

Simposio/Symposium

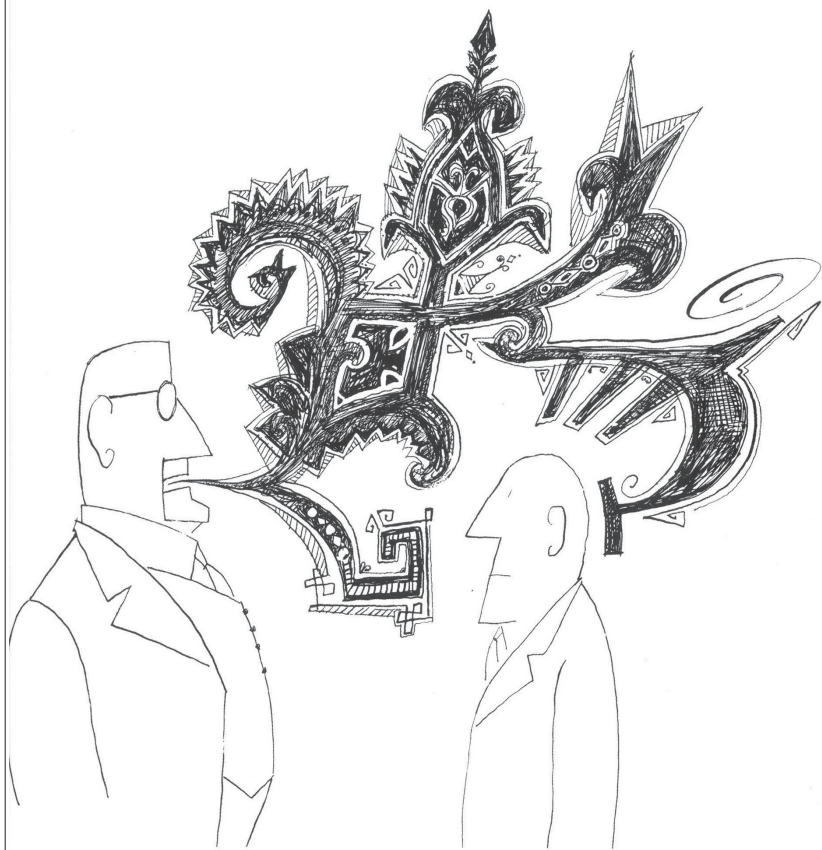
Hugo Mercier and Dan Sperber

The Enigma of Reason

A New Theory of Human Understanding

HUGO MERCIER · DAN SPERBER

The Enigma of Reason



Resumen de *The Enigma of Reason*

Hugo Mercier y Dan Sperber

The Enigma of Reason [Mercier and Sperber (2017)] se abre con un doble enigma. Muchos eruditos a lo largo de la historia han pensado en la razón como una especie de bala de plata cognitiva que permitiría a los humanos innovar, vencer sus fallos cognitivos y emocionales, resolver una amplia variedad de problemas y comprender mejor el mundo entorno. El primer enigma es, entonces, ¿por qué solo los humanos habrían de estar dotados con tal superpoder? ¿Por qué tal capacidad con sus múltiples ventajas no podría haber evolucionado en muchos otros organismos? El segundo enigma surge de la falta de encaje entre este elevado punto de vista sobre la razón y la realidad: la experiencia y los experimentos han mostrado una y otra vez que la razón humana es tan imperfecta, está tan sesgada, y es tan susceptible de cometer errores, como el resto de nuestra cognición.

Ofrecemos una doble respuesta a estos dos enigmas. En primer lugar, desarrollamos una nueva comprensión de lo que *es* la razón, así como nuevas hipótesis sobre aquello *para lo que* es (partes II y III). En segundo lugar, defendemos esas hipótesis concentrándonos en las hipótesis funcionales desarrolladas en la parte III (partes IV y V).

La parte I, “Haciendo temblar el dogma”, ofrece una breve mirada crítica sobre el “estado del arte” en el estudio de la razón humana. Desde su irrupción en los años 60 del pasado siglo, la psicología del razonamiento ha estado enredada en profundos debates –sobre lo que es la razón, cómo funciona, cuáles son los mejores instrumentos para modelarla, qué papel han de desempeñar en su estudio las consideraciones normativas y, sobre todo, en interminables debates sobre un puñado de paradigmas experimentales como el de la selección de Wason, o la falacia de la conjunción de Kahneman y Tversky. Si bien los debates son una parte saludable de la ciencia, la falta de resolución después de varias décadas de trabajo, incluso cuando se trata de la interpretación de problemas simples y bien estudiados, sugiere que algo completamente fundamental no va bien en el modo

en que la razón ha sido entendida hasta ahora. El surgimiento a lo largo de los últimos veinte años de teorías de procesamiento dual parecía prometer un marco unificado que permitiera la convergencia de diferentes perspectivas. Con todo, si bien las teorías de procesamiento dual han resultado ser un producto exportable con mucho éxito (son ahora muy populares en muchos campos, desde el razonamiento moral a economía conductual) irónicamente están en retirada dentro de la propia psicología del razonamiento [véase, por ejemplo, Bago y De Neys (2017); Melnikoff y Bargh (2018)]. Así pues, a pesar de haber obtenido durante las pasadas décadas muchos experimentos brillantes y muchas intuiciones teóricas, creemos que el estudio de la razón humana está maduro para que se lo vuelva a pensar seriamente, empezando por unos fundamentos diferentes y, esperemos que así sea, más robustos.

Una de las fortalezas de nuestra perspectiva es precisamente situar la razón junto con otros mecanismos cognitivos. Para hacer esto, la parte II, “Comprendiendo la inferencia”, desarrolla una taxonomía de mecanismos cognitivos que se resumen en la FIGURA 1. Los círculos representan nuevas categorías cada vez más estrechas de mecanismos cognitivos. El hecho de que la razón esté en el centro es un mero accidente que resulta de que la razón está en el punto focal del trabajo que se hace hoy en día, y podría trazarse un diagrama equivalente para cualquier otro mecanismo cognitivo.

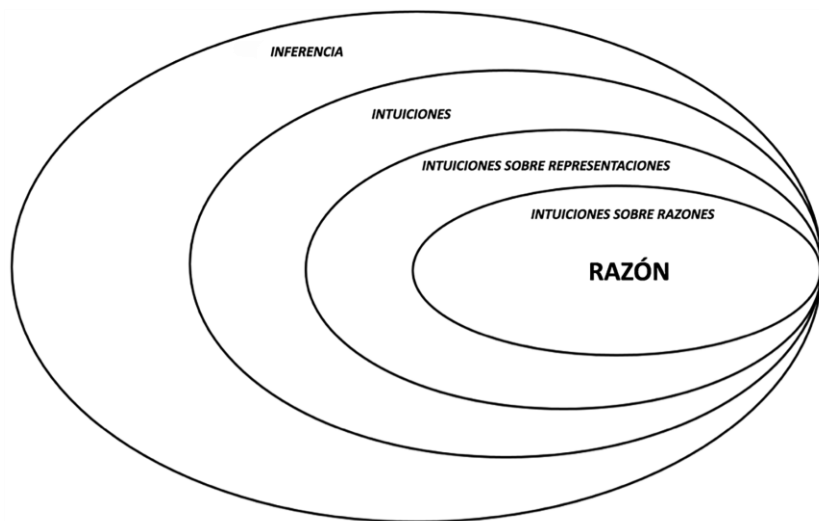


FIGURA 1. Categorías de los mecanismos mentales
(FIGURA 1 de *El enigma de la razón*)

La categoría más amplia de la FIGURA 1 es la de inferencia. Los humanos y otros animales realizan constantemente inferencias; trabajan sobre lo que ya saben para extraer nuevas conclusiones. Esas inferencias nos permiten dotar de sentido a nuestro entorno perceptivo, anticipar las consecuencias de nuestras acciones, predecir lo que otros harán o entender lo que la gente quiere decir cuando habla. En contraste con algunas perspectivas del procesamiento dual, no creemos que los humanos, o cualquier otro animal, extraiga muchas de sus inferencias por medio de alguna capacidad inferencial de tipo general. Más bien, los humanos usan toda una amplia variedad de mecanismos inferenciales especializados, cada uno de ellos dedicado a ayudarnos a resolver un problema distinto. ¿A qué distancia está un objeto en la escena visual? ¿Cuánta fuerza se necesitará para levantar algún objeto? ¿Es esto comestible? ¿Quién sería para alguien la pareja más adecuada? ¿Es peligroso este animal? ¿Hacerle frente o huir? ¿Cómo reaccionar ante esta señal emitida por alguien de mi propia especie? Y así sucesivamente. Esas capacidades inferenciales especializadas son en parte instintivas, en parte el resultado del desarrollo y el aprendizaje, y el balance entre ambos componentes varía de una capacidad a la siguiente.

Plausiblemente, todas las inferencias en los animales no-humanos, y en la gran mayoría de las inferencias en humanos, nunca alcanzan el estado consciente (piénsese en el procesamiento visual o sintáctico de nivel inferior, en el control motor, etc.). Definimos las intuiciones (o la inferencia intuitiva) como aquellas inferencias que tienen un resultado consciente junto con propiedades metacognitivas específicas, tales como un sentimiento característico de auto-confianza [tal como se discute en Thomson (2014)]. Por ejemplo, en un restaurante uno podría echar un vistazo a un plato que está en otra mesa y tener la intuición de que te gustará. Incluso si uno no tiene ni idea de por qué o de cómo ha llegado a esta conclusión, uno siente como si el proceso interno oculto que le ha llevado a extraer esta conclusión fuera suficientemente sólido.

La categoría presentada a continuación en la FIGURA 1 es la de las intuiciones sobre representaciones (o inferencias meta-representacionales). La razón no es la única peculiaridad humana. Los humanos son también los únicos animales capaces de representar representaciones de forma fluida, y, de hecho, los humanos ocupan gran parte de su tiempo haciéndolo. Siempre que estamos con otras personas, o simplemente pensamos en ellas, representamos sus estados mentales: cuáles son sus creencias, intenciones, deseos. Cuando nos embarcamos en comunicaciones efectivas

o imaginadas, procesamos diferentes niveles de representaciones [Sperber & Wilson (1995)].

Nuestra primera afirmación central es que la razón es “solo” otro tipo de inferencia. Más específicamente, la razón es un mecanismo de inferencias intuitivas meta-representacionales que descansa sobre un tipo de representaciones: las razones.

Esta afirmación se desarrolla en la parte III, “Repensando la razón”, en la que sugerimos que el estudio de la razón (en tanto que capacidad cognitiva) y de las razones (justificaciones, argumentos) debería unificarse, incluso si, como ha sucedido a menudo, han sido tratadas separadamente. Las razones son esencialmente sociales. Presentamos razones para justificar nuestras acciones o creencias y para convencer a los demás. Evaluamos las razones de los otros para decidir la cantidad de revisión de creencias que garantizan. Si podemos usar razones por nosotros mismos razonando en solitario, eso se debe sólo a que se trata de un uso derivativo. Además, de acuerdo con este punto de vista, el estudio de la razón está completamente separado del de la lógica (o de cualquier otro sistema de reglas formales). Más bien que ser el fundamento de la razón, la lógica es un instrumento teórico que nos ayuda a expresar argumentos más claramente al subrayar, y a menudo exagerar, la relación entre premisas y conclusión.

Si la razón no está guiada por la lógica, si usar la razón por nosotros mismos es un mero subproducto, ¿cuál es entonces la función de la razón? Nuestra segunda afirmación central es que la razón satisface mayormente funciones sociales, más específicamente, el intercambio de justificaciones y de argumentos.

Más que cualquier otro primate, los humanos cooperan, no solamente con sus congéneres, sino también con los que no lo son. Esta cooperación no podría sostenerse si no evaluásemos la fiabilidad de los demás en tanto que compañeros en la cooperación y, en particular, su competencia y equidad [Baumard, André y Sperber (2013)]. Cierta evidencia de fiabilidad viene proporcionada por el modo en que las personas se comportan, pero esta evidencia está limitada a las propias observaciones y está abierta típicamente a una gran variedad de interpretaciones. Así pues, dependemos en gran medida de evidencia proporcionada por otros mediante el testimonio y las habladurías. Esta evidencia social, cuando se acumula, determina la reputación de una persona. Tener una buena reputación como cooperador fiable es una condición de éxito social y, más allá de esto, de aptitud biológica. Esto nos ofrece un poderoso incentivo para intentar proteger y mejorar nuestra reputación

explicándonos y justificándonos. Como observadores, tenemos un incentivo para evaluar y posiblemente desafiar esas autojustificaciones. Presentar esas justificaciones y evaluarlas es, mantenemos nosotros, una de las dos funciones de la razón.

Otro rasgo sorprendente de los humanos es el de que hasta qué punto se comunican. Al igual que la cooperación, la comunicación lleva consigo sus propios problemas evolucionistas: cómo evitar el que se nos mienta, se nos desoriente o manipule. Para ello, calibramos nuestra confianza en los demás y estamos más inclinados a dar por buena la palabra de alguien que es considerado competente y benevolente [Mercier (2017); Sperber et al. (2010)]. Sin embargo, al descansar puramente en la confianza, corremos el riesgo de perder información que es valiosa pero que viene proporcionada por un hablante en el que no confiamos totalmente para cambiar de parecer. En esta situación, el hablante puede presentar argumentos que, a diferencia del mero testimonio, son evaluables por sus propios méritos, independientemente de la fuente de donde provengan. Aceptar tales argumentos puede llevarnos a cambiar de opinión cuando la confianza en la fuente no habría sido lo suficientemente fuerte para que lo hubiéramos hecho. Tales intercambios de argumentos dan lugar a discriminaciones de grano mucho más fino entre mensajes valiosos e inadecuados o dañinos.

En el pasado, nos hemos centrado en esta última función de la razón desarrollando una teoría argumentativa del razonamiento [Mercier (2016b), Mercier y Sperber (2011)]. Puesto que en *The Enigma of Reason* afirmamos que la razón también sirve para otra función social, a saber: la justificación, preferimos llamar interaccionista a nuestro actual enfoque de la razón.

La parte IV, “Lo que la razón puede y no puede hacer”, pasa revista a la evidencia que pertenece al funcionamiento de la razón humana y argumenta que la evidencia es más coherente con nuestra explicación interaccionista que con otras perspectivas sobre la razón. Si la explicación interaccionista es correcta, deberíamos esperar que la razón exhiba ciertos rasgos (resumidos en la tabla 1). Dado que la razón desempeña papeles muy diferentes cuando produce razones que apuntan hacia otros, que cuando evalúa razones que los demás nos adjudican, deberíamos esperar que las razones exhiban diferentes rasgos cuando desempeñan cada uno de esos papeles.

	Sesgo	Control de calidad
Producción de razones	<i>Sesgadas</i> : la gente produce, en su mayoría, razones que están a su favor	<i>Perezosa</i> : las personas no son muy exigentes respecto de sus propias razones
Evaluación de las razones de otros	<i>No-sesgadas</i> : la gente acepta razones que sean incluso desafiantes, si son suficientemente fuertes.	<i>Exigentes</i> : a las personas solo se las convence con razones suficientemente buenas.

TABLA 1. Las principales características de la razón humana
(FIGURA 19 de *The Enigma of Reason*)

Cuando presentamos razones, solemos estar fuertemente sesgados hacia nuestro punto de vista. No vamos a parecer más racionales si producimos razones de por qué lo que hicimos era estúpido; no vamos a convencer a alguien dándole argumentos a favor de sus puntos de vista o en contra del nuestro. Esto explica un rasgo problemático de la razón: el sesgo en favor mío [o sesgo de confirmación, véase Mercier (2016a)].

Podríamos también esperar que la razón fuera capaz de producir razones muy fuertes para defendernos mejor a nosotros mismos y convencer a los demás. Sin embargo, el encontrar razones fuertes es una tarea cognitivamente exigente, una tarea que se logra mejor, no mediante fuerza bruta, sino prestando atención a las reacciones que se producen. En nuestra explicación, la razón ha evolucionado tal como se usaba en contextos sociales, dialógicos. En tales contextos, si alguien no está convencido por una razón, la reacción natural es proporcionar, a su vez, otra razón, típicamente un contraargumento. Es entonces posible que los hablantes se apoyen en la reacción observada en sus interlocutores para producir mejores razones, más bien que tener que encontrar tales razones exclusivamente por uno mismo. Como resultado, esperamos que la producción de razones sea perezosa, pero en un sentido inteligente: la primera razón que presentamos puede ser completamente banal, pero debemos de ser capaces de adaptarnos a los contraargumentos planteados por nuestros interlocutores.

El modo en que evaluamos las razones de los demás –por lo menos cuando las razones tratan de cambiar nuestras opiniones– debería proceder de manera opuesta a como producimos las razones. Deberíamos de ser ca-

paces de reconocer las buenas razones, incluso si desafían nuestras creencias anteriores, y provienen de fuentes en las que no confiamos plenamente. De hecho, hacerlo así es, de acuerdo con nosotros, la verdadera finalidad de la razón. Y deberíamos de ser capaces de rechazar las razones débiles, de modo que no nos convenzamos de algo cuando el hacerlo no esté garantizado.

Estos rasgos de la razón tienen diversas implicaciones. Cuando las personas razonan por sí mismas, en su mayor parte presentan razones para defender sus opiniones preexistentes. Esto explica por qué un razonador solitario no logra, de manera general, corregir sus propias intuiciones equivocadas: a medida en que las razones se amontonan para apoyar esas intuiciones, la persona que razona se arriesga incluso a volverse más confiado. Cuando, para empezar, un razonador solitario no tiene una intuición fuerte, la razón lo empuja hacia la decisión que se justifica más fácilmente, sea o no una buena decisión, un fenómeno conocido en la literatura sobre el juicio y la toma de decisiones como elección basada en la razón.

Por contraste, cuando la gente razona entre sí, los participantes se turnan para presentar razones y evaluar las de los otros. Bajo las condiciones correctas —un pequeño grupo con algunos incentivos comunes, pero que está en desacuerdo sobre algún punto— la discusión y el intercambio de razones produce maravillas. Los que tienen las mejores ideas, o alguna intuición valiosa sobre el problema, pueden convencer al otro grupo de miembros, conduciéndolos a mejoras sensibles, algunas veces muy importantes [para una revisión de evidencias recientes, véase Mercier (2016b); véase también Cladière, Trouche y Mercier (2017)]. Que el intercambio de razones permite que las buenas ideas se diseminen y se incremente su eficacia es algo que ha sido observado en una variedad de contenidos, desde los problemas lógicos hasta la predicción de decisiones médicas.

Mientras que la parte IV descansa, en su mayoría, en evidencia reunida en el laboratorio, la Parte V, “Razón salvaje”, trata con problemas similares, pero concentrándose en otras fuentes de evidencia: la antropología, la historia o la sociología. Comenzamos mirando hacia la robustez transcultural de los principales rasgos de la razón. ¿Es sesgada y perezosa la producción de la razón en culturas diferentes? ¿Es la gente capaz en todos los sitios de formar los mejores grupos de discusión? *Grosso modo*, la respuesta es sí. Nos volvemos entonces a dos dominios en los que se ha afirmado que la razón funciona en desacuerdo con nuestro enfoque. En el dominio de la moralidad, intentamos refutar una posición pesimista de acuerdo con la cual los argumentos morales serían esencialmente inertes

para mostrar en su lugar que las buenas razones pueden cambiar las opiniones de las personas, incluso en asuntos morales y emocionales. Nos volvemos hacia la ciencia, vista a menudo como el dominio de genios singulares dando forma por sí mismos a grandes teorías y no logrando vencer de sus brillantes intuiciones a sus colegas de rango inferior, totalmente lo opuesto de lo que precedimos. Sin embargo, el trabajo en la historia y en la sociología de la ciencia presenta un cuadro diferente: uno de los científicos en intercambio constante (o, al menos, revisando intercambios pasados y anticipando otros nuevos), que empuja a otros a que desarrollen mejores argumentos y cuyas teorías pueden conquistar rápidamente todo un campo, tan pronto como se las apoya de manera suficiente.

En este libro argumentamos entonces a favor de una solución original a los dos enigmas de la razón. La razón es algo único de los humanos, puesto que ha evolucionado en respuesta a presiones por parte de la selección a las que sólo los humanos han hecho frente: cooperación y comunicación de una magnitud y complejidad sin rival. Estas presiones de la selección dieron lugar a mecanismos inferenciales dedicados al procesamiento de razones para consumo social: para intercambiar argumentos y justificaciones. La razón no es un superpoder, es solo otro mecanismo cognitivo especializado, cuyas fortalezas y debilidades pueden explicarse a la luz de su función, y que puede lograr maravillas, pero a una escala social y cultural más bien que a escala individual.

*Institut Jean Nicod, Département d'études cognitives,
ENS, EHESS, PSL University, CNRS,
Paris Francia
E-mail: hugo.mercier@gmail.com*

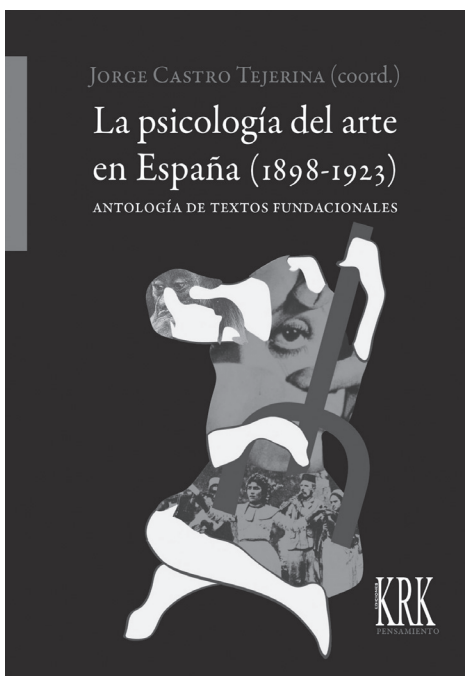
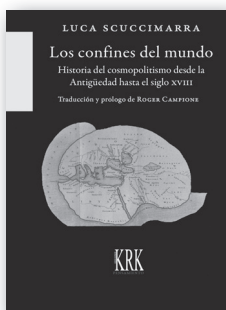
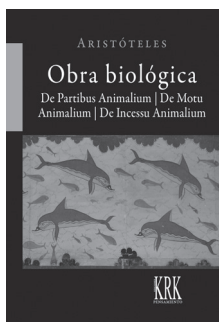
*Department of cognitive science and Department of philosophy,
Central European University, Budapest, Hungary
Institut Jean Nicod, Département d'études cognitives,
ENS, EHESS, PSL University, CNRS,
Paris Francia
E-mail: dan.sperber@gmail.com*

ACKNOWLEDGMENTS

El trabajo de HUGO MERCIER'S está patrocinado por la Agence Nationale de la Recherche, EUR FrontCog ANR-17-EURE-0017. El trabajo de DAN SPERBER'S está patrocinado por el European Research Council, bajo el programa European Union's Seventh Framework (FP7/2007-2013)/ERC grant agreement n° [609819], SOMICS.

REFERENCIAS BIBLIOGRÁFICAS

- BAGO, B. y DE NEYS, W. (2017), “Fast Logic?: Examining the Time Course Assumption of Dual Process Theory”; *Cognition*, 158, pp. 90-109.
- BAUMARD, N., ANDRÉ, J. B. y SPERBER, D. (2013), “A Mutualistic Approach to Morality: the Evolution of Fairness by Partner Choice”; *Behavioral and Brain Sciences*, 36(01), pp. 59-78.
- CLAIDIÈRE, N., TROUCHE, E. y MERCIER, H. (2017), “Argumentation and the Diffusion of Counter-Intuitive Beliefs”; *Journal of Experimental Psychology: General*, 146(7), pp. 1052-1066.
- MELNIKOFF, D. E. y BARGH, J. A. (2018), “The Mythical Number Two”; *Trends in Cognitive Sciences*, 22(4), pp. 280-293.
- MERCIER, H. (2016a), “Confirmation (or Myside) Bias”; en R. Pohl (Ed.), *Cognitive Illusions* (2nd ed., pp. 99-114). Londres: Psychology Press.
- (2016b), “The Argumentative Theory: Predictions and Empirical Evidence”; *Trends in Cognitive Sciences*, 20(9), pp. 689-700.
- (2017), “How Gullible Are We? A Review of the Evidence from Psychology and Social Science”; *Review of General Psychology*, 21(2), p. 103.
- MERCIER, H. y SPERBER, D. (2011), “Why Do Humans Reason? Arguments for an Argumentative Theory”; *Behavioral and Brain Sciences*, 34(2), pp. 57-74.
- (2017), *The Enigma of Reason*, Harvard, Mass., Harvard University Press.
- SPERBER, D., CLÉMENT, F., HEINTZ, C., MASCARO, O., MERCIER, H., ORIGGI, G. y WILSON, D. (2010), “Epistemic Vigilance”; *Mind and Language*, 25(4), pp. 359-393.
- SPERBER, D. y WILSON, D. (1995), *Relevance: Communication and cognition*. Nueva York: Wiley-Blackwell.
- THOMPSON, V. A. (2014), “What Intuitions Are... and Are Not.”; en B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 60, pp. 35-75). Burlington: Academic Press.



EDICIONES
KRK

PEDIDOS
correo@krkediciones.com
www.krkediciones.com