teorema
Vol. XXIX/3, 2010, pp. 5-30
ISSN: 0210-1602
[BIBLID 0210-1602 (2010) 29:3; pp. 5-30]

Medidas de apoyo evidencial: un análisis comparativo

Valeriano Iranzo e Ignacio Martínez de Lejarza

ABSTRACT

There are different Bayesian measures to calculate the degree of confirmation of a hypothesis *h* in relation to an experimental result *e* (where degree of confirmation is understood as an *increment* in probability). Firstly, we try to identify the basic disagreement among those measures and, also, in which situations it does matter. We claim that the crucial point here is whether the evidential support which *e* gives to *h* depends on *h*'s prior probability or not. Then, we argue a negative answer to this question and, because of this, we favour the measure $\boldsymbol{S_r}$ $(= \log[p(h \mid e) / p(h)])$.

KEYWORDS: *Scientific Testing, Confirmation, Bayesianism, Evidential Support.*

RESUMEN

Existen diversas medidas bayesianas para cuantificar el grado en que un resultado experimental *e* confirma, en el sentido de incremento de probabilidad, una hipótesis *h*. Nuestra intención es, primero, discutir en qué circunstancias son sustantivas tales diferencias e identificar cuál es su raíz. La cuestión crucial, pensamos, es si el apoyo que *e* concede a *h* debe depender de la probabilidad inicial de *h*. La última parte del artículo argumenta una respuesta negativa a esta pregunta y defiende la medida $\boldsymbol{S_r}$ $(= \log[p(h \mid e) / p(h)])$.

PALABRAS CLAVE: *contrastación de hipótesis, confirmación, bayesianismo, apoyo evidencial.*

I. GRADOS DE CONFIRMACIÓN

Un problema clásico en filosofía de la ciencia es determinar cuándo la evidencia -el descubrimiento de un fenómeno empírico particular, un resultado experimental,…– apoya una hipótesis. La teoría cuantitativa de la confirmación pretende responder a esta cuestión vinculando estrechamente el "*grado de confirmación*" de una hipótesis *h* en relación a una evidencia *e* con la noción matemática de probabilidad condicionada: $p\ (A \mid B)$. Desde sus inicios, sin embargo, la teoría de la confirmación se vio aquejada por una ambigüedad; de ahí los dos sentidos de confirmación que Carnap distinguiera en la segunda edición de *The Logical Foundations of Probability*: confirmación como "firmeza" (*firmness*) y como "incremento de firmeza" (*increase in firmness*) [Carnap (1962), prefacio]. En el primer caso h es confirmada por e cuando la probabilidad de h dado e supera cierto umbral (por ejemplo, cuando es mayor que 0'5 y por tanto, cuando es más probable h dado e que $\neg h$ dado e). En su segunda acepción, e confirma a h cuando e produce un incremento en la probabilidad de h respecto a la probabilidad que ésta tenía sin contar con e. Debe diferenciarse, entonces, entre el valor de la probabilidad de la hipótesis condicionada a la evidencia disponible, $p\ (h \mid e\ \&\ b)$, siendo b el cuerpo de enunciados aceptados, o "conocimiento de fondo", y el grado de *apoyo* que una porción de evidencia confiere a la hipótesis en cuestión, esto es, la variación que provoca e en $p\ (h \mid e\ \&\ b)$ en relación a $p\ (h \mid b)$. Dicho brevemente, cuán probable es h dado e y b, en términos absolutos, no es lo mismo que *cuánto más* (o menos) probable es h a consecuencia de e y b.

El Teorema de Bayes, una consecuencia deductiva de los axiomas y definiciones de la teoría matemática de la probabilidad, es crucial en este punto. Aunque dicho teorema versa sobre *sucesos*, en el sentido matemático del término, los filósofos de la ciencia bayesianos lo que hacen es interpretarlo "confirmacionalmente". Así, la función de probabilidad se aplica tomando enunciados (hipótesis, descripciones de la evidencia,…) como valores de sus variables:

$$p(h \mid e \,\&\, b) = \frac{p(h \mid b) \cdot p(e \mid h \,\&\, b)}{p(e \mid b)} \tag{1}$$

siempre que $p\ (e \mid b) \neq 0$, de modo que:

> $p\ (h \mid b)$ es la *probabilidad inicial* (*prior probability*) de h, o sea, la probabilidad de h al margen de la evidencia e;
>
> $p\ (e \mid h\ \&\ b)$ es la *verosimilitud* (*likelihood*) de h, esto es, la probabilidad de e si nuestra hipótesis y el conocimiento de fondo fueran verdaderos;
>
> $p\ (e \mid b)$ es la probabilidad de e de acuerdo con el conjunto de conocimientos aceptados sin contar con h;
>
> $p\ (h \mid e\ \&\ b)$ es la *probabilidad condicionada* de h, esto es, la probabilidad de h dado que $e\ \&\ b$ sea verdadero.

El *grado de confirmación*, como *"firmeza*", de h respecto a e se define así:

$$\boldsymbol{C}\ (h,e) =_{def}\ p(h \mid e \,\&\, b) \tag{2}$$

Para los bayesianos contemporáneos (2) sigue siendo una de sus señas de identidad, a pesar de que rechazan la noción de probabilidad lógica defendida por Carnap en las décadas de los 50 y 60 del pasado siglo, y de que consideran que las probabilidades no son más que grados de creencia subjetivos (*credences*). Esto último, no obstante, no les impide postular una normatividad epistémica, cuya fuente son las leyes de la probabilidad matemática. En el terreno de la ciencia, esto significa que la credibilidad racional de una hipótesis viene dada por su probabilidad condicionada. ***C*** especifica, pues, cuál es la probabilidad de *h* dado que *e* y *b* son correctos, pero al tiempo nos está dando también la *credibilidad* o *plausibilidad* que merece *h*. Dado que ***C*** es una función de probabilidad, $0 \leq \boldsymbol{C} \leq 1$. Un valor cercano a uno, por ejemplo, para $p\ (h_i \mid e\ \&\ b)$ indica que h_i es altamente plausible.

Sin embargo, a veces lo que nos interesa determinar es la repercusión que un resultado experimental particular tiene sobre la plausibilidad previa de la hipótesis. En este caso lo que se quiere calcular es el efecto neto de *e*, sea un incremento o una disminución, sobre la probabilidad de *h* sin contar con *e*. Se trata, pues, de una noción *incremental* de confirmación:

$$\boldsymbol{S}\ (h,e) = \Delta_e(p(h \mid b)) \tag{3}$$

Se han propuesto múltiples expresiones para referirse a ***S***: *increase in firmness* (R. Carnap), *factual support* (J. Kemeny y P. Oppenheim), *weight of evidence* (I. J. Good), *degree of support* (D. Gillies), *evidential support* (E. Eells y B. Fitelson), *evidential relevance* (J. Joyce), *incremental confirmation* (F. Huber)..., pero ninguna de ellas se ha impuesto. Por eso muchos autores prefieren hablar simplemente de confirmación, sobreentendiendo que se trata de confirmación incremental. Para referirnos a la propiedad que las ***S***-medidas intentan cuantificar emplearemos indistintamente 'confirmación incremental' o 'apoyo evidencial'. Lo que debe tenerse en mente en todo momento es que ***C*** y ***S*** miden cosas distintas. En el primer caso tenemos una asignación de probabilidad condicionada que, como tal, debe ser consistente con los axiomas de la teoría matemática de la probabilidad, mientras que ***S*** no cuantifica la probabilidad en sí misma; no es, por tanto, una función de probabilidad, y por eso no necesariamente ha de respetar aquéllos (por ejemplo, ***S*** puede tener un valor negativo si el efecto de *e* es disminuir en vez de incrementar la probabilidad previa de *h*).

Dejando a un lado las dificultades técnicas que puedan surgir en el proceso, para un bayesiano no hay controversia sobre el cálculo de ***C***: no hay más que aplicar el Teorema de Bayes. Por contra, hay múltiples propuestas sobre cuál es la función matemática más adecuada para medir el incremento producido al pasar de $p\ (h \mid b)$ a $p\ (h \mid e\ \&\ b)$. El problema de la pluralidad de medidas surge, por tanto, respecto a la confirmación incremental y es, en realidad, el problema de la pluralidad de las ***S***-medidas.

No faltan voces autorizadas [v. por ejemplo, Howson (2003), 183-85] que consideran que ésta no es una cuestión importante. En el marco bayesiano lo que determina, y lo que permite comparar, la credibilidad de las alternativas disponibles en un dominio, $h_1, h_2, \ldots, h_n$, es ***C***: la hipótesis más plausible es, simplemente, la que posee una probabilidad condicionada $p\ (h_i \mid e\ \&\ b)$ mayor. ¿Para qué complicarse entonces introduciendo una medida adicional?

Podría pensarse, pues que al bayesiano le basta con ***C***. Pero quienes defienden la importancia de la noción de confirmación incremental alegan por su parte que, tanto en nuestros juicios cotidianos relativos a la confirmación como en las estimaciones que hacen los científicos sobre el respaldo empírico de una hipótesis, lo que se evalúa muy a menudo no es la credibilidad de un enunciado, sino el diferencial generado por una porción de evidencia particular. Es posible que una hipótesis esté fuertemente apoyada por un resultado experimental concreto, a pesar de que su credibilidad global siga siendo baja (o a la inversa); y nada impide que *e* juegue en contra de *h*, a pesar de que la plausibilidad de *h* siga siendo, aun contando con *e*, alta. Por eso, aunque concedamos a ***C*** un papel fundamental en la dinámica de la ciencia, sosteniendo que la aceptación o el rechazo de una hipótesis quedan enteramente determinados por su probabilidad, sin que intervengan otro tipo de consideraciones[1], conviene insistir de cara a nuestra argumentación posterior, en que a menudo lo que nos interesa es precisar si la evidencia e_1 confirma a h_1 en mayor grado que a h_2, *y ello con independencia de la credibilidad que corresponda en términos absolutos a tales hipótesis*. Esta es una situación relativamente común en una investigación científica en curso. De hecho, una parte esencial de la tarea del investigador consiste en evaluar el respaldo empírico respectivo que los hallazgos experimentales aportan a hipótesis rivales. Visto así, el problema de la pluralidad de las ***S***-medidas, sobre el que muchos bayesianos pasan de puntillas, no es una cuestión baladí. La ausencia de consenso supone un quebradero de cabeza para quienes piensan que la medida de la confirmación incremental es una pieza básica de la teoría bayesiana de la confirmación. Lo que está en cuestión, a fin de cuentas, es la operatividad de la propia noción.

II. La pluralidad de medidas de confirmación incremental (***S***-medidas)

Una regla de juego asentada en la discusión contemporánea es el "criterio de relevancia", según el cual:

si $p\ (h \mid e) > p\ (h)$, $\boldsymbol{S}\ (h, e) > 0$

si $p(h \mid e) = p(h)$, $\boldsymbol{S}(h, e) = 0$

si $p(h \mid e) < p(h)$, $\boldsymbol{S}(h, e) < 0$

Este criterio es, ante todo, un requisito de normalización puramente formal[2]. De hecho, la mayoría de las ***S***-medidas sugeridas lo cumplen, o admiten versiones normalizadas que lo satisfacen. Del criterio se sigue, en fin, una definición de *confirmación incremental* en relación a los valores numéricos de referencia:

e confirma h sii $\boldsymbol{S}(h, e) > 0$

e disconfirma h sii $\boldsymbol{S}(h, e) < 0$

e es *irrelevante* para *h*, *sii e* ni confirma ni disconfirma *h*, o sea, *sii* $\boldsymbol{S}(h, e) = 0$

Aunque entre las ***S***-medidas propuestas hay más de una decena que cumplen con el criterio de relevancia [para una lista exhaustiva, v. Crupi, Tentori y González (2007)], señalamos a continuación las que más presencia han tenido en la discusión:

$$\boldsymbol{S}_d(h,e) =_{\text{def}} p(h \mid e) - p(h) \qquad (4)$$

$$\boldsymbol{S}_r(h,e) =_{\text{def}} log \frac{p(h \mid e)}{p(h)} \qquad (5)$$

$$\boldsymbol{S}_l(h,e) =_{\text{def}} log \frac{p(e \mid h)}{p(e \mid \neg h)} \qquad (6)$$

Nuestro objetivo es dirimir las discrepancias entre dichas medidas, cada una con su legión de partidarios[3], y justificar las ventajas que pudiera haber para preferir alguna de ellas. Pero primero repararemos en cuál ha sido el planteamiento predominante a la hora de abordar la cuestión.

II.1 *Cómo (no) afrontar el problema*

El objetivo del análisis carnapiano [Carnap (1962)] fue proporcionar una explicación (*explicatum*) adecuada de la noción de confirmación. A pesar de su meritorio trabajo, Carnap no llegó a ninguna conclusión firme al respecto. Sin embargo, aquél sigue siendo el objetivo perseguido por algunos de los participantes más activos en el debate sobre las ***S***-medidas. A continuación comentaremos algunas propuestas concretas en esta línea.

La estrategia seguida en Eells y Fitelson (2002) consiste en plantear una serie de simetrías relativas a la confirmación para precisar después cuáles de ellas deberían ser respetadas y cuáles no, según nuestras intuiciones preanalíticas. Las simetrías que discuten son:

— simetría evidencial (ES): e confirma h en el mismo grado que $\neg e$ disconfirma h;

— simetría conmutativa (CS): e confirma h en el mismo grado que h confirma a e;

— simetría total (TS): e confirma h en el mismo grado que $\neg e$ confirma a $\neg h$;

— simetría de hipótesis (HS): e confirma h en el mismo grado que e disconfirma $\neg h$.

Según Eells y Fitelson, solamente HS debe ser satisfecha bajo cualquier elección de e y h. A continuación comparan cinco ***S***-medidas diferentes para ver si se comportan del modo esperado respecto a las simetrías mencionadas. Centrándonos en las medidas consignadas anteriormente, $\boldsymbol{S_d}$, $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$, la segunda ha de rechazarse porque incumple HS, aunque satisface CS, mientras que $\boldsymbol{S_d}$ y $\boldsymbol{S_l}$ sí cumplen HS. Hay, por tanto, más de una alternativa; luego los criterios de simetría no bastan por sí solos para seleccionar una única ***S***-medida[4].

Más recientemente B. Fitelson ha refinado esta línea de ataque por su cuenta [v. Fitelson (2006)]. Bajo el planteamiento carnapiano de que la idea central de una lógica inductiva es la noción de confirmación, y que la función de confirmación debe ser una generalización cuantitativa de la relación de consecuencia lógica, Fitelson plantea el requisito de la "logicidad" (*logicality*) como una exigencia que debe cumplir cualquier ***S***-medida.

Según este requisito, la confirmación incremental obtenida por h a consecuencia de e, $\boldsymbol{S}\ (h, e)$ debería ser máxima cuando $e \models h$, y mínima cuando $e \models \neg h$. Sin embargo, casi ninguna medida respeta la logicidad. Así, en el caso de $\boldsymbol{S_d}$, por ejemplo, si h_1 y h_2 (o sus negaciones) son implicadas por e, puede ocurrir, no sólo que no se alcance el valor máximo, 1, (mínimo, –1), sino que además $\boldsymbol{S_d}\ (h_1, e)$ y $\boldsymbol{S_d}\ (h_2, e)$ tomen valores diferentes. Por ejemplo, sea e = sacar el as de picas, h_1 = sacar una carta negra, h_2 = sacar una pica. Entonces, $\boldsymbol{S_d}\ (h_1, e)$ = 0'5 y $\boldsymbol{S_d}\ (h_2, e)$ = 0'75. Fitelson aduce que, dejando aparte casos límite en que e y h no son contingentes (o sea, cuando $p\ (e)$ y $p\ (h)$ están entre cero y uno), de las tres medidas mencionadas solamente $\boldsymbol{S_l}$ satisface la logicidad. De ahí concluye que $\boldsymbol{S_l}$ es, de las medidas propuestas hasta ahora, la más adecuada [Fitelson (2006), p. 512].

Pero el empeño de Fitelson en introducir un criterio para filtrar las candidatas dejando sólo una no está claro que de sus frutos. Nótese que siendo p (h) contingente, y por tanto, $0 < p\ (h) < 1$, tal como asume Fitelson en su argumentación, $\boldsymbol{S_l}$ no estará definida cuando h es una consecuencia lógica de e, o sea, cuando $p\ (h \mid e) = 1$, o lo que es equivalente, cuando $p\ (\neg h \mid e) = 0$. En tales condiciones, puede comprobarse mediante el Teorema de Bayes que, si $p(\neg h \mid e) = 1$, entonces $p\ (e \mid \neg h) = 0$. Podemos suponer que $\boldsymbol{S_l}$ aumenta (disminuye) cuanto mayor (menor) es $p\ (h \mid e)$, siendo $+\infty$ y $-\infty$ sus límites asintóticos; pero, en sentido estricto, cuando h es una consecuencia lógica de e, el denominador de (6) vale cero, con lo cual $\boldsymbol{S_l}$ simplemente no está definida en tal caso y, por consiguiente, nada en absoluto puede decirnos sobre cuánto confirma e a h.

Como hemos visto, Eells y Fitelson (2002) solamente discuten cuatro simetrías, justamente las cuatro que ya fueran abordadas por Carnap [Carnap (1962), secc. 67]. En la misma línea de trabajo, Crupi *et. al.* (2007) explora exhaustivamente todas las posibles simetrías e introduce un principio semejante al criterio de logicidad de Fitelson, para determinar cuáles de aquellas han de ser satisfechas por una $\boldsymbol{S}$-medida adecuada. Sea $\boldsymbol{V}$ una función que asigna un valor positivo, el mismo, a todo argumento construido con h y e *sii* $e \models h$ un valor equivalente, pero negativo, *sii* $e \models \neg h$ y 0 en cualquier otro caso. El principio de Extensión (del dominio deductivo al inductivo) propuesto por Crupi *et al.* dice que "si $\boldsymbol{V}\ (h_1, e_1) > \boldsymbol{V}\ (h_2, e_2)$, entonces $\boldsymbol{S}\ (h_1, e_1) > \boldsymbol{S}\ (h_2, e_2)$" y garantiza que "a cualquier argumento concluyentemente confirmatorio (tal que $e \models h$) se le asigna un valor más alto que a cualquier otro que no lo sea, y cualquier argumento concluyentemente disconfirmatorio tendrá un valor más bajo que todo el que no lo sea" [Crupi *et al.*, p. 232]. Como ninguna de las medidas propuestas hasta ahora, ni siquiera $\boldsymbol{S_l}$, lo cumple, estos autores propugnan una nueva, $\boldsymbol{S_z}$[5].

Crupi *et. al.* (2007) supone una vuelta más de tuerca en una línea de trabajo cuyos resultados son controvertidos[6]. Aun cuando es posible proseguir la discusión en este sentido, conviene destacar una limitación importante. Nótese que tanto en los ejemplos intuitivos sobre qué simetrías deben respetarse o en las apelaciones al criterio de logicidad por parte de Fitelson, como en el principio de Extensión propuesto en Crupi *et al.* (2007), los casos decisivos, los que supuestamente despejan el empate entre las medidas de apoyo rivales, son aquellos en los que la evidencia implica lógicamente la hipótesis o la refuta concluyentemente, o sea, cuando $e \models h$ ó $e \models \neg h$. Lo paradójico, en nuestra opinión, es que los criterios que determinan qué medidas de confirmación son correctas se apuntalan en intuiciones sobre ejemplos en los que la conexión entre evidencia e hipótesis no es inductiva, en un sentido genuino, sino *deductiva*, cuando el sentido de introducir todo el aparato matemático-probabilístico no es sino superar las limitaciones de una teoría cualitativa clásica de la confirmación, como la hempeliana, circunscrita a las conexiones deductivas. Algo va desencaminado cuando, en el momento decisivo, las intuiciones invocadas para validar una definición cuantitativa del apoyo inductivo se extraen de ejemplos que plantean una conexión *deductiva* entre evidencia e hipótesis[7]. El aparato probabilístico introducido para abordar la problemática de la confirmación muestra su potencia justamente en los casos en los que no hay una conexión deductiva. La cuestión es, pues, cuán relevantes son los criterios invocados hasta aquí para decidir entre las ***S***-medidas[8].

La óptica más aconsejable para avanzar, en nuestra opinión, pasa por desvincular en lo posible el problema de la pluralidad de ***S***-medidas del proyecto analítico de elaborar una teoría general de la confirmación según el modelo de la "lógica" inductiva. El punto de partida debe ser identificar los contextos donde la pluralidad resulta problemática. Son variopintos los ejemplos en los que puede ser conveniente apelar a una estimación, siquiera aproximada, de la confirmación incremental: averiguar el riesgo de que una persona contraiga una enfermedad conforme vamos conociendo los factores antecedentes, justificar un veredicto de culpabilidad o inocencia en un juicio según se acumulan las pruebas, interpretar un mensaje encriptado ($\boldsymbol{S_l}$, en particular, fue propuesta por A. Turing e I. J. Good en la segunda guerra mundial, cuando colaboraban con las autoridades militares británicas en esta tarea), o dirimir el apoyo respectivo que un fenómeno empírico concede a hipótesis teóricas rivales, como la deflexión de la luz frente a las teorías newtoniana y einsteiniana de la gravitación. Sin embargo, una noción *cuantitativa* de confirmación tiene su lugar idóneo, aunque no exclusivo, en el ámbito de la ciencia. Esta es una razón de peso, en nuestra opinión, para referir el debate sobre la adecuación de las ***S***-medidas a los contextos propios de la investigación científica. A ello dedicaremos los apartados siguientes.

II.2 *Situaciones-tipo*

Como situaciones típicas en el contexto de la investigación científica donde el cálculo de ***S*** se hace aconsejable cabe reseñar las siguientes:

— se piensa que merece la pena seguir investigando sólo *h* si *e* la confirma, y que hemos de dejarla a un lado en caso contrario;

— se quiere defender *h* frente a diversas porciones de evidencia negativa; en este caso *e*, *e'*,…, disconfirman *h*, pero nos interesa saber cuál de ellas es la que tiene un impacto negativo mayor, por ejemplo, sobre *h*, para contrarrestarla de algún modo buscando otras consecuencias observacionales más favorables;

— de entre las hipótesis rivales, h, h',..., se considera que hay que concentrar esfuerzos y recursos en la que obtenga más apoyo de e;

— de entre las hipótesis rivales, h, h',..., se quiere comparar el apoyo respectivo que diversos resultados experimentales, e, e',…, les confieren.

Estas situaciones remiten a una tipología básica con su cuestión correspondiente:

T1 "Una evidencia-una hipótesis": ¿cuánto confirma e a h?

T2 "Múltiple evidencia-una hipótesis": ¿confirma e a h en mayor (igual, menor) grado que e'?

T3 "Una evidencia-múltiples hipótesis": ¿confirma e a h en mayor (igual, menor) grado que a h'?

T4 "Múltiple evidencia-múltiples hipótesis": ¿es h confirmada por e en mayor (igual, menor) grado que h' por e'?[9]

Comenzando por T1, la situación más elemental, no es de extrañar que el grado en que e confirma h varíe según se emplee una u otra $\boldsymbol{S}$-medida. Ya que se trata de funciones diferentes, siendo $\boldsymbol{S_1}$ y $\boldsymbol{S_2}$ dos medidas cualesquiera, en general ocurre que $\boldsymbol{S_1}(h, e) \neq \boldsymbol{S_2}(h, e)$. Quien piense que esto es un problema está admitiendo implícitamente que el valor exacto de aquello que se mide es decisivo aquí, pero no parece que lo sea, máxime cuando las simulaciones realizadas muestran un grado de correlación elevado entre las diversas medidas[10]. Los resultados de las $\boldsymbol{S}$-medidas son importantes, en nuestra opinión, en tanto se convierten en guía para tomar decisiones relativas al curso de la investigación. El problema, en relación a T1, se produciría, entonces, si e es evidencia positiva para h según $\boldsymbol{S_1}$ y evidencia neutra, o negativa, según $\boldsymbol{S_2}$, por ejemplo. Pero, a pesar de que en general $\boldsymbol{S_d} \neq \boldsymbol{S_r} \neq \boldsymbol{S_l}$, nunca puede ocurrir que conduzcan a un veredicto *cualitativo* discrepante. Esto se sigue del hecho de que las tres cumplen el criterio de relevancia expuesto en la sección 1.

Así, es fácil ver que $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$ lo cumplen, puesto que $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$ >/=/< 0, *sii* $p(h \mid e)$ >/=/< $p(h)$. Por tanto, $\boldsymbol{S_d}(h, e)$ >/=/< 0 *sii* $\boldsymbol{S_r}(h, e)$ >/=/< 0. $\boldsymbol{S_l}$, por su parte, se define apelando a $p(e \mid h) / p(e \mid \neg h)$, el denominado "factor de Bayes", llamémosle B, de h respecto a $\neg h$. Del teorema de Bayes se sigue que B >/=/< 1 *sii* $p(h \mid e)$ >/=/< $p(h)$, o sea que para $\boldsymbol{S_l}$ también es verdad que $\boldsymbol{S_l}$ >/=/< 0, *sii* $p(h \mid e)$ >/=/< $p(h)$. La coincidencia entre las tres medidas en los casos T1 queda garantizada. Con otras palabras, si el cálculo de $\boldsymbol{S}$ persiguiera únicamente determinar si e es evidencia positiva, negativa o neutral respecto a la hipótesis, el problema de la pluralidad de las medidas se habría disuelto. Las discrepancias sustantivas entre las $\boldsymbol{S}$-medidas surgirán, en todo caso, en T2, T3 y T4, situaciones que involucran una *comparación*, y que exigen, por tanto, un *listado ordenado* de las evidencias y/o de las hipótesis en función del apoyo que confieren o reciben, respectivamente.

II.3 *Equivalencia ordinal y desacuerdos comparativos entre las **S**-medidas*

El requisito de equivalencia ordinal (EO), dice que dos medidas S_1 y S_2 son *ordinalmente equivalentes* sólo si para todo *h*, *e*, *h'*, y *e*' se cumple:

$$\boldsymbol{S_1}(h, e) \geq \boldsymbol{S_1}(h', e') \text{ sii } \boldsymbol{S_2}(h, e) \geq \boldsymbol{S_2}(h', e')$$

Hay consenso en que las ***S***-medidas tomadas dos a dos no satisfacen EO [v. Fitelson (2001), p. 6, y las referencias incluidas en Fitelson (1999), nota 6], con lo cual queda acotado el ámbito de discrepancia entre aquéllas.

Por nuestra parte, estamos de acuerdo en que el problema de la pluralidad de las ***S***-medidas radica en las discrepancias en cuanto al ordenamiento de hipótesis y/o evidencias. Creemos, sin embargo, que el incumplimiento de EO no es decisivo para poder afirmar un conflicto sustantivo entre las medidas. Nótese que no se está imponiendo ninguna restricción sobre el contenido de la evidencia o la hipótesis. Mas, ¿qué objeto puede tener comparar el grado de confirmación entre determinados restos óseos, *e*, y la hipótesis *h* de que la bipedestación humana se originó por un cambio climático, por un lado, con el que confiere un resultado positivo de una prueba médica, *e'*, a la hipótesis *h'* de que el paciente sufre una infección vírica? Aunque nuestro problema surge en contextos comparativos, cabe remarcar que, por lo común, en el marco de la investigación científica *las comparaciones relevantes se establecen entre hipótesis rivales, o entre diversos resultados experimentales en cuanto éstos pueden apoyar diferencialmente a la hipótesis en cuestión o a sus rivales.* Una vez admitido que $\boldsymbol{S_1}(h, e) \neq \boldsymbol{S_2}(h, e)$ no es preocupante, y que $\boldsymbol{S_1}$ y $\boldsymbol{S_2}$ coinciden en cuanto al carácter confirmador o disconfirmador de *e* respecto a *h*, tal como se vio en la sección anterior, que $\boldsymbol{S_1}$ considere que unos restos óseos determinados apoyan en mayor medida la hipótesis de la bipedestación como consecuencia de un cambio climático, por una parte, de lo que el resultado positivo del test confirma la hipótesis de que un individuo tiene SIDA, por otro, mientras que para $\boldsymbol{S_2}$ ocurre al revés, no parece un gran contratiempo, aunque ello suponga violar EO. Sí parece más grave que, según $\boldsymbol{S_1}$, la deflexión de la luz medida en el eclipse solar de 1919 cuente más a favor de la Teoría General de la Relatividad (TGR) que la comprobación experimental de la dilatación del tiempo en relojes situados en la superficie terrestre, y que $\boldsymbol{S_2}$ afirme lo contrario; o que la deflexión apoye más a la mecánica newtoniana (MN) que a TGR, según $\boldsymbol{S_1}$, y que ocurra lo contrario según $\boldsymbol{S_2}$, dado que MN y TGR son lógicamente incompatibles. Estas son situaciones T2 y T3, pero el lector podrá imaginar por su cuenta otros ejemplos para el caso T4. El incumplimiento de EO no es, en definitiva, preocupante a menos que la comparación en cuestión se plantee *en un contexto de rivalidad*. Por eso pensamos que es ahí donde su satisfacción por parte de las ***S***-medidas debe ser planteada y discutida. ¿Cómo reformular entonces EO de modo que esto sea tenido en cuenta?

Intuitivamente, dos hipótesis son rivales en tanto se pronuncian sobre un dominio empírico compartido, ofrecen respuestas diferentes a las mismas preguntas, etc. Pero para analizar el comportamiento de las ***S***-medidas se necesita una caracterización formal de la rivalidad. Que h y h' sean rivales en un sentido genuino implica que son mutuamente excluyentes. No obstante, la opción de identificar rivalidad con incompatibilidad lógica no sirve. Visto así, h y $\neg h$ serían hipótesis rivales, y de hecho, EO podría incorporar este caso sencillo haciendo equivalentes h' y e' a $\neg h$ y $\neg e$, respectivamente. Mas no es éste el sentido de rivalidad que suele interesar a metodólogos y filósofos de la ciencia, ya que a menudo lo que se dirime no es la hipótesis enfrentada a su negación, sino varias hipótesis o explicaciones *alternativas* sobre unos resultados experimentales. Pensamos que puede dotarse de un contenido formal preciso a la idea de rivalidad mediante una noción capital en la teoría de la probabilidad como es la de *partición*. Dado un conjunto de hipótesis $\boldsymbol{H} = \{h_1, h_2,..., h_n\}$, se dice que $\boldsymbol{H}$ constituye una partición del espacio muestral cuando las hipótesis son lógicamente incompatibles tomadas dos a dos y si, además, la unión de todas ellas equivale a la totalidad del espacio muestral. Esto último implica que $p(h_1) + p(h_2) + \ldots + p(h_n) = 1$. Dicho esto, cualquier par de elementos (h_i, h_j) de la partición incluye hipótesis que son rivales entre sí[11].

En la práctica estadística el Teorema de Bayes se aplica comúnmente presuponiendo una partición del espacio muestral. Por eso, en vez de la expresión (1), en la que no se hace ninguna referencia a los elementos que constituyen la partición a la que pertenece h, se prefiere una formulación equivalente. La partición más sencilla de un espacio muestral contiene solamente dos elementos, h y $\neg h$, ya que $\neg(h \,\&\, \neg h)$ y, además, $p(h) + p(\neg h) = 1$. Para este caso, eliminando la referencia al conocimiento de fondo, el Teorema de Bayes se formula así:

$$p(h \mid e) = \frac{p(h) \cdot p(e \mid h)}{p(h) \cdot p(e \mid h) + p(\neg h) \cdot p(e \mid \neg h)} \tag{7}$$

En términos más generales, para un espacio muestral discreto constituido por n hipótesis, el teorema adopta la siguiente forma:

$$p(h_j \mid e) = \frac{p(h_j) \cdot p(e \mid h_j)}{\sum_{i=1}^{n} p(h_i) \cdot p(e \mid h_i)} \qquad j=1, 2,\ldots, n \tag{8}$$

Naturalmente, el valor del denominador será idéntico para cualquier h.

De acuerdo con nuestra idea de que la equivalencia ordinal sólo cabe exigirla en las situaciones de rivalidad, la discusión de las medidas puede típicamente realizarse en el contexto de una partición, encajando en ese formato los ejemplos que la ilustren En T3 y T4, los casos en los que hay múltiples hipótesis, esto obliga a determinar los puntos del espacio muestral, las hipótesis, de modo exhaustivo[12]. Además, con objeto de acotar con precisión el alcance de las discrepancias conviene reformular EO, el requisito de equivalencia ordinal introducido al comienzo de este subapartado, en estrecha correspondencia con las situaciones tipo T2, T3 y T4. Así, siendo e y e' dos porciones de evidencia cualesquiera, y $\boldsymbol{H} = \{h_1, h_2, ..., h_n\}$ una partición del espacio muestral,

— EO2: Para todo h_i, e y e', $\boldsymbol{S_1}(h_i, e) \geq \boldsymbol{S_1}(h_i, e')$ *sii* $\boldsymbol{S_2}(h_i, e) \geq \boldsymbol{S_2}(h_i, e')$

— EO3: Para todo e y todo $h_i, h_j \in \boldsymbol{H}$, $\boldsymbol{S_1}(h_i, e) \geq \boldsymbol{S_1}(h_j, e)$ *sii* $\boldsymbol{S_2}(h_i, e) \geq \boldsymbol{S_2}(h_j, e)$

— EO4: Para todo e y e' y todo $h_i, h_j \in \boldsymbol{H}$, $\boldsymbol{S_1}(h_i, e) \geq \boldsymbol{S_1}(h_j, e')$ *sii* $\boldsymbol{S_2}(h_i, e) \geq \boldsymbol{S_2}(h_j, e')$

Sería incorrecto pensar que si se cumplen EO2 y EO3, se cumplirá también EO4, aunque el resultado recíproco sí es cierto. No obstante, cabe la posibilidad de reducir las situaciones T4 (múltiple evidencia/múltiples hipótesis) a las T3 (una evidencia/múltiples hipótesis) conjuntando e y e': $\boldsymbol{S_1}(h_i, e \,\&\, e') \geq \boldsymbol{S_1}(h_j, e \,\&\, e')$ *sii* $\boldsymbol{S_2}(h_i, e \,\&\, e') \geq \boldsymbol{S_2}(h_j, e \,\&\, e')$. No tenemos por qué pensar que esto es una maniobra puramente *ad-hoc*. Al exigir que h_i y h_j sean rivales, estamos admitiendo que hay un dominio empírico compartido, y que ambas se pronuncian sobre éste. Cabe pensar entonces que no habría una posición de ignorancia total respecto a las verosimilitudes correspondientes, $p(e \,\&\, e' \mid h_i)$ y $p(e \,\&\, e' \mid h_j)$, componentes imprescindibles para calcular (8). Sería necesario, eso sí, que e y e' no fueran lógicamente incompatibles, pues, en caso contrario, $\sum p(h_i) \cdot p(e \,\&\, e' \mid h_i) = 0$, con lo cual $\boldsymbol{C}(h_i, e)$ no estaría definida, y consiguientemente, $\boldsymbol{S}(h_i, e)$ tampoco.

El lector podría apuntar aquí que esta estrategia de reducción no reconoce el papel de las contrastaciones "cruciales" en la ciencia, contrastaciones en las cuales se apela a evidencias *incompatibles* entre sí[13]. Lo que se busca en estos casos es una evidencia sobre la que las hipótesis rivales (h_i, h_j) se pronuncian en sentido opuesto. Aparentemente son situaciones T4, con la peculiaridad de que la evidencia múltiple consiste, no en dos piezas de evidencia cualesquiera, sino en e y $\neg e$, de modo que $p(e \,\&\, \neg e') = 0$, porque e y $\neg e$ son evidencias incompatibles, y $p(e \mid h_i) = 1$ y $p(\neg e \mid h_j) = 1$, y por tanto, $p(\neg e \mid h_i) = 0$ y $p(e \mid h_j) = 0$. Sin embargo, estas situaciones son, en realidad, del tipo T3, ya que las hipótesis discrepan, a fin de cuentas, *sobre la misma evidencia*, sea ésta e ó $\neg e$. Esto no supone mutilar ningún elemento esencial en tales situaciones. De hecho, el resultado de la contrastación, bien sea e ó $\neg e$, es concluyente, ya que una vez la probabilidad de una hipótesis alcanza el valor de cero, cualquier evidencia posterior no logrará modificarlo. Como puede comprobarse fácilmente en (8), si $p\ (e \mid h_j) = 0$, entonces $p\ (h_j \mid e) = 0$, y h_j ha sido refutada concluyentemente: cualquier adición posterior de evidencia, e', no podrá variar ese valor mínimo de ningún modo, ya que en el cálculo de $p\ (h_j \mid e')$ deberíamos tomar el cero como valor para $p\ (h_j)$, con lo cual el numerador será siempre cero. Y eso es lo supuestamente valioso en los "experimentos cruciales".

En resumen, aunque no podemos afirmar que *todas* las situaciones T4 interesantes, o sea las que pueden surgir respecto a hipótesis rivales, sean reducibles a T3, las observaciones anteriores bastan para mostrar que en muchos casos, entre los que cabría incluir los experimentos cruciales, es posible entender las situaciones T4 como T3 (y consiguientemente, queda justificado subsumir EO4 como un caso particular de EO3). Así las cosas, la cuestión prioritaria es, pues, si se cumplen EO2 y EO3.

(a) *La equivalencia ordinal en T2 (evidencias diversas /una hipótesis)*

$\boldsymbol{S_d}$, $\boldsymbol{S_r}$, $\boldsymbol{S_l}$ *son ordinalmente equivalentes* en estos casos, como puede verse a continuación. Sean e y e' dos resultados experimentales tales que $\boldsymbol{S_l}\ (h, e) \geq \boldsymbol{S_l}\ (h, e')$. Puede mostrarse que al cumplirse dicha condición se cumple una condición suficiente para que, según las medidas $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$, también e apoye a h en mayor medida que e'. Por la definición de $\boldsymbol{S_l}$, $\boldsymbol{S_l}\ (h, e) \geq \boldsymbol{S_l}\ (h, e')$ *sii*

$$\log\ [p\ (e \mid h) / p\ (e \mid \neg h)] \geq \log\ [p\ (e' \mid h) / p\ (e' \mid \neg h)]$$

Aplicando (7) a ambos lados de la desigualdad obtenemos:

$$\log \frac{p(h \mid e) \cdot p(\neg h)}{p(\neg h \mid e) \cdot p(h)} \geq \log \frac{p(h \mid e') \cdot p(\neg h)}{p(\neg h \mid e') \cdot p(h)}$$

Asumiendo, como es razonable, que $0 < p\ (h) < 1$, y por tanto, $p\ (\neg h) \neq 0$, el factor $p\ (\neg h)\ /\ p\ (h)$ puede simplificarse a ambos lados de la desigualdad. Y finalmente, por la monotonía de la función logaritmo podemos concluir que

$$\text{si } \boldsymbol{S_l}\ (h, e) \geq \boldsymbol{S_l}\ (h, e') \text{ entonces } \frac{p(h \mid e)}{p(\neg h \mid e)} \geq \frac{p(h \mid e')}{p(\neg h \mid e')}$$

Ahora bien, como para $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$, en atención a sus respectivas definiciones, e apoyará a h más que e' *sii* $p\ (h \mid e) \geq p\ (h \mid e')$, queda demostrado que si $\boldsymbol{S_l}\ (h, e) \geq \boldsymbol{S_l}\ (h, e')$, se cumple que $\boldsymbol{S_d}\ (h, e) \geq \boldsymbol{S_d}\ (h, e')$ y $\boldsymbol{S_r}\ (h, e) \geq \boldsymbol{S_r}\ (h, e')$.

La demostración se completa desarrollando el argumento a la inversa, es decir, mostrando que en todos los casos en los que $\boldsymbol{S_d}(h, e) \geq \boldsymbol{S_d}(h, e')$ y $\boldsymbol{S_r}(h, e) \geq \boldsymbol{S_r}(h, e')$ ocurre también que $\boldsymbol{S_l}(h, e) \geq \boldsymbol{S_l}(h, e')$. Para ello basta con deducir $log\ [p\ (e \mid h) / p\ (e \mid \neg h)] \geq log\ [p\ (e' \mid h) / p\ (e' \mid \neg h)]$ a partir de $p\ (h \mid e) / p\ (\neg h \mid e) \geq p\ (h \mid e') / p\ (\neg h \mid e')$.

Así pues, en los casos T2 el orden de las evidencias, *e*, *e'*,…, en función de la verosimilitud, por un lado, y de la credibilidad que conceden a la hipótesis, por otro, se determinan recíprocamente. En suma, la pluralidad de ***S***-medidas no es problemática en los casos "múltiple evidencia/una hipótesis" porque las tres medidas consideradas, $\boldsymbol{S_d}$, $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$, satisfacen EO2.

(b) *La equivalencia ordinal en T3 (una evidencia/diversas hipótesis).*

Supongamos la partición $\boldsymbol{H} = \{h_1, h_2, h_3\}$, siendo h_3 la hipótesis residual (v. nota 12), y los siguientes valores de partida de probabilidad iniciales y verosimilitudes:

	$p\ (h_i)$	$p\ (e \mid h_i)$
h_1	0,4	0,4
h_2	0,1	0,4
h_3	0,5	0,01

Nótese que h_1 y h_2 están empatadas en cuanto a su verosimilitud[14]. Aplicando el Teorema de Bayes, obtenemos: $p\ (h_1 \mid e)$ = 0'7804, $p\ (h_2 \mid e)$ = 0'1951, $p\ (h_3 \mid e)$ = 0'0243. Y ahora, haciendo los cálculos correspondientes:

	$p\ (h_i)$	$p\ (e \mid h_i)$	$p\ (h_i \mid e)$	$\boldsymbol{S_d}(h_i, e)$	$\boldsymbol{S_r}(h_i, e)$	$\boldsymbol{S_l}(h_i, e)$
h_1	0,4	0,4	0,7804	0,3804	0,2903	0'7269
h_2	0,1	0,4	0,1951	0,0951	0,2903	0'3388
h_3	0,5	0,01	0,0243	–0,4757	–1,3117	–1'602

Por consiguiente,

$$\boldsymbol{S_d}(h_1, e) > \boldsymbol{S_d}(h_2, e) > \boldsymbol{S_d}(h_3, e)$$

$$\boldsymbol{S_r}(h_1, e) = \boldsymbol{S_r}(h_2, e) > \boldsymbol{S_r}(h_3, e)$$

$$\boldsymbol{S_l}(h_1, e) > \boldsymbol{S_l}(h_2, e) > \boldsymbol{S_l}(h_3, e)$$

Aunque en este ejemplo no se viola, en sentido estricto, la equivalencia ordinal, se produce, sin embargo, una situación en la que la equivalencia entre h_1 y h_2, según $\boldsymbol{S_r}$, no se corresponde con la evaluación que de ambas hipotésis hacen las otras dos medidas, lo cual puede tener implicaciones a la hora de decidir entre seguir investigando en la línea de h_1, buscar evidencia adicional para desempatar… En otros ejemplos fácilmente concebibles se da una inversión estricta del orden (violación de EO3); se trata de situaciones en las que las probabilidades iniciales y las verosimilitudes de las distintas hipótesis presentan un orden inverso[15]. En cualquier caso, el incumplimiento de EO3 queda demostrado, en concreto para las parejas $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$, por un lado, y $\boldsymbol{S_l}$ y $\boldsymbol{S_r}$, por otro.

Los ejemplos ilustran, además, la fuente del desacuerdo entre las **S**-medidas. Según $\boldsymbol{S_d}$ y $\boldsymbol{S_l}$, a igualdad de verosimilitudes, *e* confirma más la hipótesis cuya probabilidad inicial es más alta (y cuando *e* es evidencia negativa, *e* penaliza más la hipótesis inicialmente menos probable). El impacto de *e* sobre *h* queda así modulado por la probabilidad inicial que se atribuye a ésta última. Sin embargo, $\boldsymbol{S_r}$ *neutraliza la probabilidad inicial de las hipótesis*. Esto es fácil de ver sustituyendo la fórmula de Bayes en las definiciones correspondientes:

$$\boldsymbol{S_d}\,(h,e) =_{\text{def}} p(h \mid e) - p(h) = \frac{p(h)\cdot p(e \mid h)}{p(e)} - p(h) = p(h)\cdot[\frac{p(e \mid h)}{p(e)} - 1] \qquad (4')$$

$$\boldsymbol{S_r}\,(h,e) =_{\text{def}} log\frac{p(h \mid e)}{p(h)} = log\frac{p(h)\cdot p(e \mid h)}{p(e)\cdot p(h)} = log\frac{p(e \mid h)}{p(e)} \qquad (5')$$

Así pues, $\boldsymbol{S_d}$ es directamente proporcional a $p\ (h)$, mientras que para $\boldsymbol{S_r}$ este factor no cuenta en absoluto. Obviamente, para todas la h_i pertenecientes a la partición en cuestión $p\ (e)$ vale lo mismo, ya que $p\ (e) = \sum p_i \cdot p\ (e \mid h_i)$. Entonces, sea cual sea la probabilidad inicial de las h_i, según (5') $\boldsymbol{S_r}\ (h_i, e)$ es directamente proporcional a $p\ (e \mid h_i)$, la verosimilitud de la hipótesis. En cuanto a $\boldsymbol{S_l}$, podría pensarse que, al tratarse de una razón de verosimilitudes, las probabilidades iniciales son irrelevantes, como ocurre con $\boldsymbol{S_r}$. Sin embargo, para una partición con *n* hipótesis:

$$\boldsymbol{S_l}\,(h_j,e) =_{\text{def}} log\frac{p(e \mid h_j)}{p(e \mid \neg h_j)} = log\frac{p(e \mid h_j)}{\dfrac{\sum_{i\neq j} p(e \mid h_j)\cdot p(h_i)}{\sum_{i\neq j} p(h_i)}} \qquad (6')$$

Suponiendo que h_1, h_2, y h_3 constituyan una partición del espacio muestral, $p(e \mid \neg h_1) = p(e \mid h_2 \vee h_3)$, $p(e \mid \neg h_2) = p(e \mid h_1 \vee h_3)$, $p(e \mid \neg h_3) = p(e \mid h_1 \vee h_2)$. Entonces, según la definición de probabilidad condicionada, $p(A \mid B) =_{\text{def}} p(A \,\&\, B) / p(B)$, obtenemos: $p(e \mid h_2 \vee h_3) = p(e \,\&\, (h_2 \vee h_3) / p(h_2 \vee h_3)$, $p(e \mid h_1 \vee h_3) = p(e \,\&\, (h_1 \vee h_3) / p(h_1 \vee h_3)$, y $p(e \mid h_1 \vee h_2) = p(e \,\&\, (h_1 \vee h_2) / p(h_1 \vee h_2)$, igualdades que contienen en su segundo miembro un denominador que es una suma de probabilidades iniciales. Nótese que, s*alvo que la distribución inicial sea uniforme, el resultado de la suma variará para cada* h_i. Por tanto, en general la distribución de probabilidad inicial sí afecta al cálculo de $\boldsymbol{S_l}$.

Desde luego, cualquier medida, incluyendo $\boldsymbol{S_r}$, puede expresarse en términos de $p(h)$. Pero sería erróneo concluir a partir de eso que el grado de confirmación de las hipótesis rivales se ve influido por las probabilidades iniciales, como acabamos de ver[16]. En realidad, cada $\boldsymbol{S}$-medida concede un peso relativo a las probabilidades iniciales. $\boldsymbol{S_r}$ es insensible por completo a la probabilidad inicial; $\boldsymbol{S_d}$, por su parte, es más sensible que $\boldsymbol{S_l}$ [v. Fitelson (2007), p. 484, donde se comenta una simulación por ordenador al respecto].

El cumplimiento de EO2 está garantizado porque el ordenamiento de verosimilitudes y probabilidades condicionadas se determina recíprocamente, sin que el factor 'probabilidad inicial' pueda influir (de hecho, T2 son situaciones "múltiple evidencia / *una hipótesis"*). Es en los casos en que hay rivalidad genuina entre hipótesis, y la distribución de probabilidad inicial no es uniforme, justamente cuando el aparato bayesiano debe exhibir sus virtualidades, cuando la equivalencia ordinal, EO3, es incumplida. La pregunta, pues, es *si en el cálculo de la confirmación incremental debe tenerse en cuenta la probabilidad inicial y, si es así, de qué modo*.

III. Probabilidad inicial y apoyo evidencial

Peter Milne, quien apuesta por $\boldsymbol{S_r}$ como la única medida correcta, ha expresado sucintamente la razón por la que la probabilidad inicial no debe ser tenida en cuenta:

> Supongamos que dos teorías implican la evidencia *e* y que se comprueba que *e* es verdadera. El hecho de que *e* ocurra no tiene capacidad para discriminar entre esas hipótesis. Ambas implican *e*, así que no hay nada *en la naturaleza misma de la evidencia* que nos de pie para diferenciarlas. Puede haber otras razones, como la simplicidad, el poder explicativo, la parsimonia de los compromisos ontológicos, para preferir una a la otra, pero la verdad de *e*, por sí mismo, no es una razón. Ya que *e* carece de poder discriminativo, sostengo que las dos hipótesis están igualmente confirmadas por *e*. Y lo que vale cuando las teorías implican *e* se puede generalizar para el caso en que $p(e \mid h_1) = p(e \mid h_2)$. La evidencia *por sí sola*, no nos da ninguna razón para discriminar entre las hipótesis [Milne (1996), pp. 22-23].

El potencial predictivo de las hipótesis se concentra en su verosimilitud. Si queremos cuantificar el efecto de e sobre la hipótesis, nada más que esto ha de ser tenido en consideración. De acuerdo con la evidencia puntual no hay ningún motivo para favorecer una u otra de las hipótesis rivales, cuando sus verosimilitudes sean idénticas. Así pues, si $p(e \mid h_1) = p(e \mid h_2)$, entonces $\boldsymbol{S}(e, h_1) = \boldsymbol{S}(e, h_2)$. Y como la única medida que respeta dicho principio es $\boldsymbol{S}_r$, ésta ha de ser la medida correcta[17]. Este es el modo de ser respetuosos al máximo con la evidencia, según Milne, porque introducir factores adicionales falsearía lo que se pretende calcular, esto es, cuánto respalda e a h.

El argumento principal en contra de $\boldsymbol{S}_r$ ha sido la objeción de "la conjunción irrelevante" (*the tacking objection*), ya discutida en su día a propósito de la teoría cualitativo-deductiva de la confirmación de Hempel. Supongamos que h implica deductivamente e y que t es un enunciado cualquiera completamente irrelevante. Entonces, $p(e \mid h) = 1$; pero como también es cierto que $p(e \mid h \,\&\, t) = 1$, ocurre que $\boldsymbol{S}_r(h, e) = \boldsymbol{S}_r(h \,\&\, t, e)$. En consecuencia, e confirma h en el mismo grado en que confirma $h \,\&\, t$, lo cual resulta un tanto extraño. Nótese que esto no tiene por qué ocurrir con el resto de medidas, pues si t es irrelevante respecto a h, $p(h \,\&\, t) = p(h) \cdot p(t)$, y si las probabilidades iniciales, $p(h)$ y $p(h \,\&\, t)$, son desiguales bajo unas condiciones mínimas –en concreto, si $p(t) < 1$ y $p(h) > 0$, entonces $p(h \,\&\, t) < p(h)$–, lo que se verá reflejado en las $\boldsymbol{S}$-medidas sensibles a la probabilidad inicial. De hecho, las soluciones al problema de la conjunción irrelevante ofrecidas hasta ahora [v. Fitelson (2002)] sostienen que $\boldsymbol{S}(e, h) > \boldsymbol{S}(e, h \,\&\, t)$, de lo que se sigue que $\boldsymbol{S}_r$ no puede afrontar dicho problema.

Nuestra respuesta aquí es que la "*tacking objection*" no constituye un problema para $\boldsymbol{S}_r$. En primer lugar hay que insistir en que en la metodología bayesiana los casos T3 se suelen plantear en condiciones estándar: en relación a una partición del espacio muestral generada por hipótesis "estructuralmente idénticas", esto es, hipótesis que asignan valores al mismo conjunto de variables aleatorias [v. Steel (2007), pp. 67 y ss.]. Piénsese en los pronósticos sobre el calentamiento global del planeta para el año 2050. El parámetro θ sobre el que se pronuncian las hipótesis rivales podría tener como rango de valores [0, 5], en grados centígrados, y e podría ser el aumento del nivel del mar, por ejemplo, en centímetros. Si optamos por una partición discreta, una posibilidad sería, por ejemplo: $h_1 = 0$, $h_2 = 1$,…., $h_6 = 5$, $h_7 = \theta \notin [0, 5]$. En una situación como ésta la objeción no surge, ya que el resultado de añadir un enunciado irrelevante a cualquiera de dichas hipótesis rompería la exigencia de identidad estructural. Desde luego, h y $(h \,\&\, t)$ no pueden pertenecer a la misma partición del espacio muestral, puesto que $h \,\&\, t \models h$, y por tanto, no son lógicamente incompatibles.

Esta réplica, cabe admitirlo, no sirve cuando hemos de vérnoslas con hipótesis teóricas de gran alcance, porque aquí la comparación no suele establecerse bajo las condiciones mencionadas. Trabajar con hipótesis estructuralmente idénticas resulta idóneo para establecer una partición del espacio muestral, pero cuando se trata de alternativas teóricas el formato es, sin duda, demasiado rígido. Y esto no es un problema específico de $\boldsymbol{S}_r$, ni siquiera de las $\boldsymbol{S}$-medidas, sino que surge también en el cálculo de $\boldsymbol{C}$, como la fórmula de Bayes deja ver. En cualquier caso, cuando las hipótesis no son estructuralmente idénticas, la objeción de la conjunción irrelevante no es un obstáculo insoslayable para $\boldsymbol{S}_r$. Siendo *e* un fenómeno empírico o un resultado experimental particular, decir que *e* confirma incrementalmente *h* en el mismo grado que *h* & *t*, no es decir que ambas son igualmente probables, o creíbles, dado *e*, puesto que eso lo decide $\boldsymbol{C}$ –o sea, $p\ (h \mid e)$ y $p\ (h \ \&\ t \mid e)$– y no las $\boldsymbol{S}$-medidas, y $\boldsymbol{C}$ sí prima, a igualdad del resto de factores, la hipótesis inicialmente más probable. Lo que se está afirmando, más bien, al decir que $\boldsymbol{S}_r\ (h, e) = \boldsymbol{S}_r\ (h \ \&\ t, e)$ es que *e* no sirve como evidencia "de desempate". Invocar ciertas particularidades de las hipótesis en conflicto, como su simplicidad, por mencionar alguna de las propiedades que a veces se incluyen en la probabilidad inicial, no parece que deba introducir diferencias en el grado en que *e*, y nada más que *e*, confirma a aquéllas.

Pero, ¿no es verdad que olvidar las probabilidades iniciales puede llevarnos a tomar decisiones abiertamente irracionales, sobre todo cuando dichas probabilidades no son meras conjeturas subjetivas? Supongamos que nuestra hija se despierta con fiebre. Por lo que hemos leído en fuentes autorizadas, sabemos que la probabilidad de tener fiebre si se ha contraído la malaria es más alta que cuando se tiene faringitis. O sea, $p\ (e \mid M) > p\ (e \mid F)$. Supongamos que *M* y *F* son hipótesis excluyentes, y que la probabilidad de que haya fiebre por una causa distinta a *M* ó *F* sea inferior a la de que la fiebre sea causada por alguna de ambas enfermedades, esto es, $p\ (e \mid \neg(M \text{ v } F) < p\ (e \mid M)$, y $p\ (e \mid \neg(M \text{ v } F) > p\ (e \mid F)$. Entonces, $\boldsymbol{S}_r\ (M, e) > \boldsymbol{S}_r\ (F, e)$ y también $\boldsymbol{S}_r\ (M, e) > \boldsymbol{S}_r\ (\neg(M \text{ v } F), e)$, pero alertar a la unidad sanitaria de enfermedades tropicales *sólo por esta razón* resultaría descabellado.

Presumiblemente, $p\ (M)$ es muy inferior a $p\ (F)$. En esto podríamos coincidir con el especialista médico, cuya estimación de las probabilidades iniciales, que suponemos bien fundada, dependerá de sus conocimientos, creencias y experiencias previas sobre casos semejantes a éste, y esto va incluido en su estimación de las probabilidades iniciales. Y es verdad que aplicando una medida sensible a las probabilidades iniciales obtendríamos una valor para $\boldsymbol{S}$ tal que $\boldsymbol{S}\ (F, e) > \boldsymbol{S}\ (M, e)$. El problema, sin embargo, no radica en $\boldsymbol{S}_r$. De hecho, la objetividad de la distribución inicial de probabilidad es irrelevante para el problema que nos ocupa [v. el ejemplo incluido en Kuipers (2000), p. 57, con una distribución inicial máximamente objetiva]. Y es que una $\boldsymbol{S}$-medida no es una función de probabilidad; sin embargo, para el bayesiano la credibilidad sí es una función de probabilidad que viene definida por el Teorema de Bayes, de modo que una probabilidad inicial muy baja puede contrarrestar una verosimilitud favorable, dando como resultado que $p\ (M \mid e)$

$< p\ (F \mid e)$. El error está, hemos de insistir, en suponer que una ***S***-medida nos dice qué hipótesis hemos de creer o aceptar, aunque sea tentativamente (respecto a lo cual la probabilidad inicial sí tiene cosas que decir). Pero $\boldsymbol{S_r}$, *como cualquier otra **S**-medida*, no se pronuncia al respecto.

Resumiendo la objeción de la conjunción irrelevante, el principal obstáculo que se aduce en contra de ***S***, no surge en la aplicación rutinaria de la metodología bayesiana a los casos T3. En las ocasiones en que la objeción puede plantearse, no está en absoluto claro que las intuiciones subyacentes respecto a lo que constituye un resultado experimental que respalda una hipótesis jueguen en contra de $\boldsymbol{S_r}$, sino más bien al contrario. Frente a la acusación de Steel (2007), donde se sostiene que el defensor de $\boldsymbol{S_r}$ comete una petición de principio al presuponer la irrelevancia de las probabilidades iniciales, las consideraciones precedentes pretenden mostrar que la carga de la prueba corresponde a quien piense que el cálculo de la confirmación incremental debe ser sensible a la probabilidad inicial de *h*, y por tanto, a quienes defienden medidas como $\boldsymbol{S_d}$ ó $\boldsymbol{S_l}$.

Nuestra argumentación a favor de $\boldsymbol{S_r}$ no pretende ser puramente negativa. Al trasladar la carga de la prueba al adversario no estamos afirmando simplemente que no hay razones para no defender $\boldsymbol{S_r}$. No obstante, puede pensarse que apenas se ha avanzado mientras no se den razones positivas a favor de $\boldsymbol{S_r}$.

Se ha sostenido que $\boldsymbol{S_r}$ es superior al resto de medidas en casos "extremos", cuando $p\ (h)$ o $p\ (e)$ valen cero, porque es posible aplicar, bien $log\ [p\ (e \mid h)\ /\ p\ (e)]$, bien $log\ [p\ (h \mid e)\ /\ p\ (h)]$ respectivamente, según (5'), con lo que la función estaría definida en ambos supuestos[18]. El caso en que $p\ (h) = 0$ ha sido objeto de discusión desde hace décadas, ya que puede plantearse cuando *h* es una generalización universal que se pronuncia sobre un número *infinito* de casos (en principio, su probabilidad seguiría siendo cero por muchos casos particulares favorables que acumuláramos). No entraremos en más detalles porque, sea como fuere, la supuesta ventaja se desvanece cuando se repara en que, según la definición de probabilidad condicionada, $p\ (e \mid h) = p\ (e\ \&\ h)\ /\ p\ (h)$, que no está definida si $p\ (h) = 0$. Y lo mismo ocurre con $p\ (h \mid e)$ cuando $p\ (e) = 0$ según el Teorema de Bayes. Si atendemos a la amplitud del dominio en que las funciones están definidas, $\boldsymbol{S_d}$ sería, pues, la medida preferible, ya que $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$ no están definidas ni para $p\ (e) = 0$, ni para $p\ (h) = 0$, mientras que $\boldsymbol{S_d}$ no está definida en el primer caso, pero sí en el segundo. De todos modos, no está claro que esta propiedad matemática sea una ventaja por sí misma. Dado que, si $p\ (h) = 0$, $\boldsymbol{S_d}\ (h, e) = 0$ para cualquier *e*, $\boldsymbol{S_d}$ nos dice que toda evidencia es *neutra* para una hipótesis cuya probabilidad sea cero, mientras que $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$ no se pronuncian en este caso, al no estar definidas. Mas, ¿por qué habría de ser, desde un punto de vista epistemológico, una cosa mejor que otra?

Estas consideraciones nos devuelven a la discusión de la sección 2. También entonces discutimos casos "extremos", pues así deben considerarse aquellos en que la conexión entre evidencia e hipótesis es deductiva ($e \models h$ ó $e \models \neg h$) porque $p(h \mid e)$ toma valores máximos, 0 ó 1, y allí sugerimos la conveniencia de plantear la polémica sobre la pluralidad de $\boldsymbol{S}$-medidas en relación a los contextos y problemas de la investigación científica. Ahora es el momento de analizar las consecuencias de optar por $\boldsymbol{S}_r$ en relación con esto[19]. Y es que de $\boldsymbol{S}_r$ se sigue de un modo natural un precepto indiscutible en las situaciones T3 como es evitar contextos de contrastación experimental en que las verosimilitudes de las hipótesis rivales coincidan. Así, partiendo de que el éxito experimental de las hipótesis en conflicto debe condicionar la estrategia investigadora a seguir, y de que queremos que el resultado de la contrastación sea informativo en este sentido, $\boldsymbol{S}_r$ nos invita a diseñar una prueba en la que las hipótesis difieran en su potencial predictivo respecto a e, lo que se traduce en verosimilitudes diferentes. Es el único modo en que el test permitirá *discriminar* entre las hipótesis en cuanto al grado de confirmación incremental, la que reciben exclusivamente por parte de e. Podemos hablar entonces del "Principio de discriminación experimental", según el cual:

> La elaboración de un diseño experimental que pretenda contrastar una hipótesis h en relación a un resultado potencial e debe plantearse de modo que las verosimilitudes de las hipótesis que rivalizan entre sí (h, h', h'', ...) difieran.

El precepto, no por obvio, es menos importante (nótese que es independiente de que la conexión entre h y e sea deductiva o no). Sería complicado para un equipo científico justificar el diseño de un experimento cuyos potenciales resultados son igualmente probables sea cual sea la hipótesis correcta de entre las rivales –en el caso más simple se exigiría que $p(e \mid h) \neq p(e \mid \neg h)$. Éste no es, desde luego el procedimiento acostumbrado en la ciencia, ya que, incluso en caso de que el espacio muestral se reduzca a h y $\neg h$, lo que el experimento pretende provocar es un fenómeno empírico que no sea indiferente al hecho de que h sea verdadera o falsa. Por lo demás, aunque de $\boldsymbol{S}_d$ o $\boldsymbol{S}_l$ no se sigue que hayamos de actuar en contra del precepto, sí que es verdad que éste no obtiene ninguna justificación de ellas. A partir de $\boldsymbol{S}_r$, en cambio, evitar empates en cuanto al potencial predictivo de las hipótesis resulta una estrategia plenamente motivada.

Alguien advertirá en este punto el eco de la preferencia popperiana por las contrastaciones experimentales *severas*. Nos parece más acertado hablar aquí de *capacidad de discriminación* de un test, que de severidad. De hecho, se trata de nociones distintas. Para Popper la severidad de un test equivale a $p(e \mid h) - p(e)$[20]. Dada una partición de un espacio muestral $\boldsymbol{H} = \{h_1, h_2,..., h_n\}$ es posible, no obstante, que e sea un test severo, en sentido popperiano, para dos hipótesis rivales, h_1 y h_2, y que $\boldsymbol{S}_r(h_1, e) = \boldsymbol{S}_r(h_2, e)$. Así, e puede ser una descripción de un fenómeno empírico que es implicada, o resulta sumamente probable, para ambas hipótesis, o sea, $p\ (e \mid h_1) = p\ (e \mid h_2) \approx 1$, pero que es altamente improbable según el estado de conocimientos actual, y por tanto, $p\ (e) \approx 0$. Como $p\ (e) = p\ (h_1) \cdot p\ (e \mid h_1 \vee h_2 \vee \neg(h_1 \vee h_2)) = p\ (e \mid h_1) \cdot p\ (h_1) + p\ (e \mid h_2) \cdot p\ (h_2) + p\ (e \mid \neg(h_1 \vee h_2)) \cdot p\ \neg(h_1 \vee h_2)$. Y ahora, si $p\ (e \mid \neg(h_1 \vee h_2))$ es muy baja dado el estado de conocimientos, y $p\ (h_1)$ y $p\ (h_2)$ también son bajas, ocurre que $p\ (e) \approx 0$, con lo cual e sería una prueba severa en sentido popperiano tanto para h_1 como para h_2, y sin embargo, e respaldaría en la misma medida ambas hipótesis, ya que $\boldsymbol{S_r}\ (h_1, e) = \boldsymbol{S_r}\ (h_2, e)$, y por tanto e sería una prueba con un poder de discriminación nulo.

Es sabido que las generalizaciones teóricas se conectan inferencialmente con las predicciones empíricas gracias al concurso de supuestos auxiliares, claúsulas *ceteris paribus* y condiciones iniciales, que no son sino instancias particulares del antecedente de la hipótesis. La historia de la ciencia proporciona bastantes ejemplos de modificaciones *ad-hoc* sobre este cuerpo adicional de enunciados, con intención de hacer encajar la hipótesis con los datos. No es tan inusual, pues, encontrarse con hipótesis teóricas rivales que mantienen una conexión deductiva con la evidencia. En el vocabulario probabilístico esto equivale a verosimilitudes máximas *e iguales* para las hipótesis rivales: $p\ (e \mid h_1\ \&\ s) = p\ (e \mid h_2\ \&\ s') = \ldots = 1$, siendo $s, s', \ldots$, el aparato de enunciados adicionales que incorpora cada una. En tales casos $\boldsymbol{S_r}$ nos dice que e no introduce ninguna diferencia entre h_1 y h_2, ya que $\boldsymbol{S_r}\ (h_1\ \&\ s \mid e) = \boldsymbol{S_r}\ (h_2\ \&\ s' \mid e)$, mientras que $\boldsymbol{S_d}$ y $\boldsymbol{S_l}$ atribuirán mayor grado de confirmación incremental al conglomerado (hipótesis + supuestos) que goce de mayor probabilidad inicial. En realidad la situación es semejante a la que se planteó antes a propósito de la "conjunción irrelevante" y nuestra respuesta aquí debe ser la misma: $(h_1\ \&\ s)$ y $(h_2\ \&\ s')$ obtiene idéntico apoyo de e. La plausibilidad que merezcan los respectivos añadidos *ad-hoc*, o la probabilidad que cabe atribuir a $(h_1\ \&\ s)$, $(h_2\ \&\ s')$, etc. dado e –que no tienen por qué coincidir aunque $p\ (e \mid h_1\ \&\ s) = p\ (e \mid h_2\ \&\ s')$ si $p\ (h_1\ \&\ s) \neq p\ (h_2\ \&\ s')$– son cuestiones aparte.

IV. Conclusiones

Las discrepancias sustantivas entre las medidas $\boldsymbol{S_d}$, $\boldsymbol{S_l}$ y $\boldsymbol{S_r}$ se producen en unos contextos determinados, a saber, cuando se compara el apoyo que un resultado experimental proporciona a diferentes hipótesis rivales (no es difícil encontrar ejemplos que muestran que $\boldsymbol{S_r}$ no es ordinalmente equivalente ni respecto a $\boldsymbol{S_d}$ ni a $\boldsymbol{S_l}$). La clave de estas discrepancias radica en el peso distinto que estas $\boldsymbol{S}$-medidas conceden a las probabilidades iniciales. Tras analizar formalmente las situaciones donde existe rivalidad entre hipótesis, entendiendo la rivalidad en términos de una partición sobre el espacio muestral compuesta por tres hipótesis al menos: h_1, h_2 y $\neg(h_1 \vee h_2)$, nos hemos decantado por $\boldsymbol{S_r}$, una medida que anula por completo el papel de las probabilidades iniciales, pues no consideramos que éstas sean relevantes para calcular cuánto apoya *e* a *h*. Respecto a las dificultades, la objeción de la "conjunción irrelevante", supuestamente el mayor inconveniente que $\boldsymbol{S_r}$ debe superar, olvida qué es lo que se espera de una $\boldsymbol{S}$-medida. En nuestra opinión la carga de la prueba corresponde a quien considere que las probabilidades iniciales son relevantes a tal efecto. Pensamos, además, que un precepto metodológico básico, al que hemos llamado "Principio de discriminación experimental", resulta justificable de un modo natural a partir de $\boldsymbol{S_r}$, mientras que no ocurre lo mismo bajo $\boldsymbol{S_d}$ y $\boldsymbol{S_l}$*.

Departamento de Lógica y Filosofía de la Ciencia
Universidad de Valencia
Avda. Blasco Ibáñez 30
E-46010 Valencia
E-mail: valeriano.iranzo@uv.es

Departamento de Economía Aplicada
Universidad de Valencia
Campus de Tarongers
Edificio Este
E-46022 Valencia
E-mail:Ignacio.Martinez-Lejarza@uv.es

NOTAS

* El presente trabajo ha recibido subvención del Ministerio de Ciencia e Innovación mediante el proyecto FFI 2008-01169/FISO. Versiones previas fueron leídas en Granada y Madrid (UNED). Conste nuestro agradecimiento a la audiencia en tales ocasiones, así como a Vincenzo Crupi, Donald Gillies, Manuel Pérez Otero y David Teira y a los revisores anónimos de **teorema** por sus atinados comentarios.

[1] Lo cual es conceder mucho, ya que no puede identificarse sin más probabilidad con aceptación. La *aceptación* de una hipótesis científica es una *decisión* que puede involucrar aspectos adicionales a su plausibilidad/probabilidad. Así, algunos bayesianos incorporarían aquí la noción de *utilidad*. V. *infra* nota 11 y también Good (1983), cap. 14, secciones 1, 2 y 4, para una discusión concisa pero iluminadora.

[2] En estricto sentido bayesiano, todas las probabilidades son probabilidades condicionadas, tal como figura en la expresión (1), pero a partir de ahora suprimiremos las referencias a *b* para evitar complicaciones innecesarias.

[3] $\boldsymbol{S_d}$: Jeffrey (1992), Earman (1992), Rosenkrantz (1977), Howson y Urbach (2006); $\boldsymbol{S_r}$ (o su variante sin *log*): Keynes (1921), Horwich (1982), Schlesinger (1995), Milne (1995) y (1996), Kuipers (2000); $\boldsymbol{S_l}$: Good (1984), Howson (2003), Jeffrey (2004) y Fitelson (2006).

[4] ¿Por qué cumplir unas simetrías y no otras? El ejemplo que Eells y Fitelson explotan es el siguiente: sea *e* = sacar un caballo y *h* = sacar una figura. Puesto que *e* implica *h*, $p(h \mid e) = 1$. Si ocurre *e*, podemos concluir la verdad de *h*, con lo cual *e* debería conferir la máxima confirmación posible a *h*. Sin embargo, si $\neg e$, no estamos disconfirmando máximamente *h*, pues si no sale un caballo, no hemos falsado concluyentemente *h*, ya que podríamos haber sacado una sota o un rey. Este razonamiento sirve para mostrar que ES no debe ser respetada en ciertas condiciones. El mismo

ejemplo permite concluir que tampoco CS y TS. En cambio, HS sí debe ser satisfecha porque, además de resultar intuitivamente plausible, los autores dicen no haber encontrado ningún contraejemplo semejante al anterior. Véase, sin embargo, el ejemplo de Kuipers (2000), p. 52 para justificar la conveniencia de respetar CS y defender $\boldsymbol{S}_r$ frente a $\boldsymbol{S}_d$. Ejemplos intuitivos, dicho sea de paso, hay para todos los gustos, como puede verse en Schlesinger (1995) y Pollard (1999).

[5] Si $p(h \mid e) \geq p(h)$, $\boldsymbol{S}_z =_{\text{def}} [p(h \mid e) - p(h) / 1 - p(h)]$; si $p(h \mid e) < p(h)$, $\boldsymbol{S}_z =_{\text{def}} [p(h \mid e) - p(h) / p(h)]$. En cuanto a las diferencias entre el requisito de logicidad de Fitelson y el principio de extensión, v. Crupi *et al.* (2007), nota 4.

[6] Lo novedoso del enfoque desarrollado en Crupi *et al.* (2007) es que no se conforma con mostrar cuál es la medida "normativamente" justificada, $\boldsymbol{S}_z$, sino que se apunta además que la corrección de una ***S***-medida también depende de que "pueda ser empleada para describir (y predecir) con exactitud la conducta y el juicio humanos" (p. 242). Las observaciones que siguen no pretenden cuestionar un planteamiento como éste, que pensamos puede arrojar interesantes resultados. Nuestra intención es, más bien, desmarcarnos de algunos de los presupuestos que subyacen a la investigación "normativa" sobre las ***S***-medidas.

[7] Y más aún cuando, de acuerdo con nuestra noción preanalítica de lo que constituye evidencia, si *e* es evidencia a favor de *h*, resulta extraño considerar que *h* puede ser evidencia, concluyente o no, a favor de *e*. V. Christensen (1999) sobre las limitaciones de la "relevancia probabilística" para dar cuenta de la noción cotidiana de apoyo evidencial.

[8] Crupi *et al.* (2007) reconocen este punto, aunque a la posibilidad de justificar $\boldsymbol{S}_z$ con ejemplos que no planteen una relación deductiva entre *e* y *h* apenas dedican un par de párrafos (v. pp. 241-242).

[9] Las respuestas a T2 y T3 son los "juicios de relevancia" (*judgments of relevance*) y los "juicios de preferencia" (*judgments of preference*) en terminología keynesiana, v. Keynes (1921), cap. 4, secc. 14.

[10] En Tentori *et. al.* (2006) se estudia la correlación entre siete ***S***-medidas sobre diez mil asignaciones al azar. Se incluyó además ***C*** (= $p(h \mid e)$). Como cabía esperar, con ***C*** se obtuvieron correlaciones por término medio (0'722) significativamente más bajas que las que se dieron entre las otras siete medidas. El valor más alto se obtuvo entre $\boldsymbol{S}_k$ y $\boldsymbol{S}_l$ (0'970). Si la simulación se restringe a $\boldsymbol{S}_d$, $\boldsymbol{S}_r$, $\boldsymbol{S}_k$, y $\boldsymbol{S}_l$, la medida que arroja valores menos correlacionados es $\boldsymbol{S}_r$, dándose la correlación inferior entre $\boldsymbol{S}_r$ y $\boldsymbol{S}_d$ (0'822), aunque sigue siendo bastante alta, a pesar de todo. En cualquier caso, estos datos son meros indicadores del comportamiento de $\boldsymbol{S}_r$, y no constituyen por sí mismos un argumento para rechazarla.

[11] Uno de los árbitros sugiere que no está justificado restringir las comparaciones a las situaciones de rivalidad entre hipótesis porque, por ejemplo, ante la decisión de qué medicamentos deben subvencionarse por la Seguridad Social es relevante comparar en qué medida los estudios experimentales efectuados sobre cada uno apoyan su efectividad. A esto cabe responder:

— las hipótesis h_1 "El medicamento A es efectivo" y h_2 "El medicamento B es efectivo", apoyadas por distintas evidencias, *e* y *e'* respectivamente, no son rivales en un sentido genuino, puesto que podrían ser verdaderas ambas a la vez;

— una *decisión* de este tipo no solamente tendrá en cuenta la probabilidad de que el medicamento sea efectivo, sino que involucrará factores adicionales, como los posibles efectos secundarios, el coste económico, la comodidad en el uso,…, esto es, las "utilidades" (*utilities*) en la jerga de la teoría bayesiana de la decisión. El conflicto surge porque nos vemos obligados a decidir, por condicionantes prácticos, entre dos cursos de acción distintos (subvencionar A o subvencionar B), pero no hay rivalidad entre h_1 y h_2, y tampoco se trata de elegir entre ellas;

— en cuanto a la información epistémica relevante para tomar la decisión, dada cualquiera de las ***S***-medidas, supóngase que $\boldsymbol{S}$ $(h_1, e) > \boldsymbol{S}$ (h_2, e'), pero p $(h_1 \mid e) < p$ $(h_2 \mid e')$. ¿Cuál es entonces el factor que deberíamos tener en cuenta? Parece que el factor crucial es la probabilidad condicionada de que cada medicamento sea efectivo, y no tanto el incremento/disminución que la evidencia provoca sobre la probabilidad inicial de la hipótesis. Visto así, las ***S***-medidas tal vez tengan muy poco que decir aquí.

Pensamos, pues, que el ejemplo ni cuestiona nuestra estrategia de evaluar la corrección de las ***S***-medidas evitando los contextos en que las hipótesis no son rivales en un sentido genuino, ni permite concluir que las situaciones en las que intervienen diferentes hipótesis apoyadas por evidencias distintas sean cruciales en nuestra investigación.

[12] Para completar el espacio muestral los bayesianos postulan una hipótesis "residual" (*catch-all hypothesis*), que es la negación de todas las demás. Supóngase que queremos comparar dos explicaciones alternativas, h_1 y h_2 de e. Aunque h_1 y h_2 sean excluyentes entre sí, p $(h_1) + p$ (h_2) no tiene por qué ser igual a 1, ya que h_2 no es simplemente la negación de h_1. Habría que considerar, pues, una tercera posibilidad, $h_3 = \neg (h_1 \text{ v } h_2)$, para que p $(h_1) + p$ $(h_2) + p$ $(h_3) = 1$.

[13] Esta objeción fue planteada por uno de los árbitros.

[14] Supongamos, a modo de ilustración, una población objeto de estudio en la que el 40 % de sus miembros tiene una formación académica media (h_1), un 10% tiene estudios universitarios (h_2), y el resto un nivel académico inferior (h_3). Así que p $(h_1) = 0{,}4$, p $(h_2) = 0{,}1$ y p $(h_3) = 0{,}5$. Consideremos el hecho de alcanzar un estatus socioeconómico elevado (e). Sabemos que el 40% de las personas con estudios medios lo adquieren; que igualmente lo alcanzan el 40% de las personas con estudios superiores; y que sólo el 10% de las personas con nivel académico inferior son capaces de obtenerlo. O sea, p $(e \mid h_1) = (p$ $(e \mid h_2) = 0'4$ y p $(e \mid h_3) = 0'1$. Nos preguntamos cuál de los niveles de formación considerados resultaría más confirmado al seleccionar como muestra un individuo de alto estatus económico. Como se verá en los cálculos que siguen, según las medidas $\boldsymbol{S_d}$ y $\boldsymbol{S_l}$ resultaría más confirmado el nivel medio de estudios, mientras que según $\boldsymbol{S_r}$ concluiríamos que el nivel medio y el universitario son confirmados por igual, precisamente porque las verosimilitudes p $(e \mid h_1)$ y p $(e \mid h_2)$ son iguales.

[15] Sea p $(h_1) = 0'4$, p $(h_2) = 0'3$, p $(h_3) = 0'3$, p $(e \mid h_1) = 0'7$, p $(e \mid h_2) = 0'71$, y p $(e \mid h_3) = 0'2$. Entonces, p $(h_1 \mid e) = 0'5063$, p $(h_2 \mid e) = 0'3851$, p $(h_3 \mid e) = 0'1084$, $\boldsymbol{S_d}$ $(h_1, e) = 0'1063$, $\boldsymbol{S_d}$ $(h_2, e) = 0'0851$, $\boldsymbol{S_d}$ $(h_3, e) = -0'1915$, $\boldsymbol{S_r}$ $(h_1, e) = 0'2357$, $\boldsymbol{S_r}$ $(h_2, e) = 0'2499$, $\boldsymbol{S_r}$ $(h_3, e) = -1'0170$, $\boldsymbol{S_l}$ $(h_1, e) = 0'4307$, $\boldsymbol{S_l}$ $(h_2, e) = 0'3796$, $\boldsymbol{S_l}$ $(h_3, e) = -1'2589$. Así pues, $\boldsymbol{S_d}$ $(h_1, e) \geq \boldsymbol{S_d}$ (h_2, e) y $\boldsymbol{S_l}$ $(h_1, e) \geq \boldsymbol{S_l}$ (h_2, e), pero $\boldsymbol{S_r}$ $(h_1, e) < \boldsymbol{S_r}$ (h_2, e). Por tanto, los pares de medidas $\boldsymbol{S_d}$ y $\boldsymbol{S_r}$, y $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$ no cumplen EO3. El ejemplo no demuestra que $\boldsymbol{S_l}$ y $\boldsymbol{S_d}$ tampoco lo hagan.

[16] Por eso es confuso decir, como se hace en Huber (2008), que $\boldsymbol{S_r}$ y $\boldsymbol{S_l}$ coinciden en que "lo que importa para la confirmación son las diferencias entre probabilidades iniciales y probabilidades posteriores [condicionadas a *e*]." (p. 416). A pesar de que

Huber concluye que $\boldsymbol{S}_r$ y $\boldsymbol{S}_l$ miden "relaciones evidenciales distintas", sorprendentemente no hace ninguna referencia a la neutralidad de $\boldsymbol{S}_r$ respecto a la probabilidad inicial.

[17] La idea que subyace a $\boldsymbol{S}_r$ guarda cierto parentesco con el planteamiento de la escuela estadística "de la verosimilitud" (*likelihoodism*). Sobre las diferencias entre la estrategia preconizada por esta escuela y lo que pretenden las ***S***-medidas bayesianas, v. Fitelson 2007.

[18] V. Kuipers (2000), sec. 3.2.1; aunque Kuipers prescinde del logaritmo, eso no afecta a la argumentación que sigue.

[19] Sobre los recursos de $\boldsymbol{S}_r$ para enfrentarse a los problemas clásicos de la confirmación como la paradoja de los cuervos de Hempel, las esmeraldas "verdules" de Goodman, etc., v. Kuipers (2000), pp. 55 y ss.

[20] Popper (1963), *addendum* 2. Sobre la relación entre la severidad popperiana y las medidas bayesianas de apoyo evidencial, v. Milne (1995) y Gillies (1998).

Referencias bibliográficas

CARNAP, R. (1950), *Logical Foundations of Probability*, Chicago, University of Chicago Press (segunda edición, 1962).

CRUPI, V., TENTORY, K. y GONZÁLEZ, M. (2007), 'On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues', *Philosophy of Science*, vol. 74, pp. 229-252.

CHRISTENSEN, D. (1999), 'Measuring confirmation', *Journal of Philosophy*, vol. 96, pp. 437-461.

EARMAN, J. (1992), *Bayes or Bust. A Critical Examination of Bayesian Confirmation Theory*, Cambridge, MIT Press.

EELLS, E. y FITELSON, B. (2000), 'Measuring Confirmation and Evidence', *Journal of Philosophy*, vol. 97, pp. 663-672.

— (2002), 'Symmetries and Asymmetries in Evidential Support', *Philosophical Studies*, vol. 107, pp. 129-142.

FITELSON, B. (1999), 'The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity', *Philosophy of Science*, vol. 66, pp. S362-S378.

— (2001), *Studies in Bayesian Confirmation Theory*, tesis doctoral no publicada, University of Wisconsin, Madison.

— (2002), 'Putting the Irrelevance Back into the Problem of Irrelevant Conjunction', *Philosophy of Science*, vol. 69, pp. 611-622.

— (2006), 'Logical Foundations of Evidential Support', *Philosophy of Science*, vol. 73, pp. 500-512.

— (2007), 'Likelihoodism, Bayesianism and Relational Confirmation', *Synthese*, vol. 156, pp. 473-489.

GILLIES, D. (1998), 'Confirmation Theory', en Gabbay, D.M. y Smets, Ph. (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. I*, Dordrecht, Kluwer, pp. 135-167.

GOOD, I. J. (1983), *Good Thinking. The Foundations of Probability and Its Applications,*Minneapolis, University of Minnesota Press.

HORWICH, P. (1982), *Probability and Evidence*. Cambridge, Cambridge University Press.

HOWSON, C. (2003), *Hume's Problem: Induction and the Justification of Belief,* Oxford, Oxford University Press.

HOWSON, C. y URBACH, P. (2006, 3ª ed.), *Scientific Reasoning. The Bayesian Approach,* La Salle, Illinois, Open Court.

HUBER, F. (2008), 'Milne's Argument for the Log-Ratio Measure', *Philosophy of Science*, vol. 75, pp. 413-420.

JEFFREY, R. (1992), *Probability and The Art of Judgment*, Cambridge, Cambridge University Press.

— (2004), *Subjective Probability: The Real Thing*, Cambridge, Cambridge University Press.

KEYNES, J. (1921), *A Treatise On Probability*, Londres, Macmillan.

KUIPERS, TH. (2000), *From Instrumentalism to Constructive Realism*, Dordrecht, Kluwer.

MILNE, P. (1995), 'A Bayesian Defence of Popperian Science?', *Analysis*, vol. 55, pp. 213-215.

— (1996), '*log [p (h/eb)/p (h/b)]* Is The One True Measure of Confirmation', *Philosophy of Science*, vol. 63, pp. 21-26.

POLLARD, P. (1999), 'Milne's Measure of Confirmation', *Analysis*, vol. 59, pp. 335-37.

POPPER, K. (1963), *Conjectures and Refutations*, Londres, Routledge and Kegan Paul.

ROSENKRANTZ, R.D. (1977), *Inference, Method and Decision*, Dordrecht, Reidel.

SCHLESINGER, G. N. (1995), 'Measuring Degrees of Confirmation', *Analysis*, vol. 55, pp. 208-212.

STEEL, D. (2007), 'Bayesian Confirmation Theory and the Likelihood Principle', *Synthese*, vol. 156, pp. 53-77.

TENTORI, K., CRUPI, V., BONINI, N., y OSHERSON, D. (2007), 'Comparison of confirmation measures', *Cognition*, vol. 103, pp. 107-119.