Anuario de Psicología
2002, vol. 33, nº 1, 25-47

Inferencias bayesianas: una revisión teórica

Diego Alonso
Universidad de Almería
Elisabet Tubau
Universitat de Barcelona

En este artículo se analizan los enfoques teóricos más relevantes sobre razonamiento bayesiano, así como sus principales apoyos empíricos. Muchos errores en estimaciones de probabilidad se atribuyen a la incapacidad de las personas para pensar en términos estadísticos cuando se enfrentan con información sobre un suceso único. Específicamente, en situaciones en las que el modelo normativo es el teorema de Bayes, se analiza el sesgo conocido como «insensibilidad a las probabilidades previas», tanto desde el enfoque de heurísticos y sesgos, de Tversky y Kahneman, como desde la hipótesis frecuentista de Gigerenzer, Cosmides et al. *También se aborda la discrepancia entre intuiciones y razonamiento matemático formal a través de los estudios sobre el problema de los tres prisioneros. Finalmente se presenta la teoría de Johnson-Laird* et al. *sobre modelos mentales en razonamiento probabilístico extensional, que explica cómo personas no expertas pueden inferir probabilidades a posteriori sin utilizar el teorema de Bayes.*
Palabras clave: *heurísticos y sesgos, razonamiento probabilístico, teorema de Bayes, probabilidades a priori, modelos mentales, razonamiento bayesiano, hipótesis frecuentista.*

This article analyzes the leading theoretical approaches to Bayesian reasoning in the literature, and its main sources of empirical support. Many errors in probability estimates are attributed to people's inability to think in statistical terms when faced with information about a single event. Specifically, in situations where the normative model is Bayes's theorem, the well-known base rate neglect is analyzed both from the heuristic and biases approach, of Tversky and Kahneman, and from the frequentist hypothesis defended by Gigerenzer, Cosmides et al. *The discre-*

Correspondencia: Diego Alonso. Área de Psicología Básica. Departamento de Neurociencia y Ciencias de la Salud. Facultad de Humanidades y CCEE. Universidad de Almería. 04120 Almería. Correo electrónico: dalonso@ual.es
Elisabet Tubau. Departament de Psicologia Bàsica. Universitat de Barcelona. Pg. de la Vall d'Hebron, 171. 08035 Barcelona. Correo electrónico: etubau@psi.ub.es

pancy between intuitions and formal mathematical reasoning is also analyzed through the studies with the three prisoners problem. Finally, we present the mental model theory of extensional probabilistic reasoning (Johnson-Laird et al., *1999) which explains how naive individuals can infer posterior probabilities without relying on Bayes's theorem.*

Key words*: Heuristics and biases, probabilistic reasoning, Bayes's theorem, base rates, mental models, Bayesian reasoning, frequentism.*

¿Posee nuestro sistema cognitivo algún mecanismo capaz de llevar a cabo estimaciones de probabilidad que se ajusten a las prescripciones del teorema de Bayes? Cualquiera que se adentre en el estudio del razonamiento probabilístico estará de acuerdo en que esta pregunta es quizás la que más controversia ha suscitado en las tres últimas décadas. Sin embargo su origen podríamos situarlo en la declaración llevada a cabo por el matemático francés Pierre Laplace (1814/1951) a principios del siglo XIX: «en el fondo, la teoría de probabilidad no es nada más que buen sentido reducido a cálculo» (p. 196), o lo que es equivalente: las herramientas que la teoría de probabilidades había desarrollado hasta ese momento eran descripciones del pensamiento probabilístico. En la década de 1970, cuando ya la teoría de probabilidades está lo suficientemente desarrollada como para ser considerada por muchos como un criterio normativo con el que poder comparar las estimaciones de las personas, algunos psicólogos cognitivos comenzaron a observar que nuestros juicios probabilísticos no se ajustaban a lo esperado según esta teoría. Kahneman y Tversky (1972), en unos influyentes estudios, llegaron a la conclusión de que las personas no somos bayesianos en absoluto, en referencia al hecho de que nuestras inferencias probabilísticas no siguen el teorema de Bayes. La idea de que nuestra mente no está diseñada para funcionar aplicando las reglas de la teoría de la probabilidad se propagó rápidamente, basada en multitud de resultados de experimentos llevados a cabo principalmente por estos autores, quienes propusieron la existencia de reglas no estadísticas tales como el «heurístico de representatividad», que explicarían sesgos como la "falacia de la conjunción», o el no tener en cuenta el tamaño de la muestra o las probabilidades previas, entre otros. Recientemente, sin embargo, han surgido explicaciones alternativas a las mantenidas por Tversky y Kahneman. En este artículo se intenta presentar una revisión de los principales enfoques teóricos y los resultados más relevantes que sobre el tema del razonamiento bayesiano ha aportado la psicología cognitiva

Distintos significados del concepto de probabilidad

¿Tiene sentido hablar de la probabilidad de un suceso único? La respuesta a esta cuestión depende del punto de vista que se asuma sobre la naturaleza de la probabilidad. En el terreno conceptual existe una fuerte polémica en cuanto a la interpretación que debe darse al término «probabilidad», sobre el que se han pro-

nunciado tanto matemáticos como filósofos y psicólogos. En el lenguaje cotidiano, la probabilidad se suele entender en un sentido amplio. Así, por ejemplo, solemos hablar de la probabilidad de que llueva mañana, aun cuando se trata de un suceso único y no hay forma de medir su frecuencia repitiendo el «experimento». Además, diferentes personas pueden asignar diferentes probabilidades al mismo evento, debido a que cada persona tiene un bagaje de conocimientos previos y creencias distintos. En estadística clásica, sin embargo, el concepto de probabilidad es más limitado. Por una parte, la corriente de opinión llamada *escuela bayesiana* considera que la probabilidad es una medida subjetiva de creencias, un grado subjetivo de confianza en la ocurrencia de un determinado evento (por ejemplo, la probabilidad de que haya vida fuera de nuestro sistema solar). Las probabilidades bayesianas son siempre probabilidades condicionales, es decir, se llevan a cabo en el contexto de algunos supuestos previos. Por ejemplo, cuando estimamos la probabilidad de que llueva el próximo fin de semana, usamos nuestro conocimiento previo sobre la lluvia en esta época del año. Desde un enfoque filosófico, la concepción bayesiana es bastante simple. Sin embargo, las dificultades pueden surgir a la hora de encontrar procedimientos matemáticos para asignar probabilidades a los distintos sucesos. Pero esto es algo similar a lo que sucede cuando hablamos de «número de peces en un lago»: comprendemos lo que quiere decir pero es difícil de calcular. De la misma manera podemos estimar que diferentes sucesos tienen diferentes probabilidades aun cuando no sepamos exactamente cómo calcularlas. Una consecuencia importante de esta concepción de probabilidad es que admite la asignación de probabilidades a sucesos únicos (la probabilidad de que mi equipo de fútbol gane el próximo partido; la probabilidad de que María apruebe un examen teniendo en cuenta los resultados de otros exámenes y las horas dedicadas a estudiar).

Por otra parte, la corriente *frecuentista* considera que la probabilidad se refiere a la frecuencia relativa (o límite de frecuencias relativas) con que ocurre un suceso, en relación con una clase de referencia (por ejemplo, probabilidad de obtener dos caras en cuatro lanzamientos de una moneda), rechazando, por tanto, la asignación de probabilidades a eventos únicos porque no tendrían frecuencia relativa. Tanto la interpretación subjetiva como la frecuentista son consistentes con los axiomas de probabilidad de Kolmogorov (1950) (ver epígrafe siguiente). Es importante la distinción *bayesiano/frecuentista* porque, como se verá más adelante, algunos autores la invocan para argumentar a favor o en contra de la existencia de sesgos sistemáticos en las estimaciones de probabilidad que llevamos a cabo las personas.

Probabilidad condicionada. Teorema de Bayes

La probabilidad se define mediante tres axiomas: (1) la probabilidad de un suceso es un número mayor o igual que cero, (2) la probabilidad del suceso seguro es igual a 1, y (3) la probabilidad de la unión de dos sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos. La probabilidad

condicionada se puede especificar en un cuarto axioma: si A y B son dos sucesos, entonces la probabilidad del suceso A, supuesto que se cumpla B (probabilidad de A condicionado a B), se representa así $p\ (A \mid B)$ y vale:

$$p\ (A \mid B) = p\ (A \text{ y } B) / p\ (B), \text{ (supuesto } p\ (B) \neq 0).$$

Con frecuencia, en muchas ramas de la ciencia, los investigadores quieren saber con qué probabilidad unos determinados datos D apoyan una hipótesis H_i, es decir, $p\ (H_i \mid D)$. Esto es lo que se conoce como probabilidad a posteriori. El teorema de Bayes es una fórmula que permite obtener esta probabilidad condicionada a partir de los valores de otras probabilidades. En su versión más simple se expresa así:

$$p\ (H_i \mid D) = \frac{p\ (H_i)\ p\ (D \mid H_i)}{p\ (D)} \quad (p\ (D) \neq 0)$$

Por tanto, la probabilidad a posteriori de una hipótesis H_i, dados unos datos D, depende de las probabilidades previas de los datos y de la hipótesis, y de la probabilidad condicionada de los datos supuesta la veracidad de la hipótesis.

En el caso de que n hipótesis $H_1, H_2, \ldots, H_n$ constituyesen una *partición* (es decir, que fuesen globalmente exhaustivas y mutuamente excluyentes), entonces $p\ (D)$ podría expresarse así:

$$p\ (D) = p\ (H_1)\ p\ (D \mid H_1) + p\ (H_2)\ p\ (D \mid H_2) + \ldots + p\ (H_n)\ p\ (D \mid H_n),$$

y, en consecuencia, la fórmula de Bayes también se podría enunciar de esta otra manera:

$$p\ (H_1 \mid D) = \frac{p\ (H_1)\ p\ (D \mid H_1)}{p\ (H_1)\ p\ (D \mid H_1) + p\ (H_2)\ p\ (D \mid H_2) + \ldots + p\ (H_n)\ p\ (D \mid H_n)}$$

Esta fórmula es un componente importante del cálculo de probabilidades, sea cual sea la interpretación –subjetivista o frecuentista- de probabilidad que una persona adopte. Así, podría aplicarse tanto para calcular la probabilidad de un suceso único como para calcular una frecuencia relativa. Desde un punto de vista subjetivista, la fórmula de Bayes permite observar el grado de racionalidad con que una persona cambia sus creencias cuando consigue nueva información. Sin esta fórmula, las probabilidades subjetivas serían meras incertidumbres subjetivas difíciles de estudiar desde un punto de vista científico.

El teorema de Bayes nos permite ver cómo las probabilidades previas pueden modificar el valor de la probabilidad de una determinada hipótesis. Ahora bien, admitiendo que la fórmula de Bayes es un modelo normativo con el que podemos comparar los juicios de probabilidad emitidos por las personas, entonces, cuando éstas tienen que emitir una estimación de probabilidad en una situación

similar a la anterior, (lo que llamamos *inferencia bayesiana*) ¿se ajustan al resultado que se esperaría según el teorema de Bayes? Ésta es la cuestión central de este trabajo. Para evitar confusiones, aclararemos que el término *bayesiano/a*, a partir de ahora, hará referencia al teorema de Bayes, y no a una concepción bayesiana –subjetivista– del concepto de probabilidad.

Insensibilidad a las probabilidades previas. El heurístico de representatividad

Los primeros estudios experimentales sobre razonamiento bayesiano (Edwards, 1968; Phillips y Edwards, 1966; Rouanet, 1961) ofrecieron resultados que se interpretaron como que las personas somos bayesianos *conservadores*, es decir, nuestras inferencias sobre probabilidades a posteriori son proporcionales a las obtenidas aplicando el teorema de Bayes, aunque menos extremas.

Sin embargo, los abrumadores resultados obtenidos por Kahneman y Tversky (1972) mostraron que las personas, cuando se enfrentan a un juicio estimativo de la probabilidad de que un determinado ejemplar pertenezca a una categoría, ignoran o infraponderan significativamente las probabilidades previas o tasas básicas de frecuencia, aun cuando se les presenten explícitamente. Además, en estos casos, los juicios parecen gobernados por consideraciones tales como el grado con que un ejemplar es representativo de una categoría. En uno de los experimentos (Kahneman y Tversky, 1973), los participantes leían breves descripciones de diferentes individuos, supuestamente extraídas al azar de un grupo constituido por 30 ingenieros y 70 abogados (o 70 ingenieros y 30 abogados). Se les pedía que evaluaran la probabilidad de que cada descripción correspondiese a un ingeniero. Los juicios de los participantes infraponderaron las probabilidades previas de cada categoría. El tamaño del efecto de estas probabilidades previas fue estadísticamente significativo pero pequeño y, en todo caso, muy lejos de lo que debiera haber ocurrido si los participantes hubieran estado usando el teorema de Bayes. Es decir, los participantes se basaban fundamentalmente en su conocimiento sobre el perfil típico de un ingeniero o abogado a la hora de hacer estas estimaciones de probabilidad. La autenticidad, robustez y generalidad de este fenómeno de insensibilidad a las probabilidades previas llegó a considerarse como un hecho claramente verificado (Bar-Hillel, 1980; Borgida y Brekke, 1981). Se propuso que, como resultado de nuestras limitadas capacidades de procesamiento, las personas tienen que calcular la probabilidad de un evento utilizando reglas no estadísticas tales como el *heurístico de representatividad*. Este término hace referencia a que ante preguntas del tipo «*¿Cuál es la probabilidad de que A pertenezca a la clase B?*», las personas evaluamos esta probabilidad a través del grado con que A es representativo de la clase B, es decir, el grado en que A se parece a B. Cuanto mayor sea el parecido de un objeto a una clase de objetos, mayor será el valor de nuestro juicio de probabilidad de que el objeto pertenezca a la clase.

Las pruebas más importantes sobre el papel que juega la representatividad en juicios predictivos provienen de los experimentos realizados usando el para-

digma metodológico denominado «de ordenación de resultados». En este paradigma, a los participantes se les proporcionan características de personalidad de un individuo y se les pide que ordenen una serie de elementos (p. ej., ocupaciones o tipos de estudios) con arreglo a ciertos criterios. En una condición el criterio es la representatividad (o sea, el grado con el que la persona se asemeja al prototipo asociado con cada elemento). En otra condición los participantes tienen que ordenar los mismos elementos con arreglo a la mayor o menor probabilidad con la que se pueden aplicar a la persona descrita. En la tercera condición, los participantes no reciben la información sobre el individuo; sólo tienen que ordenar los elementos según las tasas básicas de frecuencia con que aparecen en la población de la que el caso ha sido sacado. Los resultados de varios experimentos mostraron que las ordenaciones por probabilidad eran casi idénticas a las de representatividad (Kahneman y Tversky, 1973; Tversky y Kahneman, 1982), pero diferían bastante de las hechas con arreglo a las tasas básicas de frecuencia.

La perspectiva de que las personas ignoramos las probabilidades previas ha dominado la investigación empírica durante muchos años. La falacia de la insensibilidad a las probabilidades previas, y su explicación en términos del heurístico de representatividad, llegaron a alcanzar la categoría de hechos universalmente aceptados en la comunidad científica. Sin embargo, tras una minuciosa revisión de la literatura sobre el tema (especialmente sobre el problema de los abogados e ingenieros, y sobre experimentos con estereotipos sociales), Koehler (1996) concluye lo contrario, es decir, que las tasas básicas de frecuencia se usan casi siempre, y que su nivel de uso depende de la estructura de la tarea y de la representación mental que induzca. En el caso particular del problema de los abogados e ingenieros se han identificado algunos factores que aumentan el uso de las tasas base de frecuencias: presentar la información sobre las tasas base después de las descripciones de personalidad de los individuos (Krosnick, Li, y Lehman, 1990), variar la tasa base a través de los ensayos (Bar-Hillel y Fischhoff, 1981), y enfatizar en las instrucciones el que los participantes piensen como estadísticos (Schwarz, Strack, Hilton, y Naderer, 1991).

A lo largo de una serie de publicaciones, Gigerenzer (1991, 1994 y 1996), Gigerenzer *et al.* (1988) y Gigerenzer y Hoffrage (1995) –ver también Cosmides y Tooby, 1996– han llevado a cabo algunas críticas al enfoque de heurísticos y sesgos de Tversky y Kahneman, y en especial al hecho de que estos últimos autores califiquen como «errores» o «falacias» algunos de los efectos que encuentran (p. ej. la insensibilidad a las probabilidades previas). Las críticas se pueden dividir en tres grupos. En relación a los datos *empíricos*, la crítica se centra en el hecho de que algunos de los sesgos identificados por Tversky y Kahneman son inestables y su magnitud puede reducirse con sólo presentar la información y formular la pregunta en términos de frecuencias en vez de probabilidades (ver epígrafe siguiente para un desarrollo más amplio de esta crítica). Desde un punto de vista *teórico*, se suele criticar la ambigüedad y la poca potencia explicativa de términos tales como «representatividad». Y, finalmente, desde un enfoque *normativo* se argumenta que puede ser inapropiado caracterizar algunos de los sesgos identificados por Tversky y Kahneman como «errores» o «falacias» por varias razones, entre las que destaca el que desde una postura frecuentista no tiene

sentido asignar probabilidades a sucesos únicos, luego no se pueden comparar los juicios sobre estos sucesos con un modelo normativo probabilístico. En consecuencia, no hay tales sesgos, y los heurísticos son medios de explicar lo que no existe, según Gigerenzer (1991).

Cobos, Caño, y López (2000) proponen una explicación asociacionista de la «desestimación de frecuencias de categorías» en las tareas de categorización probabilística. Consideran que este sesgo (y también la falacia de la conjunción) se puede interpretar, en algunos casos, como consecuencia de la intervención de procesos de aprendizaje asociativo activados por los contenidos presentes en las tareas de categorización. Para estos autores, lo que subyace al razonamiento intuitivo concebido por Tversky y Kahneman (procesos rápidos, poco costosos, y activados por las propiedades de los estímulos) son los mecanismos de aprendizaje asociativo. Cuando las personas tienen que emitir un juicio de probabilidad sobre la relación existente entre dos contenidos que han jugado el papel de *clave y resultado* respectivamente en una tarea previa de aprendizaje predictivo, los mecanismos de aprendizaje asociativo generan automáticamente una cierta cantidad de expectación sobre la ocurrencia del resultado, que se interpreta como medida de la relación entre ambos contenidos, y a la que las personas tienen acceso directo y automático. El sesgo ocurriría cuando la situación experimental está diseñada de tal forma que la aceptación de este conocimiento como base para la respuesta en un juicio probabilístico viola alguno de los supuestos normativos del teorema de Bayes. Esta explicación cubre una de las deficiencias que se le han atribuido a la teoría de los sesgos y heurísticos de Tversky y Kahneman: su alto grado de imprecisión.

La hipótesis frecuentista

Una de las críticas que ha recibido el término representatividad es que es una noción vaga y mal definida (Gigerenzer y Murray, 1987; Shanteau, 1989; Wallsten, 1983; Gigerenzer y Hoffrage, 1995). Se le suele objetar que hasta ahora no se han especificado los procesos cognitivos subyacentes a este heurístico ni se han especificado las variables que pueden explicar el porqué unas veces se ignora la información contenida en las probabilidades previas y otras veces sí se tiene en cuenta (Ajzen, 1977; Bar-Hillel, 1980; Borgida y Brekke, 1981). Gigerenzer *et al.* mantienen que cuando los problemas aparecen presentados en formato frecuentista en vez de en formato probabilístico, las ilusiones cognitivas desaparecen y las personas llevamos a cabo correctamente inferencias bayesianas debido a que la información sobre frecuencias se corresponde con la manera secuencial en que los organismos adquieren información en contextos naturales (Gigerenzer y Hoffrage, 1995). Esta idea es compartida también por Cosmides y Tooby (1996). La «hipótesis frecuentista», por tanto, mantiene que «algunos de nuestros mecanismos de razonamiento inductivo incorporan aspectos de un cálculo de probabilidades, pero están diseñados para que, tanto los *inputs* como los *outputs* estén expresados en términos de frecuencias» (Cosmides y Tooby, 1996, p. 3, traducido del original).

contraintuitivos que, aunque puedan tener variaciones superficiales al ser diferentes en cuanto a la historia que cuentan y puedan presentar distintos valores sus parámetros numéricos, mantienen la misma estructura matemática: se presenta un suceso X con una determinada probabilidad, a continuación se introduce información adicional relativa a algún otro suceso del mismo espacio muestral, y al final se pregunta cuál será la nueva probabilidad del suceso X. Un problema análogo al de los tres prisioneros, que también ha merecido la atención de los investigadores es el llamado dilema de Monty Hall (Granberg, 1999; Granberg y Brown, 1995; Granberg, y Dorr, 1998; Alonso y Tubau, 2001; Tubau y Alonso, 2002).

A partir del paradójico enunciado original del problema de los tres prisioneros (Gardner, 1961; Mosteller, 1965), Shimojo e Ichikawa (1989) enunciaron una nueva versión que les permitió distinguir entre diferentes estrategias de solución que utilizan las personas. La versión original es la siguiente:

> Tres hombres, *A*, *B*, y *C* estaban en prisión. *A* sabía que uno de ellos sería puesto en libertad y que los otros dos iban a ser ejecutados. Pero no sabía quién sería el perdonado. Al carcelero, que sí lo sabía, *A* le dijo: «Puesto que dos de nosotros tres serán ejecutados, es seguro que al menos *B* o *C* lo serán. Usted no me proporciona ninguna información sobre mis propias posibilidades si me dice el nombre de un hombre, *B* o *C*, que vaya a ser ejecutado». El carcelero, después de pensarlo, aceptó este argumento y le dijo: «*B* será ejecutado». Al oír esto, *A* se sintió más feliz porque ahora o él o *C* serían liberados, luego sus posibilidades se habían incrementado de 1/3 a 1/2. Esta felicidad del prisionero *A* puede ser o no razonable. ¿Qué piensa usted?

La inmensa mayoría de las personas que se enfrentan a este problema tienden a considerar que las posibilidades de que *A* sea puesto en libertad han aumentado con esta información. Sin embargo, de acuerdo con el teorema de Bayes, las posibilidades de *A* no han cambiado: sigue siendo 1/3 (ver Apéndice II para una solución bayesiana razonada del problema). A partir de esta versión del problema, los autores derivan otra, modificando las probabilidades iniciales de ser liberado cada uno de ellos –*A*, *B*, o *C*–, de tal forma que estos tres sucesos no sean equiprobables, asignándoles los valores 1/4, 1/4, y 1/2, respectivamente, con lo que obtienen un problema más contraintuitivo que el original. Analizando las estimaciones que hacen los participantes en estos experimentos, así como los resultados de cuestionarios en los que los individuos tienen que justificar razonadamente sus respuestas, Shimojo e Ichikawa concluyen que los procesos psicológicos que dirigen los juicios intuitivos en estas tareas bayesianas son cualitativamente diferentes del razonamiento matemático. Incluso después de que los individuos hubieran comprendido el razonamiento matemáticamente correcto de estos problemas, seguían «sintiendo» que iba contra su intuición. Estas creencias intuitivas –heurísticos- sobre probabilidad (que los autores denominan «teoremas subjetivos») varían de un individuo a otro, pudiéndose categorizar en alguna de las siguientes tres principales intuiciones: (1) «Número de casos». Si el número de alternativas posibles es N, entonces la probabilidad de cada una de ellas es 1/N. (2) «Razón constante». Si una alternativa es eliminada, la razón entre las probabilidades de dos cualesquiera de las alternativas restantes no varía.

(3)»Irrelevante, luego invariable». Si es seguro que al menos una de varias alternativas (A_1, A_2, ..., A_k) será eliminada, la información que especifica qué alternativa será eliminada es irrelevante y no cambia las probabilidades de las otras alternativas (A_{k+1}, A_{k+2}, ..., A_N).

Como se puede observar fácilmente, las estimaciones de probabilidad que los individuos llevan a cabo dependen del «teorema subjetivo» que apliquen en cada caso. Los autores también afirman que los participantes no se ajustaban a las prescripciones del teorema de Bayes, por lo que sugieren «la existencia de un módulo mental de razonamiento intuitivo más o menos independiente del razonamiento matemático formal» (Shimojo e Ichikawa, 1989). Consideran también que la dificultad que los individuos tienen con este problema no se puede atribuir a la tendencia a ignorar las probabilidades previas en problemas bayesianos, ya que la utilización (por parte de muchos participantes) de los teoremas «razón constante» o «irrelevante, luego invariable» implica el uso explícito de estas probabilidades previas. En parte, la dificultad se debe –afirman Shimojo e Ichikawa– a la tendencia a ignorar el contexto en el que ocurre el evento (en este caso, la influencia de la respuesta del carcelero: éste dirá «*B* será ejecutado» cuando el liberado vaya a ser *C*, mientras que si el liberado fuese *A* tendría un 50% de probabilidad de que lo dijese, luego *C* tiene el doble de probabilidades que *A* de ser liberado).

El problema de los tres prisioneros ha sido analizado en otros estudios. En uno de ellos, Falk (1992) retoma el asunto de las «creencias espontáneas y heurísticos intuitivos», afirmando que las intuiciones más frecuentes son las de «número de casos» e «irrelevante, luego invariable», a las que llama «uniformidad» y «no-noticias, no cambio», respectivamente.

En opinión de Johnson-Laird *et al.* (1999), la carencia fundamental que presentan estos estudios sobre el problema de los tres prisioneros es la falta de una explicación en términos de representaciones mentales. Como veremos, Johnson-Laird *et al.* (1999) recogen las ideas de Shimojo e Ichikawa (1989) y Falk (1992) y las enmarcan dentro de la teoría de modelos mentales de razonamiento probabilístico, ofreciéndonos una visión más general de este tipo de razonamiento.

La teoría de los modelos mentales sobre razonamiento probabilístico de personas no expertas

Esta teoría, elaborada por Johnson-Laird, Legrenzi, Girotto, Sonino-Legrenzi, y Caverni (1999), intenta explicar cómo personas no expertas razonan sobre probabilidades. Es muy importante resaltar que los autores limitan el ámbito de aplicación de su teoría a lo que llaman «razonamiento *extensional* sobre probabilidades», donde el término *extensional* se debe entender como «inferir la probabilidad de un suceso a partir de las diferentes formas posibles en las que puede ocurrir». Este proceso sería deductivo, en contraposición al razonamiento *no-extensional* sobre la probabilidad de un suceso, que sería inductivo (p. ej. cuando consideramos que es muy probable que *A* pertenezca a la categoría *B* porque *A* es muy similar al prototipo de la categoría *B*).

La teoría de los modelos mentales sobre razonamiento probabilístico de personas no expertas está basada en la teoría de los modelos mentales (Johnson-Laird, 1983; Johnson-Laird y Byrne, 1991) desarrollada en principio para explicar la comprensión del discurso y el razonamiento deductivo. Un postulado básico de esta teoría afirma que ante tareas de razonamiento o de comprensión, las personas construimos modelos mentales de la situación. Según los autores, un modelo mental es la representación de una posibilidad que contiene lo que es común a todas las distintas formas en las que la posibilidad puede ocurrir. Esta teoría postula que los modelos mentales representan explícitamente lo que es verdadero, pero no lo que es falso. Esta representación incompleta puede explicar la aparición de errores sistemáticos en tareas de razonamiento.

Limitándonos al ámbito del razonamiento probabilístico extensional, la teoría se sustenta en cinco principios: (1) El principio de verdad. Las personas nos representamos situaciones construyendo conjuntos de modelos mentales en los que cada modelo representa lo que es verdad en una posibilidad verdadera. (2) El principio de equiprobabilidad. Cada modelo representa una alternativa equiprobable, a no ser que las personas tengan información o creencias de lo contrario, en cuyo caso les asignarán diferentes probabilidades a los distintos modelos. (3) El principio de proporcionalidad. Asumida la equiprobabilidad, la probabilidad de un suceso, A, depende de la proporción de modelos en los que ocurra el suceso. Es decir, $p\ (A) = n_A / n$, donde n_A es el número de modelos que contienen a A, y n es el número de modelos. (4) El principio numérico. Si una premisa hace referencia a una probabilidad numérica, los modelos pueden etiquetarse con su correspondiente valor numérico. Una probabilidad desconocida puede calcularse restando la suma de las $n-1$ probabilidades conocidas, de la probabilidad global de las n posibilidades en la partición. (5) El principio de subconjunto. Supuesta la equiprobabilidad, una probabilidad condicionada, p (A | B), depende del subconjunto de B que es A. La proporción de modelos de «A y B» en el conjunto de modelos de B daría el valor numérico. En los demás casos, si los modelos están etiquetados con sus frecuencias absolutas (o posibilidades), entonces la probabilidad condicionada es igual a la frecuencia (posibilidades) del modelo de "A y B" dividida por la suma de todas las frecuencias (posibilidades) de los modelos que contengan a B.

Veamos con un ejemplo el conjunto de modelos mentales que generaría un problema de probabilidad condicionada:

> Un paquete de cuatro cartas está formado por el as de oros, el as de espadas, el rey de oros, y el rey de espadas. Se extraen aleatoriamente dos cartas de ese paquete. ¿Cuál es la probabilidad de que salgan ambos ases sabiendo que ha salido al menos un as? Se trataría de hallar p (A | B), siendo A= «salir los dos ases», y B = «al menos sale un as». El conjunto de modelos mentales equiprobables de esta situación sería:

A	B
	B
	B
	B
	B
	...

donde cada fila representa una posibilidad verdadera. La teoría propone que el caso en que ambos sucesos son falsos (es decir, el caso «¬A y ¬B») se representa en un único modelo *implícito* (representado por la línea de puntos). Se trataría de un modelo sin contenido explícito, cuyo olvido explicaría la aparición de algunos errores (aunque no en este caso). El espacio muestral (nombre que recibe el conjunto de todos los sucesos posibles en un experimento aleatorio) asociado a este problema sería el siguiente:

> {(as espadas, as oros), (rey oros, as oros), (rey espadas, as oros), (rey oros, as espadas), (rey espadas, as espadas), (rey oros, rey espadas)}

La probabilidad pedida en el enunciado del problema se calcularía en virtud del principio de subconjunto, calculando la proporción entre modelos de «A y B» y modelos de B. Por tanto, obtendríamos p (A | B) = 1/5. El mismo resultado se habría obtenido si se hubieran construido los siguientes modelos numéricos:

		Frecuencias
A	B	1
	B	4
...		1

En este caso el resultado se obtendría dividiendo la frecuencia del modelo de «A y B» entre la suma de las frecuencias de B, o sea, también p (A | B) = 1/(1+4) = 1/5.

Según Johnson-Laird *et al.* (1999), las tareas de razonamiento sobre probabilidades condicionadas son especialmente difíciles para personas no expertas, por dos motivos:

En primer lugar, es difícil comprender que un determinado problema pueda requerir la utilización de probabilidades condicionadas. Un ejemplo citado por los autores para ilustrar esta dificultad es el siguiente (Bar-Hillel y Falk, 1982):

> Los Smith tienen dos bebés. Uno de ellos es una niña. ¿Cuál es la probabilidad de que el otro sea también una niña?

La respuesta más frecuente es 1/2. Pero es errónea. La teoría de los modelos mentales explica este error aduciendo que las personas interpretan que el problema pide la probabilidad de que un bebé sea «niña» y, en consecuencia, construyen los modelos:

Niño
Niña

y a partir de ahí deducen que p (niña)= 1/2. Pero, realmente lo que el problema pide es una probabilidad condicionada: p (las-2-sean-niñaslal-menos-1-es-niña).

Una sencilla justificación matemática de la solución sería la siguiente:

El espacio muestral se expresaría así:

{(niño, niño), (niño, niña), (niña, niño), (niña, niña)}

pero, al decirnos que «uno de ellos es niña», quedaría así:

{(niño, niña), (niña, niño), (niña, niña)}

por tanto (admitiendo que ambos sexos son igualmente probables), tendríamos:

p (2-niñaslal-menos-1-es-niña) = 1/3.

También en el caso del problema de los tres prisioneros mencionado antes, la teoría de los modelos mentales ofrece una explicación plausible del error que con más frecuencia suele aparecer. Los individuos suelen afirmar que la probabilidad de que el prisionero *A* sea liberado aumenta desde 1/3 hasta 1/2 al conocer la respuesta del carcelero. De acuerdo con esta teoría, los individuos comienzan con los siguientes modelos de posibilidades:

A	*B*	*C*
Liberado		
	Liberado	
		liberado

En consecuencia, al principio, la probabilidad de que el prisionero *A* sea liberado vale 1/3. La información que proporciona el carcelero («*B* será ejecutado») elimina las posibilidades de que *B* sea liberado, por tanto incorporan esta modificación a su conjunto de modelos, quedando así:

A	*C*
Liberado	
	Liberado

lo que les lleva a pensar que ambos (*A* y *C*) tienen la misma probabilidad de ser liberados, luego dirán que la probabilidad de que *A* sea liberado será 1/2 (Ver Apéndice III para una construcción correcta del conjunto de los modelos mentales de este problema).

En segundo lugar, debido a que los modelos representan sólo lo que es verdadero, se hace muy difícil distinguir entre dos relaciones de inclusión. Es decir, hallar p ($A \mid B$) supone relacionar «*A* y *B*» con *B*. El error se puede producir por una confusión a la hora de elegir el conjunto de referencia (en este caso, *B*). Si en lugar de relacionar «*A* y *B*» con *B*, lo relacionamos con A, entonces lo que obtendríamos será p (B | A). Consideremos, por ejemplo, la siguiente información:

> Al Sr. X se le ha pasado una prueba desarrollada para detectar cáncer de pulmón, y ha dado positivo. Si X no tiene cáncer, la probabilidad de que el test salga positivo es de 1 entre 10.000.

El enunciado nos dice que *p* (test-positivo|¬cáncer) = 1/10.000, (donde ¬cáncer representa la hipótesis alternativa «no tener cáncer»), y la representación de esta probabilidad condicionada es, según el principio numérico:

		Frecuencias
¬cáncer	test positivo	1
. . .		9.999

Si ahora se pregunta algo relacionado con p (¬cáncer|test-positivo), puede ser que la persona no experta la confunda con su inversa, p(test-positivo|¬cáncer) y, en consecuencia, dé como valor 1/10.000, aunque realmente el enunciado del problema no proporciona ninguna información sobre la probabilidad del conjunto de referencia (en este caso «test-positivo»).

A pesar de la dificultad intrínseca que tiene la realización de inferencias bayesianas, Johnson-Laird *et al.* (1999) consideran que su teoría ofrece una explicación convincente de por qué en algunos casos las personas no expertas pueden calcular probabilidades a posteriori sin utilizar el teorema de Bayes. Consideremos, por ejemplo, la siguiente versión esquemática del problema del diagnóstico médico:

> 4 de cada 100 personas tienen la enfermedad.
> 3 de cada 4 personas enfermas dieron positivo en el test.
> 12 de cada 96 personas sanas también dieron positivo en el test.
> Una persona seleccionada al azar ha dado positivo en el test.
> ¿Cuál es la probabilidad de que esa persona tenga la enfermedad?

Las personas no expertas podrían construir los siguientes modelos numéricos:

		Frecuencias
Enfermedad	Test positivo	3
Enfermedad		1
	Test positivo	12
. . .		84

y, a partir de aquí, aplicando el principio de subconjunto, obtendrían que la probabilidad de que esa persona tenga la enfermedad, dado el resultado positivo del test, es 3/15 (o, simplificado, 1/5). Es decir, habrían calculado la probabilidad a posteriori sin aplicar el teorema de Bayes. La teoría predice también que ante una versión probabilística del mismo problema, cuando la información se presenta en forma de «posibilidades» (por ejemplo, *«el Sr. X tiene 4 posibilidades entre 100 de tener la enfermedad»*), las personas construirían los mismos modelos mentales que con la versión frecuentista, con lo que ambas versiones tendrían un nivel de dificultad similar. Sin embargo, cuando la información numérica no facilita la construcción de los modelos mentales correctos debido a que no permite ver las relaciones de inclusión entre conjuntos, el problema será muy difícil de resolver. Los resultados obtenidos por Girotto y González (2001) confirman esta predicción, en contra de lo que cabría esperar si nuestras estimaciones

de probabilidad se ajustasen a la hipótesis frecuentista enunciada por Gigerenzer *et al.* (1995) y Cosmides *et al.* (1996).

La teoría de los modelos mentales ofrece una explicación plausible a los resultados de estudios en torno a la influencia que pueda tener el formato en el que se presenta la información: probabilístico *vs.* frecuentista. Los estudios más recientes ofrecen resultados difíciles de explicar desde la óptica frecuentista. Veamos algunos de ellos:

A lo largo de tres experimentos, Evans *et al.* (2000) han examinado las condiciones que pueden subyacer al hecho de que las personas hagan uso o no de la información contenida en las probabilidades previas. Presentando a los participantes distintas versiones del problema del diagnóstico médico llegan a la conclusión de que el formato probabilístico no implica necesariamente mayor dificultad que el formato frecuentista. Incluso en algunas ocasiones puede ser más sencillo. Afirman que las dos versiones que comparan Cosmides *et al.* (1996) no se diferencian sólo en el formato –probabilístico *vs.* frecuentista–, sino que la información numérica de esta última permite con mayor facilidad la formación de modelos mentales de inclusión de conjuntos, y es éste el factor causante del mayor porcentaje de respuestas correctas. En consonancia con esta idea, observan que en otra versión también frecuentista, pero con datos numéricos que no facilitan la formación de modelos mentales, se mantiene el mismo nivel de dificultad que con la versión probabilística clásica (con datos expresados como porcentajes). La Tabla I presenta un ejemplo de dos versiones del problema del diagnóstico médico, matemáticamente equivalentes, que ilustra la distinta dificultad en construir modelos mentales a partir del enunciado del problema.

Mientras que la Versión 1 (véase Tabla I) permite llevar a cabo fácilmente una partición de la muestra y calcular la p(enfermedad|test-positivo) mediante la división entre la frecuencia del modelo de «enfermedad y test-positivo» y la de «test-positivo» (en este caso, 3/27), la Versión 2 no sugiere el conjunto de modelos mentales tan fácilmente. Análogamente, si en la Versión 1 se cambiara la frase «*24 de cada 96 personas sanas también dieron positivo en el test*» por esta otra matemáticamente equivalente a ella, «*1 de cada 4 personas sanas también dieron positivo en el test*», ya no sería tan obvia la frecuencia de cada uno de los modelos, y, en consecuencia, aumentaría la dificultad del problema, aun cuando no se hubiera variado el formato –frecuentista– en el que la información se ha presentado. Éste es el tipo de manipulación de variables que realizan Evans *et al.* (2000) para llegar a rechazar la hipótesis de Gigerenzer *et al.* (1995) y Cosmides *et al.* (1996). Además, Evans *et al.* (2000) encuentran otro sesgo que va en contra de la hipótesis frecuentista. A pesar de que la «insensibilidad a las probabilidades previas» es el sesgo que más se menciona en la literatura especializada, estos autores observan el sesgo opuesto, consistente en dar como respuesta precisamente la probabilidad previa que se menciona en el problema, ignorando la información diagnóstica. Lo llamativo es que este sesgo lo encuentran, precisamente, cuando la información se presenta en formato frecuentista o bien cuando a los individuos se les pide que expresen la respuesta en formato frecuentista.

TABLA 1. DOS VERSIONES MATEMÁTICAMENTE EQUIVALENTES DEL PROBLEMA DEL DIAGNÓSTICO MÉDICO Y EL CONJUNTO DE MODELOS MENTALES DE ESTA SITUACIÓN. SE PUEDE OBSERVAR LA DISTINTA FACILIDAD CON QUE, PARTIENDO DE CADA UNO DE ELLOS, SE PUEDEN CONCEBIR LOS MODELOS MENTALES CORRECTOS

PROBLEMA DEL DIAGNÓSTICO MÉDICO		
Versión 1 (fácil)		Versión 2 (difícil)
– 4 de cada 100 personas tienen la enfermedad. – 3 de cada 4 personas enfermas dieron positivo en el test. – 24 de cada 96 personas sanas también dieron positivo en el test. – Una persona seleccionada al azar ha dado positivo en el test. – ¿Cuál es la probabilidad de que esa persona tenga la enfermedad?		– Una persona tiene un 4% de probabilidad de tener la enfermedad. – Hay un 75% de probabilidad de dar positivo en el test, si la persona está enferma. – Hay un 25% de probabilidad de dar positivo, si la persona no está enferma. – El Sr. X ha dado positivo en el test. – ¿Cuál es la probabilidad de que realmente esté enfermo?
		Frecuencias (o posibilidades)
Enfermedad	Test positivo	3
Enfermedad		1
	Test positivo	24
. . .		72

Los estudios llevados a cabo por Girotto *et al.* (2001) van en la misma dirección de los de Evans *et al.* (2000). Intervienen en la polémica en torno a la validez de la hipótesis frecuentista aportando resultados que contravienen esta hipótesis y la explicación evolucionista en que se basa. Es posible que tengamos un módulo que extraiga frecuencias de una *secuencia real de observaciones*, pero la mayoría de los estudios no han utilizado estas secuencias reales, sino *enunciados verbales* en los que los participantes tienen que realizar inferencias a partir de frecuencias representadas por símbolos numéricos. Por tanto, la tesis evolucionista también tendría que asumir que los símbolos numéricos que expresan frecuencias son más «naturales» que los que expresan probabilidades de sucesos únicos, y como consecuencia las personas deberíamos llevar a cabo mejor los problemas enunciados con frecuencias que con probabilidades. Sin embargo, como demuestran en su estudio Girotto *et al.* (2001), los resultados van en contra de esta suposición. Como explicación alternativa, estos autores sugieren que las personas no expertas pueden realizar inferencias bayesianas correctamente siempre que la estructura en que se presente la información permita a los individuos representarse mentalmente la partición de la muestra en subconjuntos a partir de los que se puedan establecer relaciones de inclusión, independientemente del formato –frecuentista *vs.* probabilístico– en que esté enunciada la información. Como puede observarse, lo que afirman es que los individuos podemos calcular probabilidades condicionadas sin necesidad de aplicar el teorema de Bayes, sin más que tener en cuenta el principio de subconjunto de la teoría de los modelos mentales de razonamiento probabilístico de Johnson-Laird *et al.* (1999).

Conclusiones

En el estudio de las inferencias bayesianas el enfoque que ha predominado ha sido el de los heurísticos y sesgos de Tversky y Kahneman, aunque en los últimos 15 años han ido apareciendo algunas críticas dirigidas tanto a aspectos empíricos como metodológicos o normativos. Hay un cierto grado de acuerdo en que el enfoque de Tversky y Kahneman maneja términos explicativos ambiguos (como, «representatividad», por ejemplo) y carece de un desarrollo teórico que detalle las variables subyacentes al hecho de que las probabilidades previas se utilicen o no. La hipótesis frecuentista aborda este problema y mantiene que, cuando la información se presenta en formato frecuentista, las personas hacemos uso de las tasas básicas de frecuencias. Sin embargo, los recientes estudios experimentales llevados a cabo tanto por Evans *et al.* (2000) como por Girotto *et al.* (2001) arrojan serias dudas sobre la viabilidad de esta hipótesis, pudiendo concluirse que la sustitución de los juicios de probabilidad subjetiva por estimaciones de frecuencias relativas no proporciona una panacea contra la insensibilidad a las probabilidades previas.

Los estudios de Shimojo e Ichikawa (1989) y Falk (1992) sobre el problema de los tres prisioneros aportaron un pequeño número de creencias o heurísticos que los individuos aplican de forma intuitiva, y que guardan cierto paralelismo con los principios expuestos por Johnson-Laird *et al.* (1999) en su teoría de los modelos mentales de razonamiento probabilístico.

De la revisión teórica y experimental realizada se puede afirmar que, actualmente, la teoría de los modelos mentales sobre razonamiento probabilístico de personas no expertas de Johnson-Laird *et al.* (1999) es el marco teórico más adecuado para explicar los procesos de inferencia probabilística bayesiana, si bien, como los propios autores especifican, su teoría sólo es aplicable a lo que llaman razonamiento extensional sobre probabilidades, donde el término extensional nos indica que se puede calcular la probabilidad de un suceso a partir de las diferentes maneras en que puede ocurrir, teniendo la característica de ser algorítmico y, por tanto, deductivo. En este sentido, algunos procesos de razonamiento *inductivo* relativos a probabilidades *subjetivas* en los que no se proporciona información explícita suficiente para que el individuo pueda formarse una representación mental del espacio muestral, quedan fuera de este marco explicativo. En estos casos, si se acepta la posibilidad de que a sucesos de caso único se les puedan asignar probabilidades, el enfoque de heurísticos y sesgos de Tversky y Kahneman, aunque no haya alcanzado un nivel óptimo de detalle en cuanto a los procesos mentales implicados, sigue siendo aceptado como la explicación más plausible, puesto que, como se ha dicho más arriba, la contrastación empírica de la hipótesis frecuentista ha producido resultados contradictorios.

Desde el lado práctico, se pueden extraer algunas recomendaciones pedagógicas dirigidas a proporcionar un método que facilite la realización de inferencias bayesianas y que, en general, sirva de ayuda a las personas en situaciones de toma de decisiones. La forma de mejorar las estimaciones bayesianas no es enseñar el teorema de Bayes, sino enseñar a los individuos a mejorar la representación mental de situaciones simples y sus principios intuitivos. De acuerdo con Gige-

renzer *et al.* (1995), para mejorar el razonamiento de los no expertos hay que enseñar representaciones en lugar de reglas. Con esta misma intención pedagógica, Johnson-Laird *et al.* (1999) sugieren el siguiente método para el cálculo de probabilidades a posteriori sin necesidad de utilizar el teorema de Bayes: (1) Representar la situación inicial. (2) Hacer explícitas las relaciones condicionales relevantes sobre la base de las suposiciones y procesos que generan los datos. (3) Construir un diagrama de las posibilidades equiprobables (una partición similar a un conjunto de modelos mentales). (4) Usar el principio de subconjunto.

Desde este mismo enfoque pedagógico, los autores de este artículo están estudiando la posibilidad de que se puedan modificar los modelos mentales causantes de la respuesta incorrecta que la mayoría de las personas suelen dar al problema de Monty Hall. Los resultados preliminares de este estudio permiten afirmar que con un adecuado procedimiento, donde se hacen explícitas las relaciones condicionales relevantes (p. ej. condiciones de eliminación), los individuos consiguen formar modelos más completos capaces de modificar las intuiciones iniciales (Tubau y Alonso, 2002; Tubau, Alonso y Moliner, en preparación).

Referencias

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on predictions. *Journal of Personality and Social Psychology, 35*, 303-314.

Alonso, D. y Tubau, E. (2001). Disociación en aprendizaje de probabilidades en un problema contraintuitivo. XIII Congreso de la Sociedad Española de Psicología Comparada. San Sebastián, 17-19 de septiembre de 2001.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*, 211-233.

Bar-Hillel, M. A. & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition, 11*, 109-122.

Bar-Hillel, M. & Fischhoff, B. (1981). When do base rates affect predictions? *Journal of Personality and Social Psychology, 41*, 671-680.

Billingsley, P. (1995). *Probability and Measure*. New York: Wiley Interscience.

Borgida, E. & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. En J. H. Harvey, W. J. Ickes & R. F. Kidd (Eds.), *New directions in attribution research* (pp. 66-95). Hillsdale, NJ: Erlbaum.

Casscells, W., Schoenberger, A. & Grayboys, T. (1978). Interpretation by phisicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999-1000.

Cobos, P. L., Caño, A. & López, F. J. (2000). Biases in probabilistic reasoning may be produced by associative learning mechanisms. En J. A. García-Madruga, N. Carriedo & M. J. González-Labra (Eds.), *Mental Models in Reasoning* (pp. 155-178). Madrid. UNED.

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1-73.

Edwards, W. (1968). Conservatism in human information processing. En B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.

Evans, J. St. B. T., Handley, S. J., Perham, N., Over, D. E. & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition, 77*, 197-213.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition, 43*, 197-223.

Gardner, M. (1961). *The second Scientific American book of mathematical puzzles and diversions*. New York: Simon & Schuster.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond «heuristics and biases». En W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83-115). Chichester, UK: Wiley.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). En G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-161). Chichester, UK: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, No. 3, 592-596.
Gigerenzer, G., Hell, W. & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513-525.
Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological Review, 102*, 684-704.
Gigerenzer, G. & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
Girotto, V. & González, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition, 78*, 247-276.
González-Labra, M. J. (2000). Content presentation in reasoning about base rates. En J. A. García-Madruga, N. Carriedo & M. J. González-Labra (Eds.), *Mental Models in Reasoning* (pp. 143-153). Madrid. UNED.
Granberg, D. (1999). A new version of the Monty Hall Dilemma with unequal probabilities. *Behavioural Processes, 48*, 25-34.
Granberg, D. & Brown, T. A. (1995). The Monty Hall Dilemma. *Personality and Social Psychology Bulletin, vol. 21, No. 7*, 711-723.
Granberg, D. & Dorr, N. (1998). Further exploration of two stage decision making in the Monty Hall Dilemma. *American Journal of Psychology, 111*, 561-579.
Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, England: Cambridge University Press.
Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Sonino-Legrenzi, M. & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62-88.
Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3*, 430-454.
Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.
Koehler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1-53.
Kolmogorov, A. (1950). *Foundations of probability theory*. New York: Chelsea.
Krosnick, J. A., Li, F. & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology, 59*, 1140-1152.
Laplace, P.-S. (1951). *A philosophical essay on probabilities*. New York: Dover. (Obra original publicada en 1814).
Mosteller, F. (1965). *Fifty challenging problems in probability with solutions*. Reading, MA: Addison-Wesley.
Phillips, L. D. & Edwards, W. (1966). Conservatism in a simple probability model inference task. *Journal of Experimental Psychology, 72*, 346-354.
Rouanet, H. (1961). Études de décisions expérimentales et calcul de probabilités. En *Colloques internationaux du centre national de la recherche scientifique* (pp. 33-43). Paris: Éditions du Centre National de la Recherche Scientifique.
Schwarz, N., Strack, F., Hilton, D. J. & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of «irrelevant» information. *Social Cognition, 9* (1), 67-84.
Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments and observations. *Accounting Organizations and Society, 14* (1/2), 165-177.
Shimojo, S. & Ichikawa, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the «problem of three prisoners». *Cognition, 32*, 1-24.
Tubau, E. y Alonso, D. (2002). Modelos mentales en un problema contraintuitivo: experiencias que posibilitan cambiarlos. IV Congreso de la Sociedad Española de Psicología Experimental. Oviedo, abril de 2002.
Tubau, E., Alonso, D. & Moliner, J.L. (en preparación). Learning to reason in a counterintuitive probabilistic problem.
Tversky, A. & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. En M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Tversky, A. & Kahneman, D. (1982). Judgments of and by representativeness. En D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84-98). Cambridge, England: Cambridge University Press.
Wallsten, T. S. (1983). The theoretical status of judgmental heuristics. En R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 21-39). Amsterdam: Elsevier (North-Holland).

APÉNDICE I

Solución al problema del diagnóstico médico

Llamemos D al suceso «dar positivo en el test», y sea H el suceso «tener la enfermedad». Entonces, por la información proporcionada en el texto del problema, sabemos que: $p(H) = 1/1000 = 0.001$, y $p(D \mid \neg H) = 5/100 = 0.05$ (donde $\neg H$ representa al suceso contrario de H). Si p $(H) = 0.001$, entonces, la probabilidad del suceso contrario será $p(\neg H) = 1 - 0.001 = 0.999$. Asumiendo que el test diagnostica la enfermedad a quien realmente la tiene (esta información no se proporcionaba en la versión original del problema), entonces, $p(D \mid H) = 1$. El objetivo es hallar la probabilidad de que una persona tenga la enfermedad, sabiendo que ha dado positivo en el test, luego se trata de calcular $p(H \mid D)$. Por tanto, aplicando la fórmula de Bayes, se obtiene:

$$p(H \mid D) = \frac{p(H)\, p(D \mid H)}{p(H)\, p(D \mid H) + p(\neg H)\, \mathrm{p}(D \mid \neg H)} = \frac{0.001 \times 1}{0.001 \times 1 + 0.999 \times 0.05} = 0.02$$

o sea, el 2% aproximadamente.

Otra forma de llegar al mismo resultado, razonando con frecuencias, sería la siguiente: de cada 1000 personas, sólo 1 tiene la enfermedad, luego 999 no la sufren. De estos 999, el 5%, o sea, aproximadamente 50 darán positivo en el test. Asumimos de nuevo que el test diagnostica correctamente al que tiene la enfermedad. Por tanto, de los 51 diagnosticados como enfermos, sólo uno lo estará realmente, luego la probabilidad de que una persona diagnosticada como enfermo por el test lo esté realmente, será 1/51, es decir, el 2%.

APÉNDICE II

Solución al problema de los tres prisioneros

Antes de la conversación de A con el carcelero, las posibilidades de supervivencia de cada uno de los tres hombres son iguales (ya que no hay información que nos induzca a pensar lo contrario). Es decir:

$$p(A) = p(B) = p(C) = \frac{1}{3}$$

Si representamos por b al suceso «el carcelero contesta que B será ejecutado», y asumimos que el carcelero no ha mentido y que no tiene preferencia por nombrar B o C si ambos tuviesen que ser ejecutados, entonces se cumpliría:

$$p(b \mid A) = \frac{1}{2} \qquad p(b \mid B) = 0 \qquad p(b \mid C) = 1$$

Es decir, $p\,(b \mid A)$ es la probabilidad de b supuesto que el carcelero conozca que A va a ser puesto en libertad. Su valor es 1/2 puesto que si A va a ser liberado, el carcelero tiene dos opciones: su respuesta podría ser «B» o «C». El valor de $p\,(b \mid B)$ es cero porque, si B va a ser liberado, es imposible el suceso b. Análogamente se deduce que $p\,(b \mid C) = 1$, ya que si C va a ser liberado, es seguro que se cumple b.

Por tanto, aplicando el teorema de Bayes, la probabilidad de que A sea puesto en libertad, sabiendo la respuesta que ha dado el carcelero, será:

$$p\,(A \mid b) = \frac{p\,(A)\,p\,(b \mid A)}{p\,(A)\,p\,(b \mid A) + p\,(B)\,p\,(b \mid B) + p\,(C)\,p\,(b \mid C)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1} = \frac{1}{3}$$

Es decir, la probabilidad de que A sea liberado no varía. Luego no es razonable la felicidad experimentada por A al oír la respuesta del carcelero.

APÉNDICE III

Construcción de un conjunto de modelos mentales del problema de los tres prisioneros (adaptado de Johnson-Laird *et al.*, 1999)

Inicialmente, el conjunto de modelos mentales sería así:

A	B	C
Liberado		
	Liberado	
		Liberado

Después de conocer que el carcelero ha dado como respuesta «B será ejecutado», esta información debe modificar el anterior modelo, quedando de esta forma:

A	C	Nombre que dice el carcelero	Probabilidades
Liberado		B	1/2
	Liberado	B	1
. . .			0

Veamos de dónde salen estas probabilidades. Si el liberado fuese A, entonces el carcelero podrá decir «B será ejecutado» o bien «C será ejecutado», por tanto hay dos alternativas, luego, la probabilidad de que el carcelero diga «B será ejecutado», sabiendo que A va a ser liberado, será 1/2. En el otro caso, si fuese C el liberado, entonces el carcelero sólo podría decir «B será ejecutado», no tendría ninguna otra alternativa, luego la probabilidad de que el carcelero diga «B será ejecutado», sabiendo que el liberado va a ser C, valdrá 1. El modelo implícito (representado por la línea de puntos) incluye el caso

en que el liberado fuese *B* sabiendo que el carcelero ha dicho «*B* será ejecutado», que evidentemente es un suceso de probabilidad 0. Finalmente, en virtud del principio de subconjunto, la probabilidad de que *A* sea liberado, conociendo la información proporcionada por el carcelero (es decir, $p(A \mid b)$, se obtendrá dividiendo la frecuencia del modelo de «*A* y *b*» entre la suma de las frecuencias de los modelos de «b», es decir: $P(A \mid b) = (1/2) / (1/2 + 1) = 1/3$.