

## HACIA UNA JUSTIFICACIÓN TRASCENDENTAL DE LA INDUCCIÓN

**SÍLVIO PINTO\***

**Resumen:** Con el descubrimiento de la lógica inductiva bayesiana se ha renovado el optimismo acerca de la posibilidad de justificación de la inferencia inductiva. La justificación de las reglas inductivas bayesianas se conoce como el argumento del libro de apuestas holandés (el teorema de Ramsey-de Finetti). La cuestión que divide a los teóricos de la inducción es si tal argumento es realmente capaz de justificar las reglas de condicionalización de Bayes y de Jeffrey (las reglas inductivas bayesianas). Aquí me interesa, en primer lugar, distinguir dos problemas de justificación respecto a la inferencia inductiva: por un lado, el de una justificación persuasiva de este tipo de inferencia y, por el otro, el de una justificación explicativa de tal inferencia. En segundo lugar, busco relacionar la cuestión de la justificación de las reglas de condicionalización bayesianas con el clásico problema propuesto por David Hume de la justificación persuasiva de la causalidad y de la inducción. En tercer lugar, intento mostrar que el argumento de Ramsey-de Finetti no es capaz de proporcionar una justificación persuasiva de las reglas de condicionalización bayesianas en consonancia con la tesis negativa de Hume acerca de la posibilidad de justificación no-circular de las inferencias de tipo inductivo. Finalmente, sugiero una justificación explicativa para las reglas inductivas de la teoría bayesiana en términos de un argumento trascendental de inspiración kantiana y de estilo davidsoniano.

---

\* Profesor-investigador, Universidad Autónoma Metropolitana-Iztapalapa, pint@xanum.uam.mx

PALABRAS CLAVE: BAYESIANISMO, CAUSALIDAD, CONDICIONALIZACIÓN, EXPLICACIÓN, INDUCCIÓN

**Abstract:** *With the discovery of Bayesian inductive logic optimism about the possibility of rationally justifying inductive inference has been renewed. The justification of Bayesian inductive rules is known as the Dutch-book argument (Ramsey-de Finetti theorem). The question dividing theoreticians of induction is whether such an argument can really justify Bayes' and Jeffrey's conditionalization rules (Bayesian inductive rules). In this paper, I will be interested, first, in distinguishing two senses of justification of inductive inference: on the one hand, persuasive justification of induction and, on the other, its explicative justification. Secondly, I will relate the problem of the justification of Bayesian conditionalization rules with Hume's classical problem of persuasively justifying causality and induction. Thirdly, I shall argue that the Dutch-book argument cannot persuasively justify Bayesian conditionalization rules in agreement with Hume's negative thesis concerning non-circular justification of inductive inference of a causal type. Finally, I will suggest a sort of explicative justification of Bayesian inductive rules in terms of a transcendental argument of kantian inspiration but davidsonian in its style.*

KEY WORDS: BAYESIANISM, CAUSALITY, CONDITIONALIZATION, EXPLANATION, INDUCTION

## EL PROBLEMA DE HUME

**D**avid Hume fue el primero en formular el llamado problema de la justificación de la inducción. El problema era encontrar un argumento convincente en favor del uso de inferencias de tipo inductivo, cuyo empleo es tan común en los razonamientos cotidianos y, particularmente, en los de la ciencia. La búsqueda humeana de una justificación racional de la inducción está íntimamente ligada con su discusión respecto a la posibilidad de justificación racional del principio de causalidad. Este último se podría expresar de la siguiente manera:

Cada suceso natural está relacionado con otros sucesos naturales como causas y efectos.

La primera definición humeana de causa que aparece en la *Investigación sobre el entendimiento humano* la concibe como “an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second” (Hume, 1777: sec. VII, parte II). De acuerdo con Hume, la identificación de semejanzas entre objetos o, como se diría más correctamente hoy, sucesos, y su conjunción constante requiere el hábito, un mecanismo psicológico innato que se desencadena a partir de la acumulación de la experiencia del sujeto con su mundo. Pero el hábito también se requiere, según el mismo autor, para la identificación de causas y efectos. En otras palabras, la transición de las premisas:

- (P<sub>1</sub>) un cierto grupo de sucesos —sucesos del tipo A— son semejantes y otro grupo de sucesos —sucesos del tipo B— también son semejantes;  
 (P<sub>2</sub>) ha habido una conjunción constante entre ejemplares de sucesos del tipo A y ejemplares de sucesos del tipo B;

a la conclusión:

- C) los ejemplares del tipo A son causas de los ejemplares del tipo B;

es inductiva y, de acuerdo con nuestro autor, el hábito hace posible tal inferencia. En palabras del propio Hume:

I shall add, for a further confirmation of the foregoing theory, that, as this operation of the mind, by which we infer like effects from like causes, and *vice versa*, is so essential for the subsistence of all human creatures [...] It is more conformable to the ordinary wisdom of nature to secure so necessary an act of the mind, by some instinct or mechanical tendency, which may be infallible in its operations. (Hume, 1975: 55)

Sin embargo, además de la fundamentación psicológica de la regla de inducción, ¿sería posible una fundamentación filosófica de dicha regla? La famosa tesis negativa de Hume respecto a la causalidad y la inducción consiste en que no puede darse una justificación filosófica de ninguna de las dos reglas. Veamos de nuevo a Hume en lo que se refiere a ésta su tesis crucial:

I shall content myself, in this section, with an easy task, and shall pretend only to give a negative answer to the question here proposed. I say then, that, even after we have experience of the operations of cause and effect, our conclusions from that experience are *not* founded on reasoning, or any process of the understanding. (Hume, 1975: 32)

Consideremos este punto a continuación.

## LA JUSTIFICACIÓN DE LA DEDUCCIÓN

A pesar de lo que han pensado muchos filósofos —incluso Hume— hay un problema respecto a la deducción que es análogo al de la justificación de la inducción. Se podría plantear la cuestión de la justificación de nuestras prácticas deductivas en la ciencia y la matemática, así como también en el lenguaje cotidiano. Algunos han intentado justificar, por ejemplo, la aplicación de las reglas de la lógica clásica de primer orden en la ciencia, la matemática y en el lenguaje ordinario mostrando que hay una propiedad que poseen únicamente las reglas de esta lógica y que justifica su uso en los razonamientos llamados deductivos. La validez entendida como preservación de la verdad correspondería a tal propiedad semántica; se demuestra que el sistema de la lógica clásica es el cálculo más extendido que garantiza la validez de las inferencias deductivas en un lenguaje de primer orden con identidad. Éstas son las conocidas pruebas de legitimidad y completud del cálculo de predicados de primer orden con la identidad.

El mostrar que las reglas de la lógica clásica son legítimas y en su totalidad completas dejó, quizá, satisfechos a los lógicos interesados en la cuestión metalógica de la equivalencia entre las nociones sintáctica y semántica de consecuencia lógica, pero parecen no haber complacido del todo a los filósofos preocupados con el problema de la justificación de la deducción. Muchos de estos últimos empezaron a sospechar de la existencia de circularidad en el argumento justificador. La idea es que cualquier argumento deductivo válido que se use para mostrar que las reglas de esta lógica preservan la verdad, tiene que utilizar en alguno de sus pasos una o más de las reglas que se está buscando probar que son válidas. No parece haber una manera de escapar a una circularidad de este tipo.

Tomemos como ejemplo la legitimación del *modus ponens*:

(L<sub>1</sub>) La oración “ $p \rightarrow q$ ” es verdadera

(L<sub>2</sub>) La oración “ $p$ ” es verdadera

(L<sub>c</sub>) La oración “ $q$ ” es verdadera

El mismo argumento justificador del *modus ponens* ejemplifica esta regla de inferencia en el metalenguaje. Pero si la justificación de una regla lógica tan básica no puede prescindir del empleo de la misma, entonces estamos obligados a aceptar la circularidad del argumento justificador, o bien optamos por distinguir entre dos tipos de *modus ponens*: aquél que se aplica en el lenguaje objeto (MP<sub>1</sub>) y un *modus ponens* metalingüístico (MP<sub>2</sub>). Esta segunda alternativa nos abriría la posibilidad de rechazar la circularidad apelando a una regla semejante, pero de segundo nivel. Sin embargo, su costo sería elevado: saltar al nivel del metalenguaje suscitara la cuestión de la justificación de (MP<sub>2</sub>), lo que requeriría un nuevo salto a un metalenguaje de segundo nivel y a un *modus ponens* de tercer nivel (MP<sub>3</sub>) y, a la larga, a una proliferación infinita de metalenguajes y sus respectivos *modus ponens* (MP<sub>n</sub>).

La salida menos costosa para los que continúan creyendo en la posibilidad de una justificación racional de la deducción ha sido reconocer la circularidad del argumento justificador y, al mismo tiempo, proponer —como lo ha hecho, por ejemplo, Michael Dummett (1973: 295-297)— una distinción entre argumentos persuasivos y explicativos para mostrar que dicha circularidad no siempre es viciosa. Dice Dummett:

Our problem is not to persuade anyone, not even ourselves, to employ deductive arguments: it is to find a satisfactory explanation of the role of such arguments in our use of language. An explanation often takes the form of constructing a deductive argument, the conclusion of which is a statement of the fact needing explanation: but, unlike what happens in a suasive argument, in an explanatory argument the epistemic direction may run counter to the direction of logical consequence. In a suasive argument, the epistemic direction must coincide with the consequential one. (Dummett, 1973: 296)

La idea es que en los argumentos persuasivos la dirección lógica coincide con la dirección epistemológica; es decir, se conocen inicialmente las premisas y el argumento nos lleva a conocer la conclusión. Con los argumentos explicativos pasa lo contrario porque se conoce de antemano la conclusión y se buscan premisas a partir de las cuales se pueda derivar. En el caso en cuestión, el que investiga el problema de la justificación de la deducción ya está convencido de la verdad de la conclusión del argumento justificador; a saber, la proposición de que un determinado conjunto de reglas lógicas garantiza la validez de cualquier transición que tenga la forma de alguna de ellas. Lo que busca el investigador, no es un argumento que pueda persuadirlo de la verdad de esa proposición, sino premisas que puedan explicarla. Tal vez se pueda llamar a los argumentos de esta especie *inferencias a la mejor explicación*.

Ahora bien, la razón por la cual los argumentos persuasivos circulares no son convincentes es porque asumimos en las premisas algo que queremos demostrar. En el ejemplo mencionado utilizamos el *modus ponens* para demostrar que es una buena regla de inferencia en el sentido en que, por ejemplo, es una regla correcta. Claramente, alguien que no esté aún convencido de eso no aceptará el argumento. Supongamos ahora que mi interlocutor, a quien deseo convencer de que el *modus ponens* es una buena regla ya está convencido de esto; quizá debido a su éxito en nuestra práctica deductiva cotidiana o por alguna otra razón. Sin embargo, le interesaría conocer una explicación filosófica razonable para tal corrección. En este caso, no importaría que dicha explicación —o argumento explicativo— utilizara el mismo *modus ponens* una vez que lo que está en juego no es la corrección de la regla, sino la cuestión metafísica de qué es lo que constituye su corrección.

## **LAS JUSTIFICACIONES PERSUASIVA Y EXPLICATIVA DE LA INDUCCIÓN**

Si hubiera una lógica inductiva sería más fácil formular un problema análogo al de la justificación de la deducción, si ese fuera el caso, entonces querríamos saber cómo se podría explicar la corrección de las reglas de tal lógica inductiva. Corrección aquí significa sólo que los argumentos evaluados de acuerdo con la lógica inductiva como inductivamente fuertes

—el equivalente más próximo dentro de esta lógica a la validez— conducen la mayor parte del tiempo de premisas verdaderas a conclusiones verdaderas. No estamos, por lo tanto, buscando un argumento persuasivo en favor de la corrección no deductiva de nuestras mejores prácticas inductivas, lo que nos interesa es saber cuál sería la mejor explicación de esta corrección.

Tales consideraciones respecto a la mejor explicación de la corrección de la inducción divergen, en mi opinión, de la preocupación de Hume acerca de la posibilidad de justificar filosóficamente la inducción. Según lo entiendo, Hume considera sólo la posibilidad de un argumento persuasivo que nos inclinara en favor de la inducción y éste debería ser o bien inductivo o bien deductivo. Su pesimismo acerca de la posible existencia de tal argumento se debe a que, por un lado, si fuera deductivo él probaría demasiado, mientras que, por otro lado, si fuera inductivo nos haría caer en una circularidad. Lo que está detrás de la idea de que un argumento justificador deductivo, para la corrección del razonamiento inductivo, probaría demasiado es lo siguiente. Para empezar, quien buscara un argumento de este tipo tendría que garantizar que la forma inductiva de razonar que suponemos ha sido correcta en el pasado —en el sentido ya discutido— va a continuar siéndolo referente al futuro. Pero, ¿cómo un argumento deductivo nos podría arrojar una conclusión sobre el futuro partiendo de premisas acerca del pasado?

Si el único sentido posible de la justificación de la inducción fuera el persuasivo, entonces Hume tendría razón en ser pesimista respecto a cualquier propuesta para justificar esta forma de inferencia. Puesto que un argumento justificador inductivo obviamente no convencería a nadie que no estuviera convencido de la corrección de las inferencias inductivas. Por otro lado, un argumento justificador deductivo tendría que garantizar la corrección futura de inferencias inductivas ya reconocidas como correctas en el pasado, pero esto no es de ninguna manera lo que se está buscando. Lo que se desea argumentar, según Hume, es que la probabilidad de que nuestro uso futuro de las inferencias inductivas sea correcto, puesto que lo fue en el pasado, es muy alta; esto es claramente un argumento inductivo.

El dilema humeano entre un argumento justificador necesariamente deductivo o inductivo se disuelve una vez que se ha rechazado la exigencia de que sea un argumento persuasivo. Mi propuesta, siguiendo a

Dummett, es que tal argumento debe ser explicativo, él debe proporcionar la mejor explicación para la corrección persistente en el tiempo de nuestras prácticas inductivas. Es evidente que, tal explicación no nos ofrecería una razón para creer que dichas prácticas son más correctas que otras no-inductivas. Dicho de otra manera, ella no serviría para convencer al escéptico más radical (aquel que duda de la corrección del razonamiento inductivo). Con tal escéptico sería, sin embargo, imposible cualquier intercambio lingüístico coherente. Cualquier persona mínimamente más razonable quizás aceptase la justificación explicativa de la inducción porque el estar persuadida de la corrección de la inducción no le impide cuestionarse acerca del por qué de la misma.<sup>1</sup>

## LA LÓGICA INDUCTIVA BAYESIANA

El llamado enfoque bayesiano o lógica inductiva bayesiana se ha desarrollado a partir de los trabajos del filósofo inglés Frank Ramsey (1926) y del matemático italiano Bruno de Finetti (1937), pero sus orígenes se remontan al descubrimiento del cálculo de probabilidades por Blaise Pascal y Pierre de Fermat en el siglo XVII. Esta lógica es llamada bayesiana en homenaje a Thomas Bayes, clérigo inglés del siglo XVIII quien descubrió uno de sus teoremas cruciales, que lleva su nombre. El teorema de Bayes, que utiliza el concepto de probabilidad condicionada,<sup>2</sup> es uno de los resultados que demuestran el carácter inductivo de la lógica bayesiana. Una de sus formulaciones es la siguiente:

---

<sup>1</sup> En el sentido metafísico o filosófico y no, por ejemplo, en el sentido causal.

<sup>2</sup> Esta probabilidad, que denotamos como  $P(q/p)$ , expresa la siguiente idea: sean todos los mundos posibles en que  $p$  es verdadera. La fracción de tales mundos en que  $q$  también es verdadera corresponde a la probabilidad de que  $q$  puesto que  $p$ . Se puede definir la probabilidad condicionada en términos de la probabilidad no-condicionada y viceversa. La nota 12 contiene una definición del primer tipo. Una definición del segundo tipo sería la siguiente:  $P(p) = P(p/\mathbf{T})$ , tal que  $\mathbf{T}$  es una tautología.



- (TB) La probabilidad de una determinada hipótesis  $h$  condicionada a la evidencia  $e$  es igual a la probabilidad de  $e$  condicionada a  $h$  multiplicada por la probabilidad de  $h$  divididas ambas por la probabilidad de  $e$ .

Veremos más adelante, cuando se introduzca la regla de condicionación de Bayes, que este teorema nos permite calcular la probabilidad posterior de una hipótesis cuando la probabilidad de una pieza de evidencia sube a 1, si sabemos la probabilidad anterior de la hipótesis (no condicionada), la probabilidad anterior de la evidencia y la probabilidad de la evidencia en el caso en que la hipótesis fuera verdadera.

Los matemáticos del siglo XVII estaban interesados en cuantificar y predecir la incertidumbre de los juegos de azar. Ramsey buscaba introducir una nueva interpretación de la probabilidad como una medida de los grados de creencia de los sujetos racionales en un determinado contenido proposicional. Específicamente le interesaba encontrar un método para calcular al mismo tiempo los grados de creencia y deseo de un sujeto racional, a partir de la información acerca de sus preferencias por determinados cursos de acción relevantes para la situación de decisión que se pretende investigar. El problema para el que Ramsey finalmente encontró una solución brillante es, en último análisis, el problema fundamental de la teoría de la decisión: dada información suficiente respecto a la escala de preferencias del agente ( $A$ ), por los diversos cursos de acción posibles para su situación de decisión, encontrar el patrón de deseos y creencias de  $A$  (Ramsey, 1926).

Una de las tesis fundamentales de la lógica bayesiana es que la noción de probabilidad interpretada de manera subjetivista debe ser utilizada para medir los grados de creencia de los agentes humanos. Así, a la certeza de que un determinado contenido proposicional es verdadero se le atribuye probabilidad subjetiva 1; en este caso, el agente cree con toda seguridad que dicha proposición es verdadera. Junto con la certeza de que una proposición es falsa estará asociada la probabilidad subjetiva 0, es decir, el agente cree con absoluta seguridad que la negación de esa proposición es verdadera. Según el cálculo de probabilidades, las tautologías y las contradicciones poseen probabilidades iguales a 1 y a 0 respectivamente. Todas las otras proposiciones tienen probabilidades o grados de creencias que varían entre estos dos extremos. La aplicación del cálculo

lo de probabilidades como escala de medida para los grados de creencia de los agentes humanos impone diversas restricciones al sistema de dichas creencias de un sujeto que razona más o menos de acuerdo con las reglas de tal cálculo. Por ejemplo, si un agente atribuye a una proposición la probabilidad subjetiva  $x$  y si además es coherente,<sup>3</sup> entonces debe atribuirle a su negación la probabilidad subjetiva  $1 - x$ . Otro ejemplo sería el siguiente: si el sujeto le asigna la probabilidad subjetiva  $x$  a la única premisa de un argumento deductivo válido, y si se le puede considerar coherente, debe atribuir a su conclusión una probabilidad mayor o por lo menos igual a  $x$ .

Las reglas del cálculo de probabilidades establecen cómo se debe comportar el patrón de probabilidades subjetivas de un agente supuestamente coherente en un determinado momento,<sup>4</sup> pero no nos dice nada acerca de cómo deberían cambiar tales probabilidades con el paso del tiempo —cómo debería variar la función de probabilidad en el tiempo—. La regla bayesiana más sencilla que gobierna los cambios temporales de las probabilidades subjetivas es la de condicionalización de Bayes (RCB), que afirma lo siguiente:

(RCB) Si la probabilidad subjetiva de una determinada proposición  $p$  sube a 1 y si la probabilidad condicionada a  $p$  de cualquier proposición  $q$  ( $P(q/p)$ ) no cambia como resultado del cambio temporal en la probabilidad de  $p$ , entonces se debe actualizar la probabilidad de  $q$  a  $P(q/p)$ . Si (i)  $P_i(p) = x < 1$  y  $P_i(p) = 1$  y (ii)  $P_i(q/p) = P_i(q/p)$ , entonces  $P_i(q) = P_i(q/p)$ .<sup>5</sup>

---

<sup>3</sup> La coherencia aquí no tiene que ver con la obediencia al principio de no-contradicción de la lógica clásica, sino más bien con la no violación de las reglas de la lógica inductiva bayesiana. Atribuir probabilidades subjetivas a una proposición y su negación cuya suma fuera diferente de 1 ejemplificaría la incoherencia en este segundo sentido.

<sup>4</sup> Este patrón corresponde a la función de probabilidad sobre el dominio de todas las proposiciones posibles de ser consideradas por un sujeto en un tiempo dado.

<sup>5</sup> Estas dos condiciones están claramente explicadas por Richard Jeffrey quien denomina la primera condición necesaria para la condicionalización RCB como condición de certeza y la segunda como condición de rigidez. Las dos juntas constituyen, según él, las

El principio de condicionalización de Bayes nos ofrece el ejemplo más claro de una regla inductiva, pues, supongamos que se substituye por  $p$  un determinado contenido proposicional —digamos, que el aumento de temperatura de una cierta olla exprés llena de agua hasta la mitad es acompañado por un aumento de la presión interna del vapor de agua confinado en su interior— cuya probabilidad antes del experimento es menor que 1. Supongamos, además, que se substituye por  $q$  la hipótesis de que la presión de un gas ideal a volumen constante es directamente proporcional a su temperatura, cuya probabilidad anterior al experimento en cuestión también es menor que 1. La evidencia proporcionada por este experimento y un sinnúmero de otros similares deberían conducir a un incremento de la probabilidad de la hipótesis  $q$ . La regla de Bayes describe la corrección de una inducción de este tipo del siguiente modo: sea cual fuera la probabilidad que le asigna un sujeto coherente a  $q$ , cada vez que se incorpora conocimiento nuevo que confirma la misma (cada premisa de la inducción) a su sistema de creencias, él debería reajustar la probabilidad subjetiva que le atribuye a  $q$  de acuerdo con la RCB.

Esta nueva perspectiva acerca de la inducción —es decir, la idea de asignar probabilidades ya sea a las premisas y conclusión de una inferencia inductiva, o bien, al condicional que las relaciona— disuelve un viejo problema relacionado con tales inferencias. Éste era el problema de clasificar las formas que corresponden a las inducciones correctas de manera análoga a como han sido clasificadas las formas de inferencia que corresponden a las deducciones válidas. Como en el caso de la inducción, la mera forma lógica de la inferencia no garantiza su corrección —no garantiza que se trata de un argumento inductivo fuerte— la idea de medir la fuerza de tales argumentos usando probabilidades (por ejemplo, la pro-

---

condiciones necesarias y suficientes para la aplicación de la RCB. La condición de certeza se llama así porque el tipo de condicionalización respecto de una proposición  $p$  propuesto por la RCB sólo se puede aplicar en los casos en que la probabilidad de  $p$  se cambia a 1. La condición de rigidez debe su nombre a que en el proceso de cambio de la probabilidad subjetiva asociada con  $q$ , cuando la función de probabilidad asociada con otra proposición  $p$  cambia su valor a 1 en el mismo intervalo, la probabilidad condicionada  $P(q/p)$  en este mismo intervalo de tiempo no se cambia (Jeffrey, 2002: 46-48).

babilidad de que la conclusión sea verdadera con la suposición de que las premisas son verdaderas) es perfectamente adecuada para resolver el problema de distinguir los argumentos inductivos fuertes de los débiles. Por ejemplo, la mera forma lógica de un argumento como el que sigue no nos dice nada sobre su fuerza o corrección inductiva:

(IE) <sup>6</sup>	$P_1$	El cisne $a_1$ es blanco
	$P_2$	El cisne $a_2$ es blanco
	...	...
	$P_n$	El cisne $a_n$ es blanco
	C	Todos los cisnes son blancos.

El enfoque bayesiano abandona una cierta concepción de la lógica, según la cual ésta debiera investigar las formas correctas de inferencia. Lo importante para la lógica bayesiana es la fuerza de los argumentos inductivos, ésta se mide en términos de la probabilidad subjetiva de que la conclusión sea verdadera con la condición de que todas las premisas sean verdaderas ( $P(c/p_1, p_2, \dots, p_n)$ ). Los argumentos inductivos considerados fuertes o correctos desde el punto de vista de la lógica bayesiana son aquellos cuya probabilidad  $P(c/p_1, p_2, \dots, p_n)$  es próxima a 1. Los argumentos deductivos válidos son casos límite de argumentos inductivos fuertes en que  $P(c/p_1, p_2, \dots, p_n)$  es igual a 1. A la lógica inductiva también le interesa la relación entre las probabilidades subjetivas de las premisas y de la conclusión de los argumentos inductivos, así como de los deductivos en condiciones de incertidumbre.

---

<sup>6</sup> Por inducción enumerativa. Se podría intentar capturar la forma de las inferencias inductivas enumerativas de la siguiente manera:

(FIE)	$P_1$	$Ca_1 \wedge Ba_1$
	$P_2$	$Ca_2 \wedge Ba_2$
	...	...
	$P_n$	$Ca_n \wedge Ba_n$
	C	$(\forall x)(Ca_x \rightarrow Ba_x)$

donde 'C' y 'B' son letras metalingüísticas que denotan predicados del lenguaje y ' $a_1$ ', ' $a_2$ ', ' $a_n$ ' son letras metalingüísticas que denotan nombres de objetos.

Se podría pensar que el problema de justificar la forma de inferencia ampliativa del tipo IE es distinto de justificar la regla RCB o cualquier otra regla inductiva de la lógica bayesiana. Ciertamente, una diferencia fundamental entre las reglas inductivas humeanas y las bayesianas es que sólo en las segundas se asignan probabilidades a las premisas y a la conclusión de un argumento, mientras que Hume nunca utiliza el aparato del cálculo de probabilidades para caracterizar las inferencias que él considera inductivas. Otra diferencia entre el tipo de inducción que Hume tenía en mente y la que subyace al bayesianismo es la siguiente. Si consideramos que para él la forma inductiva por excelencia es la que está dada por FIE, se podría interpretar la RCB como si incluyera la inducción enumerativa IE, además de otros tipos de inducción no-enumerativos. Una forma de comprobar esto sería tomar la FIE como una serie finita de  $n$  aplicaciones de la RCB cuando en este orden las proposiciones  $P_1, P_2, \dots, P_n$  son observadas.<sup>7</sup> Así, mientras la RCB ordena que se incremente de manera sucesiva la probabilidad subjetiva asociada con  $C$ , la IE ordena la transición ampliativa de la verdad de  $P_1, P_2, \dots, P_n$  a la verdad de  $C$ . Estrictamente hablando, la RCB representa un refinamiento y una ampliación de IE para los casos no-enumerativos. De la misma manera, como mostraré en la siguiente sección, la regla de condicionalización de Jeffrey (RCJ) es un refinamiento de la RCB. En adelante, cuando hable de la justificación de la forma inductiva de inferencia me refiriré en especial a la RCB o la RCJ.

En relación con la primera diferencia entre las inducciones humeana (IE) y bayesiana (RCB) —a saber, que en la IE no se asignan probabilidades a premisas y conclusiones mientras que en la RCB sí— no me parece que ésta sea una diferencia fundamental entre las dos. Si olvidamos por un momento la genuina diferencia entre IE y RCB —la mencionada diferencia en términos de refinamiento y ampliación— y reformulamos la IE en términos de asignación de probabilidades subjetivas a sus premisas y conclusión, esta primera diferencia desaparece.<sup>8</sup> En mi opinión, esto muestra que, mientras que la formulación de la RCB debe ser forzosamente

<sup>7</sup> En otras palabras, sus respectivas probabilidades subjetivas se cambian a 1.

<sup>8</sup> La reformulación de IE utilizando probabilidades subjetivas tomaría la siguiente forma: si las probabilidades de  $P_1, P_2, \dots, P_n$  son todas iguales a 1, entonces la probabilidad de  $C$  es alta.

metalingüística<sup>9</sup> para poder expresar los posibles infinitos grados de creencia en la verdad de sus premisas y conclusión, la IE puede tener una formulación no-metalingüística porque los dos únicos grados que intenta expresar — $S$  cree que  $p$  es verdadera y  $S$  cree que  $p$  es falsa— se pueden también expresar por medio de la afirmación de  $p$  y de la afirmación de  $\neg p$  respectivamente.

## EL INTENTO BAYESIANO DE JUSTIFICACIÓN PERSUASIVA DE LAS REGLAS INDUCTIVAS DE LA LÓGICA BAYESIANA

Se dijo que la lógica bayesiana no dispone de herramientas para proporcionar una justificación persuasiva de sus propias reglas no-inductivas e inductivas. Sin embargo, varios de los que han contribuido al desarrollo de la lógica bayesiana se han dedicado también a una justificación persuasiva de dichas reglas. Debemos a Ramsey y a de Finetti un argumento que se conoce como el teorema del libro de apuestas holandés (*the Dutch-book argument*). En este argumento se propone mostrar que el cálculo de probabilidades —el conjunto de reglas no-inductivas de la lógica bayesiana— es la representación más perfecta de la coherencia del patrón de grados de creencias en un agente humano idealizado.<sup>10</sup> El argumento explora la idea de que un apostador cuyo patrón de grados de creencia no obedece a las reglas del cálculo de probabilidades ganaría o perdería —según ocupe la posición de apuntador o la de su oponente— cualesquiera que fueran los resultados de sus apuestas. Desde la hipótesis de que un agente coherente no va a apostar sabiendo que perderá sea cual sea el resultado, su patrón de creencias debe satisfacer los axiomas y teoremas de dicho cálculo, como condición de preservación de la coherencia de sus decisiones respecto de las apuestas que hace.

---

<sup>9</sup> Por 'formulación metalingüística de una regla de inferencia' quiero decir que la formulación misma menciona explícitamente una relación semántica de las premisas y de la conclusión de la regla con el mundo, a saber, la probabilidad subjetiva de que estas proposiciones sean verdaderas.

<sup>10</sup> Un sujeto humano idealizado en el sentido de que sus grados de creencia en un determinado momento obedecen a las reglas del cálculo de probabilidades.

La estrategia usada por Ramsey y de Finetti es la siguiente: ellos tratan de demostrar que el apostador perdería o ganaría siempre, para los casos más sencillos en los que sus grados de creencia violaran los axiomas del cálculo de probabilidades. Con esto queda demostrado que lo mismo vale para cualquier violación de sus teoremas. Ilustraré la estrategia de prueba del teorema de Ramsey-de Finetti para el caso del primer axioma: el cual afirma que la probabilidad de cualquier proposición es siempre mayor o igual a 0. Supongamos que un apostador (*A*) asigna a la proposición *p* una probabilidad subjetiva menor que 0. *A* está en favor de *p*: él gana *a* pesos si *p* es verdadera y pierde *b* pesos si *p* es falsa. Su oponente (*B*) obviamente ganará *b* pesos si  $\neg p$  es verdadera y perderá *a* pesos si  $\neg p$  es falsa. Supongamos además que la probabilidad subjetiva de *p* para *A* es igual a *x*, tal que  $-1 < x < 0$ , y que la probabilidad subjetiva de  $\neg p$  es igual a  $1 - x$ . La tabla de las ganancias líquidas de *A* y las probabilidades subjetivas para esta apuesta simple son las siguientes:

APUESTA EN <i>p</i>	GANANCIA LÍQUIDA	PROBABILIDAD SUBJETIVA
V	+ <i>a</i>	<i>x</i>
F	- <i>b</i>	$1 - x$

Supongamos también que *A* considera esta apuesta justa, lo que significa que su ganancia media (GM) es igual a 0.<sup>11</sup> Tenemos entonces la siguiente ecuación:

$$GM(p) = 0, \text{ donde } GM(p) = a \times x + (-b) \times (1 - x)$$

Hay sólo dos modos de satisfacer esta ecuación: o bien *b* es menor que 0 y *a* mayor que 0, con módulo de *a* mayor que el módulo de *b*, o bien *a* es

---

<sup>11</sup> Una apuesta que no es justa puede ser o bien ventajosa para un determinado apostador cuando su ganancia media es mayor que 0, o bien desventajosa para él en cuyo caso su ganancia media es menor que 0. La ganancia media de una apuesta es igual a la suma de los productos de las ganancias de cada situación posible multiplicadas por sus respectivas probabilidades.

menor que 0 y  $b$  mayor que 0, con módulo de  $a$  mayor que el módulo de  $b$ . En la primera situación,  $A$  ganará siempre; en la segunda, perderá pase lo que pase. Razonamientos análogos se aplican a los demás axiomas del cálculo de probabilidades, incluso al que relaciona la probabilidad condicionada con las probabilidades no condicionadas.<sup>12</sup>

El uso del teorema de Ramsey-de Finetti para justificar las reglas no-inductivas de la lógica bayesiana ha sugerido, naturalmente, la siguiente cuestión: ¿habrá algún argumento semejante al del libro de apuestas holandés que pudiera justificar de manera persuasiva las reglas inductivas de esta lógica y en particular la RCB?

Ramsey y de Finetti no ofrecen ninguna justificación para esta regla inductiva. Al inicio de la década de 1970, David Lewis y Paul Teller anunciaron haber encontrado una estrategia semejante al libro de apuestas holandés para justificar la RCB. El argumento de Lewis-Teller parte de la suposición de que la probabilidad subjetiva posterior de una hipótesis  $h$  es distinta de la probabilidad anterior de  $h$  condicionada a  $e$  ( $P_f(h) \neq P_i(h/e)$ ) (Teller, 1973: 222-224). El argumento trata de exhibir una estrategia combinada de apuestas en  $h$  dado  $e$ , en  $e$ , y, finalmente, en  $h$  usando las probabilidades  $P_i(h/e)$ ,  $P_i(e)$  y  $P_f(h)$  en sus respectivos cálculos, tal que el apostador pierde todo, cualquiera que sea el valor de verdad de la condición ( $e$ ) o de la hipótesis ( $h$ ). Esto muestra, según los autores, que la RCB es la estrategia más razonable para el cambio de nuestros grados de creencia, dada la ocurrencia de una determinada evidencia.

Es posible describir la estrategia propuesta por Lewis-Teller de la siguiente manera: supongamos que el apostador ( $A$ ) cambia la probabilidad subjetiva de su creencia en la verdad del contenido proposicional  $h$  cuando se entera de la verdad de la evidencia  $e$ , y lo hace de manera distinta de la que le ordena la RCB. Supongamos además que su nueva probabilidad ( $P_f(h)$ ) sea tal que  $P_i(h) < P_i(h/e)$ . Tomemos entonces  $P_i(h/e) = x$  y  $P_i(h/e) - P_f(h) = y$ . Imaginemos ahora que el apostador acepta la siguiente secuencia de apuestas como justa. Inicialmente (en  $t_0$ ) acepta:

---

<sup>12</sup> Este axioma afirma que la probabilidad de  $q$  condicionada por  $p$  es igual a la probabilidad de  $q$  y  $p$  dividida por la probabilidad de  $p$  para los casos en que esta probabilidad está definida y es diferente de 0.



a) una apuesta en  $h$  condicionada a  $e$  tal que él gana  $P_i(\neg h/e)$  si  $h$  y  $e$  son ambas verdaderas, pierde  $P_i(h/e)$  si  $e$  es verdadera pero  $h$  es falsa y no gana ni pierde en todos los otros casos; b) una apuesta en  $e$  tal que él gana  $yP_i(\neg e)$  si  $e$  es verdadera, y pierde  $yP_i(e)$  si  $e$  es falsa; por último (en  $t_j$ ) si  $e$  es verdadera acepta c) una apuesta en  $h$  tal que él gana  $P_f(h)$  si  $h$  es falsa y pierde  $P_f(\neg h)$  si  $h$  es verdadera. La tabla de las ganancias líquidas de  $A$  es la siguiente:

$h$	$e$	g. l. $_i(h/e)$	g. l. $_i(e)$	g. l. $_f(h)$	g. l. $_{total}(h/e + e + \neg h)$
V	V	$1 - x$	$yP_i(\neg e)$	$x - y - 1$	$-yP_i(e)$
F	V	$-x$	$yP_i(\neg e)$	$x - y$	$-yP_i(e)$
V	F	0	$-yP_i(e)$	0	$-yP_i(e)$
F	F	0	$-yP_i(e)$	0	$-yP_i(e)$

La última columna de la tabla muestra de manera clara que  $A$  va a perder  $-yP_i(e)$  en todos los casos. Mediante un argumento semejante se puede mostrar que lo mismo pasaría si  $A$  hubiera atribuido a la nueva probabilidad de  $h$  un valor mayor que  $P_i(h/e)$ .

Mucho se ha escrito contra el argumento de Lewis-Teller. Colin Howson y Peter Urbach (1993: cap. 6), por ejemplo, no han aceptado que tal argumento pueda ofrecer algún tipo de justificación a la RCB. Esto porque si la condición (ii)  $P_i(q/p) = P_f(q/p)$ <sup>13</sup> es indispensable para la aplicabilidad de esta regla de condicionalización, entonces se sigue lógicamente de (ii) y de (i)  $P_f(p) = 1$  que  $P_f(q) = P_i(q/p)$ .<sup>14</sup> Tal y como la enunciamos antes, la

<sup>13</sup> La que Jeffrey llama condición de rigidez, Howson y Urbach denominan condición de invariancia.

<sup>14</sup> Si se supusiera que la condición (ii) no es indispensable, entonces, según Howson y Urbach, obviamente no sería una verdad lógica que:

RCB es un enunciado analítico y, por lo tanto, la pregunta por su justificación persuasiva no tendría sentido una vez que en este momento no está en juego alguna posible alternativa al canon de analiticidad que nos ofrece la lógica clásica. En todo caso, para lo que sí tendría sentido pedir una justificación persuasiva sería para las condiciones (i) y (ii). Iniciemos pensando cómo se podría justificar la segunda.

Consideremos el siguiente ejemplo ilustrativo de la falla de (ii).<sup>15</sup> Sea un contenido proposicional ( $h$ ) acerca de la percepción del color de algún objeto. Supongamos que un sujeto ( $S$ ) le atribuye probabilidad 1, lo que significa que tiene certeza absoluta respecto de su percepción. Supongamos además que  $S$  sospecha que posee una lesión cerebral cuyo efecto es reducir la certeza con que sostiene sus creencias algún tiempo después. Tenemos que  $P_i(h) = 1$  y  $0 < P_f(h) = r < 1$ . En relación con la probabilidad condicionada  $P(h/P_f(h) = r)$ , vamos a tener  $P_i(h/P_f(h) = r) = 1$  y  $P_f(h/P_i(h) = r) = r$ . Si usamos la RCB obtendremos como probabilidad posterior de  $h$ :  $P_f(h) = 1$ . Pero en esta situación parece más razonable cambiar la probabilidad de  $h$  al valor  $r$  cuando la proposición  $P_f(h) = r$  se torna verdadera.

El punto de Howson y Urbach es que no siempre está justificada la condición de invariancia ( $P_i(q/p) = P_f(q/p)$ ) y, además, el argumento de Lewis-Teller no ofrece ninguna justificación para ella. Según Howson y Urbach, no hay ninguna incoherencia diacrónica en violarla, aunque sí haya una incoherencia (lógica en el sentido clásico) en una estrategia de cambio de creencias que viole la condicionalización de Bayes cuando obtienen las condiciones de certeza e invariancia. Esta última condición estaría justificada solamente en aquellos casos en que la observación no afectara nuestras estimaciones condicionadas a lo que se ha observado

---

(RCB')

Si (i) entonces  $P_f(q) = P_i(q/p)$ .

Pero, el argumento de Lewis-Teller tampoco justificaría la nueva formulación de la condicionalización de Bayes (RCB') una vez que para ellos la idea misma de que pudiera haber una incoherencia diacrónica entre el valor  $P_i(q)$  y el valor  $P_f(q)$  que fuera diferente de  $P_i(q/p)$  —o sea, que violara la RCB'— no tiene el mínimo sentido.

<sup>15</sup> El ejemplo está en Howson y Urbach, 1993: cap. 6, pero es fácil imaginar otros ejemplos semejantes.

sobre la verdad de otras proposiciones. En los casos en que la aplicación de la condición de invariancia no se justifica tendríamos el problema adicional de cuál sería la manera correcta de actualizar nuestras probabilidades subjetivas; hasta donde sé, no hay una regla alternativa para estos casos.

Las situaciones en las cuales la condición (i) no se aplica —aquellas en que como resultado de la observación no se cambian las probabilidades subjetivas de los enunciados de observación  $e_1, e_2, \dots, e_n$  a 1— son más fáciles de tratar, porque ahí sí hay una regla de condicionalización alternativa. Ésta es la llamada regla de condicionalización de Jeffrey que ordena lo siguiente:

(RCJ) [...] si las probabilidades subjetivas de las proposiciones  $p_n$ , ( $n = 1, \dots, m$ ) que forman una partición del espacio de probabilidades se cambian de  $P_i(p_n)$  a  $P_f(p_n)$  y si la probabilidad condicionada a  $p_n$  de cualquier proposición  $q$  ( $P(q/p_n)$ ) no cambia como resultado del cambio temporal en la probabilidad de  $p_n$ , entonces se debe actualizar la probabilidad de  $q$  a  $P_i(q/p_1) \times P_f(p_1) + \dots + P_i(q/p_n) \times P_f(p_n)$ . Si (i)  $P_i(p_n) = x \neq y = P_f(p_n)$  y (ii)  $P_i(q/p_n) = P_f(q/p_n)$ , entonces  $P_f(q) = P_i(q/p_1) \times P_f(p_1) + \dots + P_i(q/p_n) \times P_f(p_n)$ . (Jeffrey, 1983: cap. 11)

También se ha intentado extender el argumento de Lewis-Teller a la RCJ con la finalidad de mostrar que ella representa la única estrategia coherente a seguir si se aceptan las condiciones (i) y (ii); un ejemplo de esta clase nos lo ofrece Brian Skyrms.<sup>16</sup> Aquí también se aplican las observaciones mencionadas antes de Howson y Urbach acerca de la ausencia de

---

<sup>16</sup>En Skyrms (1987). En este texto magníficamente claro, el autor revisa el argumento de Lewis-Teller para la RCB, propone un nuevo libro-de-apuestas-holandés para la RCJ y, finalmente, presenta argumentos inversos del tipo del libro-de-apuestas-holandés en favor de ambas. Un argumento de este tipo muestra que si no respeta la regla de condicionalización en la actualización de las probabilidades subjetivas asociadas con enunciados no-observacionales, un agente está sujeto a una estrategia del libro de apuestas holandés. Un argumento inverso muestra más bien que si respeta la regla de condicionalización en la actualización de las probabilidades subjetivas asociadas con estos enunciados, el agente es invulnerable a cualquier estrategia del tipo del libro de apuestas holandés.

incoherencia diacrónica en la violación de la condición (ii) en ciertas situaciones y, por lo tanto, respecto de la carencia de justificación persuasiva del argumento del tipo del libro-de-apuestas-holandés para una regla de condicionalización.

Además, alguna reflexión respecto a la estructura de preferencias de un apostador que debe decidir si acepta o no la estrategia del libro-de-apuestas-holandés diacrónico, seguramente arrojará luz acerca de si podría haber algún otro tipo de incoherencia en tal aceptación o rechazo. Limitémonos a la RCB por ser ésta la regla inductiva más simple. Se trata de una secuencia de tres apuestas, dos de las cuales el apostador debe aceptar en  $t_0$  mientras que la tercera debe ser aceptada en  $t_1$ ; esto quiere decir que cada una de ellas, por separado, es considerada justa en su respectivo momento. Sin embargo, la secuencia misma de apuestas debe ser rechazada como injusta de acuerdo con la RCB. Pero, esto parece estar en conflicto con otro principio de la teoría bayesiana que podríamos denominar principio de la aditividad de la preferencia (PAP) y que afirma lo siguiente:

(PAP) Si un agente prefiere cada una de una serie finita de apuestas *al statu quo*, entonces debe preferir también el paquete completo de apuestas de la serie *al statu quo*.

En el caso del libro-de-apuestas-holandés diacrónico, nuestro agente prefiere cada una de las tres apuestas aisladamente y, sin embargo, si condicionaliza de acuerdo con la RCB no acepta el paquete. John Earman (1992: cap. 2) observa que un conflicto con el PAP también está presente en la justificación del tipo del libro-de-apuestas-holandés sincrónico del

---

Vale la pena observar que Skyrms, Teller y también Howson y Urbach optan principalmente por discutir la justificación de la formulación RCB' de la condicionalización de Bayes. De manera análoga para la RCJ, ellos examinan de manera primordial la posibilidad de una justificación persuasiva de:

(RCJ) Si (i)  $P_i(p_n) = x \neq y = P_f(p_n)$ , entonces  $P_f(q) = P_i(q/p_1) \times P_f(p_1) + \dots + P_i(q/p_n) \times P_f(p_n)$

axioma de la aditividad.<sup>17</sup> Si Earman tiene razón, entonces todavía la obediencia a los dictámenes de la RCB es compatible con algún tipo de incoherencia en el sentido de una flagrante violación del PAP. Pero, ¿estaría tal principio justificado?

Además, hay seguramente otros presupuestos en la teoría bayesiana para los cuales se podría formular la misma pregunta. Uno de ellos es que no sería racional en el sentido bayesiano<sup>18</sup> apostar a sabiendas de que se va a perder pase lo que pase si los únicos deseos del apostador que son relevantes para la situación de decisión están representados por sus ganancias monetarias. No sería racional elegir este curso de acción en lugar de otros en los que uno no perdería dinero en todos los casos (por ejemplo, aquél en que decidiera no aceptar esta apuesta tan desventajosa) simplemente porque un agente que actuara de esta manera estaría violando el principio básico de la teoría de la decisión bayesiana: el llamado principio de la maximización de la utilidad esperada, que podríamos expresar como el siguiente imperativo:

(PMUE) Actúa de tal manera que la función de utilidad media asociada al curso de acción que elegiste entre los varios cursos de acción posibles en una dada situación de decisión sea máxima.

Podríamos preguntarnos respecto de la justificación de este principio bayesiano de racionalidad lo siguiente: ¿por qué debemos aceptar la no-

---

<sup>17</sup> Uno de los cuatro axiomas del cálculo de probabilidades propuesto por Andrei N. Kolmogorov en 1950. Este axioma afirma que la probabilidad de la disyunción de dos proposiciones  $p$  y  $q$  es igual a la suma de las probabilidades de cada una de ellas.

<sup>18</sup> Este sentido de racionalidad corresponde a la concepción instrumentalista, según la cual ésta se explica en términos de una relación de adecuación entre medios y fines de un agente humano. En este sentido, un agente racional sería aquél que busca los medios más adecuados para satisfacer sus fines. La expresión más precisa de esta concepción de la racionalidad humana está dada por el principio de maximización de la utilidad esperada que será discutido más adelante. Dicha concepción de la racionalidad, que ya está presente en Aristóteles, es la que se encuentra implícita en la mayor parte de la tradición empirista.

ción de racionalidad bayesiana (que se encuentra, digamos, parcialmente caracterizada por el principio de maximización de la utilidad esperada) en lugar de otras concepciones alternativas de la acción humana?<sup>19</sup>

La pregunta por la justificación persuasiva del principio de maximización de la utilidad esperada se puede extender de manera obvia a todos los principios de la teoría bayesiana del razonamiento: los que con anterioridad denominé principios de coherencia del patrón de creencias del sujeto, e inclusive, los principios de coherencia más clásicos, como la no-contradicción, el tercero excluido y otros principios de la lógica clásica que también son parte constitutiva de la teoría bayesiana. Pero si la cuestión de la justificación de los principios básicos del bayesianismo nos lleva de manera directa a la noción misma de su racionalidad, caracterizada de manera implícita por el conjunto de tales principios, ¿cómo podríamos justificar persuasivamente la elección de dicha noción frente a sus rivales sin caer en la circularidad viciosa denunciada por Hume con suma claridad desde hace tanto tiempo?

## **LA JUSTIFICACIÓN EXPLICATIVA DE LOS PRINCIPIOS BÁSICOS DE LA LÓGICA BAYESIANA**

Una posible salida sería apelar a la respuesta que Immanuel Kant ofrece al problema humeano de la justificación persuasiva del principio de causalidad. En consonancia con la sugerencia de Dummett respecto de la circularidad no-viciosa de un argumento justificador explicativo, Kant propone que la mejor explicación filosófica de la objetividad del contenido de nuestros juicios sintéticos empíricos descansa en que ésta requiere la intervención de principios organizadores tales como el de causalidad y el de la permanencia de la sustancia; lo que significa que nosotros no podríamos hablar con verdad o falsedad de objetos o sucesos relacionados de

---

<sup>19</sup> Aclaro que en este artículo no discutiré las posibles concepciones de racionalidad alternativas a la instrumental, porque esto rebasaría los límites del mismo. Además, en lo que toca a la misma racionalidad instrumental tampoco la puedo examinar en detalle, pues mi objetivo es, modestamente, proponer una justificación explicativa del razonamiento inductivo.

manera causal, o de cosas que permanecen a pesar del cambio de sus accidentes, si no poseyéramos estas y otras reglas de organización de la multiplicidad de nuestras sensaciones. Respecto de la causalidad, por ejemplo, Kant dice lo siguiente:

Therefore it is only because we subject the sequence of appearances and thus all alteration to the law of causality that experience itself, *i.e.*, empirical cognition of them, is possible; consequently they themselves, as objects of experience, are possible only in accordance with this law. (KrV: B 234)

Según Kant, una aplicación inmediata de los principios organizadores —como la causalidad— a la generación del contenido de nuestros juicios sintéticos empíricos nos la ofrece la física newtoniana, donde el principio de causalidad funciona como una regla para buscar causas del movimiento de un cuerpo material; Kant denomina esta aplicación del principio de causalidad en la mecánica: el principio de inercia. Esto es lo que se puede apreciar en el siguiente pasaje de los *Principios metafísicos de la ciencia de la naturaleza*:

Every change in matter has an external cause [...] Matter as mere object of the external senses has no other determinations than those of external relations in space and hence undergoes no changes except by motion. With regard to such change, insofar as it is an exchange of one motion with another, or of motion with rest, and *vice versa*, a cause of such change must be found [...] But this cause cannot be internal, for matter has no absolutely internal determinations and grounds of determination. Hence, all change of a matter is based upon an external cause. (Kant, 1970: 104-105)

En tanto regla constitutiva del contenido objetivo de nuestro conocimiento científico, la causalidad es considerada acertadamente por Kant como un principio de la razón teórica en la medida en que hace posible la existencia del contenido causal de los juicios sintéticos de experiencia. La discusión de su justificación está en la sección de la primera *Crítica* llamada “Análítica Trascendental” que a su vez es parte de la “Lógica Trascendental”. La justificación kantiana de la causalidad es netamente circular, aunque no viciosa, pues tiene la forma de una inferencia a la mejor expli-

cación del hecho, según Kant universalmente aceptado, de que nuestros juicios empíricos son objetivos. La objetividad de estos juicios se explica a su vez en términos de la posibilidad de verdad y falsedad del contenido de ellos y estos contenidos se constituyen a partir de la operación de los principios organizadores de la multiplicidad sensible como la causalidad.

La justificación explicativa de la aplicación de tales principios para constituir el contenido de los juicios sintéticos *a posteriori*, tanto de la ciencia de la naturaleza, como del conocimiento cotidiano, toma la forma, según Kant, de una argumentación trascendental, es decir: se conciben estos principios como condiciones de posibilidad de la experiencia, o como se dijo antes, condiciones de posibilidad del contenido objetivo de todos los juicios empíricos. La circularidad de la justificación explicativa kantiana de la causalidad estriba en que la experiencia requiere de causalidad para su constitución, mientras que la causalidad requiere de la experiencia para su correcta aplicación. La virtud de la circularidad de esta argumentación consiste en que la justificación kantiana es más bien explicativa. Quien acepta tal explicación está convencido de la corrección del empleo de la causalidad al mundo de la experiencia; lo que busca es simplemente una respuesta filosófica al por qué de dicha corrección.

Vamos a ilustrar el argumento justificador kantiano para la causalidad por medio de un ejemplo que aparece en los *Prolegómenos a toda metafísica futura*. Tomemos el juicio de percepción —en la terminología kantiana, un juicio cuyos elementos de contenido se relacionan sólo de manera subjetiva (Kant, 1997: § 18)— con el siguiente contenido *si el sol brilla sobre la piedra ella se calienta*. Para que tal contenido como un todo se pueda concebir como objetivo y, por lo tanto, pueda ser verdadero o falso, es necesaria la intervención de la regla de la causalidad que lo transforma en el siguiente contenido: *la iluminación del sol es la causa del calentamiento de la piedra*. Así lo expresa Kant:

To have a more easily understood example, consider the following: If the sun shines on the stone, it becomes warm. This judgement is a mere judgement of perception and contains no necessity, however often I and others also have perceived this; the perceptions are only usually found so conjoined. But if I say: the sun *warms* the stone, then beyond the perception is added the understanding's concept of cause, which connects *necessarily* the concept of



sunshine with that of heat, and the synthetic judgement becomes universally valid, hence objective, and changes from a perception into experience. (Kant, 1997: § 20, nota \*\*)

En este artículo no me interesa profundizar demasiado en la terminología kantiana, por ejemplo, acerca de la conexión necesaria entre conceptos o la validez universal de juicios. Me basta que se me conceda la interpretación según la cual un juicio de percepción, si bien puede incluir contenidos conceptuales, no tiene todavía un contenido proposicional objetivo —esto es, que sea verdadero o falso—. El punto de Kant es que se requiere de principios organizadores como la causalidad y la permanencia de la sustancia para constituir tal contenido proposicional. Sin duda, hay que reconocer el enorme esfuerzo kantiano para mostrar, apelando a la forma lógica de cualquier juicio, que estos principios operan confirmando contenido a todos los juicios con una cierta forma lógica. Por ejemplo, el principio de causalidad daría un contenido causal a juicios de la forma hipotética y el de permanencia de la sustancia otorgaría el contenido de la relación sustancia-accidente a los juicios de la forma categórica —esto es, de la forma sujeto-predicado.

Sin embargo, es muy probable que hoy en día nadie acepte que la totalidad de los principios del entendimiento propuestos por Kant como los responsables del contenido objetivo de nuestros juicios empíricos lo sean realmente, aunque algunos de ellos, como la causalidad, se puedan reconocer todavía como principios constitutivos de nuestra cognición e interacción con el mundo. De manera más específica, se rechazaría la tesis de que la causalidad pudiera fungir como principio semántico para todo un conjunto de proposiciones empíricas de una determinada forma. La idea de un principio semántico, que aquí estaría en juego, es la que se conoce desde Alfred Tarski; en este sentido, los principios semánticos determinarían las condiciones de verdad y falsedad de todas las proposiciones de este conjunto.<sup>20</sup> Por otro lado, también se rechazaría la idea de que

---

<sup>20</sup> En una semántica tarskiana un teorema semántico expresaría la condición de verdad de una oración del lenguaje para la cual se está explicitando su semántica; en esta misma, un axioma daría, por ejemplo, las condiciones de satisfacción de los predicados primitivos

un principio semántico deba operar sobre un contenido ya dado (una relación meramente subjetiva entre los contenidos ya establecidos) para generar un contenido proposicional objetivo. Esto significaría concebir la teoría del contenido de nuestras representaciones —conceptos y juicios— como descansando en algún tipo de contenido más primitivo: una de las maneras, como diría Wilfrid Sellars (1956), de caer en el mito de lo dado.

A pesar de todas estas críticas y muchas otras que se han dirigido a la justificación explicativa kantiana de la causalidad, pienso que se podría aprovechar de la siguiente manera. Podríamos afirmar que la regla de la causalidad y formulaciones adecuadas de las reglas inductivas son principios constitutivos de la cognición y acción humanas, en tanto que son indispensables para la construcción de una teoría de la mentalidad y del contenido lingüístico de un hablante, en una situación radical en que tenemos a nuestra disposición únicamente evidencia en términos de las conductas lingüísticas y no-lingüísticas del hablante —en resumen, la situación de interpretación radical—. Esto lo han sugerido Willard Van Orman Quine<sup>21</sup> y de manera más clara algunos de sus discípulos —por ejemplo, Donald Davidson—; ellos han afirmado que sin tales principios no podríamos atribuir contenido a las palabras de nuestros semejantes y tampoco motivación a sus conductas intencionales.<sup>22</sup> En otras palabras, no podríamos interpretarlos como hablando con sentido y actuando de modo intencional a la luz de sus deseos y creencias. Lo dice Davidson en los siguientes pasajes:

The question whether a creature “subscribes” to the principle of continence, or to the logic of the sentential calculus, or to the principle of total evidence for inductive reasoning, is not an empirical question. For it is only by interpreting

---

del lenguaje. La causalidad no podría fungir como principio semántico ni en el sentido de un teorema de la teoría semántica ni tampoco en el sentido de un axioma de la misma (*cfr.*, Tarski, 1933: seccs. 2, 3 y 4).

<sup>21</sup> Quine lo dice por primera vez en los capítulos iniciales de su monumental obra *Palabra y objeto*, (1960).

<sup>22</sup> Davidson, por ejemplo, lo dice en varios textos. Uno de ellos es Davidson, 1985. A continuación, se mencionarán otros textos de él en que se afirma lo mismo.

a creature as largely in accord with these principles that we can intelligibly attribute propositional attitudes to it. (Davidson, 1985: 352)

I should never have tried to pin you down to an admission that you ought to subscribe to the principles of decision theory. For I think everyone does subscribe to those principles, whether he knows it or not. This does not imply, of course, that no one ever reasons, believes, chooses, or acts contrary to those principles, but only that if someone does go against those principles, he goes against his own principles. (Davidson, 1985: 351)

No me detendré a discutir muchos de los principios que Davidson considera constitutivos de una teoría de la interpretación radical como, por ejemplo, los mencionados en la cita anterior, a saber, los principios de continencia y de evidencia total.<sup>23</sup> Es importante insistir en que las reglas inductivas de la teoría bayesiana están claramente incluidas entre estos principios. También sería conveniente decir algo en torno a la afirmación, muy de moda hoy día, de que nosotros de manera sistemática violamos los principios de la teoría de la decisión cuando actuamos a la luz de deseos y creencias, y que, por lo tanto, dicha teoría es falsa; esta afirmación entra en conflicto con lo que defiende Davidson en la anterior cita. El espacio de este ensayo no conforma una discusión detallada y abarcadora de la verdad o falsedad empírica ni del carácter *a priori* o *a posteriori* de tal teoría. Sin embargo, la cuestión de la posibilidad de someter los principios de la teoría bayesiana al tribunal de la experiencia es demasiado relevante para que la dejemos de lado en la presente discusión.

Tomemos como ilustración el principio de la transitividad de la noción de preferencia: si el agente prefiere la apuesta *A* a la *B* y también la *B* a la

---

<sup>23</sup>El principio de continencia sería equivalente al principio de la maximización de la utilidad esperada. Alguien que violara este principio estaría actuando en contra de su mejor juicio —éste sería un caso claro de debilidad de la voluntad (*akrasia*)—. El principio de evidencia total ordenaría al agente a aceptar la hipótesis que esté corroborada por la totalidad de evidencia a su disposición. Se podría interpretar este principio como la formulación de sentido común de los principios de condicionalización bayesianos (RCB, RCB', RCJ y RCJ').

C, y además es consistente, debe preferir la primera a la tercera. Ahora bien, varios experimentos psicológicos cuidadosos han sido propuestos para mostrar que los agentes sistemáticamente violan este principio.<sup>24</sup> Confrontado entonces con experimentos tales como los llevados a cabo por Amos Tversky, ¿estaría Davidson obligado a conceder que la teoría bayesiana ha sido refutada o que alguno de sus principios es falso y por lo tanto debe ser eliminado? De ninguna manera, una vez que, según él, todavía están abiertas las siguientes posibilidades al intérprete: 1) reinterpretar los resultados experimentales de manera que se pueda eliminar la aparente intransitividad; 2) proponer que los agentes siempre cambiaban de idea.<sup>25</sup> La razón por la cual los principios básicos de la teoría bayesiana no pueden ser refutados por la experiencia es porque son parcialmente constitutivos de los conceptos de creencia, deseo y preferencia; sin estos principios sería imposible discernir dichos elementos fundamentales de la mentalidad humana. En todo caso, se podría intentar rechazar la teoría bayesiana como un todo. Pero, si Davidson tiene razón, no hay una teoría alternativa que permita igualmente identificar los patrones de creencias y de deseos de un agente en relación causal y racional con su acción intencional.<sup>26</sup>

Por otro lado, es necesario enfatizar que, según Davidson, no hay una lista fija de principios constitutivos del significado y en general de la mentalidad humana. De lo cual se desprende, sin embargo, que el método interpretativo considerado como mejor explicación para la constitución y atribución de contenido, y para la motivación de las acciones de nuestros

---

<sup>24</sup> Uno de los más conocidos está reportado en Tversky, 1969.

<sup>25</sup> Estas dos posibilidades son mencionadas en Davidson, 1974 y 1976.

<sup>26</sup> Insisto en que habría mucho que decir acerca de su carácter empírico o *a priori*. Mi intención aquí es simplemente llamar la atención respecto de la afirmación davidsoniana de que los principios de la teoría de la decisión son constitutivos de los conceptos mentales. Esto de ninguna manera implica que la teoría psicológica como un todo sea *a priori*, más bien, según lo entiendo, Davidson propone que tales principios definen implícitamente los conceptos mentales en el contexto de la teoría bayesiana, la cual sólo en conjunción con la psicología popular se torna una teoría empírica (*cf.*, Davidson, 1976: 272-274 y 1970: 220-221).

semejantes requiere de ciertos principios, como el de la coherencia del patrón de creencias de un agente/hablante, el principio de caridad (la máxima que nos ordena atribuir a nuestros semejantes un patrón de creencias verdaderas lo más parecido posible con el nuestro), el antes mencionado principio de maximización de la utilidad esperada y otros.

Para empezar, quizá sea indiscutible que ciertos principios lógicos como la no-contradicción deberían entrar necesariamente en esta lista. Si no fuera así, ¿cómo podríamos tener éxito en la atribución de valores de verdad a las oraciones de nuestros interlocutores? Al respecto dice Quine:

Or consider the familiar remark that even the most audacious system-builder is bound by the law of contradiction. How is he really bound? If he were to accept contradiction, he would so readjust his logical laws as to insure distinctions of some sort; for the classical laws yield all sentences as consequences of any contradiction. But then we would proceed to reconstrue his heroic novel logic as a non-contradictory logic, perhaps even as familiar logic, in perverse notation. (Quine, 1960: 59)

Las reglas de la lógica deductiva jugarían el papel de constitutivas de las relaciones semánticas entre los contenidos proposicionales de las creencias o de las oraciones del agente-hablante. Pero también principios que no pertenecen a la lógica deductiva deben entrar en la lista de las reglas constitutivas de una teoría interpretativa de la mente y de la acción. Davidson, por ejemplo, ha sugerido que la teoría de la interpretación radical debe incluir una teoría del significado de los enunciados declarativos del lenguaje de los interpretados en términos de sus condiciones de verdad y una teoría del patrón de deseos y creencias de los mismos en los moldes de la teoría bayesiana de la decisión descubierta por Frank Ramsey y elaborada por Richard Jeffrey. La teoría de la verdad daría estructura a los contenidos proposicionales de las oraciones y actitudes del interpretado, mientras que la teoría de la decisión sistematizaría las intensidades de deseos y creencias del mismo.

Un principio constitutivo de esta segunda teoría debe ser una de las reglas bayesianas de condicionalización,<sup>27</sup> la cual permite al intérprete actualizar las probabilidades subjetivas de todas las oraciones no-observacionales como consecuencia de los cambios de las probabilidades subjetivas asociadas a las oraciones de observación (relacionadas lógicamente con las primeras como evidencia e hipótesis) que él identifica en la conducta lingüística del interpretado. La interpretación de las oraciones observacionales del interpretado procede mediante la observación de los cambios de las probabilidades subjetivas asociadas a tales proposiciones, como resultado de las situaciones en el mundo que el intérprete supone que están causando esos cambios. Pero, la empresa interpretativa empieza por la identificación de las constantes lógicas en el lenguaje del interpretado. Veamos de manera resumida cómo describe Davidson esta empresa<sup>28</sup> y qué papel constitutivo le atribuye, por un lado, a las reglas de condicionalización bayesianas y, por otro, al principio de causalidad en este proceso:

Jeffrey's version of decision theory, applied to sentences, tells us that a rational agent cannot prefer both a sentence and its negation to a tautology, or a tautology to both a sentence and its negation. This fact makes it possible for an interpreter to identify, with no knowledge of the meanings of the agent's sentences, all of the pure sentential connectives, such as negation, conjunction and the biconditional.<sup>29</sup> This minimal knowledge suffices to determine the subjective probabilities of all of the agent's sentences [...] and then, in turn, to fix the relative values of the truth of those sentences [...] The subjective probabilities can then be used to interpret the sentences. For what Quine calls observation sentences, the changes in probabilities provide the obvious clues to first-order interpretation when geared to events and objects easily perceived

---

<sup>27</sup> RCB o su refinamiento: la mencionada regla de condicionalización de Jeffrey (RCJ). Se podría también incluir sus respectivas versiones sin la condición de rigidez o invariancia.

<sup>28</sup> Descripciones más detalladas del tema se encuentran en Davidson, 1980b y 1984.

<sup>29</sup> Una descripción más detallada del proceso de identificación de las constantes lógicas utilizando el aparato de la teoría de Jeffrey nos lo ofrece Davidson (1990) en el apéndice de su artículo en torno a la verdad.

simultaneously by interpreter and the person being interpreted. Conditional probabilities and entailments between sentences, by registering what the speaker takes to be evidence for his or her beliefs, provides the interpreter with what is needed to interpret more theoretical terms and sentences. (Davidson, 1995: 9-10)

La regla de condicionalización es indispensable para el método interpretativo antes descrito, en la medida que hace posible la interpretación de las oraciones no-observacionales y de los términos teóricos del lenguaje a ser interpretado. Por otro lado, el principio de causalidad es pieza indispensable para la interpretación de las oraciones observacionales de tal lenguaje. De acuerdo con Davidson, para el caso de esas oraciones, el intérprete debe observar cuáles son los sucesos y objetos fácilmente percibidos simultáneamente por él y por el interpretado que, según el primero, deben estar causando los cambios en las probabilidades subjetivas asociadas con dichas oraciones observacionales.

Nos queda la pregunta de por qué deberíamos aceptar esta teoría del contenido y de la mente como correctas. Su justificación es circular aunque no viciosa. Como vimos, su defensor la presenta como la mejor explicación filosófica de la posibilidad de atribución de contenido a nuestras palabras, como también de la posibilidad de una teorización referente a nuestra acción intencional. Nosotros ya estamos convencidos de manera independiente<sup>30</sup> de que hay contenido intencional y mentalidad. Sin embargo, en tanto filósofos buscamos una explicación que nos satisfaga respecto a la cuestión de cómo el contenido intencional y la mentalidad son posibles. Apelamos a los principios constitutivos de la mentalidad para mostrar de qué manera se podría generar una mentalidad de este tipo por medio de los principios en cuestión; por otro lado, apelamos al hecho mismo de la existencia de la mentalidad humana para señalar la posible aplicación de sus principios constitutivos. Como argumento explicativo, la justificación davidsoniana de los principios básicos de la teoría de la interpretación radical comparte con otros argumentos parecidos la caracterís-

---

<sup>30</sup> Si no estuviéramos convencidos de ello, no trataríamos a los otros como si tuvieran mentalidad y tampoco esperaríamos que los otros nos trataran como si la tuviéramos.

tica de que su dirección lógica —de los principios constitutivos de la teoría de la interpretación radical del discurso y de la acción como premisas al contenido intencional y a la mentalidad como conclusión— no coincide con su correspondiente dirección epistemológica, la cual parte del conocimiento de que hay mentalidad y contenido lingüístico para acceder al conocimiento de que los principios de constitución de tal mentalidad y contenido deben ser los que permiten construir una teoría de la interpretación radical de un agente-hablante.

## CONCLUSIÓN

Seguramente hay algunos otros candidatos a teorías del significado y de la acción, pero en el espacio de este ensayo no puedo examinarlos. Sin embargo, si lo que hemos discutido antes es correcto, la teoría de la interpretación radical davidsoniana tiene la ventaja de ofrecernos una justificación en el sentido explicativo para el equivalente contemporáneo de la causalidad e inducción humeanas: el principio de causalidad y las reglas de condicionalización bayesianas (RCB' o su refinamiento, la RCJ').

## BIBLIOGRAFÍA

- Davidson, Donald (1970), "Mental events", en Donald Davidson (1980a), pp. 207-227.
- Davidson, Donald (1974), "Psychology as philosophy", en Donald Davidson (1980a), pp. 229-239.
- Davidson, Donald (1976), "Hempel on explaining action", en Donald Davidson (1980a), pp. 261-275.
- Davidson, Donald (1980a), *Essays on Actions and Events*, Oxford, Estados Unidos, Oxford University Press.
- Davidson, Donald (1980b), "A unified theory of thought, meaning, and action", *Grazer Philosophische Studien*, vol. XI, pp. 1-12.
- Davidson, Donald (1984), "Expressing evaluations", en *Lindley Lecture monograph*, Kansas, Estados Unidos, University of Kansas, pp. 3-20.
- Davidson, Donald (1985), "Incoherence and irrationality", *Dialéctica*, vol. XXXIX, núm. 4, pp. 345-354.
- Davidson, Donald (1990), "The structure and content of truth", *The Journal of Philosophy*, vol. LXXXVII, núm. 6, pp. 279-328.



- Davidson, Donald (1995), "Could there be a science of rationality", *International Journal of Philosophical Studies*, vol. III, núm. 1, pp. 1-16.
- de Finetti, Bruno (1937), "La prevision: ces lois logiques, ces sources subjectives", *Annales de l'Institut Henri Poincaré*, vol. VII, núm. 1, pp. 1-68.
- Dummett, Michael (1973), "The justification of deduction", en Michael Dummett (1978), pp. 290-318.
- Dummett, Michael (1978), *Truth and Other Enigmas*, Londres, Reino Unido, Duckworth.
- Earman, John (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, Cambridge, Londres, Reino Unido, The MIT Press.
- Howson, Colin y Peter Urbach (1993), *Scientific Reasoning*, Chicago, Estados Unidos, Open Court.
- Hume, David (1777), *Essays and Treatises on Several Subjects*, vol. II, Londres, Reino Unido, T. Cadell.
- Hume, David (1975), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford, Reino Unido, Clarendon Press.
- Kant, Immanuel (1970), *Metaphysical Foundations of Natural Science*, Nueva York, Estados Unidos, The Bobbs-Merrill Company.
- Kant, Immanuel (1997), *Prolegomena to any Future Metaphysics*, Cambridge, Reino Unido, Cambridge University Press.
- Kant, Immanuel (1998), *Critique of Pure Reason* (edición B), Cambridge, Reino Unido, Cambridge University Press.
- Jeffrey, Richard (1983), *The Logic of Decision*, Chicago, Estados Unidos, The University of Chicago Press.
- Jeffrey, Richard (2002), *After Logical Empiricism*, Lisboa, Portugal, Colibri.
- Quine, Willard von Orman (1960), *Word and Object*, Cambridge, Estados Unidos, The MIT Press.
- Ramsey, Frank (1926), "Truth and probability", en David. Mellor (1990), *F. P. Ramsey Philosophical Papers*, Cambridge, Reino Unido, Cambridge University Press, pp. 52-94.
- Sellars, Wilfrid (1956), "Empiricism and the philosophy of mind", en Herbert Feigl y Michel Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. I: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minneapolis, Estados Unidos, University of Minnesota Press, pp. 253-329.
- Skyrms, Brian (1986), *Choice and Chance. An Introduction to Inductive Logic*, Belmont, Estados Unidos, Wadsworth Publishing Company.
- Skyrms, Brian (1987), "Dynamic coherence and probability kinematics", *Philosophy of Science*, vol. LIV, núm. 1, pp. 1-20.
- Tarski, Alfred (1933), *Pojecie prawdy w jezykach nauk dedukcyjnych*, Varsovia, Polonia, Warsaw Academy of Sciences.
- Tarski, Alfred (1956), "The concept of truth in formalised languages", en Alfred Tarski (1956), pp. 152-278.

Tarski, Alfred (1956), *Logic, Semantics, Metamathematics*, Oxford, Reino Unido, Oxford University Press.

Teller, Paul (1973), "Conditionalization and observation", *Synthese*, vol. XXVI, pp. 218-258.

Tversky, Amos (1969), "Intransitivity of preferences", *Psychological Review*, vol. LXXVI, núm. 1, pp. 31-48.

**Sílvio Pinto:** es profesor titular del Departamento de Filosofía de la Universidad Autónoma Metropolitana-Iztapalapa, desde 2002. Ha trabajado en la Universidad Autónoma del Estado de Morelos como profesor titular y en Instituto de Investigaciones Filosóficas-UNAM como investigador posdoctoral. Obtuvo su doctorado en 1998, en el King's College de la Universidad de Londres. Sus principales intereses de investigación giran en torno a la metafísica, epistemología, filosofía del lenguaje, de la mente y de las matemáticas.

D. R. © Sílvio Pinto, México, D.F., julio-diciembre, 2007.