

Convergencias teóricas en la psicología del razonamiento contrafáctico

Susana SEGURA

Universidad de Málaga (España)

Resumen

El razonamiento contrafáctico consiste en comparar la realidad con alternativas simuladas. Este artículo presenta una revisión teórica a través de un recorrido histórico por las principales teorías en la psicología del razonamiento contrafáctico y establece dos reflexiones centrales desde una visión de conjunto: (a) la investigación en el razonamiento contrafáctico ha contribuido al desarrollo y convergencia de teorías relevantes que defienden un concepto de racionalidad no normativa de la mente: la teoría de los heurísticos y sesgos (junto con la teoría normativa) y la teoría de los modelos mentales; y (b) estos tres planteamientos suscitaron a su vez la aparición de las principales teorías actuales en la psicología del razonamiento contrafáctico: la teoría funcionalista del razonamiento contrafáctico y la teoría de la imaginación racional. La convergencia y complementariedad de estas teorías permite concluir que la racionalidad no normativa y la imaginación desempeñan una relevante función de aprendizaje.

Abstract

Counterfactual thinking is the comparison of reality with simulated alternatives. This paper presents a theoretical revision of theories on the psychology of counterfactual thinking going through history, establishing two main reflexions from an overall view: (a) Research on counterfactual thinking has contributed to the development and convergence between significant theories that maintain a non-normative rationality concept of mind: the Heuristics and Biases Theory (together with the Normative Theory) and the Mental Models Theory, and (b) These three approach yield the emergence of the two main current theories in the Psychology of counterfactual thinking: the Functional Theory of Counterfactual Thinking and the Rational Imagination Theory. The convergence and complementarities of these theories allows for conclude that non-normative rationality and imagination accomplish a relevant function of learning.

Razonamiento contrafáctico significa, literalmente, razonamiento “contrario a los hechos” y consiste en comparar una o varias alternativas imaginadas con unos acontecimientos reales. Por lo general, esta comparación se lleva a cabo mediante la simulación de escenarios o representaciones mentales de situaciones que son contrarias o diferentes a hechos pasados, por lo que el razonamiento contrafáctico ha sido denominado como el pensamiento sobre “lo-que-podría-haber-sido” (Byrne 2005; Roese, 1997; Kahneman, y Tversky, 1982). Los pensamientos contrafácticos tienen unas características específicas que han determinado su inclusión como objeto de estudio de la psicología científica a lo largo de las cuatro últimas décadas. Las revisiones teóricas que se han realizado hasta ahora sobre el razonamiento contrafáctico han estado focalizadas solamente en fenómenos psicológicos concretos o han sido

realizadas desde una perspectiva exclusivamente social o exclusivamente cognitiva (Byrne, 2002; Byrne, 2016; Miller, Turnbull, y McFarland, 1990; Roese y Olson, 1995; Roese y Morrison, 2009). Este artículo presenta una revisión teórica que muestra un nuevo punto de vista al analizar una visión de conjunto de las perspectivas y teorías existentes a lo largo de la historia de la psicología del razonamiento contrafáctico.

A continuación, se presentarán las primeras teorías que se formularon con objeto de analizar los razonamientos sobre lo-que-podría-haber-sido: la *teoría de los heurísticos y sesgos*, especificada con más detalle para el razonamiento contrafáctico en la *teoría normativa*, y la *teoría de los modelos mentales*. Estas tres teorías promovieron la aparición de las principales teorías existentes en la actualidad que definen la razón y constitución de los pensamientos sobre lo-que-

Dirección de la autora: Departamento de Psicología Básica. Facultad de Psicología. Campus de Teatinos. 29071 Málaga. *Correo electrónico:* s_segura@uma.es

Recibido: octubre de 2016. *Aceptado:* noviembre de 2016.

podría-haber-sido. Las dos primeras originaron la aparición de la *teoría funcionalista del razonamiento contrafáctico*, mientras que la tercera enmarca la *teoría de la imaginación racional*. La evolución de todas estas teorías junto con la consideración de sus aportaciones a la psicología suscita nuevas reflexiones sobre la investigación en la psicología del razonamiento contrafáctico.

Teorías sobre el pensamiento humano y el razonamiento contrafáctico

La teoría de los heurísticos y sesgos

El razonamiento inductivo es un proceso mental que consiste en llegar a conclusiones generales a partir de casos específicos. Se convierte en la primera fuente de conocimiento del mundo porque las personas comienzan su aprendizaje encontrándose con datos del entorno que acaban por agrupar en sus mentes en forma de abstracciones aplicables a todas las personas, cosas o situaciones. Para llevar a cabo este proceso de forma óptima, las personas deberían seguir las normas de la Estadística. Sin embargo, en la vida diaria no se puede tener seguridad sobre la fiabilidad, representatividad y suficiencia de los datos recopilados. Además no se conocen, no se aplican o se aplican incorrectamente las leyes de la estadística. Esto fue lo que demostraron Amos Tversky y Daniel Kahneman (1974) desde un enfoque cognitivo-social, desarrollando la *teoría de los heurísticos y sesgos*, por la que recibieron el premio Nobel en economía en el año 2002. A los principios que realmente siguen las personas en la vida diaria al llevar a cabo razonamientos inductivos los denominaron *heurísticos*. Los *heurísticos* son reglas generales que sirven de atajos mentales y se aplican al realizar juicios sociales. Son pragmáticos y económicos porque requieren de pocos recursos cognitivos pero, debido a su falta de rigor al considerar información, pueden llevar a sesgos, es decir, a errores graves y sistemáticos respecto a la norma, que en este caso es la estadística.

El razonamiento contrafáctico depende del heurístico de simulación. Las personas siguen este heurístico cuando imaginan un escenario alternativo o hipotético porque idean lo que podría haber sido en función del escenario que está más disponible en la memoria y no en función de la verdadera probabilidad estimada de que un escenario ocurra, que debería ser la regla estadística a seguir en estas situaciones. Además, las personas realizan una comparación de ese escenario alternativo simulado con el real y perciben que es más probable que ocurra el escenario alternativo que el real, simplemente también por esa disponibilidad en la memoria. Este heurístico lo aplican las personas en aquellos casos en los que tienen que realizar juicios para estimar probabilidades de aparición de determinados acontecimientos, realizar atribuciones causales o formarse impresiones generales acerca de hechos acontecidos. La aplicación del

heurístico de simulación tiene consecuencias emocionales de arrepentimiento, tristeza o frustración. Estas consecuencias afectivas aparecen de manera tan evidente que, en los primeros estudios, los investigadores las tuvieron en cuenta como medidas del razonamiento contrafáctico. Es decir, en lugar de preguntar a los participantes por escenarios hipotéticos alternativos respecto a los escenarios presentados en las tareas utilizadas durante los experimentos, les preguntaban por la frustración o el arrepentimiento que sentían en situaciones en las que era fácil imaginar esos escenarios alternativos (Kahneman y Tversky, 1982). Por lo tanto, presuponían que la existencia de emociones reflejaba la aplicación del heurístico y esta estrecha asociación entre emociones y pensamientos sobre lo-que-podría-haber-sido es una de las peculiaridades distintivas del razonamiento contrafáctico.

La teoría normativa

Posteriormente, en la *teoría normativa*, Kahneman y Miller (1986) precisaron el mecanismo por el que las personas generan alternativas a los acontecimientos que les suceden: las personas imaginan situaciones alternativas a hechos acontecidos cuando experimentan sorpresa ante esos hechos reales. La sorpresa es consecuencia de una gran discrepancia entre los hechos acontecidos y lo que cabría esperar teniendo en cuenta los antecedentes de esos hechos reales. Lo que cabría esperar en esas condiciones es lo que las personas consideran lo normal, pero establecen esa normalidad inmediatamente después de que hayan ocurrido los hechos, exclusivamente para esos hechos y en función de la mutabilidad del acontecimiento que antecede a esos hechos.

La mutabilidad consiste en la facilidad con que puede cambiarse mentalmente un evento antecedente a un hecho o unos hechos específicos. Cuando los antecedentes son mutables, los hechos acontecidos son percibidos como anormales, por lo que se generan una o varias alternativas contrafácticas. Por el contrario, cuando los antecedentes se perciben como normales o poco mutables, los hechos acontecidos se perciben como un resultado normal, incluso inevitable, y las personas no generan alternativas a los hechos acontecidos. En conclusión, la facilidad con que se generan alternativas en la mente depende de las características de los acontecimientos antecedentes a un hecho y son los antecedentes que se alejan de la norma los que se perciben más mutables.

La focalización de esta teoría en las características de los antecedentes de los eventos para explicar cómo las personas imaginan lo-que-podría-haber-sido orientaron todos los estudios que se realizaron en las dos primeras décadas. Los investigadores descubrieron que las personas tienden a imaginar alternativas a los hechos acontecidos cuando consideran antecedentes de esos hechos que implican acción como, por ejemplo, cuando se cambia una opción

una vez se ha seleccionado (Landman, 1987; Kahneman y Tversky, 1982). Este fenómeno fue denominado “efecto de acción”. Asimismo, descubrieron que las personas tienden a imaginar alternativas a los hechos acontecidos cuando consideran antecedentes de esos hechos que implican acciones controlables y dependen de la libertad de elección como, por ejemplo, entrar a tomar algo en un bar (Giroto, Legrenzi y Rizzo, 1991), lo que se denominó “efecto de controlabilidad”. Finalmente, demostraron que las personas tienden a imaginar alternativas a los hechos acontecidos cuando consideran antecedentes de esos hechos que están situados en el primer lugar de una cadena causal como, por ejemplo, cuando un individuo se encuentra con un obstáculo en el camino que genera las condiciones para que aparezca un segundo obstáculo (Wells, Taylor y Turtle, 1987) o en el último lugar de una serie temporal en la que los acontecimientos son independientes como, por ejemplo, cuando dos individuos tiran dos monedas a cara o cruz intentando emparejar el resultado (Miller y Gunasegaram, 1990). Los autores denominaron a estos fenómenos “efecto de primacía causal” y “efecto de orden temporal”, respectivamente. Todos estos estudios, que se llevaron a cabo de manera poco sistemática, estuvieron más centrados en describir estos efectos concretos que en desarrollar la *teoría normativa*, por lo que aunque descubrieron la mutabilidad percibida de los antecedentes de un suceso, la idea de partida de cómo la normalidad se establece en función de esa mutabilidad en el momento y a posteriori quedó como un concepto teórico con insuficiente apoyo empírico.

Por otra parte, estos autores profundizaron en el origen de la particularidad descubierta por la *teoría de los heurísticos y sesgos* que consiste en que determinadas emociones como el arrepentimiento, la decepción o el alivio están precedidas por los razonamientos contrafácticos, por lo que se denominan “emociones contrafácticas”. Investigaciones posteriores en el marco de la *teoría funcionalista del razonamiento contrafáctico* sustentaron esta afirmación y mostraron que lo que diferencia a unas emociones específicas de otras es el tipo de antecedente seleccionado al construir escenarios alternativos. Por ejemplo, mientras que a la vergüenza le precede la focalización en antecedentes que tienen que ver con uno mismo, a la culpa le precede la focalización en aquellos antecedentes que tienen que ver con los demás (Roese y Olson, 1995).

Las teorías sobre el razonamiento contrafáctico en la lógica

El conocimiento adquirido a través del razonamiento inductivo se concreta en una información sobre la que se llevan a cabo razonamientos deductivos, porque las deducciones son procesos mentales en los que una afirmación general aprendida se aplica a un caso específico desconocido al que se le atribuyen las propiedades que

contiene esa abstracción. Uno de los razonamientos deductivos más comunes que suelen realizar las personas son los razonamientos condicionales. Para llevar a cabo los razonamientos condicionales como para cualquier tipo de razonamiento deductivo de forma óptima, las personas deberían seguir las leyes de la lógica. Según la lógica se pueden relacionar dos proposiciones cualesquiera “p” y “q” a través de las conectivas lógicas: “si, entonces”, de lo que resulta un enunciado o afirmación general: “si p, entonces q”. De tal manera que para cualquier situación concreta en la que puedan estar presentes o ausentes “p” o “q”, se puede validar una conclusión consultando unas tablas de verdad previamente establecidas. Por lo tanto, si las personas parten de un enunciado que es verdad en el mundo real y realizan razonamientos válidos siguiendo estas tablas, pueden tener la certeza de que la conclusión también será verdad en el mundo real. Sin embargo, a diferencia de estos condicionales ordinarios, los condicionales contrafácticos toman la forma: “si hubiese ocurrido p, entonces habría ocurrido q” por lo que es difícil categorizarlos como argumentos deductivos, ya que las afirmaciones se refieren a situaciones que ciertamente no existen. Y si no existen, no se puede asegurar que sean verdaderas en el mundo real ni puede aplicárseles las tablas de verdad. De ahí que en filosofía se hable de esta cuestión como “el problema de los contrafácticos” (Nute, 1984) y, por este motivo, algunos autores propusieron que su validación debería establecerse en términos de las conexiones causales entre las proposiciones (Chisholm, 1946; Goodman, 1973), idea que se mantiene en el área de la filosofía de la ciencia (Barker, 2003; Woodward, 2004), o en términos de su significado (Lewis, 1973; Stalnaker, 1968).

Esta es otra particularidad de los razonamientos contrafácticos, que reside en el hecho de que la imaginación forma parte del razonamiento pero no puede validarse a través de la consideración del mundo real, por lo que una explicación desde la psicología en términos de cómo y para qué se llevan a cabo estos pensamientos, resulta probablemente más necesaria para el estudio de los razonamientos contrafácticos que para el caso de otro tipo de deducciones. Y probablemente, también por esta indefinición en la filosofía, el estudio de los razonamientos contrafácticos fue abordado por una teoría, la de los heurísticos y sesgos, que tradicionalmente ha dado explicación a la psicología del razonamiento inductivo y por otra teoría, la de los modelos mentales, que tradicionalmente ha dado explicación a la psicología del razonamiento deductivo.

La teoría de los modelos mentales

Phil Johnson-Laird y Ruth M.J. Byrne (1991, 1993) desarrollaron en Europa, desde un enfoque computacional en el área de la psicología cognitiva, uno de los modelos teóricos que mejor ha explicado los principios que rigen el

razonamiento deductivo de las personas: la *teoría de los modelos mentales*. Esta teoría, que se encuentra alejada de los modelos de la psicología social, determina que cada razonamiento humano se basa en la aparición parsimoniosa en la memoria operativa de modelos del mundo, llamados “modelos mentales”. Estos modelos están formados a partir de la búsqueda ordenada de ejemplos y contraejemplos sobre hechos reales o imaginados. Según esta teoría, se puede afirmar que a mayor exhaustividad en este proceso de búsqueda, más probable será que las personas obtengan conclusiones válidas conforme a las tablas de verdad de la lógica. Este proceso de razonamiento está influido por diversos factores como el nivel de conocimiento de las personas o la tarea concreta pero, especialmente, por los límites de la memoria operativa. La sobrecarga de esta memoria suele impedir que se contemple toda la información posible, lo que conduce a las personas a cometer sesgos o errores graves y sistemáticos respecto a la norma, que en este caso es la lógica.

Consecuentemente, al igual que para la *teoría de los heurísticos y sesgos*, para la *teoría de los modelos mentales* la disponibilidad de la información también es una variable esencial en el razonamiento humano, puesto que la facilidad con que se traen modelos a la mente es un factor clave en el proceso de completar representaciones del mundo. Probablemente, por este motivo, el propio autor, Johnson-Laird (1994), sostuvo que la explicación de los modelos mentales es compatible con la idea de estimar probabilidades a modo de escenarios propuesta por los autores de la teoría de los heurísticos y sesgos. De este modo, gracias al estudio de la psicología de los pensamientos sobre lo-que-podría-haber-sido, Johnson-Laird tendió un puente entre estas dos teorías.

De hecho, la *teoría de los modelos mentales* proporcionó una explicación para los primeros sesgos o efectos analizados en relación con el razonamiento contrafáctico, proponiendo que se producen por el fallo en la recuperación de la información debido a los límites cognitivos, aunque también precisó que se debe a las características del razonamiento humano y de este razonamiento en particular. Como se verá más adelante, esta explicación la ofreció la *teoría de los modelos mentales* en los inicios del desarrollo de la *teoría de la imaginación racional*, ya que esta segunda teoría se encuentra enmarcada en la primera.

Por consiguiente, una vez que la *teoría de los heurísticos y sesgos* definió que el razonamiento contrafáctico se establece a partir de la facilidad para traer escenarios a la mente, la *teoría de los modelos mentales* analizó específicamente cuáles, cuántos, en qué orden y de qué depende esa generación de escenarios en la memoria operativa, demostrando a través del estudio de estos razonamientos que estas dos perspectivas relevantes no son irreconciliables y permitiendo que se perciban las similitudes entre ellas, tal y como se analizará a continuación.

Convergencia de teorías en el estudio de la psicología del pensamiento: la racionalidad no normativa

La *teoría de los heurísticos y sesgos* ha explicado básicamente los sesgos en el razonamiento inductivo mientras que la *teoría de los modelos mentales* ha dado explicación especialmente a los sesgos en el razonamiento deductivo. Aunque el tipo de razonamiento estudiado por estas dos teorías difiere, ambas tienen un punto de partida que no es esencialmente diferente en la explicación de los errores sistemáticos en el razonamiento contrafáctico. Lo hacen basándose en la falta de información en la memoria operativa, que lleva a razonamientos poco rigurosos, según la *teoría de los heurísticos y sesgos*; o incompletos, según la *teoría de los modelos mentales*.

Igualmente, las dos teorías sostienen que las personas realizan procesos cognitivos que no siguen los patrones de ninguna norma reglada conocida como la estadística o la lógica. En lugar de eso, las personas se guían por principios psicológicos o patrones propios, aunque difieren en la especificación de estos principios pues estarían definidos en términos de reglas en lugar de por normas de estadísticas, según la *teoría de los heurísticos y sesgos*; o en términos de modelos mentales, en lugar de por normas lógicas, según la *teoría de los modelos mentales*.

Además de esta convergencia en el planteamiento teórico, las dos teorías desarrollaron posteriormente una concepción de la mente a distintos niveles que apoya esta idea. Por un lado, la *teoría de los heurísticos y sesgos* propuso que las personas tienen dos sistemas de pensamiento. En un primer término, poseen la forma de pensar heurística no rigurosa mencionada anteriormente, que utilizan cuando razonan o toman decisiones deprisa y que es tan intuitiva como emocional. En un segundo término, poseen una forma de pensamiento deliberativo para cuando razonan o toman decisiones más despacio (Kahneman, 2011). Por otro lado, la *teoría de los modelos mentales* especificó cómo los razonamientos de las personas pueden ser más o menos exhaustivos en función de lo amplia que sea su capacidad de imaginar posibilidades o modelos del mundo. En consecuencia, una misma persona puede comportarse de modo más o menos racional en función de esta capacidad de generar modelos en un momento dado. En otras palabras, el ser humano posee una racionalidad variable (Johnson-Laird y Byrne, 1991). Posteriormente, Johnson-Laird (2006) propuso que, tanto si el razonamiento depende de modelos como si depende de reglas, la variable fundamental que determina el nivel de razonamiento es la memoria operativa, pues es el sistema del que depende especialmente el nivel de profundidad en los pensamientos. Recientemente, en el marco de esta teoría, Gubbins y Byrne (2015) defienden que existen dos formas de razonamiento: una más intuitiva, emocional, rápida y otra más reflexiva o controlada.

En la actualidad, Jonathan Evans (2014) que mantiene una teoría que comparte esta visión de la racionalidad, centrada en una mente humana dividida en dos formas diferenciadas de mecanismos mentales (intuitivos frente a racionales), ha señalado que coincide con las teorías aquí expuestas y otras perspectivas relevantes, que incluyen enfoques de carácter más biológico. Todas estas teorías que defienden con suficiente apoyo empírico que las personas tienen en primer lugar una mente más experiencial, emocional o elemental e intuitiva y en segundo lugar otra más evolucionada, racional o lógica.

En conclusión, tanto la *teoría de los modelos mentales* como la *teoría de los heurísticos y sesgos* confluyeron en el estudio de los pensamientos sobre lo-que-podría-haber-sido. Además, aportaron a la psicología la idea, hoy extendida, de que los principios cognitivos que subyacen al razonamiento en la vida diaria no son normativos y están sujetos a sesgos, por lo que las personas necesitan realizar un esfuerzo cognitivo para poder superar las restricciones naturales de la mente en caso de querer alcanzar las mejores conclusiones posibles en sus razonamientos.

Principales teorías en la psicología del razonamiento contrafáctico

La teoría funcionalista del razonamiento contrafáctico

Neil J. Roese (1994, 1997) en los EEUU y Canadá, dentro del área de la psicología social, desarrolló la *teoría funcionalista del razonamiento contrafáctico*, que asegura que existen dos mecanismos básicos que subyacen a la mayoría de las consecuencias del razonamiento contrafáctico en la vida cotidiana: los efectos de contraste, producidos al comparar la realidad con lo que podría haber sido, y los efectos de las inferencias causales, que se producen cuando las personas interpretan que de no haber sucedido un hecho, no existiría la situación actual y, consecuentemente, ese hecho es percibido como la causa de la situación presente. Es pertinente destacar en primer lugar, que los efectos de contraste están claramente inspirados en los mecanismos explicados por la *teoría normativa* ya que, tal y como afirman Roese y Olson (1995), es una teoría de la sorpresa, centrada en la discrepancia entre las expectativas de la persona y la realidad actual. De este modo, adolece de los mismos problemas de indefinición de la *teoría normativa*: no especifica adecuadamente cuáles son los mecanismos mentales que están detrás de las inferencias causales que se producen cuando se contemplan pensamientos sobre lo que podría haber sido y no presenta suficiente apoyo empírico a ninguno de los dos postulados.

La *teoría funcionalista del razonamiento contrafáctico* además recoge el interés por la primera singularidad estudiada por la *teoría normativa*: la especial relación que los razonamientos contrafácticos tienen con las emociones,

puesto que los pensamientos sesgados, más intuitivos, rápidos o simplemente menos elaborados, son aquellos en los que tienen un mayor peso las emociones o, al menos, se consideran asociados a estas. De este modo, la teoría funcionalista del razonamiento contrafáctico analizó de forma más sistemática si las emociones eran consecuencias del razonamiento (por ejemplo, Roese y Olson, 1995, Roese, y Morrison, 2009), o si las emociones son determinantes de estos razonamientos (por ejemplo, Roese y Hur, 1997). Asimismo, estos psicólogos dan explicación a determinadas emociones tan ligadas estrechamente a los pensamientos hipotéticos que de otro modo probablemente no habían sido previamente evaluadas de forma tan apropiada como son el arrepentimiento, la vergüenza, la responsabilidad o la culpa. Por este motivo, se puede afirmar que la investigación en los pensamientos sobre lo que podría haber sido resultó útil para ampliar el conocimiento de la relación entre cognición y emoción.

La *teoría normativa* influyó en la *teoría funcionalista del razonamiento contrafáctico*, orientando su interés hacia la sorpresa en coincidencia con diferentes teorías relevantes en el área de la psicología del aprendizaje que comenzaron su desarrollo durante los mismos años y en la actualidad tienen substancial apoyo empírico. Estos modelos afirman que la sorpresa es necesaria para que aparezca el aprendizaje. Dicho en otras palabras, solo se aprende de lo que sorprende (por ejemplo, Mackintosh, 1975; Rescola y Wagner, 1972). Consecuentemente, de forma más relevante para la psicología social, la *teoría funcionalista del razonamiento contrafáctico* señala a distinto nivel explicativo que estos razonamientos aparecen debido a que cumplen dos funciones. Por un lado, la función preparativa o de aprendizaje que aparece cuando las personas imaginan cómo podría haber sido mejor (contrafácticos ascendentes), pues esta comparación les indica nuevos caminos que tomar en el futuro. Y, por el otro, la función emocional de consuelo, que aparece cuando las personas imaginan cómo podría haber sido peor (contrafácticos descendentes).

Aunque las dos funciones sirven al objetivo de la supervivencia, los razonamientos sobre lo que podría haber sido sirven, principalmente, para adquirir experiencias de mejora para el futuro. Roese y Morrison (2009) exponen que esta afirmación se sostiene por varias razones: espontáneamente se realizan más razonamientos contrafácticos ascendentes que descendentes, tienden a ocurrir después de un fallo más que después de un éxito, están enfocados a resolver problemas, en sí mismos mejoran el aprendizaje y distintas investigaciones han demostrado un correlato psicofisiológico entre del razonamiento contrafáctico y el aprendizaje.

Otro argumento de peso, que se puede añadir aquí para apoyar la idea de que el aprendizaje es la principal función del razonamiento contrafáctico, lo proporciona el hecho de que los contrafácticos descendentes no cumplen una función de consuelo ni bienestar ante emociones de culpa, vergüen-

za, arrepentimiento, decepción y tristeza ni ante el estrés psicológico (Mandel, 2003) y, sin embargo, sí que aparece la función de aprendizaje en estos casos y se convierten en una llamada de atención que hace que las personas estén motivadas a cambiar y a mejorar (McMullen y Markman, 2000). Por ejemplo, en el contexto de las organizaciones, Morris y Moore (2000) han demostrado estos efectos de aprendizaje pues han confirmado que las personas aprenden de los accidentes o errores cuando imaginan los costes vitales o importantes que podrían haber ocurrido, es decir, cuando llevan a cabo contrafácticos descendentes.

Epstude y Roese (2008, 2011) definen definitivamente la *teoría funcionalista del razonamiento contrafáctico*, que explora cómo el aprendizaje se produce en el momento existente para orientar conductas concretas posteriores en el sentido esperado, por lo que también muestra el papel de estos razonamientos en la motivación. Además, la teoría aclara la conexión entre los pensamientos sobre el pasado con los pensamientos sobre el presente y el futuro por lo que el aprendizaje generado por estos pensamientos sirve para definir el sentido de las conductas de las personas y situarlos en su transcurso vital (por ejemplo, Waytz, Hershfield y Tamir, 2014). En conclusión, la *teoría funcionalista del razonamiento contrafáctico* es una teoría que ha demostrado ampliamente la función para el aprendizaje humano de los pensamientos sobre lo-que-podría-haber-sido.

La teoría de la imaginación racional

Ruth M.J. Byrne (2002, 2005) y sus colaboradores desde el área de la psicología cognitiva y en el marco de la *teoría de los modelos mentales*, han determinado los mecanismos cognitivos que subyacen a los razonamientos sobre lo-que-podría-haber-sido. Planteándose los asuntos propios de la filosofía anteriormente expuestos, estos investigadores demostraron que los mecanismos que subyacen al razonamiento con condicionales ordinarios son los mismos que los que subyacen al razonamiento con condicionales contrafácticos (Byrne y Tasso, 1999) y los que subyacen al razonamiento causal también son los mismos que los que subyacen al razonamiento contrafáctico (McEleney y Byrne, 2006). De este modo, la *teoría de la imaginación racional* inicialmente se centra en la segunda singularidad de los pensamientos sobre lo-que-podría-haber-sido: el problema que conlleva el no poder validar con tablas de verdad la información que procede de la imaginación. Esta dificultad no es vigente en el área de la psicología del pensamiento según estos autores, pues demuestran que el razonamiento humano no sigue las reglas de la lógica cuando lleva a cabo deducciones. Además, ofrecen una explicación más articulada que la *teoría funcionalista del razonamiento contrafáctico* sobre la relación entre razonamiento causal y razonamiento contrafáctico. Asimismo,

aportan una evidencia experimental a aquellas perspectivas que tienen más peso en la actualidad dentro de la filosofía y que, tal y como se ha explicado, sostienen que al no tener validación posible solo pueden entenderse en término de las inferencias causales que permiten.

Estos principios cognitivos explican los fenómenos mencionados relacionados con los determinantes del razonamiento contrafáctico analizados al comienzo: “el efecto de acción” (Byrne y McEleney, 2000; Feeney y Handley, 2006), “el efecto de controlabilidad” (McCloy y Byrne, 2000) así como “el efecto de primacía causal” y “el efecto de orden temporal” (Byrne, Segura, Culhane, McAlinney, Tasso y Berrocal, 2000; Segura, Fernandez-Berrocal y Byrne, 2002). De este modo, los autores formularon la *teoría de la imaginación racional*. El hecho de ofrecer una explicación unitaria de los sesgos estudiados hasta entonces desde un enfoque teórico global, estructurado y con relevante apoyo científico (*Mental Models in Thinking and Reasoning*, 2016), convierten a esta teoría en la explicación más especificada y elaborada para dar cuenta de los mecanismos cognitivos que subyacen al razonamiento contrafáctico.

Según la *teoría de la imaginación racional*, estos mecanismos cognitivos están basados en el siguiente proceso: las personas representan parsimoniosamente modelos mentales que simbolizan estados del mundo y comienzan por un modelo que constituye un escenario que se corresponde con la situación actual o hechos acontecidos, para después continuar representando modelos que constituyen escenarios posibles imaginados. Este desarrollo de escenarios viene definido principalmente por los límites de la memoria operativa. Debido a estos límites, las personas representan solo uno o algunos de todos los mundos alternativos posibles lo que les lleva respuestas sesgadas, como las observadas al analizar los efectos mencionados arriba. Los autores ofrecen apoyo empírico a esta explicación llevando a cabo experimentos en los que demuestran que las personas dejan de cometer estos sesgos y se imaginan más alternativas posibles en aquellas condiciones que implican acceder a una mayor disponibilidad de la información porque les permite representar un mayor número de modelos mentales y no por una mera accesibilidad a esa información (Para una revisión, Byrne, 2002, 2005).

Byrne (2002, 2005) ha puesto de relieve que este proceso y los principios que guían la representación de alternativas imaginadas en la memoria operativa son los mismos que los que rigen tanto al pensamiento lógico (que subyace a cualquier razonamiento deductivo), como los que guían al pensamiento creativo (que subyace al *insight* necesario para la solución de problemas). En conclusión, la *teoría de la imaginación racional* es una teoría que ha demostrado cuál es la relación entre la imaginación y la racionalidad no normativa en la mente humana a través del estudio de los pensamientos sobre lo que podría haber sido.

Racionalidad, imaginación y aprendizaje en el razonamiento contrafáctico

La *teoría funcionalista del razonamiento contrafáctico* derivó de la *teoría de los heurísticos y sesgos* y de la *teoría normativa*. La *teoría de la imaginación racional*, derivó de la *teoría de los modelos mentales*. A pesar de que estas teorías se han desarrollado de forma independiente, sus aportaciones son complementarias pues proporcionan apoyo empírico a distintos aspectos del razonamiento contrafáctico. La *teoría funcionalista del razonamiento contrafáctico* ha explicado la relación entre los pensamientos sobre lo-que-podría-haber-sido y determinadas emociones que están intrínsecamente asociadas a estos pensamientos; pero especialmente el estudio del razonamiento contrafáctico desde esta teoría ha aportado a la psicología la idea de que estos pensamientos cumplen una función de aprendizaje para el ser humano. La *teoría de la imaginación racional* ha analizado la utilidad de evaluar la constitución de estos pensamientos sobre lo-que-podría-haber-sido explicando los principios cognitivos que los sustentan, en lugar de determinar su cuestionable validez como ocurre en la filosofía; pero especialmente ha aportado a la psicología la idea de que la racionalidad y la imaginación no se rigen por normas formales como las de la lógica. La evolución de las teorías aquí presentadas y su aportación a la psicología pueden observarse en la figura 1.

Un recorrido histórico y una visión de conjunto sobre la psicología del razonamiento contrafáctico permiten concluir

que la investigación en este campo ha contribuido especialmente al desarrollo de la concepción de una racionalidad humana no normativa y a la consideración de que esta forma de racionalidad junto con la imaginación juegan un papel de aprendizaje para la supervivencia en la vida cotidiana de las personas.

Esta conclusión se robustece si tenemos en cuenta que la relevancia del estudio de los pensamientos contrafácticos y su aplicación en la vida diaria se ha extendido a distintos ámbitos.

Aplicación en distintos ámbitos de la psicología y en otras disciplinas

Por un lado, los investigadores han demostrado su influencia en la *psicología clínica*, especificando, por ejemplo, un papel central en el insomnio (Schmidt y Van der Linden, 2009) y la rumiación (Leithy, Brown y Robbins, 2006).

En el ámbito legal y penal, los psicólogos encuentran que los juicios y atribuciones tanto causales como de responsabilidad están basados en los hechos pero también en lo-que-podría-haber-sido (Catellani y Millesi, 2001).

En la *psicología de las organizaciones*, los razonamientos contrafácticos intervienen en el análisis de errores y generan aprendizaje grupal (para una revisión, ver Segura y Morris, 2005).

Más recientemente, también los psicólogos han descubierto la influencia del razonamiento contrafáctico en otras

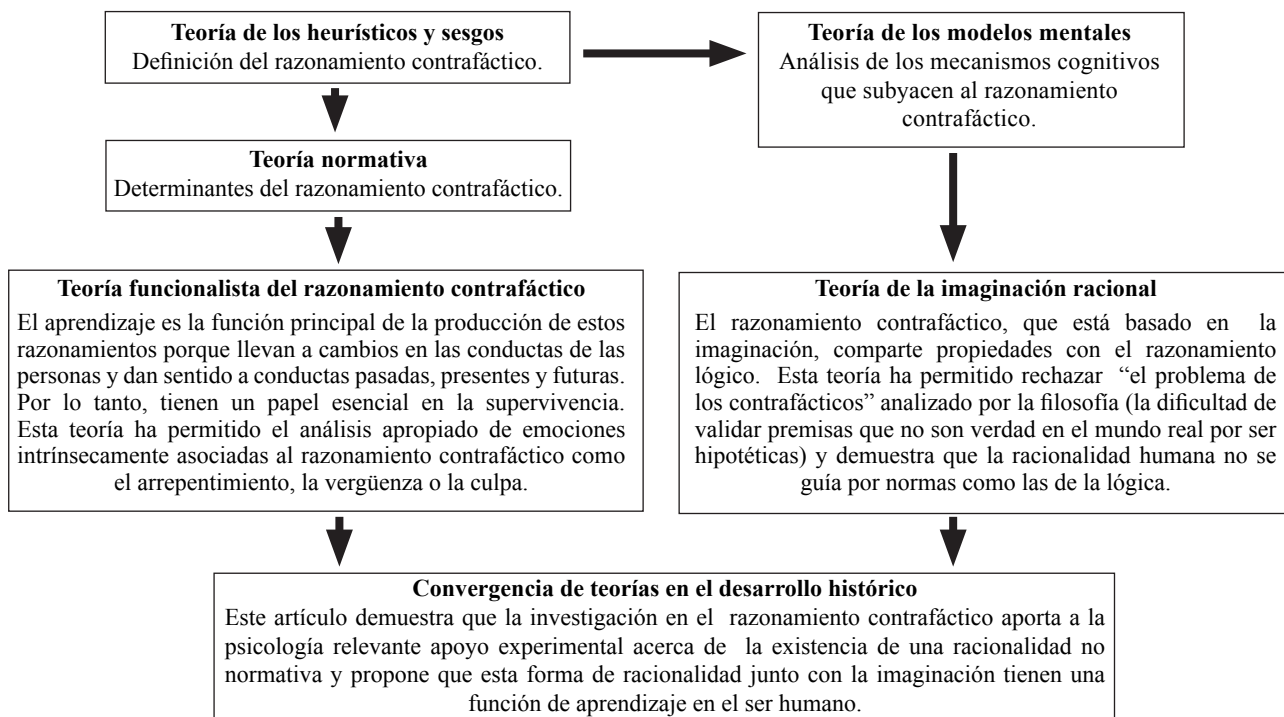


Figura 1. Convergencia de las principales teorías en la historia de la psicología del razonamiento contrafáctico.

áreas entre las que destacan la *psicología de la educación*, mostrando, por ejemplo, su influencia en el éxito académico (Petrocelli, Seta, Seta, y Prince, 2012), o en la *psicología del desarrollo*, explicando cómo estos pensamientos dan significado a la vida de las personas (Kray, George, Liljenquist, Galinsky y Tetlock, 2010).

Al mismo tiempo, la psicología de los pensamientos sobre lo-que-podría-haber-sido ha contribuido crecientemente con sus investigaciones al desarrollo de diversas disciplinas, lo que muestra que también es útil a nivel grupal y a nivel científico.

Es destacable especialmente la influencia de los razonamientos contrafácticos en los análisis de la *historiografía*, donde es necesario prestar atención a la reconstrucción imaginada de los hechos para realizar inferencias causales, por lo que estos pensamientos se convierten en una herramienta muy poderosa (para una revisión, ver Nolan, 2013).

De igual forma, distintos investigadores han analizado su función en las *ciencias políticas* precisamente porque permiten establecer un análisis causal entre un antecedente concreto del pasado y la realidad actual (Catellani y Covelli, 2013).

En la *economía*, se han desarrollado estudios en el área del marketing, centrados sobre todo en el estudio del arrepentimiento anticipado del consumidor, lo que ha llevado al desarrollo de la *teoría de la regulación del arrepentimiento* por su papel en la toma de decisiones (Zeelenberg y Pieters, 2007).

La diversidad de estas disciplinas y el hecho de que los conocimientos que se ellas se generan estén interconectados, han convertido al estudio del razonamiento contrafáctico en un asunto multidisciplinar (Maielli y Booth, 2008).

Conclusiones

En primer lugar, la investigación en el razonamiento contrafáctico ha contribuido al desarrollo y convergencia de dos teorías esenciales bien diferenciadas en la psicología del pensamiento humano: la *teoría de los heurísticos y sesgos* y la *teoría de los modelos mentales*. Estas teorías han coincidido en demostrar experimentalmente y defender la idea, aceptada en la actualidad, de que las personas actúan en función de una racionalidad no normativa. Esta racionalidad presenta dos niveles de pensamiento: uno intuitivo y otro racional.

En segundo lugar, estas dos teorías originaron a su vez el desarrollo de las dos principales teorías explicativas del razonamiento contrafáctico: la *teoría de los heurísticos y sesgos*, junto a la *teoría normativa*, influyó en la *teoría funcionalista del razonamiento contrafáctico*, mientras que la *teoría de los modelos mentales* enmarcó la *teoría de la imaginación racional*. La consideración conjunta de ambas lleva a una conclusión final: la imaginación y la racionalidad no guiada por reglas formales, que son cualidades distintivas e intrínsecamente humanas y que

comparten propiedades con el pensamiento lógico, cumplen una función de aprendizaje para la supervivencia, proporcionando nuevas enseñanzas para el futuro.

Referencias

- Barker, S. (2003). A dilemma for the counterfactual analysis of causation. *Australasian Journal of Philosophy*, 81 (1), 62-77.
- Byrne, R.M.J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, 6, 426-431.
- Byrne, R.M.J. (2005). *The rational imagination. How people create alternatives to reality*. Cambridge, MA.: MIT Press.
- Byrne, R.M. J. (2016). Counterfactual Thought. *Annual Review of Psychology*, 67, 135-57
- Byrne, R.M.J. y McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1318-1331.
- Byrne, R.M.J. y Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition*, 27, 726-740.
- Byrne, R.M.J., Segura, S., Culhane, R., McAlinney, P., Tasso, A. y Berrocal, P. (2000). Cognitive processes in the temporality effect for counterfactual thinking about what might have been. *Memory & Cognition*, 28, 264-281.
- Catellani, P. y Milesi, P. (2001). Counterfactual and roles: Mock victims' and perpetrators' accounts of judicial cases. *European Journal of Social Psychology*, 31, 247-254.
- Catellani, P. y Covelli, V. (2013). The strategic use of counterfactual communication in politics. *Journal of Language and Social Psychology*, 32 (4), 480-489.
- Chisholm, R.M. (1946). The contrary to fact conditional. *Mind*, 55. 189-307.
- Epstude, K. y Roese, N.J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, 12, 168-192.
- Epstude, K. y Roese, N.J. (2011). When goal pursuit fail. The functions of counterfactual thought in intention formation. *Social Psychology*, 42, 19-27.
- Evans, J.B.T. (2014). Two minds rationality. *Thinking & Reasoning*, 20 (2), 129-146.
- Feeney, A. y Handley, S.J. (2006). Mental models, comparisons, and the action effect in judgments of regret. *Memory & Cognition*, 34, 1422-1430.
- Giroto, V., Legrenzi, P. y Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, 78, 111-133.
- Gubbins, E. y Byrne, M.J.R. (2014). Dual processes of emotion and reason in judgments about moral

- dilemmas. *Thinking & Reasoning*, 20 (2), 245-268.
- Goodman, N. (1973). *Fact, fiction and forecast*. Nueva York: Bobbs-Merril.
- Johnson-Laird, P.N. (1994). Mental Models and Probabilistic Thinking. *Cognition*, 50 (1), 189-209.
- Johnson-Laird, P.N. (2006). *How we reason*. Oxford: Oxford University Press.
- Johnson-Laird, P.N. y Byrne, R.M.J. (1991). *Deduction*. Hove: Lawrence Erlbaum Associates.
- Johnson-Laird, P.N. y Byrne, R.M.J. (1993). Precip of Deduction. *Behavior and Brain Sciences*, 16 (2), 323-380.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Nueva York: Farrar, Straus and Giroux.
- Kahneman, D. y Miller, D.T. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136-153.
- Kahneman, D. y Tversky, A. (1982). The simulation heuristic. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 201-208). Nueva York: Cambridge University Press.
- Kray, L.J., George, L.G., Liljenquist, K.A., Galinsky, A.D. y Tetlock, P.E. (2010). From what have been to what must have been: Counterfactual thinking creates meaning. *Journal of Personality and Social Psychology*, 98, 106-118.
- Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin*, 13, 524-536.
- Leithy, S.E., Brown, G.P. y Robbins, I. (2006). Counterfactual thinking and posttraumatic stress reactions. *Journal of Abnormal Psychology*, 115, 629-235.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82 (4), 276-298.
- Maielli, G. y Booth, C. (2008). Counterfactual history, management and organizations: Reflections and new directions. *Management & Organizational History*, 3 (1), 49-61.
- Mandel, D.R. (2003). Counterfactual, emotions, and context. *Cognition & Emotion*, 17, 139-159.
- McCloy, R. y Byrne, R.M.J. (2000). Counterfactual thinking about controllable events. *Memory & Cognition* 28, 1071-1078.
- McEleney, A. y Byrne, R.M.J. (2006). Spontaneous counterfactual thoughts and causal explanations. *Thinking & Reasoning*, 12, 235-255.
- McMullen, M.N. y Markman, K.D. (2000). Downward counterfactuals and motivation: The wake-up call and the Pangloss effect. *Personality and Social Psychology Bulletin*, 26, 575-584.
- Mental Models in Thinking and Reasoning (2016). *Mentalmodelsblog* [<https://mentalmodelsblog.wordpress.com>].
- Miller, D.T. y Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, 59, 1111-1118.
- Miller, D.T., Turnbull, W. y McFarland, C. (1990). Counterfactual thinking and social perception: Thinking about what might have been. En M.P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23) (pp. 305-331). Nueva York: Academic Press.
- Morris, M.W. y Moore, P.C. (2000). The lesson we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, 45, 737-765.
- Nolan, D. (2013). Why historians (and everyone else) should care about counterfactuals. *Philosophical Studies*, 163 (2), 317-335.
- Nute, D. (1984). Conditional logic. En D. Gabbay y F. Guenther (Eds.), *Handbook of Philosophical Logic*. Dordrecht: Reidel Publishing Company.
- Petrocelli, J.V., Seta, C.E., Seta, J.J. y Prince, L.B. (2012). "If only I could stop generating counterfactual thoughts": When counterfactual thinking interferes with academic performance. *Journal of Experimental Social Psychology*, 48 (5), 1117-1123.
- Rescorla, R.A. y Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2, 64-99.
- Roese, N.J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66, 805-818.
- Roese, N.J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121, 133-148.
- Roese, N.J. y Morrison, M. (2009). The Psychology of Counterfactual Thinking. *Historical Social Research*, 34 (2), 16-26.
- Roese, N.J. y Hur, T. (1997). Affective Determinants of Counterfactual Thinking. *Social Cognition*, 15, 274-290.
- Roese, N.J. y Olson, J.M. (1995). Counterfactual thinking: A critical overview. En N.J. Roese y J.M. Olson (Eds.), *What might have been: The Social Psychology of Counterfactual Thinking* (pp. 1-59). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schmidt, R.E. y Van der Linden, M. (2009). The aftermath of rash action: Sleep-interfering counterfactual thoughts and emotions. *Emotion*, 9, 549-553
- Segura, S., Fernández-Berrocal, P. y Byrne, R.M.J. (2002). Temporal and causal order effects in thinking about what might have been. *Quarterly Journal of Experimental Psychology*, 55, 1295-1305.

- Segura, S. y Morris, M.W. (2005). Scenarios simulations in learning: Forms and functions at the individual and organizational levels. En D.R. Mandel, D.J. Hilton y P. Catellani (Eds.), *The Psychology of Counterfactual Thinking* (pp. 94-109). Nueva York: Routledge.
- Stalnaker, R.C. (1968). A Theory of Conditionals. En N. Rescher (Ed.), *Studies in Logical Theory*. Oxford: Basil Blackwell.
- Tversky, A. y Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Wells, G.L., Taylor, B.R. y Turtle, J.W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421-430.
- Woodward, J. (2004). Counterfactuals and causal explanation. *International Studies in the Philosophy of Science*, 18 (1), 41-72.
- Waytz, A., Hershfield, H.E. y Tamir, D. (2014). Mental simulation and meaning in life. *Journal of Personality and Social Psychology*, 108 (2), 336-355.
- Zeelenberg M. y Pieters R. (2007). A Theory of Regret Regulation 1.0. *Journal of Consumer Psychology*, 17, 3-18.