

# Inter-relaciones entre los niveles computacional y algorítmico (Inter-relations between computational and algorithmic levels)

ANTONIO CAÑO, FRANCISCO J. LÓPEZ, PEDRO L. COBOS  
Y JULIÁN ÁLMARAZ  
*Universidad de Málaga*



*.... y como éstos {semejanza, contigüidad y causación} son los únicos lazos de nuestros pensamientos, son ellos para nosotros el auténtico cemento del universo, y todas las operaciones de la mente deben, en gran medida, depender de ellos.*

David Hume (1888/1992)

El estudio de los procesos de inferencia causal basada en la observación de acontecimientos ha ocupado históricamente un lugar destacado en la literatura sobre los procesos cognitivos básicos. El artículo de Perales y cols. viene a ofrecernos una visión teórica actualizada de un aspecto de ese estudio, esencial para entender cómo se llevan a cabo estos procesos. Nos referimos al análisis computacional (Marr, 1982), es decir, un análisis de qué tipo de información resulta relevante en el procesamiento causal y por qué ese tipo de información es relevante.

El análisis computacional que realizan los autores está inspirado, casi totalmente, en consideraciones de tipo normativo o racionales (ver Cheng, 1997, 2000 para una posición similar). Sin embargo, hay que reseñar que la aproximación racional que realizan de la caracterización de las relaciones causales, siguiendo a Cheng, es muy parcial. Está basada en criterios exclusivamente estadísticos, obviándose una serie de principios al menos igualmente definitorios de causalidad. Por ejemplo, en un análisis racional de la inferencia causal, Anderson (1991) menciona la importancia de la contigüidad espacial y temporal y de la similitud entre causas y efectos. Es importante subrayar que el aprendizaje asociativo es sensible a estos principios. Además, la similitud entre causas y efectos puede considerarse un caso particular de un principio más general necesario para caracterizar relaciones causales: la relación entre la naturaleza de los acontecimientos. Incluso manteniendo la misma relación estadística con un efecto, dos candidatos causales tienen diferente probabilidad de generarlo, dependiendo de su naturaleza (es razonable que un sabor sea más fácilmente causa del dolor de

---

*Agradecimientos:* Las investigaciones descritas en el presente trabajo han sido financiadas por la Junta de Andalucía (Grupo HUM0105) y por el Ministerio de Ciencia y Tecnología (Proyecto BSO2000-1216).

*Correspondencia con los autores:* Dr. Antonio Caño González. Departamento de Psicología Básica. Facultad de Psicología. Campus de Teatinos. Universidad de Málaga. Málaga 29071. E-mail: canyo@uma.es

estómago que lo sea una descarga). Nuevamente, el aprendizaje asociativo es sensible a este principio, que incluso recibe el nombre de *relevancia causal*. Dado que el aprendizaje asociativo es sensible a los principios mencionados, incluyendo la relación estadística entre estímulos (Mackintosh, 1983), parece que el aprendizaje asociativo permite captar razonablemente bien la esencia de las relaciones causales. Nuestra postura es que un análisis computacional sobre relaciones de causalidad podría beneficiarse de la consideración de los principios que rigen el aprendizaje asociativo.

Ciertamente, los autores realizan algunas otras consideraciones relevantes para el análisis computacional más allá del análisis normativo (p. ej., conocimientos previos de los individuos sobre el rol causal de los acontecimientos que intervienen en la relación), pero en la práctica estos otros aspectos quedan excluidos del análisis.

Con todo, el análisis computacional que Perales y cols. realizan es un buen punto de partida para conocer cuáles son los objetivos computacionales que el sistema cognitivo debe satisfacer a la hora de realizar inferencias causales. Como señalan los autores, es de esperar, desde un punto de vista evolutivo, que nuestra capacidad de procesamiento favorezca una adaptación eficaz a nuestro medio, y dicha adaptación parece pasar, en parte al menos, por distinguir nexos verdaderamente causales de correlaciones espurias entre acontecimientos. Así, en adelante daremos por bueno el análisis computacional-racional-normativo realizado desde esta perspectiva estadística para centrarnos en otros aspectos del análisis realizado.

Incluso aceptando el punto de vista exclusivamente estadístico del análisis racional, el hecho de que las consideraciones normativas expuestas por los autores constituyan un buen punto de partida del análisis computacional no significa que sean suficientes. La exposición de esa insuficiencia y el análisis de algunas de sus implicaciones, en particular la relación del análisis computacional con el nivel de explicación algorítmico, constituirán el objetivo central del presente comentario (véase López, Cobos, Caño y Shanks, 1998a para una exposición más detallada). Si bien, como ya hemos mencionado, los autores reconocen la existencia de factores ajenos al análisis normativo que realizan, su análisis computacional es independiente de restricciones procedentes de los propios mecanismos psicológicos (nivel algorítmico) que presumiblemente son evocados en tareas de aprendizaje causal.

Considerar estas inter-relaciones entre diferentes niveles de análisis o explicación (como sugería el propio Marr, 1982) es una tarea necesaria si se quiere ofrecer un panorama completo e integrador (como reza en el título del artículo diana) de cuáles son los objetivos computacionales que el sistema cognitivo satisface en situaciones de inferencia causal.

Las consecuencias más dramáticas de la desconsideración de estas inter-relaciones las podemos ver cuando las inferencias causales que realizan los individuos se alejan del análisis normativo, es decir, cuando se producen sesgos o errores en las inferencias causales. En estas circunstancias, el análisis computacional (normativo) no nos permite conocer cuáles son los objetivos que el sistema cognitivo satisface con su comportamiento, por lo que deja de ser útil para la explicación de aquél. Así, no es de extrañar, como reconocen Perales y cols., que aquellos autores que mantienen una desconsideración equivalente (p. ej., Cheng, 1997, 2000) tiendan a ignorar o rechazar la existencia de tales sesgos.

La existencia de una determinada desviación en relación con un determinado análisis normativo no ha de interpretarse necesariamente como un sesgo o error. Es posible que dicha desviación no sea tal si se tiene en cuenta un nuevo y más

sofisticado análisis normativo. Un ejemplo de esta posibilidad lo tenemos en lo que hasta la aparición de la Teoría de la Potencia Causal (Cheng, 1997) se consideraba un sesgo de las estimaciones causales, a saber, el peso diferencial otorgado a las diferentes casillas de la matriz de contingencias  $2 \times 2$  (p. ej., Wasserman, Dorner y Kao, 1990). Resultó que aquello que se consideraba un sesgo o error de acuerdo con el análisis normativo de referencia (el Modelo de Contraste Probabilístico de Cheng y Novick, 1992), en realidad no era más que una consecuencia natural de haber realizado estimaciones causales en perfecto acuerdo con un análisis normativo más sofisticado, la Teoría de la Potencia Causal.

No obstante, existen otras ocasiones en las que un determinado sesgo o desviación de las inferencias causales no parece que venga a satisfacer ningún tipo de requerimiento normativo. Estas ocasiones son, sin duda, aquellas en las que se hace especialmente relevante atender a las restricciones que la arquitectura del sistema impone. Una buena forma de hacerlo es atendiendo al comportamiento, caracterizando sus regularidades, y derivando la función que el comportamiento pueda desempeñar. Esto no quiere decir que el sujeto realice tal función de forma óptima, pero sirve para identificar la función, que es en definitiva la tarea del análisis computacional. Que el comportamiento del sujeto sea funcional, aunque no necesariamente óptimo, y que al tiempo obedezca a las restricciones que la arquitectura del sistema impone no parece una aseveración arriesgada. Después de todo, desde un punto de vista evolutivo, es de esperar que nuestra adaptación haya servido para desarrollar un sistema cognitivo que satisfaga nuestras necesidades de supervivencia (p. ej., inferir relaciones causales) contando con las restricciones del organismo biológico en que se implementa. Así, el estudio de las desviaciones o sesgos en la realización de inferencias causales puede servir para conocer objetivos computacionales que el análisis normativo no contempla. En este sentido, entendemos que el nivel de explicación algorítmico, tal y como es definido por Marr, puede resultar muy útil para conocer los objetivos computacionales del sistema concreto que es nuestro organismo biológico, al resultar un compromiso entre el nivel computacional (que atiende exclusivamente a la función) y el nivel implementacional (que atiende a las restricciones del sistema). Por tanto, también en este sentido la aproximación computacional de Perales y cols. nos parece insuficiente, al restringir el análisis computacional al análisis normativo.

En lo que sigue, describiremos algunos ejemplos a través de los que se ilustrará la utilidad de considerar la inter-relación entre niveles (computacional y algorítmico) para conocer cuáles son los objetivos computacionales del sistema, precisamente en aquellas ocasiones en las que el comportamiento de los individuos resulta sesgado. Comenzaremos con los efectos de orden de los ensayos. López, Shanks, Almaraz y Fernández (1998b) mostraron que aunque las relaciones estadísticas programadas entre claves y resultados a lo largo de una tarea no variasen, si el orden de presentación de los ensayos variaba (en cada ensayo se informa al participante de la/s clave/s presentes y el resultado apropiado), las inferencias sobre el valor predictivo de las claves eran sensibles a dicho orden de presentación.

De acuerdo con un análisis normativo (ya sea el Modelo de Contraste Probabilístico o Teoría de la Potencia Causal), resulta difícil entender el origen de esta sensibilidad al orden de presentación de los ensayos pues dicho orden no altera las relaciones estadísticas programadas entre claves y resultados. Estos resultados parecen, por tanto, ir más allá de los objetivos computacionales derivados de esta clase de análisis normativo. A pesar de ello, si atendemos a los procesos psicológicos concretos que explican, al menos en parte, el comportamiento de los indi-

viduos en este tipo de experimentos, podremos entender el origen de esta sensibilidad y a qué clase de objetivos computacionales pueden estar sirviendo dichos procesos.

Los mecanismos asociativos de aprendizaje han sido propuestos en la literatura (p. ej., Dickinson, Shanks y Evenden, 1984 o véase Shanks, Holyoak y Medin, 1996 para una revisión más reciente) como explicación del comportamiento inferencial de los individuos en este tipo de tareas. Los participantes (y por ende, los mecanismos asociativos de aprendizaje) se han de enfrentar con el problema de determinar ensayo a ensayo el resultado correcto a partir de la/s clave/s presentes. Si los ensayos que constituyen la tarea de aprendizaje no están perfectamente entremezclados (como ocurría en López y cols., 1998b), una actuación óptima sólo puede conseguirse a través de un sistema que sea capaz de actualizar su conocimiento sobre el valor predictivo de una clave de acuerdo con los cambios que en el medio se produzcan en torno a dicho valor predictivo. Los mecanismos asociativos de aprendizaje cumplen precisamente con esta clase de requisito o de objetivo de actualización del conocimiento [de hecho, tradicionalmente estos modelos han cumplido *en exceso* con este requisito, como muestra el fenómeno de la interferencia catastrófica (McCloskey y Cohen, 1989). No obstante, existen algoritmos asociativos de aprendizaje con los que no se produce dicho problema sin que deje de satisfacerse el objetivo de adaptación flexible a cambios en el medio (ver Lewandowsky, 1994 para una revisión)].

Un segundo ejemplo del modo en que el conocimiento sobre los algoritmos nos puede ayudar a entender los objetivos computacionales que el individuo satisface lo tenemos en aquellos experimentos en los que se manipula el orden causal de la tarea (orden predictivo: causa-efecto; orden diagnóstico: efecto- causa). Contrariamente a las consideraciones que Perales y cols. realizan sobre este asunto, sí existe evidencia empírica que muestra que los juicios inferenciales de los participantes son esencialmente simétricos en situaciones predictivas y diagnósticas en relación con manipulaciones de la competencia entre claves. Es decir, fenómenos equivalentes de competencia entre claves se producen cuando los participantes realizan inferencias predictivas e inferencias diagnósticas (Caño, Cobos, López, Luque y Almaraz, 1997; Cobos, Caño, López, Luque y Almaraz, 2000; Cobos, López, Caño, Almaraz y Shanks, 2001; Matute, Arcediano y Miller, 1996; Price y Yates, 1995; Shanks y López, 1996; Waldmann, en prensa, Experimento 1; Waldmann y Holyoak, 1992, Experimento 2).

Desde el punto de vista de un análisis computacional inspirado exclusivamente en consideraciones de carácter normativo, dichas simetrías constituyen sesgos o desviaciones de dichas consideraciones (ver p. ej., Waldmann, 2000 ó Waldmann y Holyoak, 1992 para una explicación). El conocimiento previo de los individuos sobre el hecho de que las causas originan los efectos y no viceversa debería llevarlos a producir juicios inferenciales sensibles a determinadas manipulaciones de la competencia entre claves cuando dichas claves desempeñan el papel de causas (tareas predictivas o causa-efecto) pero no cuando las claves desempeñan el papel de efectos (tareas diagnósticas o efecto- causa).

Este comportamiento sesgado es predecible, sin embargo, si asumimos de nuevo la operación de mecanismos asociativos de aprendizaje. Si la salida (*output*) de estos mecanismos de aprendizaje es el factor determinante de los juicios inferenciales de los individuos, los juicios predictivos y diagnósticos serán igualmente sensibles a las manipulaciones de la competencia entre claves. Los mecanismos asociativos de aprendizaje son exclusivamente sensibles al orden temporal en el que el individuo conoce los acontecimientos sobre los que se define la relación (las claves se conocen antes y sirven para anticipar la presen-

cia del resultado), pero no así al orden causal. En otras palabras, los algoritmos asociativos de aprendizaje son *cognitivamente impenetrables* por el conocimiento que los individuos pueden tener sobre la interpretación causal de claves y resultados. De ahí que podamos esperar un comportamiento fundamentalmente simétrico en situaciones predictivas y diagnósticas a partir de la operación de esta clase de algoritmos.

El hecho de que existan datos empíricos a favor de la asimetría (Van Hamme, Kao y Wasserman, 1993; Waldmann y Holyoak, 1992, Experimentos 1 y 3; Waldmann, 2000) no resta importancia al argumento que mantenemos, más bien todo lo contrario. La existencia de asimetrías viene a mostrar que los individuos en determinadas circunstancias son capaces de mostrar su competencia causal. La cuestión está en saber cuáles son las consecuencias que esas condiciones procedimentales conllevan desde el punto de vista de los mecanismos psicológicos (nivel algorítmico) que se ponen en marcha en una u otra circunstancia. Es decir, de nuevo nos permitimos subrayar la importancia de los algoritmos en la medida en que sólo hipótesis teóricas definidas en este nivel de análisis nos permitirán conocer *a priori* cuáles son las condiciones en las que se activarán procesos psicológicos sensibles al rol causal desempeñado por claves y resultados y en qué circunstancias no se activarán (ver a continuación algunas explicaciones tentativas sobre cuáles pueden ser estas circunstancias o Perales y cols. para sugerencias similares).

A diferencia de lo analizado con ocasión de la sensibilidad al orden de los ensayos, en esta ocasión no es evidente el objetivo computacional que viene a satisfacer el sesgo descrito. Sin embargo, a partir del contraste entre algunos detalles procedimentales de los experimentos realizados en nuestro laboratorio y los realizados en otros laboratorios con resultados opuestos en relación a la mencionada asimetría entre inferencias predictivas y diagnósticas, podremos entender algunas razones para la activación de mecanismos de aprendizaje insensibles al rol causal desempeñado por claves y resultados, y para la no intervención de mecanismos de aprendizaje sensibles a este conocimiento. También, para finalizar, ofreceremos algunas razones por las que el comportamiento de los participantes, si bien no es óptimo o normativo, resulta adaptativo.

En estos experimentos, los participantes se enfrentan a una doble tarea. En primer lugar, a través de una serie de ensayos, deben aprender a acertar cuál es el resultado correcto a partir de la/s clave/s presentes en el ensayo en cuestión. Esta descripción es válida tanto para la tarea predictiva como para la diagnóstica. En segundo lugar, una vez terminada esta primera parte, los participantes se enfrentan a una tarea en la que deben emitir juicios inferenciales ya sean predictivos o diagnósticos. El contenido concreto de estos juicios inferenciales hace referencia a las claves y resultados que formaron parte de la tarea realizada en primer lugar. La operación de un algoritmo asociativo de aprendizaje asegura un comportamiento óptimo de los participantes durante la primera tarea. Sin embargo, la base de conocimiento adquirida a través de este tipo de algoritmos no contiene los elementos necesarios para responder adecuadamente (normativamente) a la tarea de los juicios inferenciales. Nuestra propuesta concreta sobre cómo los participantes han resuelto la doble tarea experimental es suponer que se han activado exclusivamente mecanismos asociativos de aprendizaje. Existen algunas razones para pensar que se han decantado por una opción no óptima, normativamente inadecuada para la segunda tarea, pero sí adaptativa si tenemos en cuenta algunas consideraciones sobre la tarea y sobre nuestro sistema cognitivo.

La primera consideración tiene que ver con la complejidad de la primera tarea. Típicamente en los experimentos en los que se muestra la simetría predic-

tivo-diagnóstico se emplean tareas complejas (p. ej., tareas en las que se han de establecer múltiples relaciones entre diversas claves y resultados).

La segunda consideración tiene que ver con el alto coste, cognitivamente hablando, que resultaría de la intervención de mecanismos de aprendizaje más elaborados (Evans y Over, 1996; Sloman, 1996), capaces de mostrar sensibilidad al tipo de conocimiento previo que requeriría un comportamiento normativo durante la segunda tarea. Muy a menudo (probablemente casi siempre) no disponemos del tiempo o de los recursos cognitivos necesarios para usar estos mecanismos más elaborados a la hora de aprender relaciones entre acontecimientos o de predecir acontecimientos relevantes para la supervivencia. En estos casos, los mecanismos asociativos de aprendizaje pueden ser lo suficientemente rápidos, poco costosos y fiables como para proporcionar respuestas al menos útiles (véase López y cols., 1998a para una exposición más detallada).

Como decía Hume (1777/1993):

...ni puede una operación de tan inmensas consecuencias para la vida, como esa de inferir efectos a partir de causas, ser confiada a los inciertos procesos del razonamiento y la argumentación (p. 71).

Como conclusión, hemos intentado a lo largo de estas páginas mostrar, a través de dos ejemplos, las limitaciones inherentes a la perspectiva mantenida por Perales y cols. de desconsiderar las inter-relaciones entre los niveles computacional y algorítmico. Hemos mantenido que no se pueden conocer de manera exhaustiva cuáles son los objetivos computacionales que el sistema cognitivo humano viene a satisfacer sin tener en cuenta consideraciones procedentes del nivel algorítmico de análisis. Si bien son necesarias las consideraciones normativas o racionales *a priori* sobre qué información es relevante y por qué lo es, en modo alguno consideramos que en sí mismas agoten el nivel computacional de análisis. En concreto, se ha mantenido que los mecanismos asociativos de aprendizaje pueden desempeñar un papel relevante tanto en la explicación de la sensibilidad del orden de presentación de los ensayos como en la explicación de los juicios que los individuos emiten en situaciones de inferencia causal y que, en este sentido, pueden desempeñar también un significativo papel a la hora de definir cuáles son los objetivos computacionales que el sistema satisface.

## Referencias

- ANDERSON, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14, 471-517.
- CAÑO, A., COBOS, P. L., LÓPEZ, F. J., LUQUE, J. L. y ALMARAZ, J. (1997). Sesgos en el aprendizaje de relaciones causales. *IX Congreso de la Sociedad Española de Psicología Comparada*. Salamanca.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405.
- CHENG, P. W. (2000). Causality in the mind: estimating contextual and conjunctive causal power. En F. C. Keil y R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227-253). Cambridge, MA: The MIT Press.
- CHENG, P. W. y NOVICK, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365-382.
- COBOS, P. L., CAÑO, A., LÓPEZ, F. J., LUQUE, J. L. y ALMARAZ, J. (2000). Does the type of judgement required modulate cue competition? *Quarterly Journal of Experimental Psychology*, 53B, 193-207.
- COBOS, P. L., LÓPEZ, F. J., CAÑO, A., ALMARAZ, J. y SHANKS, D. R. (2001). *Mechanisms of predictive and diagnostic causal induction*. Manuscrito enviado para publicación.
- DICKINSON, A., SHANKS, D. R., y EVENDEN, J. L. (1984). Judgment of act-outcome contingency: the role of selective attribution. *Quarterly Journal of Experimental Psychology*, 36A, 29-50.
- EVANS, J. ST B. T. y OVER, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- HUME, D. (1888/1992). *A treatise of human nature*. Oxford: Oxford University Press.
- HUME, D. (1777/1993). *An enquiry concerning human understanding*. Indianapolis, IN: Hackett.

- LEWANDOWSKY, S. (1994). On the relation between catastrophic interference and generalization in connectionist networks. *Journal of Biological Systems*, 2, 307-333.
- LÓPEZ, F. J., COBOS, P. L., CAÑO, A. y SHANKS, D. R. (1998a). The rational analysis of human causal and probability judgement. En M. Oaksford y N. Chater (Eds.), *Rational models of cognition* (pp. 314-352). Oxford: Oxford University Press.
- LÓPEZ, F. J., SHANKS, D. R., ALMARAZ, J. y FERNÁNDEZ, P. (1998b). 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.
- MACKINTOSH, N. J. (1983). *Conditioning and associative learning*. Oxford: Oxford University Press.
- MARR, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.
- MATUTE, H., ARCEDIANO, F. y MILLER, R. (1996). Test question modulates cue competition between causes and between effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 182-196.
- MCCLOSKEY, M. y COHEN, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. En G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109-165). Nueva York: Academic Press.
- PRICE, P. C. y YATES, J. F. (1995). Associative and rule-based account of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1639-1655.
- SHANKS, D. R. y LÓPEZ, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory and Cognition*, 24, 511-522.
- SHANKS, D. R., HOLYOAK, K. J. y MEDIN, D. L. (Eds.) (1996). *The psychology of learning and motivation* (Vol 34: *Causal Learning*). San Diego, CA.: Academic Press.
- SLOMAN, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- 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 and Cognition*, 21, 802-808.
- WALDMANN, M. R. (2000). Competition among causes but of effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 53-76.
- WALDMANN, M. R. (En prensa). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review*.
- WALDMANN, M. R. y HOLYOAK, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121, 222-236.
- WASSERMAN, E. A., DORNER, W. W. y KAO, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 509-521.