PRINCIPIA 22(2): 229–249 (2018) doi: 10.5007/1808-1711.2018v22n2p229
Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

¿EXISTE CONOCIMIENTO EPISTÉMICAMENTE IRRACIONAL?

IS THERE EPISTEMICALLY IRRATIONAL KNOWLEDGE?

MANUEL PÉREZ OTERO
Departamento de Filosofía. Facultad de Filosofía. Universidad de Barcelona, SPAIN
perez.otero@ub.edu

Abstract. I present an epistemological puzzle about perceptual knowledge and its relation to the evaluation of probabilities. It involves cases, concerning a given subject S and a proposition P in a determinate context, where apparently: S has perceptual knowledge of P; the epistemic justification S has for believing Not-P is much greater than her epistemic justification for believing P. If those two theses were true, the following very plausible epistemological principle would fail: If S knows P, then the epistemic justification S has for believing Not-P is not greater than her epistemic justification for believing P. I offer a solution to the puzzle, which is compatible with basic intuitions and theses of orthodox Bayesianism.

Keywords: Base-rate fallacy • Bayesian epistemology • perceptual warrant • rational credence • lotteries

RECEIVED: 17/10/2017 REVISED: 20/01/2018 ACCEPTED: 28/03/2018

Introducción

La percepción sensorial tal vez constituya la fuente de conocimiento menos controvertida de todas. No obstante, también hay — por supuesto — discrepancias entre los filósofos sobre el papel que dicha fuente desempeña. El propósito principal de este artículo es presentar un problema epistemológico relacionado con la percepción y la evaluación de probabilidades. La tensión entre diferentes postulados aparentemente verdaderos es tal que el problema pudiera merecer ser clasificado como paradoja. Sin embargo, por lo que yo sé, nadie se ha referido a él en la literatura. Entre los objetivos secundarios de mi trabajo está comentar de forma breve algunas de las posibles posiciones ante ese problema e indicar qué respuesta me parece preferible.

El caso que describiré involucra a un sujeto, S, que en una situación perfectamente normal (normal en lo concerniente a las condiciones para el funcionamiento de su sistema visual) parece tener conocimiento perceptivo de cierta proposición, P. De ese modo, resulta intuitivamente muy plausible esta tesis:

(C) S sabe que P

Por otra parte, también es muy verosímil el siguiente postulado que establece un vínculo entre el saber y la justificación epistémica:

(CJ) Si x sabe que P, entonces la justificación epistémica que tiene x para creer No-P no es mayor que la justificación epistémica que tiene para creer P.

Sin embargo, cuando reflexionamos teniendo en cuenta los datos que voy a ofrecer, también estamos fuertemente inclinados a sostener esta otra tesis sobre el sujeto S:

(NJ) La justificación epistémica que tiene S para creer No-P es mucho mayor que la justificación epistémica que tiene para creer P.

El problema obvio, o paradoja, consiste en que esas tres tesis, (C), (CJ) y (NJ), son incompatibles entre sí.

Antes de presentar y desarrollar la discusión detallada de este problema, conviene abordar ya una cuestión fundamental. Tanto en nuestras comunicaciones cotidianas como en la investigación teorética sistemática (en la ciencia y en la filosofía) usamos a veces un concepto cualitativo, *todo-o-nada*, de creencia y en otras ocasiones un concepto cuantitativo. Este segundo concepto se manifiesta, por ejemplo, cuando tras decir (usando el concepto cualitativo) que creo la proposición Q1 y también creo la proposición Q2, añado que mi creencia en Q1 es más profunda o más arraigada que mi creencia en Q2 (es decir, que creo Q1 en mayor grado que Q2), o cuando afirmo que los indicios apoyan racionalmente adoptar respecto a Q1 un grado mayor de creencia que el grado de creencia respecto a Q2 que apoyen adoptar. En este artículo voy a usar el concepto cuantitativo de creencia. Sin embargo, en (CJ) y en (NJ) se invoca — aparentemente — el concepto cualitativo de creencia. Para que las consideraciones probabilísticas subsiguientes (concernientes al sentido cuantitativo) se vinculen de forma apropiada con (CJ) y (NJ), se requiere algún supuesto sobre la conexión entre ambos sentidos de creer. El supuesto podría ser un principio puente, o bien una propuesta reductiva que caracterice uno de ellos en términos del otro. Como describe con precisión Leitgeb (2013), casi todas las opciones que se han elaborado a este respecto son problemáticas (cf. también Buchak 2014); aunque Leitgeb ofrece su propia definición reductiva, que permitiría caracterizar la creencia cualitativa en términos de creencia cuantitativa.

En lo concerniente al problema que discutiré, considero viables dos opciones sobre esta cuestión; aunque me parece que no necesito pronunciarme por una de ellas: (i) Adoptar el enfoque reductivo de Leitgeb (que no detallo aquí). (ii) Reformular el enigma sustituyendo (CJ) y (NJ) por tesis análogas en las cuales se reemplaza el concepto cualitativo de creencia por un concepto cuantitativo, sin comprometerme por ello con un principio puente general o una definición reductiva del concepto todo-o-nada de creencia (aplicable a otros casos en los que se invoca ese concepto cualitativo). Esta opción (ii) podría conllevar estas reformulaciones de (CJ) y (NJ):

(CJ*) Si x sabe que P, entonces los indicios epistémicos globales de que dispone x no apoyan o justifican adoptar respecto a No-P un grado mayor de creencia que el grado de creencia respecto a P que apoyen o justifiquen adoptar.

(NJ*) Los indicios epistémicos globales de que dispone x apoyan o justifican adoptar respecto a No-P un grado mayor de creencia que el grado de creencia respecto a P que apoyan o justifiquan adoptar.

Por simplicidad, continuaré usando las formulaciones originales, (CJ) y (NJ). Pero el lector preocupado por esta cuestión, puede consultar la definición de Leitgeb (2013), o bien sustituir tales formulaciones por las versiones cuantitativas (CJ*) y (NJ*).[1]

En la sección 1 presento el caso, al que denominaré CANICA. Como señalo en la sección 2, el caso CANICA tiene semejanzas con otros casos en los cuales si el sujeto creyera cierto dato observacional (el resultado de una prueba médica, o la visión de un coche) incurriría en la denominada *falacia de desestimación de la proporción de base*. En la sección 3 comento algunas posibles reacciones ante el enigma que considero insatisfactorias o insuficientes. Ofrezco una solución al enigma (sección 4) que, sin apartarse del principio (NJ) ni del bayesianismo epistemológico, preserva la intuición de que el sujeto del caso CANICA tiene conocimiento perceptivo. En la sección 5 destaco algunas diferencias entre ese caso y aquellos otros que involucrarían realmente la falacia antedicha, enfatizando el papel desempeñado por nuestras estimaciones intuitivas de probabilidades.

1. El enigma epistémico

Nos aproximaremos al caso problemático o paradójico que discutiré partiendo de otro caso similar, que no involucra ningún tipo de tensión teorética. Supongamos que tenemos un enorme montón de canicas. Algunas son blancas, todas las otras son rojas. Se realiza un sorteo para extraer, por azar, una canica. Cierto sujeto, Ildefonso, dispone de todos esos datos. La canica que ha resultado seleccionada se le muestra a Ildefonso, quien — en condiciones óptimas de visibilidad — ve que es una canica roja. A Ildefonso le parece estar viendo una canica roja. ¿Está Ildefonso justificado en creer que la canica es roja? ¿Tiene Ildefonso conocimiento perceptivo de que la canica es roja? Ambas preguntas se responden positivamente, y no parece haber motivo para recelar ante este caso.

Consideremos ahora un caso emparentado con ése, pero en el cual se añaden algunas especificaciones. A este caso lo llamaremos CANICA. La desproporción entre canicas de uno y otro color es enorme. Por cada canica roja, hay un millón de canicas blancas. Así, la probabilidad de sacar aleatoriamente una canica roja es muy baja. Ildefonso está informado sobre esas cantidades relativas de canicas rojas y blancas.

Como en el caso anterior, se extrae una canica al azar, roja, y se le entrega a Ildefonso. Cuando éste mira la canica, le parece estar viendo una canica roja. En esta nueva situación: ¿también está Ildefonso justificado en creer que la canica es roja? ¿tiene conocimiento perceptivo de que la canica es roja? Antes de explorar la viabilidad de una u otra respuesta, complementemos la descripción del caso CANICA con algunos datos adicionales.

Respecto al caso anterior, he indicado que Ildefonso *tenía todos los datos* expuestos en el ejemplo. Debemos ser más precisos, especialmente ahora. Con frecuencia, los individuos que colocamos en ciertas situaciones imaginarias carecen de cierta información crucial que nosotros, los diseñadores y/o evaluadores del caso, sí tenemos. Sucede así, paradigmáticamente, con los casos Gettier. Los ejemplos comentados en este artículo no son de ese tipo. Sin embargo, tampoco sería exacto afirmar que esos sujetos (sobre cuya racionalidad vamos a preguntarnos) tienen los mismos datos que nosotros. Aquí estipulamos las proporciones entre canicas rojas y blancas, y *estipulamos* que la canica vista por Ildefonso es roja. Pero no es así como Ildefonso, el sujeto que habita dicha situación imaginaria estipulada, adquiere esa información. Hemos de suponer que Ildefonso accede a tales datos como accedería alguien en una situación real: se le ha dicho cuál es la proporción entre unas y otras canicas, así como el resto de información que enseguida indicaré; y, por supuesto, no sabe por estipulación que se le ha entregado una canica roja (eso sería imposible), sino que mira la canica y ésta le parece muy claramente roja.[2]

Imaginemos que las ciencias cognitivas han experimentado avances extraordinarios y se dispone de información empírica precisa sobre la fiabilidad de los sentidos. Se conoce también la fiabilidad, en particular, del sistema visual de Ildefonso. En la circunstancias específicas en que Ildefonso mira la canica (dadas las condiciones de visibilidad y su competencia visual) es muy alta la fiabilidad de su visión: la probabilidad de que a Ildefonso le parezca estar viendo una canica roja cuando está viendo realmente una canica roja es aproximadamente r. Naturalmente, r es un número positivo muy cercano a 1. Supongamos, por ejemplo, que $r = 0'9999$. Nada importante dependerá de esa elección concreta del valor de r, que he seleccionado meramente para ilustrar la problemática. Si alguien creyera que ese valor es demasiado bajo, deberá entender que podríamos elevarlo y, para presentar el enigma, elevar al mismo tiempo la proporción de canicas blancas que suponemos existe en el montón. (No hay enigma si postulamos que $r = 1$. Pero sobre ese punto, cf. nuestra sección 3.) Si alguien creyera que ese valor es demasiado alto, eso no afecta significativamente al enigma, que resulta reforzado.

Por otra parte, también se sabe que la probabilidad de que a Ildefonso le parezca estar viendo una canica roja cuando está viendo realmente una canica blanca es aproximadamente s. Este número debe ser menor que $1 - r$. Si s fuera $1 - r$, eso significaría que todas las ocasiones excepcionales en que Ildefonso mira una canica blanca pero le parece de otro color son ocasiones en que la canica le parece roja. En algunas situaciones de canica blanca y percepción errónea el color que Ildefonso creerá ver será azul, amarillo, etc. Cierto, Ildefonso asume que si la canica no es blanca, su color será rojo. Eso quizá ejerza alguna influencia y en casos de percepción errónea tal vez la apariencia de canica roja sea más frecuente que otras apariencias. Pero no es plausible que la *penetración cognitiva* sea tal que s llegue hasta $1-r$. (Hay penetración cognitiva cuando factores extra-epistémicos, como — pongamos por caso — las expectativas previas de un sujeto, influyen en la apariencia sensorial que dicho sujeto experimenta. Cf., por ejemplo, Siegel 2012.) Supongamos que $s = 0'00005$. De nuevo, nada importante dependerá de esa elección concreta del valor de s. Si alguien creyera que ese valor es demasiado alto, podríamos reducirlo y elevar al mismo tiempo la proporción de canicas blancas que suponemos existe en el montón. Si alguien creyera que ese valor es demasiado bajo, eso no afecta significativamente al enigma, que resulta reforzado.

Finalmente, Ildefonso también está informado respecto a dichas magnitudes, r y s.

Expresemos los diferentes datos empleando recursos conceptuales sencillos y poco controvertidos de la epistemología bayesiana. Usaremos la expresión ‘roja’ para abreviar ‘la canica resultante del sorteo es roja’ y la expresión ‘parece-roja’ como abreviatura de ‘la canica que estoy viendo, resultante del sorteo, me parece roja’ (usaremos ‘blanca’ de forma análoga). Conforme a la descripción del caso CANICA:

$$\mathrm{Prob}_i(\text{roja}) = 0'000001$$
$$\mathrm{Prob}_i(\text{blanca}) = 0'999999$$
$$\mathrm{Prob}(\text{parece-roja}\,|\,\text{roja}) = 0'9999$$
$$\mathrm{Prob}(\text{parece-roja}\,|\,\text{blanca}) = 0'00005$$ [3]

El sujeto, Ildefonso, dispone de esos datos (con la salvedad especificada en el tercer párrafo de esta sección). Asumiré que es lícito modelar el grado de apoyo o justificación epistémica para sus creencias que tienen los individuos (Ildefonso, en particular) mediante probabilidades. Esa estrategia, característica del bayesianismo epistemológico y de otros enfoques emparentados, no está exenta de dificultades. Por ejemplo, conlleva suponer que tenemos máxima justificación para creer cualquier verdad lógica, incluso si cierta verdad lógica es compleja y no sabemos reconocerla como tal (por no ser lógicamente omniscientes). Pero la discusión en este trabajo no concierne a esos problemas clásicos de la epistemología bayesiana; trato de identificar *otro* problema, diferente. Por decirlo así, aunque tuviéramos una solución satisfactoria para la cuestión de la omnisciencia lógica, permanecería la dificultad que estoy exponiendo. En ese sentido, el interés de presentar y discutir el caso no requiere — a mi juicio — un posicionamiento previo respecto a otros problemas ya conocidos del bayesianismo.

Con objeto de resaltar esa reinterpretación de las probabilidades en términos de grado de creencia racional por parte de Ildefonso, usaré 'Cr' para expresar ese concepto. Tenemos entonces:

$\mathrm{Cr}_i(\text{roja}) = 0'000001$
$\mathrm{Cr}_i(\text{blanca}) = 0'999999$
$\mathrm{Cr}(\text{parece-roja}\,|\,\text{roja}) = 0'9999$
$\mathrm{Cr}(\text{parece-roja}\,|\,\text{blanca}) = 0'00005.$

Así pues, $0'000001$ sería el grado de confianza que es racional que Ildefonso deposite en la proposición de que la canica resultante del sorteo será roja (antes de verla). Y $0'9999$ sería el grado de confianza que es racional que Ildefonso deposite en la proposición de que la canica resultante del sorteo parecerá roja bajo el supuesto de que es roja. Procederemos a calcular Cr(roja|parece-roja), es decir, el grado de confianza que es racional que Ildefonso deposite en la proposición de que la canica resultante del sorteo es roja bajo el supuesto de que le parece roja. Como ese supuesto es real (a Ildefonso le parece estar viendo una canica roja), esta magnitud, Cr(roja|parece-roja), es precisamente la que interesa averiguar: $\mathrm{Cr}_f(\text{roja})$.

Recordemos nuestras preguntas principales. En la situación descrita, Ildefonso está mirando una canica que, en buenas condiciones de visibilidad, le parece claramente roja. ¿Está justificado en creer que es roja? ¿Sabe que es roja? Aplicando el Teorema de Bayes, obtenemos los siguientes resultados:

$\mathrm{Cr}(\text{roja}\,|\,\text{parece-roja}) = \mathrm{Cr}_i(\text{roja}) \times \mathrm{r} \;/\; \mathrm{Cr}_i(\text{parece-roja}).$
$$\mathrm{Cr}_i(\text{parece-roja}) = [\mathrm{Cr}(\text{parece-roja}\,|\,\text{roja}) \times \mathrm{Cri}(\text{roja})] + [\mathrm{Cr}(\text{parece-roja}\,|\,\text{blanca}) \times \mathrm{Cr}_i(\text{blanca})].$$

Por consiguiente:

$$(*)\quad \mathrm{Cr}(\text{roja}\,|\,\text{parece-roja}) = 0'000001 \times \mathrm{r} \;/\; [\mathrm{r} \times 0'000001] + [s \times 0'999999] = 0'01961.$$ [4]

Así pues, usando 'no-roja' para abreviar 'la canica resultante del sorteo no es roja', tenemos:

$\mathrm{Cr}(\text{no-roja}\,|\,\text{parece-roja}) = 0'98039.$

Ildefonso tiene en su mano la canica. Está mirándola y le parece claramente roja. La situación es perfectamente normal y típica en lo concerniente a las condiciones para el desempeño de la percepción visual de Ildefonso (no está drogado, ni mareado; la iluminación es óptima; etc.). Aunque, ciertamente, es atípico un factor extrínseco a tales condiciones: la enorme desproporción entre canicas blancas y rojas. El razonamiento bayesiano dictamina que dicho factor extrínseco — al ser conocido por Ildefonso — interfiere, impidiendo que el sujeto tenga respaldo racional para creer que la canica es roja. Dadas las circunstancias del caso (también, en particular, el hecho de que Ildefonso comparte con nosotros — en el sentido que se aclaró — todos los datos sobre la situación), parece muy natural asumir que la justificación epistémica de Ildefonso es correlativa del respaldo racional, medido por el concepto Cr. Así, conforme al bayesianismo, (NJ) es válido, y por tanto, si mantenemos nuestra adhesión al principio (CJ), debemos renunciar a (C):

(NJ) La justificación epistémica que tiene Ildefonso para creer que la canica no es roja es mucho mayor que la justificación epistémica que tiene para creer que la canica es roja.

(CJ) Si x sabe que P, entonces la justificación epistémica que tiene x para creer No-P no es mayor que la justificación epistémica que tiene para creer P.

(C) Ildefonso sabe que la canica es roja.

Es una situación teorética problemática, o quizá incluso paradójica. Muchas personas aceptamos la plausibilidad de las premisas que han conducido a ese resultado, al mismo tiempo que tenemos una fuerte intuición en favor de (C). Pero los tres desiderata, (NJ), (CJ) y (C), son incompatibles entre sí. (En adelante usaré '(NJ)' y '(C)' para referirme a estas tesis concretas relativas a Ildefonso, aunque en la Introducción las usé para expresar formulaciones más abstractas, sin especificación del sujeto epistémico.)

Quien propusiera mantener (C) y (NJ) renunciando a (CJ) estaría postulando la existencia de un conocimiento que podría describirse como epistémicamente irracional. Pero las etiquetas que usemos para catalogar los diversos posicionamientos posibles no serán de gran importancia. Lo relevante es comprender el enigma, e intentar ofrecer para el mismo una solución satisfactoria, que idealmente requeriría saber qué tesis debe sacrificarse y por qué — pese a ello — parece claramente verdadera.

2. La falacia de desestimación de la proporción de base

La epistemología bayesiana tiene una serie de problemas, relativamente bien conocidos (he recordado algún ejemplo en la sección anterior). Creo que esos inconvenientes no contrarrestan las ventajas teóricas generales del enfoque. No obstante, podría existir la tentación de reaccionar ante el enigma descrito en la sección anterior optando por aferrarse a las tesis (C) y (CJ) y rechazar la tesis distintivamente bayesiana, (NJ), pensando que seguramente el caso CANICA es una variante de alguno de esos problemas típicos del bayesianismo; o que ilustra tal vez un nuevo problema — desconocido hasta ahora — para los bayesianos. (Si esto último fuera así, ello ya justificaría suficientemente el interés en el estudio del caso.)

Esa reacción podría ser precipitada. Los presupuestos bayesianos usados en la descripción del enigma son muy sencillos y poco controvertidos. Encuentro arriesgado atribuir el problema a otros inconvenientes del bayesianismo, o suponer que hemos tenido la fortuna de haber identificado un nuevo inconveniente. Con vistas a reforzar esta idea, intentaré mostrar que CANICA parece ejemplificar cierto tipo de patrón argumentativo que los bayesianos llaman *falacia de desestimación de la proporción de base* ('base-rate fallacy' es la denominación más frecuente en inglés). Para ser precisos, si Ildefonso siguiera lo que su percepción visual le indica y creyera que la canica es roja, estaría incurriendo en dicha falacia (al menos si asumimos el análisis realizado hasta ahora). Sin embargo, no he visto ninguna referencia a problemas como el que ilustra CANICA ni en la literatura concerniente a la falacia de desestimación de la proporción de base ni en textos sobre epistemología de la percepción. Por esa razón, he considerado que merecía la pena describirlo aquí y proceder a su análisis.

Conviene ser más específico sobre este punto, concerniente a la originalidad de la paradoja ilustrada mediante el caso CANICA. Por supuesto, algunos aspectos de la situación descrita (el hecho de tener lugar un sorteo; que la epistemología bayesiana afronte dificultades relativas al conocimiento derivado de la percepción; que se incurra — aparentemente — en la falacia de desestimación de la proporción de base; que el caso quizá ejemplifique un conflicto entre los razonamientos probabilísticos y las opiniones preteóricas de sentido común sobre el saber; etc.) son abordados en muchos textos preexistentes. Lectores de una versión previa de este trabajo me han mencionado algunas publicaciones en las cuales se tematizan cuestiones que — a su juicio — pudieran tener cierto grado de parentesco: Christensen (1992); Dawes, Faust y Meehl (1989); Pollock y Cruz (1999, cap.4); Titelbaum (2013a); Williamson (2009). Debo decir que — pese a las obvias conexiones temáticas — en ninguno de esos textos se anticipan consideraciones suficientemente cercanas al enigma aquí descrito. No puedo descartar que en *otros* trabajos, de los que no tengo noticia, se hayan descrito y problematizado casos similares al que he planteado. Pero hasta el momento no he hallado ningún indicio concreto de que sea así.

Veamos algunos ejemplos con los cuales suele ilustrarse el fenómeno de la falacia de desestimación de la proporción de base. Al primer caso lo denominaremos HOSPITAL (este ejemplo y también el siguiente, TAXI, proceden de Bar-Hillel 1980). Cierta persona, Matilde, se ha sometido a un test médico diseñado para averiguar si se padece una determinada enfermedad, E. El resultado ha sido positivo, es decir, el test indica que Matilde tiene la enfermedad E. Usaré 'E' para representar (también) la proposición expresada por 'Matilde tiene E' y 'parece-E' para 'el test indica que Matilde tiene E'. Es una enfermedad bastante minoritaria: sólo una persona de cada mil tiene E. Así pues, $\mathrm{Prob}_i(\mathrm{E}) = 0'001$. Respecto a la fiabilidad del test, se dispone de estos dos datos: la probabilidad de que el test ofrezca un resultado positivo aunque la persona no tenga E (un *falso positivo*) es del 3 %; la probabilidad de que el test dé positivo cuando la persona tiene E es del 98 %. Por tanto, Prob(parece-E|no-E) $= 0'03$ y Prob(parece-E|E) $= 0'98$. Añadamos, por último, el supuesto de que Matilde está al tanto de toda esa información (por tanto, podemos sustituir 'Prob' por 'Cr' en todos los comentarios sobre esta situación). ¿Estaría Matilde justificada en creer que tiene la enfermedad E? Aplicando el Teorema de Bayes, de forma análoga a como se ha mostrado con el caso CANICA, la respuesta es negativa. Si al conocer el resultado del test Matilde creyera padecer E, estaría cometiendo la falacia en cuestión, por desestimar indebidamente una información relevante: cuál es la (baja) proporción de base de E en la población general ($0'1$ %). Especifiquemos el cálculo:

$$\begin{aligned}
&\mathrm{Prob}(\mathrm{E}|\text{parece-E}) = \mathrm{Prob}_i(\mathrm{E}) \times \mathrm{Prob}(\text{parece-E}|\mathrm{E}) \;/\; \mathrm{Prob}_i(\text{parece-E}).\\
&\mathrm{Prob}_i(\text{parece-E}) = [\mathrm{Prob}(\text{parece-E}|\mathrm{E}) \times \mathrm{Prob}_i(\mathrm{E})]\\
&\qquad\qquad + [\mathrm{Prob}(\text{parece-E}|\text{no-E}) \times \mathrm{Prob}_i(\text{no-E})]\\
&\qquad = [0'98 \times 0'001] + [0'03 x 0'999] = 0'03095.
\end{aligned}$$

En consecuencia,

$$\mathrm{Prob}_f(\mathrm{E}) = \mathrm{Prob}(\mathrm{E}|\text{parece-E}) = 0'001 \times 0'98/0'03095 = 0'03166397.$$

Conforme a ese resultado, la magnitud Prob(no-E|parece-E) es mucho mayor que la magnitud Prob(E|parece-E), porque es aproximadamente $0'968$.

Supongamos que Matilde tiene la enfermedad E. Eso no implica ninguna diferencia respecto al resultado obtenido, sobre su falta de justificación para creer dicha proposición verdadera. Siendo así, además, si Matilde creyera que padece E, esta creencia — asumiendo la tesis (CJ) — no constituiría conocimiento.

El caso que describiré a continuación, TAXI, tiene mayor interés para nuestra discusión, por ser más cercano al caso CANICA. Un taxi ha atropellado a una persona y luego ha huido. Pedro ha sido testigo del accidente; le ha parecido que el taxi era verde. En esa ciudad el 85 % de los taxis son azules y el 15 % restante son verdes. Se ha examinado con precisión la capacidad de Pedro para distinguir entre taxis azules y taxis verdes en las condiciones de visibilidad nocturnas del momento del accidente, dada la distancia a la que se encontraba. Dicho examen proporciona, en términos aproximativos, estos cuatro datos (usaré 'verde' para 'el taxi es verde' y 'parece-verde' para 'el taxi le parece verde a Pedro'; análogamente con 'azul' y 'parece-azul'): Prob(parece-azul|azul) $= 0'8$; Prob(parece-verde|verde) $= 0'8$; Prob(parece-azul|verde) $= 0'2$; and Prob(parece-verde|azul) $= 0'2$.[5] Imaginemos que a Pedro se le han comunicado esos datos sobre la fiabilidad de su visión (en tales condiciones) y sobre la proporción de taxis verdes y azules. ¿Sabe Pedro que el taxi que vio aquella noche era verde? ¿Está, al menos, justificado en creerlo? De nuevo, procedamos a calcular las respuestas bayesianas:

$$\text{Prob(verde|parece-verde)} = \text{Prob}_i\text{(verde)} \times \text{Prob(parece-verde|verde)} / \text{Prob}_i\text{(parece-verde)}.$$
$$\begin{aligned}\text{Prob}_i\text{(parece-verde)} &= [\text{Prob(parece-verde|verde)} \times \text{Prob}_i\text{(verde)}] \\ &\quad + [\text{Prob(parece-verde|no-verde)} \times \text{Prob}_i\text{(no-verde)}] \\ &= [0'8 \times 0'15] + [0'2x0'85].\end{aligned}$$

Por consiguiente:

$$\text{Prob(verde|parece-verde)} = 0'15 \times 0'8/(0'12 + 0'17) = 0'41.$$
$$\text{Prob(no-verde|parece-verde)} = 0'59.$$

También en ese caso Pedro cometería la falacia si creyera que el taxi era verde. Asumiendo que no tiene apropiada justificación para dicha creencia, y dado (CJ), no sabría que el taxi era verde aunque así lo creyera.

Hay dos diferencias importantes entre el caso CANICA y el caso TAXI. La primera es que las condiciones de visibilidad en que el sujeto del primer caso, Ildefonso, mira la canica que le parece roja son óptimas; sería un ejemplo paradigmático de uso del sistema visual. Por el contrario, Pedro ve el taxi a cierta distancia y de noche. No obstante, esa diferencia ya queda recogida en los distintos datos sobre la fiabilidad de su visión en las condiciones dadas: las magnitudes Cr(parece-roja/roja) $= 0'9999$ y Prob(parece-verde/verde) $= 0'8$, respectivamente. Además, esa diferencia queda "compensada" — podríamos decir — por la otra diferencia importante, concerniente a las proporciones de ítems minoritarios en las respectivas muestras: en TAXI los taxis verdes son 15 de cada 100; en CANICA las canicas rojas son sólo una por cada millón. Como consecuencia, en ambas situaciones creer la proposición apoyada por los sentidos sería una opción epistémicamente no racional, al menos según el bayesianismo (y dado el análisis propuesto hasta el momento). Respecto a ese resultado final, CANICA y TAXI no son casos relevantemente distantes: Cr(roja|parece-roja) $= 0'01961$ y Prob(verde|parece-verde) $= 0'41$, respectivamente (no es una diferencia relevante *favorable a que Ildefonso tenga mayor justificación que Pedro*; más bien sería al contrario).

En definitiva: TAXI es un ejemplo típico con el cual se ilustra la falacia de desestimación de la proporción de base (aunque los ejemplos más típicos son similares a HOSPITAL). El razonamiento bayesiano respecto a CANICA es análogo al razonamiento respecto a TAXI. Así pues, todo parece indicar que si Ildefonso creyera que la canica es roja también incurriría en dicha falacia.

3. Algunas respuestas al problema

Como ya he señalado, mi objetivo principal en este trabajo era presentar el enigma que he descrito, sobre el caso CANICA. Un objetivo secundario es ofrecer una defensa de la tesis (C) (Ildefonso sabe que la canica es roja). Esa defensa tiene los siguientes rasgos: permite rechazar (NJ) (La justificación epistémica que tiene Ildefonso para creer que la canica no es roja es mucho mayor que la justificación epistémica que tiene para creer que la canica es roja) sin renunciar al bayesianismo ortodoxo ni a (CJ); es compatible con negar el conocimiento en los casos HOSPITAL y TAXI (Matilde no sabe que padece E; Pedro no sabe que vio un taxi verde). Presento mi defensa en la próxima sección. Antes, en esta sección, comentaré algunas posibles reacciones ante el enigma que considero insatisfactorias o insuficientes, incluyendo reacciones que también comportan alinearse con la tesis (C).

Algunas personas tienen una doble inclinación (respecto al caso CANICA): creer que debe rechazarse la tesis (C); y creer, incluso, que ese rechazo es la respuesta obvia, tan obvia que no hay aquí ningún enigma teorético, pues — una vez se reflexiona sobre los diversos datos — no hay motivación racional para mantener (C). Ante esa posición bayesiana radical, yo respondería con dos comentarios o peticiones muy diferentes entre sí, ninguno de los cuales es propiamente una objeción. Primero, sugiero al acérrimo bayesiano que explore las intuiciones de otras personas sobre el caso CANICA. Mi experiencia ha sido la de encontrar una proporción significativa de filósofos con fuerte tendencia a preservar (C), incluso si no se dispone de una buena solución al enigma (una solución que explique por qué (CJ), o bien (NJ), debe ser rechazado).[6] Pareciera que descartar (C) sin ninguna elucidación adicional hace poca justicia al enigma. En segundo lugar, otra petición es que se evalúe mi propia re-elaboración del problema y la subsiguiente defensa de (C), en la sección 4.

La tesis (C) puede ser apoyada desde motivaciones diferentes. Quiero mencionar dos que considero poco convincentes. Según algunos filósofos, cuando las condiciones de uso de nuestros sentidos son paradigmáticas, las correspondientes creencias perceptivas serían infalibles. Eso les permite rechazar (NJ) en el caso CANICA, alegando que Cr(roja|parece-roja) es simplemente 1, y no $0'01961$. Otro grupo de epistemólogos (que quizá se solapen parcialmente con los antedichos *infalibilistas*) también niegan (NJ), sosteniendo — en concreto — que he aplicado incorrectamente el aparato teórico bayesiano: Ildefonso no debería "condicionar" respecto a [*conditionalize on*] la proposición de que la canica le parece roja, sino respecto a la proposición de que la canica es roja; es decir, la observación le garantiza el dato de que la canica es roja, no el dato de que parece roja. Siendo así, al intentar averiguar la magnitud Crf(roja), en lugar de preguntarnos por Cr(roja|parece-roja) debemos preguntarnos por Cr(roja|roja), cuyo valor — obvio — es 1.

La opción infalibilista tiene dificultades tradicionales sobre las cuales no voy a extenderme, salvo para recordar que — en principio — las condiciones para la visión óptimas o paradigmáticas son compatibles con el error perceptivo. Si alguien objeta a esa tesis, puede reinterpretar el caso CANICA asumiendo que las condiciones son típicamente buenas, aunque sin estipular que sean "óptimas o paradigmáticas" en el presunto sentido en que, según ese objetor, ello implica infalibilidad. El enigma resurge y la opción infalibilista parece desvanecerse.

Tanto los infalibilistas sobre la percepción paradigmática como — sobre todo — los partidarios de que Ildefonso condicione respecto a la proposición de que la canica es roja afrontan una especie de dilema. Hemos constatado (al final de la sección previa) que la diferencia entre el caso TAXI y el caso CANICA es gradual. El dilema es el siguiente: unos y otros deben (a) sostener que dicha diferencia cuantitativa no afecta al conocimiento del sujeto, y también Pedro sabe que el taxi era verde, o bien (b) sostener que Pedro no tiene dicho conocimiento y explicar qué factor marca la diferencia entre uno y otro caso.

El cuerno (a) requiere — asumiendo (CJ) — negar que TAXI ejemplifique la falacia de desestimación de la proporción de base. Es un coste muy considerable.

También el cuerno (b) es problemático: ¿a cuánta distancia tendría que estar el taxi para que Pedro pudiera legítimamente — como se supone que podría hacer Ildefonso — condicionar respecto a la proposición de que el taxi es verde, en lugar de condicionar respecto a la proposición de que le parece verde (o para que su situación fuera un caso de percepción paradigmática infalible)? En contraposición con mi tratamiento en la sección 4, estas propuestas no permiten una correspondencia apropiada entre el carácter gradual, cuantitativo de la diferencia TAXI-CANICA y el carácter cualitativo (no gradual) de condicionar respecto a una u otra proposición (o de asignar a Cr(roja|parece-roja) el valor 1).

4. Incertidumbre sobre la información previa

Voy a describir la que — desde mi punto de vista — es la mejor motivación para sostener (C). Mi propuesta tiene estas características: no se aparta de las intuiciones y tesis básicas del bayesianismo ortodoxo; rechaza (NJ); preserva (CJ); mantiene el rechazo a atribuir conocimiento en los casos HOSPITAL y TAXI, en los cuales los sujetos efectivamente tienen mayor justificación para creer la negación de la proposición principal (las negaciones de ‘padezco E’ y ‘el taxi era verde’, respectivamente).

Imaginemos, de nuevo, la situación de Ildefonso en el caso CANICA. Tratemos de ponernos en su piel, para intentar ofrecer una descripción realista de su vivencia. ¿Cómo reaccionaríamos si — teniendo los datos que se le han proporcionado — cuando se realiza el sorteo y nos entregan la canica seleccionada la miramos y nos parece claramente roja? Una reacción bastante natural es plantearnos si no nos habremos precipitado al aceptar algunos de los datos que se nos transmitieron. En concreto: quizá pensemos que (por algún error, o de forma intencionada) la proporción de canicas rojas no era tan baja como se nos dijo, o tal vez el sorteo no ha sido realmente aleatorio.

Veamos cómo articular ese tipo de sospecha que resulta natural atribuir a un sujeto en tales circunstancias. Dado el modo en que hemos aplicado el aparato bayesiano para analizar el caso CANICA, la opción que estamos contemplando ahora queda descartada de antemano. Para tomarla en consideración se requiere rectificar parcialmente dicho análisis (sin apartarse del bayesianismo estándar u ortodoxo).[7] Siguiendo el tratamiento estándar de situaciones similares a ésta, hemos asumido que Ildefonso aceptaba sin ningún titubeo todos los datos del caso, incluyendo estas dos tesis cruciales:

TV: Es verídico el dato sobre la proporción de canicas rojas en la muestra.

SA: El sorteo ha sido aleatorio; no estaba trucado.

TV está por "testimonio verídico". Podemos imaginar que Ildefonso recibió por testimonio esa información (no contó él mismo las canicas). En cierto sentido, también el dato SA puede haberlo recibido por algún tipo de testimonio, quizá indirecto. Así, usemos 'T' para nombrar a la conjunción de ambas tesis, TV y SA. Que Ildefonso aceptara T sin mayores problemas significa que asignó 1 a dicha proposición. Es decir, en nuestro análisis un punto de partida ha sido la ecuación $\mathrm{Cr}_i(\mathrm{T}) = 1$.[8] Pero si ésa era la expectativa racional inicial respecto a T, el enfoque bayesiano estricto no permite cambiarla en el futuro (con la salvedad indicada en la nota previa). Si queremos representar como una opción racional la reacción ulterior de Ildefonso (al ver la canica que le parece roja) de albergar dudas sobre la veracidad de T, se requiere que previamente Ildefonso no hubiera asignado 1 a T. Constataremos, primero, que asignar 1 a $\mathrm{Cr}_i(\mathrm{T})$ es la condición que permite sustentar la ecuación $\mathrm{Cr}_i(\mathrm{roja}) = 0'000001$, la cual es crucial para que el valor Cr(roja/parece-roja) sea tan bajo: $0'01961$. Luego exploraremos la posibilidad de rectificar nuestro análisis, asignando a $\mathrm{Cr}_i(\mathrm{T})$ un valor diferente a 1.

Si desglosamos explícitamente la magnitud Cri(roja), lo que tenemos es esto:

(#) $\mathrm{Cr}_i(\mathrm{roja}) = [\mathrm{Cr}_i(\mathrm{roja}|\mathrm{T}) \times \mathrm{Cr}_i(\mathrm{T})] + [\mathrm{Cr}_i(\mathrm{roja}|\text{no-T}) \times \mathrm{Cr}_i(\text{no-T})]$

Bajo el supuesto $\mathrm{Cr}_i(\mathrm{T}) = 1$, el valor $\mathrm{Cr}_i(\mathrm{roja})$ es, en efecto, $0'000001$, tal y como habíamos asumido (lo cual conduce a que $\mathrm{Cr}_f(\mathrm{roja})$, es decir Cr(roja|parece-roja), sea $0'01961$, según la ecuación (*) de nuestra sección 1). Pero el valor $\mathrm{Cr}_i(\mathrm{roja})$ será más bajo si descartamos dicho supuesto. Imaginemos, por ejemplo, que la convicción de Ildefonso en T hubiera sido sólo de $0'997$; es decir, imaginemos que $\mathrm{Cr}_i(\mathrm{T}) = 0'997$. (Elijo ese valor para ilustrar la cuestión. Después — en ésta y en la próxima sección — comentaré posibles asignaciones de valores diferentes.) Cabe hacer, entonces, dos consideraciones.

En primer lugar, el hecho de que $\mathrm{Cr}_i(\mathrm{T})$ difiera de 1 (sea su valor $0'997$ o cualquier otro valor) ya no permite establecer un cálculo preciso de Cr(roja|parece-roja), porque para calcular Cr_i(roja), conforme a (#), se requiere un dato desconocido: Cr_i(roja|no-T). Si la proporción de canicas roja no era la que se le dijo a Ildefonso o (lo fuera o no) el sorteo estaba trucado, ¿qué probabilidad había de que saliera una canica roja? Lo ignoramos. T puede ser falso porque el sorteo estuviera trucado (¿quizá Cr_i(roja|no-T) sería, entonces, un valor muy alto?), o porque hubiera un error involuntario al contar las canicas (¿quizá Cr_i(roja|no-T) sería, entonces, un valor cercano a $0'000001$, por tratarse de un error pequeño?), o porque se engañara intencionalmente a Ildefonso sobre la proporción de canicas (¿había muchas más canicas rojas? ¿sólo había canicas rojas?). Un corolario crucial es que hemos resuelto el enigma inicial, pues éste se ha disuelto, ha desaparecido. Ya no podemos afirmar legítimamente que Cr(roja|parece-roja) sea inferior a $0'5$. Por tanto, no tenemos base para sostener la tesis (NJ). Así, la tesis que queríamos preservar, (C), no está amenazada por un riesgo concreto que podamos especificar.

¿Qué razones hay para suponer que Ildefonso no debió asignar 1 a T? Obviamente, no es buena respuesta contestar que eso resuelve nuestro enigma. Pero, teniendo en cuenta la significación de asignar 1 a T, me parece lícito afirmar que se requerirían más razones para justificar dicha asignación que para justificar una asignación inferior a 1. Cabe añadir lo siguiente. T no pertenece a ninguna de las dos clases de proposiciones a las cuales — en el marco del bayesianismo — podría proponerse asignar probabilidad 1. T no es una verdad lógica o matemática. Por otro lado, tampoco es una proposición aprendida directamente a partir de la experiencia, que la regla de "condicionalidad" estricta obligue a asignar probabilidad 1; estoy asumiendo que sólo proposiciones como *parece roja* son de este tipo (cf. nuestras notas 8 y 10); T ni siquiera es una proposición "observacional" al estilo de *la canica es roja*. (Pronto, en ésta y en la próxima sección, añadiré algunas reflexiones sobre la posibilidad de que Ildefonso no asigne 1 a T, pero sí un valor muy cercano, mucho mayor que $0'997$.)

La segunda consideración es que quizá valga la pena continuar nuestro ejercicio de especulaciones y calcular Cr_i(roja) basándonos en alguna hipótesis sensata sobre cuál podría ser el valor desconocido Cr_i(roja|no-T). Tal vez no existe tal hipótesis mínimamente plausible. Eso nos deja de vuelta en la consideración anterior, con el enigma resuelto. A mi juicio, *si* existe una hipótesis plausible al respecto, estaría — al menos en parte — inspirada por alguna versión del *principio de indiferencia*.[9] Si tiene sentido asignar un valor a Cr_i(roja|no-T), creo que debería ser $0'5$ o mayor que $0'5$. Hagamos los cálculos asumiendo que $\mathrm{Cr}_i(\text{roja}|\text{no-T}) = 0'5$.

$$\begin{aligned}\mathrm{Cr}_i(\text{roja}) &= [\mathrm{Cr}_i(\text{roja}|\text{T}) \times \mathrm{Cr}_i(\text{T})] + [\mathrm{Cr}_i(\text{roja}|\text{no-T}) \times \mathrm{Cr}_i(\text{no-T})] \\ &= [0'000001 \times 0'997] + [0'5 \times 0'003] = 0'001500997\end{aligned}$$

La probabilidad inicial de que salga una canica roja también es muy pequeña, pero no tanto como bajo el supuesto $\mathrm{Cr}_i(\text{T}) = 1$. La diferencia resulta decisiva para el valor crucial, Cr_f(roja). En efecto:

$$\mathrm{Cr(roja|parece\text{-}roja)} = \mathrm{Cr}_i\mathrm{(roja)} \times \mathrm{r} \,/\, \mathrm{Cr}_i\mathrm{(parece\text{-}roja)}.$$
$$\mathrm{Cr}_i\mathrm{(parece\text{-}roja)} = [\mathrm{Cr(parece\text{-}roja|roja)} \times \mathrm{Cr}_i\mathrm{(roja)}] + [\mathrm{Cr(parece\text{-}roja|blanca)} \times \mathrm{Cr}_i\mathrm{(blanca)}].$$

Por consiguiente:

$$\begin{aligned}\mathrm{Cr}_f\mathrm{(roja)} &= \mathrm{Cr(roja|parece\text{-}roja)}\\ &= 0'001500997 \times \mathrm{r} \,/\, [\mathrm{r} \times 0'001500997] + [s \times 0'998499003]\\ &= 0'9678063.\end{aligned}$$

Explicitemos los resultados obtenidos. Hemos explorado la hipótesis de que Ildefonso no hubiera otorgado confianza plena a la proposición T (la conjunción de TV y SA), es decir que no le hubiera asignado 1. Siendo así, desconocemos el valor Cr_i(roja|no-T), y por consiguiente no podemos calcular Cr(roja|parece-roja); con lo cual carecemos de base para suponer que es inferior a 0′5. Siempre de conformidad con (CJ), la tesis (NJ) no tiene fundamento y no hay motivo para rechazar (C). En ese sentido, nuestro enigma inicial se ha disipado. Por otra parte, también hemos especulado partiendo de dos datos hipotéticos: que Cr_i(T) fuera 0′997 y que Cr_i(roja|no-T) fuera 0′5. Bajo esos supuestos, la tesis (NJ) es falsa: la justificación de Ildefonso para creer que la canica es roja es mayor que su justificación para creer que no es roja. De nuevo, no habría buenas razones contra (C).[10]

Mi defensa de la tesis (C) depende de que el sujeto no asigne probabilidad 1 a la información previa que se le proporciona. Para decirlo con mayor precisión, esa defensa requeriría que no le asignara 1, ni tampoco un valor "demasiado" próximo a 1, suponiendo que entendemos así esa noción: en las condiciones descritas, z es demasiado próximo a 1 si y sólo si (si $\mathrm{Cr}_i(\mathrm{T}) = z$, entonces Cr(roja|parece-roja) es menor que 0′5). Si Ildefonso asigna a T el valor 1 o un valor "demasiado" próximo a 1, entonces el enfoque bayesiano estándar volverá a ofrecer el resultado presentado en la sección 1: se cumple (NJ) y — asumiendo (CJ) — en tal caso Ildefonso no tiene conocimiento perceptivo de que la canica es roja, se infringe (C).

Mi propuesta en favor de (C) está globalmente mejor fundamentada si aceptamos — como yo acepto — esta otra tesis: asignar a T un valor "demasiado" próximo a 1 es una opción epistémicamente poco prudente por parte de Ildefonso. Cuando antes nos preguntábamos qué razones hay para suponer que Ildefonso no debió asignar 1 a T, contesté con dos tipos de consideraciones. A mi juicio, pueden reiterarse — en lo esencial — si nos preguntamos qué razones hay para suponer que Ildefonso no debió asignar a T un valor "demasiado" próximo a 1. Primero, me parece más apropiado preguntarse ¿qué razones habría para suponer que Ildefonso debió asignar a T un valor "demasiado" próximo a 1? Segundo, T no es una proposición relevantemente similar a una verdad lógica o matemática, ni a una proposición aprendida directamente a partir de la experiencia (si lo fuera y fuera correcto asignar a éstas probabilidad 1, podría quizá justificarse que T recibiera una asignación "demasiado" próxima a 1). Añado una reflexión adicional: como se verá por la discusión en la próxima sección, mi descripción general de las relaciones entre el caso CANICA y el caso TAXI pretende mostrar, en realidad, algunas conexiones entre nuestros valoraciones intuitivas de estos casos y nuestros juicios sobre si las asignaciones de probabilidades de un sujeto son o no suficientemente distantes de 1.[11]

Algunos filósofos quizá querrían sostener una defensa más radical de la tesis (C). No quiero comprometerme con la siguiente posición: *conviene preservar (C), incluso aunque Ildefonso hubiera asignado a T un valor "demasiado" próximo a 1*. La propuesta que he presentado no sirve para sustentar ese posicionamiento. Pero pueden hacerse unas reflexiones adicionales que apuntarían en su dirección. Si, como pretendo, seguimos aferrados al principio (CJ), la cuestión es si Ildefonso tiene buena justificación epistémica para creer lo que le parece percibir. En otras palabras, dados sus compromisos epistémicos (incluyendo haber asignado a T un valor "demasiado" próximo a 1) ¿debería Ildefonso creer que la canica es roja, o debería creer que no es roja? Surge la posibilidad de invocar alguna distinción entre varios sentidos de justificación — o proponer alguna no coincidente con las existentes en la literatura — para atribuir alguna ambigüedad a (CJ), y afirmar que conforme al sentido más relevante para el conocimiento, incluso en este caso Ildefonso tiene justificación para su creencia perceptiva y por consiguiente tiene conocimiento, preservándose (C).

A este respecto, también podría ser útil una analogía. Supongamos que una norma moral prohibe incumplir nuestras promesas y otra norma moral prohibe hacer R. Eva promete hacer R. ¿Debería hacer R? ¿Tiene justificación (moral, en este caso) para hacer R? No parece insensato decir lo siguiente: en cierto sentido, Eva debería hacer R, pues así lo ha prometido; aunque, en otro sentido, no debería hacer R. Todavía menos controvertido sería decir que Eva no debió haber prometido hacer R; y quizá por ello no debería hacer R, pese a haberlo prometido. Un epistemólogo podría mantener lo siguiente: Ildefonso no debería haber asignado a T un valor "demasiado" próximo a 1; aunque esa asignación le compromete a rechazar la creencia perceptiva, también está comprometido — en otro sentido — con aceptar dicha creencia perceptiva (porque no debió hacer esa asignación epistémicamente "imprudente"); este segundo compromiso es prioritario; por ello, tiene justificación, contra (NJ), y tiene también conocimiento perceptivo. Como he sugerido, no evaluaré la plausibilidad de esa posición, pues tiene muchas ramificaciones y complicaría innecesariamente la discusión principal. Para mis propósitos, me basta aquí con exponer — tras haber presentado el enigma en la sección 1 — una forma coherente de defender (C) sin renunciar al bayesianismo ni al principio (CJ).

5. Estimaciones intuitivas de probabilidades

Varias cuestiones requieren todavía ser comentadas. Una es la arbitrariedad del valor escogido para $\mathrm{Cr}_i(\mathrm{T})$, $0'997$. Otras conciernen a las diferencias entre el caso CANICA y el caso TAXI.[12] Ambos tipos de cuestiones están emparentadas. Con objeto de preparar el terreno para algunas de las reflexiones subsiguientes, consideremos otro caso, al que podemos llamar TORRES:

Jenny está en Nueva York el 9-11-2001 y ve un avión chocar contra una de las Torres Gemelas. Considera dos hipótesis alternativas: (I) ha sido un choque intencionado; (A) ha sido un accidente. Tal vez — dado su conocimiento "de trasfondo", basado en datos generales cotidianos — le parece bastante más probable la hipótesis (I) (aunque un avión accidentado sobrevolara N. York a esa altura, parece mucha casualidad ir a estrellarse contra la torre), pero la probabilidad que asigna — quizá implícitamente — a (A) también es significativa. Pocos minutos después, ve un segundo avión chocar contra la otra torre. Inmediatamente, la convicción de Jenny respecto a (I) aumenta muchísimo (Jenny cree que ambos choques han sido intencionados, y coordinados) y la probabilidad que asigna a (A) desciende hasta hacerse casi irrelevante. Asumo como dato poco controvertido que no hay nada anómalo ni irracional en esa sucesión de creencias de Jenny sobre (I) y (A). Por otra parte, cuando se le pregunta qué probabilidades, aunque fueran meramente aproximadas, había asignado a cada hipótesis relevante, Jenny no ofrece ninguna respuesta muy informativa. Tras reflexionar, alega que la probabilidad de un primer choque accidental contra la torre, (A), incluso tras observar el impacto, le parecía baja, pero la probabilidad de un primer choque accidental seguido de otro minutos después contra la otra torre (ya fuera éste accidental o intencionado) le parecía muchísimo menor, tan baja que la consideró irrelevante y quedó convencida de la hipótesis alternativa, (I). Todo ello sin tener remota idea de si una y otra probabilidad es del orden de una entre un millón, o una entre un billón, o ...

Traigo a colación el caso TORRES sólo para ilustrar otro dato que también me parece poco controvertido: como sujetos epistémicos, con frecuencia razonamos competentemente basándonos, como Jenny, en estimaciones intuitivas de probabilidades — por ejemplo, haciendo estimaciones comparativas — incluso cuando nuestra ignorancia sobre los valores de las probabilidades es enorme.

Volvamos a los casos CANICA y TAXI. Las diferentes importantes entre ellos son meramente cuantitativas, de carácter gradual. Pero esas diferencias permiten sostener (C) y rechazar (NJ), respecto al caso CANICA, al tiempo que se rechaza que la creencia perceptiva que tuviera el sujeto, Pedro, en el caso TAXI (el taxi era verde) constituya conocimiento; si creyera eso, Pedro incurriría en falacia de desestimación de la proporción de base, porque tiene mayor justificación para creer que el taxi no era verde.

He indicado, respecto a CANICA, que el hecho de que Ildefonso asigne a T un valor no “demasiado” próximo a 1 — por ejemplo, 0′997 — permite, al hacer los cálculos, que su creencia perceptiva esté justificada. Si no está justificada la creencia perceptiva de Pedro, ¿es porque asigna a la información previa (sobre la proporción de taxis verdes) un valor mucho mayor que 0′997? (No podríamos preguntarnos si Pedro asigna un valor “demasiado” próximo a 1, pues dicho concepto está definido solo para el caso CANICA.) No. No hay motivos para suponer que la confianza previa de Pedro en esa supuesta información deba ser muy diferente a la confianza previa de Ildefonso en T. (Si suponemos que Pedro asigna también 0′997 a esa información previa y aplicamos un principio de indiferencia de forma análoga a cómo lo hemos aplicado en CANICA, el resultado no difiere sustancialmente: Pedro sigue teniendo más justificación para creer que el taxi no era verde. Dejo para el lector la comprobación de este dato.) Las diferencias epistémicas relevantes entre ambos casos no derivan de esa confianza previa, que puede ser similar, sino de las otras diferencias ya mencionadas: los diversos grados de fiabilidad de su visión en las condiciones dadas; las diversas proporciones de canicas rojas y taxis verdes en las respectivas muestras.

Hemos seleccionado el valor 0′997 arbitrariamente. Con esa selección (tanto para T como para el dato sobre la proporción de taxis verdes), se preservan nuestras intuiciones sobre el caso TAXI y el caso CANICA: suponiendo que ambos sujetos forman la correspondiente creencia perceptiva, el primero (Pedro) comete la falacia y carece del correspondiente conocimiento perceptivo, mientras que el segundo (Ildefonso) no es falaz y tiene conocimiento perceptivo. Si Pedro o Ildefonso asignan un grado de confianza previa en la información recibida diferente a 0’997 pero dentro de lo que resulta “sensato” o “prudente” epistémicamente, se mantendrá ese resultado.

¿Qué base tenemos para afirmar eso? La misma que permite a Jenny formar sus creencias en el caso TORRES incluso sin tener idea — ni siquiera aproximada — de cuáles puedan ser las probabilidades respectivas de las distintas hipótesis. También la misma base que nos permite juzgar que Jenny no incurre en ninguna anomalía o irracionalidad aunque tampoco nosotros tengamos idea de esas cifras. Esa estimación intuitiva, informal de probabilidades también nos permite entender (habida cuenta de las diferencias entre CANICA y TAXI) que para que Pedro estuviera justificado en creer que el taxi era verde, se requeriría que su confianza previa en la información recibida fuera considerablemente baja, más baja de lo que parece prudente epistémicamente. Por supuesto, podemos inventar casos intermedios entre TAXI y CANICA en los cuales nuestras apreciaciones intuitivas serán mucho más inestables y — consiguientemente — más necesitadas de ser complementadas por el cálculo preciso.

Referencias

Bar-Hillel, M. 1980. The Base Rate Fallacy in Probability Judgments. *Acta Psychologica* **44**: 211–33.

Buchak, L. 2014. Belief credence and norms. *Philosophical Studies* **169**: 285–311.

Christensen, D. 1992. Confirmational Holism and Bayesian Epistemology. *Philosophy of Science* **59**: 540–57.

Dawes, F.; Meehl. 1989. Clinical versus actuarial judgment. *Science* **243**: 1668–74.

Jeffrey, R. C. 1965. *The Logic of Decision*. Nueva York: McGraw-Hill.

Leitgeb, H. 2013. Reducing Belief Simplicitar to Degrees of Belief. *Annals of Pure and Applied Logic* **164**: 1338–89.

Lewis, D. 1981. Causal Decision Theory. *Australasian Journal of Philosophy* **59**: 5–30. Reimpreso en sus *Philosophical Papers, vol. II*, pp. 305-339. Oxford: Oxford University Press, 1986.

Pollock, J. L.; Cruz, J. 1999. *Contemporary Theories of Knowledge*. 2^a edición. Maryland: Rowman y Littlefield.

Siegel, S. 2012. Cognitive Penetrability and Perceptual Justification. *Noûs* **46**: 201–22.

Titelbaum, M. G. 2013a. Ten Reasons to Care About the Sleeping Beauty Problem. *Philosophy Compass* **8**: 1003–17.

———. 2013b. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford: Oxford University Press.

Williamson, T. 2009. Probability and Danger. *The Amherst Lecture in Philosophy* **4**: 1–35. http://www.amherstlecture.org/williamson2009/

Notas

[1]Estoy agradecido a un/a anónimo/a evaluador/a de *Principia*, quien ha destacado la importancia del asunto comentado en los dos últimos párrafos.

[2]Un/a anónimo/a evaluador/a de *Principia* me ha hecho ver la conveniencia de modificar una versión previa del artículo, en la cual podía haber ambigüedad sobre este punto.

[3]Obviamente, Prob_i(roja) cuantifica la probabilidad *inicial* de que la canica sea roja; es decir, previamente a que Ildefonso tenga la experiencia sensorial de ver la canica. La probabilidad *final*, Prob_f(roja), es la magnitud que interesa averiguar. Esta Prob_f(roja) será Prob(roja|parece-roja). Para simplificar la notación, no uso subíndices cuando cierta probabilidad inicial, Prob_i(X), coincide con la correspondiente probabilidad final, Prob_f(X). Simplificaré análogamente respecto a Cr(X).

[4]Asumo la regla de "condicionalidad" estricta [*strict conditionalization rule*]: $\mathrm{Cr}_f(\mathrm{H}) = \mathrm{Cr}_i(\mathrm{H}|\mathrm{E})$, siendo *E* algo aprendido mediante la experiencia (algo tal que $\mathrm{Cr}_f(\mathrm{E}) = 1$). Aunque lo aprendido por la experiencia no es que la canica es roja, sino que parece roja. Cf. también las notas 8 y 10.

[5]Por las razones ya indicadas en la sección anterior — cuando mencionábamos el fenómeno de la penetración cognitiva —, esos datos sólo son verosímiles si se toman como meras aproximaciones. Es poco problemática la equivalencia entre Prob(parece-azul|azul) y Prob(parece-verde|verde). Pero si Prob(parece-azul|azul) es $0'8$, no resulta realista suponer

que Prob(parece-verde|azul) es exactamente $0'2$. En algunos casos de error perceptivo ante la visión de un taxi azul, dicho taxi no le parecerá verde, sino de algún tercer color (incluso asumiendo que Pedro espera ver sólo taxis azules o verdes).

[6]Según un lector de una versión previa de este texto, esa afirmación (que describe mi experiencia sobre la reacción al enigma de diversos filósofos) no tiene "base comprobable alguna". Salvo que por "base comprobable" entendamos algo muy alejado de lo que se exige a las hipótesis y las observaciones empíricas científicas, incluyendo las pertenecientes a las ciencias humanas, la tesis de ese lector es errónea. La primera parte de la experiencia en cuestión (presentar el enigma a filósofos) puede ser replicada (en el sentido de ser repetida) por otros investigadores, de forma que — en líneas generales — es comprobable si el resultado se asemeja suficientemente al descrito.

[7]Esa imposibilidad deriva de la regla de "condicionalidad" estricta. Pero hay opciones que se apartan de esa ortodoxia bayesiana. Destaca, a este respecto, el tratamiento de Titelbaum (2013b), que posibilita la *pérdida de la certeza* [certainty-loss] previa.

[8]Estoy asumiendo varias simplificaciones, en aras de la claridad. Conviene mencionar ahora dos de ellas. (1) En nuestro análisis, también hemos supuesto que Ildefonso asignaba 1 a otros datos de la información previa, sobre la fiabilidad de su sistema perceptivo (sobre los valores r y s) (análogamente respecto a los casos HOSPITAL y TAXI). Pero variar dichas asignaciones no repercute relevantemente en la discusión principal, por lo que mantendré tales supuestos. (2) He asumido, asimismo, que Ildefonso asignaba 1 a la proposición "la canica me parece roja" (análogamente respecto al caso TAXI). Un estudio más preciso y detallado de todo el enigma requeriría cuestionar también dicha asignación, pues parecería implicar la infalibilidad del conocimiento de nuestros datos sensoriales, negada por muchos epistemólogos. Pero ello complicaría mucho toda la exposición y, de nuevo, sin modificar significativamente el núcleo del problema y el sentido de mi enfoque. (Cf. la próxima nota 10.)

[9]Una formulación (de una versión restringida) del principio, apta para nuestros propósitos, podría ser la siguiente. Si Q es un enunciado del castellano y no se conoce razón alguna para considerar más probable la verdad de Q que la verdad del enunciado resultante de anteponer a Q la locución "No es el caso que", entonces debe asignarse la misma probabilidad a cada una de esas dos alternativas (las alternativas expresadas, respectivamente, por ambos enunciados).

[10]Destaco algunas diferencias entre el tratamiento bayesiano no estándar desarrollado por Jeffrey (1965) y las hipótesis que propongo. Yo sugiero no asignar 1 (certeza) a la información previa recibida por el sujeto. La "condicionalización" de Jeffrey propone no asignar siempre 1 a los datos resultantes de la observación. En mi análisis, ya he asumido no asignar 1 al dato observacional ("la canica es roja"); aunque — como indico en la nota 8 —, por razones de simplicidad, sí asigno 1 al dato proximal, expresado por "la canica parece roja".

[11]Según Lewis (1981, p.316) siempre sería epistémicamente irracional asignar valor 1 a una proposición, con la posible excepción de proposiciones derivadas del testimonio de los sentidos. Que tales proposiciones constituyan una excepción es muy controvertido; cf., por ejemplo, Christensen (1992) y Pollock y Cruz (1999, p.102). Nuestro enfoque se aleja de dicho supuesto.

[12]Algunas de las consideraciones que haré sobre nuestra apreciación implícita de probabilidades en el caso TAXI se aplican también a otros casos paradigmáticos de falacia de desestimación de la proporción de base, como HOSPITAL.

Reconocimientos

El tema de este trabajo se discutió en un grupo de lectura LOGOS sobre *Conocimiento, modalidad y probabilidades*, que coordiné en la Universidad de Barcelona durante el curso 2014-15. He presentado las ideas principales en dos conferencias invitadas: en el Seminario EPISOC (Univ. Autónoma de Madrid, 14-2-2017) y en el *Tercer Coloquio sobre Conceptos y Percepción* (Universidad Nacional de Córdoba, Argentina, 17-11-2017). Agradezco a los participantes sus comentarios y sugerencias. Estoy también en deuda con otras varias personas que han intercambiado conmigo sus puntos de vista sobre el problema epistemológico aquí analizado, especialmente Juan Comesaña, José A. Díez, Carl Hoefer, Daniel Kalpokas, Manolo Martínez, Fernando Martínez Manrique, Claudia Picazo, Daniel Quesada, Sven Rosenkranz, Ernesto Sosa, Ignacio Vicario y Agustín Vicente. Financiación: Proyectos "Objetividad-subjetividad en el conocimiento y en la representación singular" (FFI2015-63892-P), (MINECO, AEI/FEDER, UE) y "Perspectival Thoughts and Facts: New Questions" (FFI2016-81858-REDC; Red CONSOLIDER 2016, Acción de Dinamización, Red de Excelencia) (Gobierno de España/Unión Europea). / Grupo de investigación consolidado *LOGOS* (2017SGR63), AGAUR (Gobierno catalán).