tiempo leyendo en las dos últimas semanas. Esto supone un grave problema para la interpretación que habíamos hecho de nuestro estudio. Si los participantes tratados con la sustancia se diferencian del grupo control no sólo en el hecho de haber recibido el tratamiento, sino también en el tiempo dedicado a la lectura, entonces ya no podemos estar tan seguros de que sea realmente la sustancia la responsable de las diferencias en la inteligencia. Sería al menos igual de plausible suponer que la lectura fue la que estimuló el desarrollo de la inteligencia. Para poder concluir que la sustancia afecta a la inteligencia es necesario que no haya ninguna diferencia entre los dos grupos de participantes salvo el hecho de haber recibido o no el tratamiento con la sustancia a estudiar.

Según el modelo de contrastes probabilísticos, propuesto por Cheng y Novick (1990, 1992), en condiciones naturales la inducción causal es sensible a este tipo de consideraciones. A la hora de saber si un factor tiene influencia sobre un efecto determinado los seres humanos trataríamos de comparar la probabilidad del efecto en presencia y en ausencia de dicho factor *manteniendo constantes todos los demás factores que puedan influir en el efecto*. Supongamos que una persona dispone de información sobre la probabilidad del efecto cuando están presentes dos factores causales, A y B, cuando sólo está presente A y también cuando ninguno de los dos factores está presente. Si esta persona quiere saber qué influencia tiene A sobre el efecto tendrá que comparar la probabilidad del efecto cuando sólo está presente A con la probabilidad del efecto cuando ni A ni B están presentes. La información sobre lo que sucede cuando A y B están presentes no debería tomarse en consideración porque si se compara la probabilidad del efecto cuando A y B están presentes con la probabilidad del efecto cuando ni A ni B están presentes, las diferencias podrían deberse tanto a A como a B. La información sobre lo que sucede con A y B podría tenerse en cuenta si se dispusiera de información sobre la probabilidad del efecto cuando sólo está presente B. En este caso, podríamos obtener información sobre la relación entre A y el efecto comparando la probabilidad del efecto en presencia de A y B con la probabilidad del efecto en presencia de B solo. Dado que la única diferencia entre

ambas condiciones es que A está presente en una y ausente en otra, cualquier diferencia en la probabilidad del efecto tiene que deberse a A y solamente a A.

En esta breve introducción no podemos dar mucha información adicional sobre la perspectiva teórica de Cheng (aunque el lector interesado puede consultar el capítulo de Luque y cols. en el presente volumen para una descripción más formal y completa del modelo de contrastes probabilísticos). Sí debemos destacar, no obstante, que este modelo permite explicar algunos fenómenos que anteriormente se consideraban compatibles únicamente con modelos asociativos. Por ejemplo, por aquella época era de sobra sabido que los juicios sobre la relación entre una causa potencial y un efecto no son necesariamente bajos en toda condición de contingencia cero. Si se muestran emparejamientos de la causa potencial A y el efecto, y también presentaciones aisladas del efecto, la contingencia es baja (la ocurrencia del efecto es probable tanto en presencia como en ausencia de A) y, lógicamente, esto da lugar a juicios bajos sobre el grado de relación entre A y el efecto. Sin embargo, si en las ocasiones en las que se presenta el efecto sin la causa potencial A se introduce otro predictor, por ejemplo una causa alternativa B, entonces los juicios sobre la relación entre A y el efecto vuelven a ser altos (Shanks, 1986). En otras palabras si en lugar de entremezclar los emparejamientos A→efecto con presentaciones aisladas del efecto, se entremezclan ensayos A→efecto con ensayos B→efecto, los juicios sobre A vuelven a ser altos. Esto no tiene sentido si se analiza sin prestar atención a las sugerencias de Cheng, puesto que tanto las presentaciones aisladas del efecto como los emparejamientos B→efecto incrementan la probabilidad del efecto en ausencia de A, lo que a su vez debería reducir la contingencia percibida entre A y el efecto en el mismo grado. Sin embargo, si uno tiene en cuenta los principios del modelo de contrastes probabilísticos es fácil entender el resultado. Cuando se entremezclan ensayos A→efecto con presentaciones del efecto solo, el participante puede calcular la contingencia entre A y el efecto manteniendo la presencia de B sistemáticamente ausente. Esta contingencia es baja y por tanto se juzga que no hay mucha relación entre A y el efecto. Sin embargo, si se entremezclan ensayos A→efecto con ensayos B→efecto, los ensayos B→efecto no pueden tenerse en cuenta al computar la contingencia entre A y el efecto: mezclar la información que proviene de situaciones en las que B está ausente (ensayos A→efecto) con la información que proviene de situaciones en la que B está sistemáticamente presente (ensayos B→efecto) supondría una violación del principio de control. Consecuentemente, sólo los ensayos A→efecto se tienen en cuenta a la hora de juzgar la efectividad causal de A. Por tanto, a la luz del modelo de contrastes probabilísticos, este fenómeno puede explicarse perfectamente como resultado de un proceso de razonamiento estadístico.

En 1997, Cheng propuso una extensión de su teoría a la que denominó Power PC (Cheng, 1997). La principal aportación de esa teoría reside en que supone que los seres humanos no sólo son sensibles al principio de control, sino que también tienen en cuenta el *efecto techo* a la hora de realizar sus inducciones causales. Imagine que queremos comprobar si participar en actividades deportivas tiene

algún efecto positivo sobre el bienestar psicológico. Para llegar a una conclusión firme lo que deberíamos hacer es tomar dos grupos de participantes tan semejantes como sea posible y adjudicarlos a dos condiciones experimentales diferentes: a uno de los grupos se le debería hacer participar en actividades deportivas y al otro no. Si tras unos meses ambos grupos difieren en su grado de bienestar psicológico, cabe suponer que esta diferencia se deba a que unos han realizado las actividades y los otros no. Pero imagine que realizamos este estudio y que nos encontramos con que no hay diferencias entre ambos grupos. Los participantes que han realizado actividades deportivas manifiestan un alto grado de bienestar psicológico. Sin embargo, los participantes que no han realizado esas actividades manifiestan un nivel de bienestar psicológico igualmente alto. ¿Quiere esto decir que las actividades físicas no tienen efecto alguno sobre el bienestar? Tal vez. Pero una explicación alternativa es que nuestro experimento no permite detectar este efecto simplemente porque hemos utilizado los participantes erróneos. Nuestro grupo de control indica que los participantes que hemos elegido para nuestro estudio son ya de por sí muy felices. Tal vez sean tan felices que las actividades deportivas poco o nada puedan hacer para mejorar su bienestar. Pero eso no quiere decir que las actividades deportivas no puedan tener un efecto más visible en personas con un nivel más bajo de bienestar. En otras palabras, es posible que las actividades sí que tengan influencia causal sobre el bienestar pero que en nuestro experimento nos hayamos encontrado con un efecto techo: el nivel de bienestar es tan alto que ya no podemos incrementarlo más.

Según la teoría Power PC, las inferencias causales que realizan los seres humanos de forma natural y espontánea también muestran sensibilidad a este efecto techo. Si una situación no es la adecuada para extraer conclusiones sobre la eficacia causal de un factor debido a que el efecto a estudiar ya se presenta en un grado muy alto (efecto techo), los participantes se mostrarán reacios a realizar inferencia causal alguna: mostrarán incertidumbre o juicios de causalidad reducidos (Wu y Cheng, 1999).

TEORÍA DEL MODELO CAUSAL

Decíamos en la sección anterior que el desafío a los modelos asociativos provenía de dos fuentes, una de ellas protagonizada por Cheng y la otra por Waldmann. Veamos ahora la aportación de este segundo autor. Según Waldmann, una limitación adicional de los modelos asociativos es que no contemplan la posibilidad de que los participantes utilicen conocimientos causales previos a la hora de aprender nuevas relaciones causales. Por ejemplo, todas las personas saben que en condiciones naturales las causas preceden a sus efectos. El hecho de que los modelos asociativos no tengan en cuenta este conocimiento les lleva a postular que el aprendizaje obedecerá las mismas leyes independientemente de que en cada ensayo se presente primero la causa

potencial y luego el efecto o que en cada ensayo se presente primero el efecto y luego la causa potencial. Según Waldmann, esto no sucede así en la realidad. Aunque se distorsionara el orden natural de los eventos presentando primero el efecto y luego la causa, los participantes reorganizarían la información para valorar la relación causal siempre en la dirección causa-efecto y nunca en la dirección inversa.

Esta hipótesis se puso a prueba por primera vez en 1992 con un experimento clásico realizado por el mismo Waldmann con la colaboración de Keith Holyoak (Waldmann y Holyoak, 1992). En aquel estudio se utilizaba un diseño de bloqueo con dos fases. En una de las fases una clave iba siempre seguida de un resultado (A→resultado). En la segunda fase, esa misma clave aparecía con una segunda clave seguida del mismo resultado (AB→resultado). Según lo que hemos visto en secciones anteriores, este procedimiento debería dar lugar al bloqueo del aprendizaje sobre la clave que se presenta por primera vez en la segunda fase (B). Dicho de otra manera, los participantes de este experimento deberían juzgar que hay, si acaso, una débil relación entre la segunda clave y el resultado. Los modelos asociativos predicen que esto es lo que debería pasar en todos los casos. Sin embargo, la teoría del modelo causal de Waldmann predice que la obtención de bloqueo debería depender de cómo se interpreten causalmente las claves y el resultado. Si se induce a los participantes a pensar que las claves son las causas potenciales del resultado, entonces cabe suponer que los participantes realmente mostrarán bloqueo. Es decir, una vez comprobado que la causa potencial A va seguida siempre del efecto, tenderán a asumir que en los ensayos en los que el efecto aparece precedido tanto de A como B la verdadera causa del efecto es A y no B. Sin embargo, este efecto de bloqueo no debería aparecer si se induce a los participantes a percibir las claves como efectos de las causas que se presentan después. Por ejemplo, podemos decir a los participantes que lo que ven en cada ensayo en primer lugar (las claves) son los síntomas que presenta un paciente imaginario y que lo que ve después (el resultado) es la enfermedad que ha provocado esos síntomas. Según la teoría del modelo causal, en este caso no debería haber bloqueo. Si en una primera fase vemos que un síntoma está originado por una enfermedad (A←enfermedad) y después, en una segunda fase, vemos que ese síntoma aparece junto con un segundo síntoma provocado por la misma enfermedad (AB←enfermedad), ¿concluiremos que hay una relación causal entre el segundo síntoma y la enfermedad? Según Waldmann, sí deberíamos percibir en este caso una relación causal entre el síntoma B y la enfermedad, puesto que la enfermedad es la única causa posible de B. Los experimentos realizados por Waldmann y Holyoak (1992) respaldan esta predicción: el bloqueo desaparece si las claves se interpretan como efectos y los resultados como causas potenciales (véanse también Waldmann, 2000, 2001), aunque, como veremos a continuación, la generalidad de este efecto ha sido ampliamente cuestionada.

LA “REBELIÓN” ASOCIATIVA

Gran parte de los estudios publicados en los 90 fueron intentos más o menos directos de contrastar las predicciones de estos nuevos modelos causales con las de las teorías asociativas. Es cierto que los datos aportados por Cheng y por Waldmann suponían un reto para las explicaciones asociativas del aprendizaje causal. Sin embargo, aún se podía cuestionar la validez y la generalidad de estas críticas. ¿Realmente son los seres humanos siempre sensibles al principio de control, a la evitación del efecto techo, a la interpretación causal de claves y resultados? Y cuando lo son, ¿realmente no pueden explicarse estos fenómenos de otra manera? Además estos modelos novedosos, principalmente los de Cheng, volvían a plantear la imagen del ser humano como un estadístico intuitivo que, teniendo ya almacenados muchos datos, detectaba relaciones de causa y efecto realizando complejos cálculos probabilísticos. Pero no se detenían a analizar la naturaleza del proceso de aprendizaje que permitía obtener la información necesaria para realizar esos cálculos (Vadillo, Miller, y Matute, 2005). Surgieron así múltiples reacciones contra estos modelos y se cuestionó la explicación del aprendizaje causal como un proceso de razonamiento estadístico y también la presunta racionalidad de la inducción causal humana.

Sería imposible revisar aquí todos los estudios que pusieron en tela de juicio las interpretaciones de estos nuevos modelos y que parecían indicar que, al menos en muchos casos, el aprendizaje causal seguía patrones que podían explicarse mejor en términos de procesos asociativos. Sin entrar en mucho detalle, podemos mencionar que se comprobó, por ejemplo, que el orden en que se proporcionaba la información tendía a sesgar los juicios de causalidad de formas difíciles de explicar para los modelos de Cheng (López, Shanks, Almaraz, y Fernández, 1998), que los juicios de causalidad se iban incrementando progresivamente dando lugar a una curva de aprendizaje inexplicable para los modelos de razonamiento estadístico (López, Almaraz, Fernández, y Shanks, 1999), que los fenómenos de competición de claves, como el bloqueo, también podían tener lugar cuando las claves se interpretaban como efectos y no como causas (Cobos, López, Caño, Almaraz, y Shanks, 2002; Shanks y López, 1996; Matute, Arcediano, y Miller, 1996), y que los participantes no controlaban siempre el efecto techo de la forma predicha por el Power PC (Lover y Shanks, 2000). Todos estos resultados, fácilmente explicables desde los modelos asociativos, cuestionaron la adecuación de las explicaciones propuestas por el modelo de contrastes probabilísticos, por la teoría Power PC y por la teoría del modelo causal.

Por otra parte, las interpretaciones asociativas volvieron a ponerse de moda porque se descubrieron algunos fenómenos que inicialmente suponían un problema para los modelos asociativos tradicionales, pero que finalmente pudieron explicarse introduciendo algunos cambios en las teorías (véase también el capítulo de Pineño y cols. en el presente volumen). Una de las limitaciones de los modelos

asociativos clásicos, como el de Rescorla y Wagner (1972), es que habían asumido que sólo se aprendía sobre una causa potencial en los ensayos en los que esa causa estaba físicamente presente. Sin embargo, hacia finales de los 90 había ya muchas pruebas de que los juicios sobre una causa potencial podían cambiar a la luz de evidencia obtenida en ausencia de esa causa (p. ej., Shanks, 1985b). En concreto, lo que una persona opine sobre una causa puede cambiar si posteriormente recibe información sobre otras causas. A los fenómenos que ilustraban este cambio de opiniones sobre estímulos ausentes se los denominó colectivamente fenómenos de reevaluación retrospectiva. Uno de los efectos de reevaluación retrospectiva más ampliamente estudiados es el llamado bloqueo hacia atrás. En un experimento de este tipo los participantes ven primero que dos causas potenciales van seguidas de un mismo efecto (AB→efecto). En una segunda fase observan que una de esas causas presentada de forma aislada también va seguida por el efecto (A→efecto). El resultado normal de estos experimentos es que, como fruto de los emparejamientos de una de las causas con el efecto en la segunda fase, automáticamente se reducen los juicios de causalidad sobre la otra causa potencial (en este caso, B).

La respuesta que se dio a este tipo de fenómenos fue revisar los modelos asociativos de tal modo que contemplaran la posibilidad de que se aprendiera sobre causas ausentes. Los dos modelos revisados más influyentes, la revisión del modelo de Rescorla y Wagner realizada por Van Hamme y Wasserman (1994) y la revisión del modelo SOP de Wagner realizada por Dickinson y Burke (1996), explicaron la reevaluación retrospectiva asumiendo que las causas ausentes pero esperadas cambiaban su fuerza asociativa en la dirección opuesta a los cambios sufridos por las causas presentes. Es decir, en el experimento de bloqueo hacia atrás cuando una de las dos causas se empareja con el efecto en la segunda fase su fuerza asociativa aumenta y, consecuentemente, la fuerza asociativa de la clave ausente debe disminuir. Este tipo de modelos reavivó el interés por las teorías asociativas y dio lugar a varios experimentos de gran interés que además proporcionaban datos poco compatibles con modelos más "racionalistas" como el modelo de contrastes probabilísticos (p. ej., Larkin, Aitken, y Dickinson, 1998; Dickinson y Burke, 1996).

ÚLTIMAS TENDENCIAS

Las investigaciones más recientes sobre aprendizaje causal no pueden entenderse sin tener en cuenta la evolución de estos estudios durante las décadas previas. Acabamos de ver que la psicología del aprendizaje causal entra en el nuevo milenio en un estado más que saludable: el desarrollo de nuevos modelos durante los 90 ha dado lugar a un intenso debate teórico y empírico, generando un conocimiento mucho más profundo de los procesos de aprendizaje causal.

Son muchos los frentes en los que la investigación sobre inducción causal ha avanzado entre el año 2000 y el 2006. La obra del investigador belga Jan De Houwer y sus colaboradores es tal vez una de las fuentes de debate más fructíferas de estos últimos años. En toda una serie de estudios recientes, estos autores tratan de demostrar que, independientemente de que en el aprendizaje causal puedan intervenir procesos asociativos, los resultados de los experimentos están siempre o casi siempre contaminados por procesos mentales de orden superior. Muchas demostraciones de bloqueo y otros fenómenos de competición de claves se han interpretado tradicionalmente desde una perspectiva asociativa. Sin embargo, estos autores han demostrado en repetidas ocasiones que estos fenómenos deberían explicarse en términos de procesos de razonamiento más complejos (controlados, conscientes, cognitivamente demandantes, etc.). Por ejemplo, el bloqueo puede potenciarse si a los participantes se les hace pensar que el efecto utilizado en el experimento no se presenta nunca en su máxima intensidad. Imagine que cuando un tanque atraviesa un campo de minas se produce una explosión de una intensidad 10 en una escala que va de 0 a 20. Cuando el tanque va por ese mismo campo de minas y además le disparamos un proyectil la intensidad de la explosión sigue siendo 10 de 20. Esta información sobre la intensidad del efecto nos da muchas más pistas para descartar que el proyectil tenga efecto alguno. Si el proyectil tuviera alguna influencia, al dispararlo, la intensidad de la explosión debería ser mayor que cuando sólo están presentes las minas. Los experimentos realizados por De Houwer y sus colaboradores (De Houwer, Beckers, y Glautier, 2002) muestran que, efectivamente, los participantes utilizan este tipo de información sobre la intensidad del efecto, resultado que parece indicar que este bloqueo se debe más a un proceso deliberado de razonamiento que a la acción de un mecanismo asociativo automático. En otra serie de experimentos, estos mismos autores demostraron que fenómenos como el bloqueo se reducen si se les pide a los participantes que realicen simultáneamente una tarea distractora que consume recursos atencionales y les impide razonar con calma (De Houwer y Beckers, 2003). Un valor añadido de esta orientación teórica es que algunas de sus hipótesis se han replicado en paradigmas de condicionamiento clásico con animales, lo que sugiere que incluso los procesos mentales que subyacen al condicionamiento clásico en ratas podrían ser más sofisticados de lo que se había pensado hasta ahora (Beckers, Miller, De Houwer, y Urushirara, 2006).

Dentro de esta misma corriente, cabe destacar los trabajos de Peter Lovibond y sus colaboradores. Estos autores se han centrado en mostrar que el bloqueo puede verse potenciado por medio de instrucciones o preentrenamientos que inviten a los participantes a analizar la información recibida de acuerdo con ciertas reglas (Lovibond, Been, Mitchell, Bouton, y Frohardt, 2003; Mitchell y Lovibond, 2002). Por ejemplo, el bloqueo se ve potenciado si a los participantes se les recuerda que el efecto combinado de dos causas debería ser mayor que el de cualquiera de ellas por separado. Utilizando el ejemplo anterior del tanque, esto implica enseñar a los participantes que si dos armas son efectivas y se disparan a la vez, los daños sufridos por el tanque deberían ser mayores que los que sufriría

si se disparara sólo una de esas armas. Esta forma de interpretar la información predispone a los participantes a mostrar bloqueo, ya que lleva a pensar que si la intensidad de la explosión es la misma cuando el tanque simplemente atraviesa el campo de minas (en la primera fase) que cuando atraviesa el campo de minas y además le disparamos un proyectil (en la segunda fase), el proyectil no ha debido ser muy efectivo. De lo contrario, su influencia sobre la explosión se habría sumado a la que producen las minas por sí solas y, por tanto, la explosión debería haber sido mayor. De nuevo, el hecho de que un fenómeno de aprendizaje tan básico como el bloqueo parezca verse afectado por este tipo de interpretaciones sugiere que buena parte de los fenómenos de aprendizaje está influida por procesos cognitivos de orden superior.

Otro interesante desarrollo de los estudios de aprendizaje causal ha sido la incorporación de nuevos modelos normativos de inducción causal conocidos como modelos de redes Bayesianas. Estos modelos Bayesianos surgen en principio en el ámbito de la inteligencia artificial y la filosofía con el fin de conseguir algoritmos que permitan detectar relaciones de causalidad de una manera lo más racional posible (Glymour, 2001; Pearl, 2000). Lo que en principio tratan de describir estos modelos es el modo óptimo en que deberían realizarse las inferencias causales en condiciones ideales. Sin embargo, algunos autores (p. ej., Gopnik, Glymour, Sobel, Schulz, Kushnir, y Danks, 2004) han dado un paso más y han supuesto que estos modelos proporcionan también una buena descripción de la forma en que los seres humanos realizan inferencias causales en su vida cotidiana. Es más, estudios recientes parecen sugerir que estos modelos también explican algunos fenómenos de condicionamiento en ratas que son muy difíciles de interpretar desde los modelos asociativos (Blaisdell, Sawa, Leising, y Waldmann, 2006). Por supuesto, esta hipótesis es discutible y cabe esperar que dé lugar a un intenso debate teórico y empírico a lo largo de los próximos años.

Una última tendencia que querríamos mencionar antes de terminar el presente capítulo es el renovado interés por algunas líneas de investigación que atrajeron una gran atención durante los 80 pero que habían caído aparentemente en el olvido hasta fechas muy recientes. Como hemos visto, gran parte de la investigación sobre aprendizaje causal realizada entre los años 60 y los 80 se realizó con procedimientos en los que los participantes debían inferir si su propia conducta (p. ej., pulsar un botón) tenía algún efecto a la hora de producir un resultado (p. ej., ganar un punto). Este tipo de tareas experimentales fue progresivamente abandonado en favor de otros procedimientos en los que el experimentador controla la presentación de los dos eventos de la relación causal, de modo que el participante debía inferir si existía una relación causal entre dos eventos que estaban más allá de su control (p. ej., inferir si determinada medicina puede provocar efectos secundarios no deseados). Actualmente, unos cuantos estudios comienzan a recuperar aquellas viejas tareas en las que los participantes debían ejecutar una respuesta y observar después si esta respuesta provocaba un resultado. Lo más interesante de estos estudios es que permiten estudiar las condiciones bajo las cuales los seres humanos perciben que, mediante sus

respuestas, pueden controlar eventos objetivamente incontrolables. Este fenómeno, comúnmente denominado ilusión de control (Alloy y Abramson, 1979; Langer, 1975), está recibiendo nuevas explicaciones, algunas de ellas más o menos compatibles con las interpretaciones asociativas más tradicionales (Matute, 1996; Msetfi y cols., 2005; Msetfi, Murphy, y Simpson, 2007) y otras formuladas en términos radicalmente novedosos (Allan y cols., 2007).

A MODO DE CONCLUSIÓN

Somos plenamente conscientes de que una introducción completa a las teorías y estudios actuales sobre aprendizaje causal requeriría un espacio mucho mayor del disponible en este capítulo y de que, por tanto, existen importantes lagunas en la exposición que aquí concluimos. Sin embargo, esperamos que estas páginas hayan permitido al lector atisbar al menos el curso que han seguido estos estudios en las últimas décadas. También nos gustaría que esta breve introducción sirviera para alejar de la opinión del lector todos los prejuicios que, como hemos mencionado al principio de este capítulo, se asocian con frecuencia a las teorías del aprendizaje. Lejos de ser una mera reliquia de formas antiguas de hacer ciencia psicológica, las investigaciones clásicas y modernas sobre aprendizaje contienen un valioso acervo de ideas aplicables a múltiples ámbitos, de los cuales la inducción causal es sólo un pequeño ejemplo.

AGRADECIMIENTOS

Este trabajo ha sido financiado por un proyecto de investigación del Ministerio de Educación y Ciencia (Ref. SEJ2007-63691/PSIC) y otro de la Junta de Andalucía (Ref. SEJ-406). Los autores desean dar las gracias a Fernando Blanco y Cristina Orgaz por sus valiosos comentarios a una versión previa de este capítulo.

REFERENCIAS

- Allan, L. G., y Jenkins, H. M. (1980). The judgment of contingency and the nature of response alternatives. *Canadian Journal of Psychology*, 34, pp. 1-11.
- Allan, L.G., y Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, 14, pp. 381-405.
- Allan, L. G., Siegel, S., y Hannah, S. (2007). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology*, 60, pp. 482-495.
- Alloy, L. B., y Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, pp. 441-485.
- Beckers, T., Miller, R. R., De Houwer, J., y Urushihara, K. (2006). Reasoning rats: Forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General*, 135, pp. 92-102.
- Blaisdell, A. P., Sawa, K., Leising, K. J., y Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, 311, pp. 1020-1022.
- Chatlosh, D. L., Neunaber, D. J., y Wasserman, E. A. (1985). Response-outcome contingency: Behavioral and judgmental effects of appetitive and aversive outcomes with college students. *Learning and Motivation*, 16, pp. 1-34.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, pp. 367-405.
- Cheng, P. W., y Novick, L.R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58, pp. 545-567.
- Cheng, P. W., y Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, pp. 365-382.
- Cobos, P. L., López, F. J., Caño, A., Almaraz, J., y Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, pp. 331-346.
- De Houwer, J., y Beckers, T. (2003). Secondary task difficulty modulates forward blocking in human contingency learning. *Quarterly Journal of Experimental Psychology*, 56B, pp. 345-357.
- De Houwer, J., Beckers, T., y Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, 55A, pp. 965-985.
- Dickinson, A., y Burke, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, 49B, pp. 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, pp. 29-50.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, pp. 380-417.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., y Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111, pp. 3-32.

Hume, D. (1964). *A treatise of human nature*. (L. A. Selby-Bigge, Ed.). London. Oxford University Press. (Trabajo original publicado en 1739.)

Jenkins, H. M., y Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, 79, pp. 1-17.

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.

Langer, E. J. (1975). Illusion of control. *Journal of Personality and Social Psychology*, 32, pp. 311-328.

Larkin, M. J. W., Aitken, M. R. F., y Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, pp. 1331-1352.

Lober, K., y Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, 107, pp. 195-212.

López, F. J., Almaraz, J., Fernández, P., y Shanks, D. R. (1999). Adquisición progresiva de conocimiento sobre relaciones predictivas: Curvas de aprendizaje en juicios de contingencia. *Psicothema*, 11, pp. 337-349.

López, F. J., Shanks, D. R., Almaraz, J., y 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, pp. 672-694.

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, pp. 133-142.

Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychological Science*, 7, pp. 289-293.

Matute, H. (2004). Investigación con humanos en aprendizaje asociativo. En R. Pellón y A. Huidobro (Eds.), *Inteligencia y aprendizaje* (pp. 261-304). Barcelona: Ariel.

Matute, H., Arcediano, F., y Miller, R. R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, pp. 182-196.

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, pp. 311-329.

Msetfi, R. M., Murphy, R. A., Simpson, J., y Kombrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: The effect of the context and intertrial interval. *Journal of Experimental Psychology: General*, 134, pp. 10-22.

Msetfi, R. M., Murphy, R. A., y Simpson, J. (2007). Depressive realism and the effect of inter-trial interval of judgments of zero, positive and negative contingencies. *Quarterly Journal of Experimental Psychology*, 60, pp. 461-481.

Papini, M. R., y Bitterman, M. E. (1990). The role of contingency in classical conditioning. *Psychological Review*, 97, pp. 396-403.

Pearl, J. (2000). *Causality*. New York: Oxford University Press.

Peterson, C. R., y Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68, pp. 29-46.

Rescorla, R. A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. *Psychonomic Science*, 4, pp. 383-384.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, pp. 1-5.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43, pp. 151-160.

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: Appellon-Century-Crofts.

Seligman, M. E. P. (1975). *Helplessness: On depression, development and death*. San Francisco: W. H. Freeman.

Shanks, D. R. (1985a). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition*, 13, pp. 158-167.

Shanks, D. R. (1985b). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, 37B, pp. 1-21.

Shanks, D. R. (1986). Selective attribution and the judgment of causality. *Learning and Motivation*, 17, pp. 311-334.

Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation*, 18, pp. 147-166.

Shanks, D. R., y Dickinson, A. (1987). Associative accounts of causality judgment. En G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 21 (pp. 229-261). San Diego, CA: Academic Press.

Shanks, D. R., y Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Memory & Cognition*, 19, pp. 353-360.

Shanks, D. R., y López, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory & Cognition*, 24, pp. 511-522.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, pp. 165-173.

Vadillo, M. A., Miller, R. R., y Matute, H. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior*, 33, pp. 172-183.

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, pp. 127-151.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, pp. 53-76.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*, 8, pp. 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, pp. 222-236.

Ward, W. C., y Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, 19, pp. 231-241.

Wasserman, E. A., Chatlosh, D. L., y Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. *Learning and Motivation*, 14, pp. 406-432.

Wu, M., y Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, 10, pp. 92-97.