

Aprendizaje predictivo y causal: ¿Evidencia de diferentes niveles? (Predictive and causal learning: Is there evidence that they are at different levels?)

LEYRE CASTRO, SONIA VEGAS Y HELENA MATUTE
Universidad de Deusto



El trabajo de Perales, Catena y Maldonado aborda un reto necesario y complicado. Muchos autores han insistido, desde perspectivas teóricas muy diferentes (p.ej., Cheng, 1997; Miller y Matute, 1996), en la necesidad de diferenciar entre aprendizaje predictivo y causal, pero por el momento no ha sido posible sistematizar de manera clara las relaciones entre ambos. Así, el intento de Perales y cols. por establecer un marco de referencia para el estudio riguroso y sistemático del aprendizaje predictivo y causal se convierte en una interesante y atractiva propuesta que permite impulsar el desarrollo de nuevas investigaciones en este campo.

Consideramos, no obstante, que muchas de las afirmaciones que los autores dan por probadas no lo son tanto, y reflejan únicamente uno de los lados de la polémica existente en la literatura actual. De los dos grandes bloques teóricos que existen en la actualidad, el modelo asociativo y el modelo normativo, los autores parecen decantarse por este último obviando el hecho de que muchos de los efectos que aquí se discuten se explican mejor desde una perspectiva asociativa. Tal y como ya han mencionado López, Cobos, Caño y Shanks (1998), mientras los modelos normativos indican cuál debería ser el resultado final del aprendizaje, los modelos asociativos explican mejor cuál es en realidad el proceso por el que se produce el aprendizaje, incluyendo los sesgos y errores que se observan a menudo en el aprendizaje humano y animal, y que no pueden ser explicados por los modelos normativos. Por tanto, a pesar del valor heurístico que tiene sin duda la propuesta de Perales y cols., se hace necesario mencionar que muchas de sus afirmaciones teóricas no tienen suficiente justificación empírica. En este comentario intentaremos señalar aquellos aspectos del artículo que consideramos más problemáticos a la luz de los datos existentes en la actualidad.

UNIDIRECCIONALIDAD VS. BIDIRECCIONALIDAD

Perales y cols. presentan un modelo en el que las relaciones de contingencia, de predicción y de causalidad están organizadas jerárquicamente en función de la

Agradecimientos: Este trabajo ha sido realizado en el marco del proyecto de investigación PI-2000-12 del Departamento de Educación, Universidades e Investigación del Gobierno Vasco. Durante la realización de este trabajo L.C. disfrutaba de una beca FPU del Ministerio de Educación, Cultura y Deporte (Ref. AP99, 14605555), y S.V. de una beca FPI del Departamento de Educación Universidades e Investigación del Gobierno Vasco (Ref. BFI00.138).

Correspondencia con los autores: Helena Matute. Departamento de Psicología. Universidad de Deusto. Apartado 1. 48080 Bilbao. E-mail: matute@orion.deusto.es

dirección en que se produce el aprendizaje. Dentro de ese marco, los autores establecen que el aprendizaje de las relaciones predictivas se caracteriza por su unidireccionalidad, del evento predictor al evento predicho. Según Perales y cols., no sería posible aprender del evento predicho al predictor o, por ejemplo, dentro del condicionamiento clásico, no sería posible aprender una asociación que fuera del estímulo incondicionado (EI) al estímulo condicionado (EC). Esto plantea diversos problemas empíricos.

En primer lugar, diferentes estudios con animales han demostrado que los sujetos pueden aprender asociaciones EC-EI y EI-EC independientemente del orden temporal en que se presentan los estímulos, aunque, como es lógico, para que se dé una *respuesta condicionada* (RC) es necesario que el EC prediga el EI durante la fase de prueba (Barnet, Arnold y Miller, 1991; Matzel, Held y Miller, 1988; Miller y Barnet, 1993). Así, según los trabajos de Miller y sus colaboradores, no habría tanto una dificultad por establecer un aprendizaje cuando el EC se presenta después del EI, tal y como Perales y cols. concluyen, sino que sería la respuesta condicionada, y no el aprendizaje de la relación, lo que dependería del orden temporal. Por tanto, o bien el condicionamiento clásico no entra dentro del nivel de representación de las relaciones predictivas, que es donde lo sitúan Perales y cols., sino dentro del nivel de las relaciones causales, o bien el aprendizaje de las relaciones predictivas no es unidireccional.

En segundo lugar, también en aprendizaje predictivo con humanos hay evidencia de bidireccionalidad. Partiendo de que la direccionalidad puede no ser inherente al proceso de aprendizaje sino al proceso de respuesta, Gerolin y Matute (1999) presentaron a los sujetos un color (clave) seguido de una determinada figura (consecuencia). Al observar después la figura en una fase posterior de prueba los sujetos eran capaces de responder cuál era el color predictor. Por tanto, durante un experimento de aprendizaje predictivo, los sujetos pueden adquirir asociaciones bidireccionales y utilizar la información tanto en una dirección como en otra. Asimismo, Arcediano, Escobar y Miller (2000) obtuvieron en una preparación conductual también con humanos evidencia de que el condicionamiento hacia atrás daba lugar a la formación de asociaciones temporales hacia atrás; es decir, asociaciones consecuencia → clave cuando durante el entrenamiento sólo se habían presentado emparejamientos en la dirección clave → consecuencia.

Por otro lado, Perales y cols. sostienen que la disociación entre el orden temporal y el orden causal es una característica exclusiva de las relaciones causales. Según estos autores, el aprendizaje causal es sensible a la manipulación de la dirección causal, es decir, se aprende siempre en dirección causa → efecto, independientemente de que las causas (C) se presenten antes que los efectos (E), o los efectos antes que las causas. Esta afirmación la apoyan en dos supuestos. En primer lugar, los autores apoyan esta afirmación en una serie de experimentos en los que no se han observado efectos de competición de claves cuando la clave objetivo y la clave competidora se definen como posibles efectos de una causa común (Van Hamme, Kao y Wasserman, 1993; Waldmann y Holyoak, 1992, Exp. 1 y 3; Waldmann, 2000). Sin embargo, y a pesar de estos resultados fallidos, es sabido que muchos otros estudios han mostrado que es posible obtener competición entre efectos de una causa común, tanto con animales (Esmoris-Arranz, Miller y Matute, 1997; Miller y Matute, 1998) como con humanos (Chapman, 1991; Matute, Arcediano y Miller, 1996; Price y Yates, 1993, 1995; Shanks, 1991; Shanks y López, 1996; Cobos, Caño, López, Luque y Almaraz, 2000; Waldmann y Holyoak, 1992, Exp. 2); y tanto cuando los efectos toman el papel de claves, como cuando toman el papel de resultados, o incluso cuando son presentados al

mismo tiempo que la posible causa. No parece, por tanto, que pueda concluirse de los estudios que no han hallado competición entre efectos, que ésta no pueda tener lugar.

En segundo lugar, Perales y cols. consideran como evidencia de la disociación entre el orden causal y el orden temporal en el aprendizaje de relaciones causales el que éste es sensible a que los efectos de una causa común tienden a correlacionar, pero que no tienden a correlacionar las causas de un efecto común. A este respecto, la prueba empírica aportada es una serie de experimentos recientes no publicados de Perales, Catena y Maldonado (2001). En el Experimento 1, dos sustancias químicas se presentan, en fases separadas, como efectos de una enfermedad común ($E1 \rightarrow C$ en la primera fase; $E2 \rightarrow C$ en la segunda fase), y los resultados muestran correlación entre ambos efectos. En el Experimento 2, dos sustancias químicas se presentan como causas de una enfermedad común ($C1 \rightarrow E$ en la primera fase; $C2 \rightarrow E$ en la segunda fase), y en este caso no se observa correlación entre ambas causas. El Experimento 3 integra las condiciones de los dos anteriores y cambia ligeramente el escenario causal: para la mitad de los sujetos dos microorganismos ficticios se presentan como posibles causas de la aparición de cierta sustancia química en el aire de una ciudad ($C1 \rightarrow E$ en la primera fase; $C2 \rightarrow E$ en la segunda fase); para la otra mitad, los dos microorganismos se describen como efectos de la sustancia química ($E1 \rightarrow C$ en la primera fase; $E2 \rightarrow C$ en la segunda fase). Los resultados replican lo anterior. Es decir, se encuentra correlación entre los dos microorganismos cuando estos se describen como efectos de la sustancia química pero no cuando se describen como sus causas. El resultado, apoya, por tanto, la afirmación de que hay correlación entre efectos pero no entre causas. Sin embargo, el efecto descrito pudiera ser debido, no a diferencias conceptuales entre causas y efectos, sino al escenario causal utilizado, el cual activa cierta experiencia previa respecto a cómo determinados efectos correlacionan entre sí y determinadas causas no lo hacen. ¿Qué hubiera pasado en un escenario causal muy distinto en el que las causas no fueran independientes sino que pudiera haber correlación entre ellas? Por ejemplo, si depresión y ansiedad fueran presentados como causa del bajo rendimiento laboral de una persona, probablemente se observaría correlación entre depresión y ansiedad. En cualquier caso, esta sería una cuestión que, sin duda, requiere de una mayor aportación de evidencia empírica.

Otro problema que observamos en la afirmación de Perales y cols. de que siempre se aprende en la dirección causa \rightarrow efecto independientemente del orden temporal de los estímulos proviene de estudios que muestran que los sujetos experimentales pueden aprender ciertas asociaciones sin tener conciencia o estando confundidos respecto a lo que es causa y lo que es efecto. Así, cuando los efectos se conocen antes que las causas, las personas tienden a confundir la probabilidad del efecto dada la causa, $p(E|C)$, con la probabilidad de la causa dado el efecto, $p(C|E)$, puesto que saben que las causas ocurren antes que los efectos (Einhorn y Hogarth, 1996). Por ejemplo, un resultado positivo en una mamografía predice la existencia de un cáncer. La prueba se utiliza para diagnosticar la enfermedad, aunque es la existencia de la enfermedad la que causa el resultado positivo en la prueba. La causa precede al efecto, pero el efecto es conocido antes y se utiliza para diagnosticar la causa. En un caso como éste, los sujetos confunden la probabilidad de tener cáncer habiendo dado un resultado positivo en la mamografía, con la probabilidad de dar un resultado positivo en la prueba teniendo cáncer (Eddy, 1982). Por tanto, los humanos no siempre mostrarían la capacidad de disociar el orden temporal del orden causal.

En conjunto, creemos que la evidencia disponible hasta la fecha no permite diferenciar entre relaciones predictivas y causales en función de su direccionalidad. Nuestra interpretación de la literatura en este punto difiere bastante de la interpretación de Perales y cols.: tanto en el aprendizaje predictivo como en el aprendizaje causal la relación entre los eventos *ocurre* en la naturaleza siempre en una misma dirección (causa → efecto; evento predictor → evento predicho), pero en ambos casos es también posible presentar la información en sentido contrario en una manipulación experimental, y en ambos casos pueden los sujetos utilizar después la información aprendida de manera flexible, tanto hacia delante como hacia atrás. También en ambos casos puede ocurrir competición entre claves y entre resultados, independientemente de si estos se definen como causas o efectos, o como eventos predictores o eventos predichos. Por tanto, a diferencia de Perales y cols., concluiríamos que, independientemente del orden en el que se presenten los estímulos y de cómo se interpreten, los sujetos pueden utilizar la información en ambas direcciones, en función de las demandas de la tarea. No parece ser ésta, desde nuestro punto de vista, una diferencia que permita discriminar entre aprendizaje predictivo y causal.

FUNCIONALIDAD DEL APRENDIZAJE DE RELACIONES CAUSALES

Coincidimos con Perales y cols. en que es importante para la supervivencia de cualquier animal diferenciar entre relaciones predictivas y relaciones causales. Sin embargo, no creemos que el valor adaptativo de la distinción entre relaciones de predicción y relaciones causales se desprenda necesariamente de los argumentos que Perales y cols. utilizan. En primer lugar, señalan que discriminar entre relaciones causales y de predicción es funcional si se pretende generalizar lo aprendido a otros contextos, puesto que las relaciones causales son estables a través de diferentes contextos y las predictivas no lo son. Sin embargo, al afirmar esto no se tiene en cuenta que determinadas causas pueden producir determinados efectos en un contexto y no en otro: por ejemplo, una cerilla prenderá en un ambiente con oxígeno, pero no cuando éste falte; ni que los cambios de contexto pueden llevar a que causas alternativas resulten más salientes, perdiendo así efectividad la causa objetivo (Mackie, 1974): si alguien en su casa bebe un vaso de leche y vomita, atribuiremos el vómito al mal estado de la leche; pero si esto le ocurre a una persona hospitalizada, atribuiremos probablemente el vómito a las condiciones de esa persona, y no al mal estado de la leche. Asimismo, hay abundantes datos en la literatura animal de que el contexto afecta a la manifestación de las asociaciones tanto predictivas (condicionamiento clásico) como causales (condicionamiento instrumental) (p. ej., Bouton, 1993; Spear, 1971) y estos datos han sido también replicados en tareas predictivas y causales con humanos (p. ej., Pineño y Matute, en prensa; Rosas, Vila, Lugo y López, 2001). Por tanto, no parece ser ésta una diferencia que justifique el mayor valor adaptativo del aprendizaje de relaciones causales frente al de relaciones predictivas.

Otra ventaja adaptativa que proponen Perales y cols. sería que el aprendizaje de relaciones causales permite manipular y controlar el medio, lo cual no lo permiten las relaciones predictivas. Si bien esto es cierto, podría resultar cuestionable si nos introducimos en el tipo de relaciones que Perales y cols. consideran predictivas. Los autores parecen asumir que en el aprendizaje animal no hay aprendizaje de relaciones causales, sino tan sólo aprendizaje de relaciones predictivas. Es cierto que la RC en el condicionamiento clásico sólo nos informa de que

el animal es capaz de predecir la ocurrencia del EI, y la predicción no implica causación. Sin embargo, en el condicionamiento instrumental el animal hace algo más que predecir el reforzador: da una respuesta que le permite obtenerlo y, por tanto, controlarlo. Además, no sería tan descabellado sugerir que si los humanos somos capaces de aprender relaciones causales, los animales probablemente tengan también esa capacidad, ya que tanto animales como humanos hemos afrontado los mismos problemas a lo largo de la evolución y hemos estado sometidos a las mismas exigencias por parte del entorno: predecir y controlar. En cualquier caso, como señalan Miller y Matute (1996) la cuestión de si los sujetos expuestos a un proceso de condicionamiento clásico aprenden simplemente que el EC predice el EI, o aprenden que el EC causa el EI, es una cuestión empírica. Por ejemplo, si los sujetos intentaran que ocurriera el EC para obtener el EI (asumiendo que la valencia del EI es positiva) en una prueba instrumental posterior, esto sería un indicativo de aprendizaje causal. En esta línea, un estudio de Killeen (1981) mostró que una paloma en una situación instrumental es capaz de distinguir entre la aparición de una luz producida por su propio comportamiento de la aparición de la luz producida por otros factores. Killeen interpretó estas observaciones como indicativo de que las palomas tenían un conocimiento de qué era lo que causaba la aparición de la luz. También podría argumentarse que las palomas desarrollaban ciertas estrategias comportamentales que les permitían resolver la tarea sin inferir causalidad; sin embargo, si este fuera el caso, también podría defenderse que la causalidad humana no es sino un reflejo de la aplicación de ese mismo tipo de estrategias (Young, 1995).

De todo lo aquí comentado, la conclusión más importante es que estamos tratando con cuestiones todavía no resueltas. El valor heurístico contenido en la propuesta de Perales y cols. se revela como una opción fundamental para impulsar el avance del conocimiento sobre el aprendizaje de las relaciones entre los acontecimientos, pero la propuesta concreta está aún lejos de ser avalada por los datos existentes y tendrá que ser la investigación de los próximos años la que determine la validez del modelo propuesto.

Referencias

- ARCEDIANO, F., ESCOBAR, M. y MILLER, R. R. (2000). Codificación temporal, asociaciones hacia atrás, y bidireccionalidad de las asociaciones en humanos. En *Libro de Resúmenes del XII Congreso de la Sociedad Española de Psicología Comparada (84)*. Granada: Universidad de Granada.
- BARNET, R. C., ARNOLD, H. M. y MILLER, R. R. (1991). Simultaneous conditioning demonstrated in second-order conditioning: Evidence for similar associative structure in forward and simultaneous conditioning. *Learning and Motivation*, 22, 253-268.
- BOUTON, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, 114, 80-99.
- CHAPMAN, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 837-854.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405.
- COBOS, P. L., CAÑO, A., LÓPEZ, F. J., LUQUE, J. L. y ALMARAZ, J. (2000). Does the type of judgement required modulate cue competition? *Quarterly Journal of Experimental Psychology*, 53B, 193-207.
- EDDY, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). Nueva York: Cambridge University Press.
- EINHORN, H. J. y HOGARTH, R. M. (1996). Judging probable cause. *Psychological Bulletin*, 99, 3-19.
- ESMORÍS-ARRANZ, F. J., MILLER, R. R. y MATUTE, H. (1997). Blocking of antecedent and subsequent events. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 145-156.
- GEROLIN, M. y MATUTE, H. (1999). Bidirectional associations. *Animal Learning & Behavior*, 27, 42-49.

- KILLEEN, R. R. (1981). Learning as causal inference. En M. L. Commons y J. A. Nevin (Eds.), *Quantitative analyses of behavior (Vol. 1): Discriminative properties of reinforcement schedules* (pp. 89-112). Cambridge, MA: Ballinger.
- LÓPEZ, F. J., COBOS, P. L., CAÑO, A. y SHANKS, D. R. (1998). The rational analysis of human causal and probability judgment. En M. Oaksford y N. Chater (Eds.), *Rational models of cognition* (pp. 314-352). Oxford: Oxford University Press.
- MACKIE, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.
- MATUTE, H., ARCEDIANO, F. y MILLER, R. R. (1996). Test Question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 182-196.
- MATZEL, L. D., HELD, F. P. y MILLER, R. R. (1988). Information and expression of simultaneous and backward associations: Implications for contiguity theory. *Learning and Motivation*, 19, 317-344.
- MILLER, R. R. y BARNET, R. C. (1993). The role of time in elementary associations. *Current Directions in Psychological Science*, 2, 106-111.
- MILLER, R. R. y MATUTE, H. (1996). Animal analogues of causal judgment. En D. R. Shanks, K. J. Holyoak y D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning* (pp. 133-166). San Diego, CA: Academic Press.
- MILLER, R. R. y MATUTE, H. (1998). Competition between outcomes. *Psychological Science*, 9, 146-149.
- PERALES, J. C., CATENA, A. y MALDONADO, A. (2001). *Outcome mediated contingency learning is sensitive to causal directionality*. Manuscrito enviado para publicación.
- PINEÑO, O. y MATUTE, H. (En prensa). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology*, 13.
- PRICE, P. C. y YATES, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, 21, 561-572.
- PRICE, P. C. y YATES, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1639-1655.
- ROSAS, J. M., VILA, J. N., LUGO, M. y LÓPEZ, L. (2001). Combined effect of context change and retention interval upon proactive interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 153-164.
- SHANKS, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 433-443.
- SHANKS, D. R., y LÓPEZ, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory & Cognition*, 24, 511-522.
- SPEAR, N. E. (1971). Forgetting as retrieval failure. En W. K. Honig y P. H. R. James (Eds.), *Animal memory* (pp. 45-199). Nueva York: Academic Press.
- VAN HAMME, L. J., KAO, S-F. y WASSERMAN, E. A. (1993). Judging interevent relations: From cause to effect and from effect to cause. *Memory & Cognition*, 21, 802-808.
- WALDMANN, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 53-76.
- WALDMANN, M. R. y HOLYOAK, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222-236.
- YOUNG, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review*, 2, 83-104.