

# EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS V JORNADAS

1995

Alberto Moreno

Editor



ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA  
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES  
UNIVERSIDAD NACIONAL DE CÓRDOBA



Esta obra está bajo una Licencia Creative Commons atribución NoComercial-SinDerivadas 2.5 Argentina



## CAUSACION DE LEWIS: UN EJEMPLO CASI FATAL†

### INTRODUCCION

El presente trabajo intenta examinar algunos aspectos de la teoría de la causación propuesta por David Lewis a partir de una crítica de Eduardo Flichman retomada por Dorothy Edgington.

### HISTORIA SUMARIA DEL PROBLEMA

D. Lewis<sup>1</sup> retoma uno de los análisis de Hume de la causación: "un objeto seguido de otro...donde, si el primero no hubiera sido, el segundo no habría existido"; y lo reformula a la luz de las teorías de los condicionales contrafácticos disponibles<sup>2</sup> como base de una teoría contrafáctica de la causación.

Es preciso hacer notar desde un comienzo que la teoría que Lewis desarrolla sólo intenta elucidar el concepto de causación *entre eventos singulares que efectivamente ocurren*<sup>3</sup>.

En lo que sigue indicaremos los eventos con letras minúsculas en bastardilla: *a, b, c*, etc. y los enunciados que afirman que estos eventos ocurren con las mayúsculas correspondientes: *A, B, C*, etc. La teoría de Lewis puede sintetizarse en unos pocos puntos de la siguiente manera:

---

†Este trabajo fue realizado en el marco del equipo que dirige E. H. Flichman y codirige el suscrito en la cátedra de Introducción al Pensamiento Científico a cargo del primero, Ciclo Básico Común, Universidad de Buenos Aires. El autor agradece a Eduardo Flichman, Jorge Paruelo y Hernán Miguel las sugerencias aportadas.

<sup>1</sup> D. Lewis, 'Causation', *Journal of Philosophy*, 70, 1973, 556-67 (reproducido en *Philosophical Papers*, vol. II, Oxford University Press, 1986).

<sup>2</sup> Lewis cita en especial la suya propia [D. Lewis, *Counterfactuals*, Blackwell, (1973)], y la de Stalnaker [R. Stalnaker, 'A Theory of Conditionals' en N. Rescher, ed., *Studies in Logical Theory* (Oxford, 1968)]. Es de notar que, si bien da a entender que su análisis de la causación es independiente de cuál teoría de contrafácticos se prefiera, y por lo tanto podría apoyarse en cualquiera de ellas, en el presente trabajo se verá que hay dificultades que no parecen tener solución si se usa la teoría de Stalnaker.

<sup>3</sup> Para satisfacer con más precisión las necesidades de esta teoría, Lewis elaborará luego su propia concepción de la noción de 'evento' en 'Events', (1986). (en *Philosophical Papers*, Vol. II, Oxford University Press, 1986)

a) Diremos que  $b$  depende causalmente de  $a$  cuando es verdadero el contrafáctico 'si  $a$  no hubiera ocurrido,  $b$  no habría ocurrido'; simbólicamente:

$$\neg A \square \rightarrow \neg B$$

b)  $a$  es una causa de  $b$  si  $b$  depende causalmente de  $a$ .

c) si  $a$  es una causa de  $b$ , y  $b$  es una causa de  $c$ , entonces  $a$  es una causa de  $c$ .

Esta teoría debería mostrar su capacidad para resolver las numerosas dificultades que se han encontrado a otros análisis de la causación. Entre ellos, la no distinción entre causas y efectos (llamado en la literatura problema de los efectos), y otros como el problema de los epifenómenos, el de la sobredeterminación, etc. Además deberá ser lo menos vulnerable posible a los contraejemplos: situaciones en que nadie querría admitir que  $a$  es una causa de  $b$  y para la teoría resulte serlo, o viceversa. Cada contraejemplo tiende a sugerir que no estamos elucidando realmente el concepto que pretendíamos.

Para mostrar que su teoría evita el problema de los efectos, Lewis<sup>4</sup> analiza el ejemplo que estudiaremos en la siguiente sección. Flichman<sup>5</sup> muestra que el propio análisis de Lewis de ese ejemplo lo convierte en un contraejemplo para la teoría. De ésta y otras críticas al análisis de Lewis (y a los de otros autores), Flichman infiere la no analizabilidad del concepto de causación, y se suma a quienes lo desechan (al menos en el ámbito de las ciencias naturales) por considerarlo esencialmente antropomórfico. Dorothy Edgington<sup>6</sup> se suma con entusiasmo a la crítica de Flichman y adhiere a la idea de la no analizabilidad del concepto de causación, aunque en su particular visión éste es un concepto que no deberíamos abandonar pero al mismo tiempo es "demasiado fundamental para que pueda ser analizado en términos que no lo presupongan".

Aquí intentaremos establecer en qué medida puede hacerse un análisis de este ejemplo que evite las críticas de Flichman y salve la teoría de Lewis. Veremos que ello puede hacerse aunque con algunos costos de importancia.

## ¿UN EJEMPLO FATAL?

Presentamos ahora una versión del ejemplo cuestionado: Se analiza el comportamiento de cierto barómetro. Llamaremos  $p$  al evento de que la presión atmosférica es de 1000 milibares;  $i$  al evento de que el barómetro indica 1000 mb; y  $b$  al evento de que el barómetro está funcionando bien. Por lo tanto los enunciados de ocurrencia de estos eventos serán:

$P$  = 'La presión es de 1000 mb'.

$B$  = 'El barómetro funciona bien'.

$I$  = 'El barómetro indica 1000 mb'

En el ejemplo, todas estas afirmaciones son verdaderas: Dicho de otro modo, en el mundo efectivo ocurren los eventos  $p$ ,  $b$ , e  $i$ . Dado que sabemos que la presión es de 1000

---

<sup>4</sup> 'Causation' (ver nota 1), fin de la sección 'Counterfactual versus Nomic Dependence'

<sup>5</sup> E. H. Flichman, 'The Causalist Program, Rational or Irrational Persistence?', *Crítica*, XXI, No. 62 (1989), 29-53.

<sup>6</sup> D. Edgington, 'Explanation, Causation and Laws', *Crítica*, XXII, No. 66 (1990), 55-73.

mb, y que el barómetro, que funciona bien, indica 1000 mb, diríamos que  $p$  es una causa de  $i$ : que la presión sea de 1000 mb es, obviamente, causa de que un barómetro en buenas condiciones indique una presión de 1000 mb. La teoría de Lewis avala este resultado: puesto que  $p$  e  $i$  son sucesos ocurrentes,  $p$  es causa de  $i$  por ser verdadero el condicional contrafáctico

$$\neg P \square \rightarrow \neg I \quad (1)$$

es decir:

*'si la presión no hubiera sido de 1000 mb, el barómetro no habría indicado 1000 mb'* (1.a)

Aquí debe observarse que en el contexto normal de enunciación de un condicional tal como éste se admitiría que, en vista de que el barómetro funciona bien, no tiene sentido pensar en situaciones (mundos posibles) en que ello no ocurra, dado que además la suposición contrafáctica acerca de la presión no parece requerirlo. Un contrafáctico como 'Si la presión no hubiera sido de 1000 mb, el barómetro habría estado roto y habría marcado 1000 mb de todos modos', sería rechazado sin lugar a dudas.

Hasta aquí la teoría de Lewis parece marchar sobre ruedas y da resultados con los que la intuición concuerda. Pero Lewis debe mostrar que la teoría es útil para dilucidar problemas que otras teorías dejaban sin aclarar. En este caso se trata del así llamado *problema de los efectos*: ¿Se puede distinguir efecto de causa? O, dicho de otro modo, ¿hemos encontrado un análisis *asimétrico* de la causación? Lewis asegura que sí, y es para probarlo que introduce este ejemplo. Si todo anda bien, la teoría no debe avalar la afirmación simétrica de (1):

$$\neg I \square \rightarrow \neg P \quad (2)$$

es decir:

*'si el barómetro no hubiera indicado 1000 mb, la presión no habría sido de 1000 mb'* (2.a)

Lewis asegura que esta afirmación debe considerarse falsa, basándose en consideraciones acerca del contexto o "resolución de vaguedad" que él considera que debe aplicarse para mantener la irreversibilidad<sup>7</sup> -lo cual parece un criterio ad-hoc en extremo- y argumentando que lo que deberíamos considerar, es que *si el barómetro no hubiera indicado 1000 mb, estaría funcionando mal*, (ya que la presión es de 1000 mb).

Este es un argumento sumamente discutible, y ha sido efectivamente objetado<sup>8</sup>. Pero es al mismo tiempo un notable traspie de Lewis, ya que parece haber hecho a la teoría

<sup>7</sup> 'Causation', nota 10.

<sup>8</sup> Así, D. Edgington sugiere que el llamado *efecto mariposa* (la posibilidad de que el aleteo de una mariposa en un lugar del mundo puede resultar en alteraciones de los sistemas de tormentas en otro

susceptible a críticas más serias que aquellas que aquéllas que aquéllas que evitar, ya que da lugar a la crítica de Flichman que aquí examinamos, y que señala lo siguiente: de aceptarse la afirmación recién citada de Lewis, habría que dar por verdadera la siguiente afirmación:

$$\neg I \square \rightarrow \neg B \quad (3)$$

es decir:

*'si el barómetro no hubiera indicado 1000 mb, no habría estado funcionando bien'*

(3.a)

)

y resultaría de acuerdo a la teoría de Lewis, que el hecho de que el barómetro indique 1000 mb es una causa de que funcione bien, contrariamente a toda intuición.

Por ende es la propia evaluación de los contrafácticos en juego hecha por Lewis la que convierte al ejemplo en un contraejemplo de su teoría.

En resumen, la crítica de Flichman (que a su vez Dorothy Edgington considera lapidaria) sugiere que Lewis se ha puesto por sí solo entre la espada y la pared<sup>9</sup>: si avala (2) no puede evitar el problema de los efectos, y si avala (3) se ve obligado a aceptar causas totalmente contrarias a la intuición.

### ¿HASTA QUE PUNTO QUEDA INDEFENSA LA TEORIA DE LEWIS?

Deseamos analizar aquí si hay algún modo de salvar la teoría de Lewis de estas críticas: ¿podría cambiarse la evaluación de los contrafácticos arriba mencionados, contra la opinión por él mismo expresada, de forma no caer en ninguno de los riesgos mencionados?

Para que ello sea posible es preciso que podamos encontrar un criterio que nos lleve a evaluar como falsos tanto a (2) como a (3).

Aquí se vuelve importante la cuestión de qué teoría de condicionales contrafácticos decidamos adoptar. Tomemos por ejemplo la teoría de contrafácticos debida a Stalnaker. Esta teoría como algunas otras, asigna validez al así llamado principio del tercero excluido condicional para los contrafácticos.

$$A \square \rightarrow B \vee A \square \rightarrow \neg B \quad (4)$$

que en consecuencia permitirá inferir la verdad de  $A \square \rightarrow \neg B$ , cada vez que sea falso  $A \square \rightarrow B$ .

---

lugar) podría obligarnos a invertir el argumento: el estado del tiempo es muy inestable, mientras que los barómetros pueden hacerse con la solidez que se desee.

<sup>9</sup> En realidad Flichman se limita a mostrar que las consecuencias para la teoría de Lewis de aceptar la verdad de (3) son tan inaceptables como las de aceptar (2). No está sugerida en su artículo la posibilidad de algún análisis que no valide ni (2) ni (3).

En nuestro caso resultará que si debemos rechazar por falso a  $\neg I \square \rightarrow \neg P$  (para evitar el problema de los efectos), en la teoría de Stalnaker deberemos considerar verdadero a  $\neg I \square \rightarrow P$ . Esto es: 'Si el barómetro no hubiera indicado 1000 mb, igualmente la presión habría sido de 1000 mb'. En ese caso es inevitable considerar que el barómetro no habría estado funcionando bien, cayéndose por lo tanto en avalar (3). Por lo tanto, usando la teoría de Stalnaker o cualquier otra que contenga al tercero excluido condicional, la opinión de D. Edgington es sólida: no sólo Lewis es susceptible a la crítica de Flichman en la medida en que mantenga su desafortunada evaluación del ejemplo mencionado, sino que no hay evaluación posible que evite al mismo tiempo los dos problemas señalados.

Ahora bien, la teoría de contrafácticos del propio Lewis no avala el tercero excluido condicional. Lewis evalúa los contrafácticos de acuerdo a una semántica de mundos posibles, según la cual el condicional 'si fuera  $A$  sería  $C$ ' es verdadero cuando en los mundos más parecidos al mundo efectivo en que se cumple  $A$  también se cumple  $C$ <sup>10</sup>. Una razón por la que no vale el tercero excluido condicional es que podría haber "empates": a veces hay mundos igualmente próximos al mundo efectivo, en algunos de los cuales se cumple  $C$  y en otros  $\neg C$ . En esos casos no vale ninguno de los dos condicionales arriba mencionados; en cambio se puede enunciar un así llamado condicional tipo "might"<sup>11</sup>: 'Si fuera el caso que  $A$ , podría ser el caso que  $B$ ' (y podría ser el caso que  $\neg B$ ).

Podemos intentar entonces una salida: hagamos una evaluación de la similaridad de mundos posibles que nos permita negar (2) y (3) y por lo tanto afirmar:

*'Si el barómetro no indicase 1000 mb, podría ser el caso que la presión atmosférica no fuese de 1000 mb; y podría ser el caso que el barómetro estuviese funcionando mal'.* (5)

Es decir, que entre los mundos en que el barómetro no indica 1000 mb, haya empate en similaridad entre aquellos en que la presión atmosférica es otra, y aquellos en los que el barómetro está funcionando mal. La situación parece ser entonces la siguiente: Consideremos todos los mundos posibles (al menos lo bastante cercanos como para que exista en ellos nuestro barómetro, etc.). Estos mundos se pueden dividir en ocho grupos según que sea en ellos verdadero o falso cada uno de los enunciados  $P$ ,  $B$ , e  $I$  (a la manera de las tablas de verdad--ver cuadro)

<sup>10</sup> Por brevedad presento aquí una versión intuitiva e imprecisa de la formulación de Lewis.

<sup>11</sup> Condicionales que expresan "posibilidad" contrafáctica; distinguir de los condicionales "would" que expresan "necesidad" contrafáctica.

Los mundos que nos interesa examinar cuando hacemos la suposición contrafáctica 'si el barómetro no hubiera indicado 1000 mb...' son aquellos en que *I* es falso. Es decir los de los grupos [2], [4], [6] y [8]. El mundo efectivo pertenece al grupo [1]. Los mundos del grupo [2] son mundos imposibles: no es digna de ser tomada en cuenta la posibilidad de que el barómetro funcione bien y al mismo tiempo indique una presión diferente a la que efectivamente hay. Los mundos del grupo [8] son seguramente menos semejantes al efectivo que los de los grupos [4] y [6] ya que en ese grupo hay simultáneamente dos alteraciones respecto del mundo efectivo que en [4] y [6] aparecen de a una: que la presión atmosférica sea diferente y que el barómetro esté descompuesto, sin que ello sea necesario para lograr que el barómetro indique otro valor. Nuestro problema se reduce entonces a comparar los mundos más cercanos de los grupos [4] y [6]. Si los de [4] son más cercanos que los de [6], valdrá (3); si es a la inversa valdrá (2). Si ninguno de los dos grupos contiene mundos más cercanos que todos los del otro, no vale ni (2) ni (3) y vale (5).

Grupo	P	B	I
[1]	V	V	V
[2]	V	V	F
[3]	V	F	V
[4]	V	F	F
[5]	F	V	V
[6]	F	V	F
[7]	F	F	V
[8]	F	F	F

Esta operación salvaría a la teoría de Lewis de las dificultades específicas señaladas anteriormente. Queda ésta sin embargo malparada en virtud de varios flancos débiles que esta solución presenta:

a) Habría que encontrar argumentos que justifiquen para este ejemplo el mencionado ordenamiento de mundos. ¿Cómo determinar claramente que los mundos más cercanos del tipo [4] no son ni un poco más ni un poco menos lejanos que los del tipo [6]?

b) Si se encuentra una solución al caso del ejemplo, ¿por qué debemos pensar que lo mismo ocurrirá para cada ejemplo análogo que se pueda presentar? Parecía cuando Lewis publicó su *Counterfactuals*, que tendríamos criterios razonables e intuitivos para ordenar los mundos posibles, y ello nos daría una base para evaluar contrafácticos. Resultaría ahora que es forzoso aceptar ciertos ordenamientos muy precisos para evitar que los contrafácticos se nos insubordinen. Tenemos que afirmar que ambos grupos de mundos son igualmente cercanos, pero no por observar propiedades del mundo sino porque no nos queda otro remedio.

c) Si recordamos que el ordenamiento de los mundos depende del contexto de enunciación de cada contrafáctico, es de presumir que si hay empate la situación queda tan inestable que cualquier mínimo cambio de contexto nos podrá volcar hacia [2] o hacia [3]. En tal caso deberemos admitir que hay muchas instancias en que la relación de causación es contexto-dependiente. Malas noticias para quienes, como Lewis, habrían deseado detectar dicha relación entendida como un atributo de la realidad y no como una característica del lenguaje.

## LAS OBJECIONES DE DOROTHY EDGINGTON

Escrito esto tropiezo con un trabajo inédito de Dorothy Edgington<sup>12</sup> donde se manejan ideas muy similares a las aquí expuestas; señalo a continuación algunas semejanzas y diferencias.

i) Dorothy Edgington sugiere que "si el barómetro no hubiera indicado esa presión, podría haber estado roto o podría haber sido distinta la presión atmosférica"; no está claro si la frase representa un contrafáctico del tipo "might" con consecuente disyuntivo, o, como considero más probable, está dicha con la imprecisión del lenguaje natural, lo que permitiría dudar sobre la interpretación formal que se le puede dar<sup>13</sup>. Simbolizando, como hace Lewis, a un condicional del tipo 'si fuera el caso que A, podría ser el caso que B' -posibilidad contrafáctica-como

$$A \Diamond \rightarrow B$$

podría pensarse en las siguientes formalizaciones para la frase de Edgington:

- a)  $(\neg I \Diamond \rightarrow \neg B) \vee (\neg I \Diamond \rightarrow \neg P)$
- b)  $\neg I \Diamond \rightarrow (\neg B \vee \neg P)$
- c)  $(\neg I \Diamond \rightarrow \neg B) \& (\neg I \Diamond \rightarrow \neg P)$

Ahora bien, suponer en este ejemplo que el barómetro no indica 1000 mb conlleva que esté descompuesto o que la presión no tenga ese valor. De allí, tanto la opción a) como b) serían analíticas, y por ende han de valer, no importa si los mundos más similares son los que dice Lewis o no, es decir: cualquiera sea la postura que tomemos en cuanto a la verdad de (2) o de (3). Dicho de otro modo, es bastante inmediato inferir a) y b) tomando (2) o bien (3) como premisa<sup>3/4</sup>pero de lo que se trataba era de encontrar una alternativa que nos permitiera rechazar (2) y (3), que son las afirmaciones que han traído problemas.

Lo que sería pertinente entonces para salvar la teoría de Lewis sería validar la opción c), que es la formalización de nuestra (5). Entendida así, la propuesta de Edgington coincide con la maniobra sugerida en este trabajo.

ii) Edgington, por otra parte, sostiene que este tipo de solución sería plausible para nuestro ejemplo si diéramos como condición de verdad de los contrafácticos "might" a la siguiente:

*A  $\Diamond \rightarrow C$  es verdadero cuando hay mundos A & C que no son demasiado lejanos o rebuscados*

que desde ya no es la condición de verdad de Lewis para tales contrafácticos.

---

<sup>12</sup> D. Edgington, 'David Lewis, Counterfactual Dependence and Causation', inédito, 1990, 18 págs.

<sup>13</sup> Una comparación con la profusa literatura dedicada al problema de los aparentes contrafácticos con antecedente disyuntivo puede dar una idea de que en enunciados de este estilo la determinación de la forma lógica implícita no es trivial. Ver H. Abeledo, E. H. Flichman y H. Miguel, 'Contrafácticos y antecedentes disyuntivos: una cuestión de privilegio', III Jornadas de Epistemología e Historia de la Ciencia, U. N. de Córdoba, dic. 1992



Aceptar esta sugerencia, que a primera vista parece bastante intuitiva pues hace posibles unos empates de "grano grueso" sin requerir la precisión en la evaluación de semejanzas que exige la teoría de Lewis, traería sin embargo consigo algunos problemas:

a) se renuncia a la interdefinibilidad de los contrafácticos "might" y "would".

Recordemos que para Lewis es:

$$A \Diamond B \equiv_{df} \neg(A \Box \rightarrow \neg B)$$

o, alternativamente:

$$A \Box B \equiv_{df} \neg(A \Diamond \rightarrow \neg B)$$

Esto es bastante intuitivo en algunas situaciones, y por otra parte conserva un paralelismo deseable con el caso de los operadores de posibilidad y necesidad de la lógica modal.

b) se trataba de mostrar que son verdaderos los "might" para justificar que no son válidos los "would" asociados; al renunciar a la interdefinibilidad podemos afirmar los "might" que aparecen en (5) sin que por ello se pierda la verdad de (2) o de (3), por lo tanto seguiríamos obligados a admitir relaciones causa-efecto antiintuitivas, y no se resuelve el problema.

c) si ahora, teniendo en cuenta a) y b), volvemos a la interdefinibilidad, pero partiendo ahora del "might" entendido como Edgington lo propone, obtenemos una nueva teoría de condicionales contrafácticos, distinta de la de Lewis y probablemente de las que han propuesto otros autores, que habría que estudiar en todos sus consecuencias para determinar si es adecuada en una gran cantidad de ejemplos en que las teorías ya desarrolladas han resultado adecuadas. Como primera impresión, diría que se estaría volviendo a los condicionales estrictos. (Correspondientes a esferas de accesibilidad restringidas.) Lewis ha presentado argumentos convincentes en contra de la interpretación de los contrafácticos como condicionales estrictos.

iii) una forma de repicar y estar en la procesión sería aceptar condicionales especiales del tipo de los mencionados en ii-c), como condicionales distintos a los contrafácticos, pero a ser usados específicamente en lugar de éstos en la teoría de la causación. Esto también debería ser estudiado, pero de todos modos es de notar que se trataría de una nueva teoría de la causación<sup>3</sup> no la de Lewis, que se basa explícitamente en los contrafácticos.

## CONCLUSIONES

Creemos haber mostrado que desde el punto de vista formal la objeción de Flichman (validar (3) para no validar (2) es introducir un problema para evitar otro) no es lapidaria: hay formas de no validar ni (2) ni (3). Hemos visto también, sin embargo, que la situación en que esta maniobra deja a la causación contrafáctica no parece convincente, ya sea

(a) porque en caso de aceptemos la maniobra sin la reinterpretación sugerida por Edgington nos queda una noción de causación contexto-dependiente por un lado, y que por otro, siendo difícil de justificar para un ejemplo en particular corre peligro de tornarse insostenible en algún otro ejemplo, o

(b) porque, si seguimos las sugerencias de Edgington, corre el peligro de desbarrancarse la teoría de los contrafácticos, que parece sostenerse bien si no se le impone la carga de sostener esta teoría de la causación.

No es seguro por otra parte que la causación contrafáctica goce de buena salud aun si adoptamos alguna de estas vías de escape, ya que son numerosas las críticas que ha recibido por una cantidad de aspectos independientes del tema de este trabajo.