CRÍTICA, *Revista Hispanoamericana de Filosofía*
Vol. XXVII, No. 81 (diciembre 1995): 79–100

LEWIS'S CAUSATION: AN ALMOST FATAL EXAMPLE*

HORACIO ABELEDO
Sociedad Argentina de Análisis Filosófico
Universidad de Buenos Aires

Introduction

The present paper intends to examine some aspects of the theory of causation suggested by David Lewis, starting from an objection put forward by Eduardo Flichman and endorsed by Dorothy Edgington.

Brief History of the Problem

D. Lewis[1] takes up one of Hume's analyses of causation ("an object followed by another. . . where, if the first object had not been, the second never had existed"), reformulates it from the standpoint of the available theories of counter-

* This paper is a result of research work done as part of the program of the team of "Introducción al Pensamiento Científico", Ciclo Básico Común, Universidad de Buenos Aires, directed by E.H. Flichman and co-directed by the author of this paper. The author wishes to thank Eduardo Flichman, Jorge Paruelo and Hernán Miguel for their useful suggestions. Thanks are also due to Sandra Girón for drafting a first English version of the Spanish original.

[1] D. Lewis, "Causation", *Journal of Philosophy*, no. 70, 1973, pp. 556–567 (reprinted in *Philosophical Papers*, vol. II, Oxford University Press, 1986).

factual conditionals[2] and uses it as a basis for a counterfactual theory of causation.

To begin with, it is important to recall that the theory developed by Lewis only intends to explicate the concept of causation *between particular events*. In this paper, moreover, we shall be concerned only with his account of causation between *actually occurring particular events*.[3]

From here on, I shall use lower case italic letters (a, b, c, etc.) to indicate events, and their upper case version (A, B, C, etc.) to symbolize the sentences that state that such events occur. Lewis's theory can be summarized in some few items in the following way:

a) We shall say that b depends causally on a when the counterfactual "if a had not occurred, b would not have occurred" is true; symbolically:

$$\neg A \ \Box\!\!\rightarrow \neg B$$

b) If b depends causally on a then a is a cause of b.

c) If a is a cause of b, and b is a cause of c, then a is a cause of c.

Now, in order to be acceptable, this theory should be able to solve the several difficulties that have arisen in other analyses of causation. Among them, the lack of distinction between causes and effects (the so-called problem of

[2] Lewis mentions especially his own (D. Lewis, *Counterfactuals*, (Oxford, Blackwell, 1973)), and Stalnaker's (R. Stalnaker, "A theory of Conditionals, in N. Rescher, ed., *Studies in Logical Theory* (Oxford, 1968)). It should be noted that, although Lewis suggests that his analysis of causation is independent of the particular choice of a theory of counterfactual conditionals —and, therefore, it could use any of them as a basis— it will be shown in this paper that there seem to be insoluble difficulties if Stalnaker's theory, for example, is adopted.

[3] In order to satisfy more accurately the requirements of this theory, Lewis elaborates later his own conception of "event" in "Events", (in *Philosophical Papers*, vol. II, Oxford University Press, 1986).

effects), and others, such as the problem of epiphenomena, of overdetermination, etc. And it should not be vulnerable, as far as possible, to counterexamples: situations in which nobody would admit that *a* is cause of *b*, whereas the theory considers it is, or vice versa. Counterexamples tend to suggest that we are not really dealing with the concept we were trying to explicate.

In order to show that his theory avoids the problem of effects, Lewis[4] analyzes the example we shall examine in the following section. Flichman[5] shows that precisely Lewis's analysis of this example turns it into a counterexample for his own theory. From this and other objections to Lewis's analysis, as well as to those of other authors, Flichman infers the non-analyzability of the concept of causation, and joins the group of those who dismiss this concept (at least in the domain of natural sciences) considering it essentially anthropomorphic. Dorothy Edgington[6] has enthusiastically endorsed Flichman's criticism and also agrees to the idea of non-analyzability of the concept of causation; in her view, however, it is a concept that we should not abandon though it is at the same time "too fundamental to be analyzed in terms which do not presuppose it".

My aim here is to establish to what extent, with a different analysis of this example, Flichman's criticism can be avoided and Lewis's counterfactual theory of causation preserved. We shall see that this can be done; but there are considerable prices to be paid for it.

[4] "Causation" (see note 1 above), end of the Section "Counterfactual versus Nomic Dependence".

[5] E.H. Flichman, "The Causalist Program, Rational or Irrational Persistence?", *Crítica*, vol. XXI, no. 62, 1989, pp. 29–53.

[6] D. Edgington, "Explanation, Causation and Laws", *Crítica*, vol. XXII, no. 66, 1990, pp. 55–73.

A version of Lewis's example follows: The behavior of a barometer is analyzed. Let us call p the event that the atmospheric pressure is 1000 mb; r the event that the barometer reads 1000 mb; and b the event that the barometer is working properly. Therefore, the statements of occurrence of these events are the following:

P = "The atmospheric pressure is 1000 mb."

B = "The barometer is working well."

R = "The barometer reads 1000 mb."

In this example the three statements are true. From the standpoint of a possible worlds approach we would say that the events p, b, and r occur in the "actual" world. Since we know that the pressure is 1000 mb, and that the barometer works well and reads 1000 mb, we would say that p is a cause of r: that the pressure is 1000 mb is, obviously, a cause that a barometer in good condition reads 1000 mb. Lewis's approach supports this result: since p and r are occurring events and the following counterfactual is true, p is a cause of r:

$$\neg P \ \Box\!\!\rightarrow \neg R \tag{1}$$

in other words,

"If the atmospheric pressure had not been 1000 mb, the barometer would not have read 1000 mb." (1a)

It should be noted here that, in the usual context of a conditional of this sort intuition would judge that, since the barometer works well, it does not make sense to consider situations in which this would not hold, since the counterfactual assumption does not seem to require it. A counterfactual such as "*If the pressure had not been*

1000 mb, then the barometer would have been out of order and would have read 1000 mb anyway" would, in a normal context, certainly be rejected.

Up to here, it seems, Lewis's theory works perfectly, and its results are coincident with intuition. However, Lewis needs to show that his approach is useful for clarifying problems that other theories leave unsolved. Such is the case of the so-called *problem of effects*: is it possible to distinguish effects from causes? Or, in other words, have we found an *asymmetrical* analysis of causation? Lewis claims he has. Precisely, he introduces his example in order to prove it. If everything works out, the theory should not render true, together with (1), the symmetrical counterfactual:

$$\neg R \ \Box\!\!\rightarrow \neg P \qquad (2)$$

in other words,

> *"If the barometer had not read 1000 mb, the pressure would not have been 1000 mb"* (2a)

Lewis claims that this affirmation should be considered false, basing his idea on considerations about the context or "resolution of vagueness" that he believes should be applied to maintain irreversibility[7] —which seems to be an extremely *ad hoc* criterion— and arguing that what should be considered true is that *if the barometer had not read 1000 mb, it would have been malfunctioning* (because the pressure is 1000 mb, and envisaging a change of the weather conditions implies considering a world much more different from the actual one than the breakdown of a barometer).

[7] "Causation", footnote 10.

This is a highly questionable argument and it has indeed been questioned.[8] But it is at the same time a remarkable *faux pas*, because it has made Lewis's theory vulnerable to criticisms more serious than those he intended to avoid; it now falls victim to Flichman's objection, to wit: if Lewis's just mentioned consideration is accepted, we should then accept as true the following affirmation:

$$\neg R \ \Box\!\!\rightarrow \neg B \qquad (3)$$

in other words,

> *"If the barometer had not read 1000 mb, then it would not have been working properly."* (3a)

which, according to Lewis's theory, implies that the fact that the barometer reads 1000 mb is a cause of its working properly. This is completely against intuition.

Thus, it is Lewis's own evaluation of the counterfactual conditionals involved that turns the example into a counterexample of his theory.

In brief, Flichman's objection (which Dorothy Edgington considers definitive) suggests that Lewis is between the devil and the deep blue sea:[9] if he endorses (2), his theory cannot avoid the problem of effects; and if he endorses (3), he is forced to accept causes totally contrary to intuition.

[8] Thus, Edgington suggests that the so-called "Butterfly Effect" (a butterfly stirring the air in one place of the world can transform the storm systems in another place) can force us to invert the argument: the weather is unstable, whereas barometers can be made as sturdily as we please.

[9] In fact, Flichman confines himself to showing that the consequences for Lewis's theory of causation of accepting the truth of (3) are no more acceptable than those of accepting (2). The possibility of an analysis validating neither (2) nor (3) is not suggested in his paper.

Is There a Way Out for Lewis's Theory?

Our object now is to analyze whether it is possible to save Lewis's theory from this criticism. That is: would it be possible to modify our evaluation of the above mentioned counterfactuals, against Lewis's opinion, so as to avoid the problems?

In order to do so, we should find a criterion which makes both (2) and (3) false.

At this point, which particular theory of counterfactuals we are to adopt becomes important. Take Stalnaker's theory of counterfactuals. This theory assigns validity to the so-called Conditional Excluded Middle law for counterfactuals :

$$A \mathrel{\Box\!\!\rightarrow} B \ \vee \ A \mathrel{\Box\!\!\rightarrow} \neg B \tag{4}$$

that, consequently, allows inferring the truth of $A \mathrel{\Box\!\!\rightarrow} \neg B$, whenever $A \mathrel{\Box\!\!\rightarrow} B$ is false.

In our case, if we consider $\neg R \mathrel{\Box\!\!\rightarrow} \neg P$ false and reject it (in order to avoid the problem of effects), according to Stalnaker's theory we must consider true $\neg R \mathrel{\Box\!\!\rightarrow} P$, that is, "*if the barometer had not read 1000 mb, the pressure would still have been 1000 mb*". Hence we must consider that the barometer has not been working properly, ending up by endorsing (3). Therefore, Edgington's opinion is indisputable if Stalnaker's theory (or any other theory that makes Conditional Excluded Middle valid) is to be used: in that case Lewis's theory is either vulnerable to Flichman's criticism if his unfortunate evaluation of the mentioned example is maintained, or else, if it is not, cannot avoid the problem of effects. There is no possible evaluation at all that avoids both problems simultaneously.

Yet in Lewis's theory of counterfactuals Conditional Excluded Middle is not valid. Lewis evaluates counterfactuals

from the standpoint of a possible worlds semantics according to which the counterfactual "would" conditional:

> "If it were the case that A, then it would be the case that B"

is true in a world w (henceforward, *base world*) if B holds in all the possible worlds most similar to w in which A holds.[10] One reason for the invalidity of Conditional Excluded Middle is that there may be "ties": there are sometimes (among the worlds most similar to the base world in which A holds), worlds in which B holds which are just as similar to the base world as others in which $\neg B$ holds. In such cases, neither of the conditionals appearing in (4) is true: not *all* the antecedent worlds most similar to the base world are B-worlds; and not *all* of them are $\neg B$-worlds.[11] Instead of the "would" counterfactuals, so-called "might" counterfactual conditionals could be asserted: "If it were the case that A, it might be the case that B", and also, "if it were the case that A, it might be the case that $\neg B$". A brief and imprecise version of the truth conditions for "might" counterfactuals may be stated thus:

> "If it were the case that A, then it might be the case that B"

is true in the base world if B holds in at least one of the possible worlds most similar to w in which A holds.

[10] For the sake of brevity, I am presenting here an intuitive and imprecise version of Lewis's formulation. A more accurate description should mention the possibility of *vacuously true* counterfactuals, as well as the possibility of a failure of the so-called *limit assumption*. Lewis's complete theory can be found in *Counterfactuals*, sections 1.1 to 1.7 (see note 2 above); for a formulation based on comparative similarity, see section 2.3.

[11] In the improbable case of failure of the limit assumption, a more complicated situation may give rise to the falsity of (4).

Equivalently, the "might" counterfactual may be defined in terms of the "would" counterfactual:

$$A \diamondsuit\!\!\rightarrow B \equiv_{\mathrm{df}} \neg(A \mathrel{\square\!\!\rightarrow} \neg B) \qquad (5a)$$

Or alternatively the "would" counterfactual may be defined in terms of the "might" counterfactual through:

$$A \mathrel{\square\!\!\rightarrow} B \equiv_{\mathrm{df}} \neg(A \diamondsuit\!\!\rightarrow \neg B) \qquad (5b)$$

In brief, in Lewis's theory it is possible to deny both disjuncts in (4); and, in view of (5a) and (5b), affirm the corresponding "might" counterfactuals. This means we can now try a possible way out: perhaps we can evaluate the similarity ordering of possible worlds in a manner that allows us to deny (2) *and* (3) and therefore affirm:

> *"If the barometer had not read 1000 mb, it might have been the case that the pressure was not 1000 mb; and it might have been the case that the barometer was malfunctioning."* (6)

In other words, let us consider the possibility that among the worlds (most similar to the base world) in which the barometer does not read 1000 mb there is a "tie" between those in which the pressure is not 1000 mb and those in which the barometer is not working properly. The situation can be described as follows: Let us consider all possible worlds (at least those similar enough to have a barometer like ours, and so forth). These worlds can be divided into eight groups, according to the truth or falsity of P, B, and R in each one of them (see table).

The worlds we should examine when we make the counterfactual supposition "*if the barometer had not read 1000 mb. . .* " are those in which R is false. That is, groups [2], [4], [6] and [8]. The actual world belongs to group [1]. The worlds of group [2] are impossible worlds: the supposition

that the barometer might work well and the reading still be higher or lower than the prevailing pressure cannot be entertained. The worlds of group [8] are surely less similar to the actual world than those of groups [4] and [6], because in group [8] there are simultaneously two alterations from the actual world that in groups [4] and [6] appear separately: the pressure is different, *and* the barometer is malfunctioning, where only one of them would suffice to produce a different reading. Thus, we should confine our attention to the more similar worlds of groups [4] and [6]. If those of group [4] are more similar than those of [6], (3) shall be valid; in the opposite case, (2) shall be valid. If none of the groups contains worlds that are nearer than all worlds in the other one, neither (2) nor (3) are valid, and (6) is valid.

Group	P	B	R
[1]	T	T	T
[2]	T	T	F
[3]	T	F	T
[4]	T	F	F
[5]	F	T	T
[6]	F	T	F
[7]	F	F	T
[8]	F	F	F

This strategy could save Lewis's theory from the specific difficulties mentioned above. However, the theory is still impaired, because this solution presents several weak points:

a) Arguments should be found to justify, for this example, this particular ordering of worlds. How are we going to determine clearly whether the most similar worlds of group [4] are not a little more or less different than those of group [6]?

b) Even if we find a solution to the example, why are we so sure that we will find one for any analogous example we run into? When Lewis published *Counterfactuals*, it seemed we could have reasonable and intuitive criteria for ordering possible worlds; this could then serve as a basis for the evaluation of counterfactuals. However, it seems now we are forced to accept certain very precise orderings to prevent counterfactuals from getting out of hand: we must insist that both groups of worlds are equally near, not as a consequence of an examination of the properties of the base world, but because we do not have any other way out.

c) If we recall that the ordering of worlds depends on the context of enunciation of each counterfactual, it is likely that, if there is a "tie", the situation can become so unstable that the slightest change of context would lead us to (2) or (3). In that case, we should admit that there are many instances in which the relation of causation is context-depending. That is bad news for those who, together with Lewis, wish to understand this relation as an attribute of reality and not as a feature of language.

Dorothy Edgington's Objections

In this section I include some comments suggested by another paper of Dorothy Edgington (which so far as I know has remained unpublished), where ideas quite similar to mine are proposed with the purpose of tackling this problem.[12]

i) Dorothy Edgington suggests taking another line that consists in accepting "*if the reading had been different, then it might have been that the pressure was different, or*

[12] D. Edgington, "David Lewis, Counterfactual Dependence and Causation", unpublished, 1990, 18 pages.

it might have been that the barometer was malfunctioning". It is not clear whether this sentence represents (as it would seem if taken literally) a "might" counterfactual with a disjunctive consequent, or, more likely, is a natural language phrase that can be construed formally in more than one way.[13] Hence, one could think of the following formalizations:

a) $(\neg R \diamondsuit\!\!\rightarrow \neg B) \vee (\neg R \diamondsuit\!\!\rightarrow \neg P)$

b) $\neg R \diamondsuit\!\!\rightarrow (\neg B \vee \neg P)$

c) $(\neg R \diamondsuit\!\!\rightarrow \neg B)$ & $(\neg R \diamondsuit\!\!\rightarrow \neg P)$

Yet, in our example, supposing that the barometer does not read 1000 mb entails supposing that either it is out of order or that the pressure is not 1000 mb. Thus, both a) and b) would be analytic, and therefore should be valid, quite regardless of which are the most similar worlds, that is, whatever our position regarding the truth of (2) or (3). This can be seen readily if one notices that it is quite immediate to infer a) and b) when either (2) or (3) is taken as a premiss. But the point was to find an alternative that could allow us to reject both (2) and (3), which are the affirmations that yield problems.

The relevant option then, in order to save Lewis's theory, is c), which is the formalization of our (6). If it could be understood in this manner, then, Edgington's proposal would be coincident with the strategy suggested here.

[13] Just a look into the profuse literature devoted to the problem of (seeming) counterfactuals with disjunctive antecedent can show that the determination of the implicit logical form of sentences of this kind is not at all trivial. See H. Abeledo, E.H. Flichman and H. Miguel, "Contrafácticos y antecedentes disyuntivos: una cuestión de privilegio", III Jornadas de Epistemología e Historia de la Ciencia, Universidad de Córdoba, December 1992.

ii) Edgington claims that this kind of solution would be plausible for our example, if we accepted the following truth condition for "might" counterfactuals:[14]

> *$A \diamondsuit\!\!\rightarrow B$ is true if there are A & B worlds that are not too far-out or far-fetched.* (7)

which, of course, is not Lewis's truth condition for such counterfactuals.

This suggestion seems in a sense to be quite intuitive because it enables "coarse-grain" ties by comparing similarities without the precision required by Lewis's theory: since it does not require $A\&B$ worlds to be among the A worlds most similar to the base world, one could accept both $A \diamondsuit\!\!\rightarrow B$ and $A \diamondsuit\!\!\rightarrow \neg B$ without having to prove that the most similar $A\&B$ worlds are just as similar as the most similar $A\&\neg B$ worlds. However, to accept the proposal would yield some problems:

a) As mentioned above, Lewis's "might" and "would" counterfactuals may be defined one in terms of the other through formulas (5a) and (5b). This is quite intuitive in some situations, and the parallel with the modal possibility and necessity operators is attractive. Since accepting (7) means abandoning Lewis's truth conditions for the "might" counterfactual, it also means laying aside Lewis-style interdefinability.

b) For our present concern, the point was to show that the "might" counterfactuals are true in order to justify that the corresponding "would" counterfactuals are not valid; but with truth condition (7) we may affirm the "might" counterfactuals that appear in (6) without necessarily denying the truth of (2) or (3). Thus, either we are still forced to admit anti-intuitive cause-effect relations, or we must deny

[14] Strictly following D. Edgington's wording it ought to be considered an assertability condition rather than a truth condition.

(2) and (3), which is equivalent to affirming the Lewis-style "might" counterfactuals.

c) If, on the other hand, bearing in mind difficulties *a*) and *b*), we decide to go back to accepting (5a) and (5b), but taking as a starting point Edgington's truth condition for the "might" counterfactual and using (5b) to define the "would" counterfactual, we shall obtain then a new semantics for counterfactual conditionals, different from Lewis's and from other authors' proposals. This theory (and its consequences) should be carefully examined so as to determine its performance in the great number of examples where those other theories have proven adequate. My impression is that we would be returning to strict conditionals for restricted spheres of accessibility.[15] Lewis has argued convincingly against the interpretation of counterfactuals as strict conditionals.

iii) A way of having the cake and eating it could be to accept a special kind of conditionals such as those mentioned in ii–*c*), i.e., conditionals distinct from Lewis-style counterfactuals, meant to be used in their place in the theory of causation. These are strict conditionals or at least similar to them; and as suggested above should not be considered counterfactuals. This idea should be carefully studied, but, should it be considered adequate, it would be a theory of causation distinct from Lewis's, which is explicitly based on counterfactuals.

Conclusions

I have shown that, from a formal point of view, Flichman's objection (that validating (3) so as not to validate (2) is tan-

[15] Since (5b) would imply that for $A \Box\!\!\rightarrow B$ to be true $A \Diamond\!\!\rightarrow \neg B$ would have to be false; that is, all $A\&\neg B$ worlds would have to be too far-out or far-fetched; hence all A-worlds that are not outside a certain sphere must be B-worlds.

tamount to solving a problem by creating another) is not quite conclusive: there are ways of considering both (2) and (3) false. However, the consequences are still unfortunate for Lewis's counterfactual theory of causation because either:

(a) we adopt the strategy without the reinterpretation suggested by Edgington, and hence obtain a context-depending notion of causation, and, moreover, since it is difficult to justify for a specific example, it is in danger of being indefensible for other examples; or

(b) we follow Edgington's suggestions, in which case the theory of counterfactuals, which seems to be quite sound if it is not supposed to support the theory of causation, is left in danger of collapsing.

It should also be remarked that it is not at all certain whether counterfactual causation could be successful even if any of the strategies here considered is deemed acceptable, since it has received many other objections regarding several aspects outside the scope of this paper.

Recibido: 14 de diciembre de 1995

RESUMEN

En el trabajo se examina una dificultad del análisis de la causalidad (en términos de condicionales contrafácticos) propuesto por David Lewis en "Causation" (*Journal of Philosophy*, vol. 70; incluido en Lewis, *Philosophical Papers*, vol. II, OUP, 1986). La dificultad fue señalada por Eduardo Flichman (Universidad de Buenos Aires) en una objeción que ha sido considerada decisiva por Dorothy Edgington (University of London). En el presente artículo se examinan algunas posibles vías de solución.

El análisis lewisiano de la causalidad

D. Lewis se ocupa de la causalidad [*causation*] entre eventos particulares. En este trabajo, el autor sólo se ocupa de la teoría de Lewis acerca de eventos que realmente ocurren. Se introduce la convención de usar letras minúsculas para eventos y las mayúsculas correspondientes para enunciados que afirman que tales eventos ocurren ('*A*' representa una oración que afirma la ocurrencia del evento *a*, etc.). Las tesis siguientes sintetizan la propuesta de Lewis:

(i) *b* depende causalmente de *a* cuando es verdadero el contrafáctico "Si *a* no hubiera ocurrido, *b* no habría ocurrido". En símbolos: $\neg A \mathrel{\Box\!\!\rightarrow} \neg B$.

(ii) Si *b* depende causalmente de *a*, *a* es una causa de *b*.

(iii) La relación *Ser una causa de* es transitiva.

Para que la teoría resulte aceptable, debe resolver adecuadamente algunas dificultades que se han presentado en otros análisis. Por ejemplo, debe evitar la confusión entre causas y efectos (el llamado "problema de los efectos") brindando un análisis asimétrico de la causalidad. Se analiza a continuación un ejemplo del mismo Lewis para ver cómo se comporta su teoría en este respecto.

¿Un ejemplo fatal?

Se supone que son verdaderas en el mundo real [*actual*]* las tres oraciones siguientes, en las que se habla de un cierto barómetro y la presión atmosférica en un momento dado:

P = "La presión atmósferica es de 1000 mb."

B = "El barómetro está funcionando bien."

R = "El barómetro marca 1000 mb."

Como las oraciones son verdaderas, los eventos p, b y r acaecen en el mundo real. Dado este supuesto, sería intuitivo concluir que p es una causa de r.

El análisis de la causalidad de Lewis, suplementado con la teoría de los condicionales contrafácticos del mismo autor, apoya la conclusión intuitiva recién mencionada. De acuerdo con tal análisis, por las tesis (i)–(ii), se probaría que p es una causa de r si se estableciera el enunciado:

(1) $\neg P \mathrel{\square\!\!\rightarrow} \neg R$,

o, en palabras,

(1a) "Si la presión atmosférica no hubiera sido de 1000 mb, el barómetro no habría marcado 1000 mb."

De acuerdo con la teoría sobre contrafácticos de Lewis, (1) (o 1a) es verdadero. Según tal teoría, un contrafáctico como (1) es verdadero en un mundo posible w sii en todos los mundos posibles más similares a w en que se cumple el antecedente, también se cumple el consecuente.** Si P, B y R son verdaderos en el mundo real, los mundos posibles más similares al real en que valga $\neg P$ conservarán la verdad de B y en ese caso el barómetro no marcará 1000 mb y valdrá $\neg R$ en tales mundos. Por lo tanto, cuando w = el mundo real, la aplicación de la teoría conduce a la verdad de (1) (o 1a).

* Flichman y Abeledo traducen 'actual' por 'efectivo' en contextos como el presente. Yo prefiero 'real' por razones que sería difícil resumir aquí.

** El autor del trabajo aclara que ésta es una versión intuitiva e imprecisa de la formulación de Lewis, cuya teoría completa puede consultarse en Lewis, *Counterfactuals*, Blackwell, 1973.

Lewis pretende que su teoría no presenta "el problema de los efectos" y da realmente un análisis asimétrico de la causalidad. En ese caso la teoría debe arrojar la falsedad de

(2) $\neg R \ \Box\!\!\rightarrow \neg P$

o, en palabras,

(2a) "Si el barómetro no hubiera marcado 1000 mb, la presión no habría sido de 1000 mb."

Lewis cree que, en efecto, (2a) debe juzgarse falso por consideraciones sobre el contexto o la "resolución de la vaguedad" relevante y observa que el contrafáctico que debiera considerarse verdadero es más bien

(3) $\neg R \ \Box\!\!\rightarrow \neg B$

o, en palabras,

(3a) "Si el barómetro no hubiera marcado 1000 mb, no habría estado funcionando bien",

ya que los mundos posibles más parecidos al real en que se cumple $\neg R$ diferirán más bien en el funcionamiento del barómetro que en la situación climática (porque en el mundo real vale P, y un cambio en el clima supone una desemejanza más grande que la falla de un barómetro).

Pero Eduardo Flichman ("The Causalist Program, Rational or Irrational Persistence?", *Crítica*, vol. XXI) hace notar que la aceptación de (3) (o 3a), en vista de las tesis (i)–(ii), implica que el hecho de que el barómetro marque 1000 mb es una causa de que funcione bien. Esto es contraintuitivo.

Lewis se encuentra, pues, ante un dilema: si acepta (2), su teoría presenta el problema de los efectos, y si acepta (3), su análisis conduce a la afirmación de un enunciado causal completamente antiintuitivo.

¿Hay una salida para la teoría de Lewis?

La respuesta a este problema depende en parte de qué teoría de contrafácticos se use junto con el análisis de la causalidad de Lewis, quien menciona la posibilidad de usar la teoría de Stalnaker. Pero en ella vale la ley del tercero excluido condicional:

(4) $A \ \Box\!\!\rightarrow B \vee A \ \Box\!\!\rightarrow \neg B$

Se sigue de (4) que si $A \ \Box\!\!\rightarrow B$ es falso, $A \ \Box\!\!\rightarrow \neg B$ es verdadero. Aplicando esta consideración a los ejemplos anteriores, se sigue que debe aceptarse que (2) es verdadero o (3) lo es, y de este modo se llega al dilema mencionado. En efecto, si no se acepta (2) ($\neg R \ \Box\!\!\rightarrow \neg P$), debe aceptarse $\neg R \ \Box\!\!\rightarrow P$, que en palabras dice: "Si el barómetro no hubiera marcado 1000 mb, la presión habría sido de todos modos 1000 mb." Pero esto último es sostenible si se piensa que en caso de no marcar 1000 mb, el barómetro habría funcionado mal. Y esto es aceptar (3).

Sin embargo, la ley del tercero excluido condicional no vale en la teoría de contrafácticos de Lewis, y el autor de este trabajo precisa que esto puede proporcionar una escapatoria del dilema. La razón por la cual (4) no vale para Lewis es que la condición de verdad de $\Box\!\!\rightarrow$ permite la falsedad de ambos disyuntos. En efecto, para que sea cierto $A \ \Box\!\!\rightarrow B$ en w, en *todos* los mundos posibles más similares a w en que vale A debe valer B, y para la verdad de $A \ \Box\!\!\rightarrow \neg B$ se requiere que en *todos* esos mundos debe valer $\neg B$. Pero podría haber "empate": podrían existir en la clase de mundos mencionada, mundos y, z, igualmente similares al real, pero tales que en y vale B y en z vale $\neg B$. Los dos disyuntos de (4) son falsos en este caso. En cambio, dos condicionales contrafácticos de otro tipo son verdaderos en el caso descrito: "Si hubiese sido el caso que A, podría haber sido el caso que B" y "Si hubiese sido el caso que A, podría haber sido el caso que $\neg B$". Las simbolizaciones son $A \Diamond\!\!\rightarrow B$ y $A \Diamond\!\!\rightarrow \neg B$. Llamaremos "contrafácticos posibles" a los de este tipo, distinguiéndolos de los "contrafácticos necesarios" que se construyen con $\Box\!\!\rightarrow$.* Se definen así:

(5a) $A \Diamond\!\!\rightarrow B =_{\text{def}} \neg(A \ \Box\!\!\rightarrow \neg B)$

Esta definición arroja una condición de verdad similar a la del $\Box\!\!\rightarrow$ pero reemplazando "en todos los mundos posibles. . . "

* Lewis llama "would counterfactuals" a los construidos con $\Box\!\!\rightarrow$ y "might counterfactuals" a los construidos con $\Diamond\!\!\rightarrow$. La traducción literal de su terminología es espantosa, y por ello introduzco las expresiones entrecomilladas de la última oración del texto. Pero el lector debe advertir que 'posible' no quiere decir aquí 'lógicamente posible'; lo mismo ocurre con 'necesario'.

por "al menos en uno de los mundos posibles...". También se puede introducir primero $\Diamond\!\!\rightarrow$ con esta condición de verdad y luego definir $\Box\!\!\rightarrow$ así:

(5b) $A \Box\!\!\rightarrow B =_{\text{def}} \neg(A \Diamond\!\!\rightarrow \neg B)$

Se sugiere entonces en el trabajo que en la teoría de Lewis es posible negar el tercero excluido condicional y los enunciados (2) y (3), afirmando en cambio:

(6) "Si el barómetro no hubiera marcado 1000 mb podría haber sido el caso que la presión no fuera de 1000 mb; y podría haber sido el caso que el barómetro no estuviera funcionando bien."

A continuación, el autor hace un análisis detallado del tipo de ordenamiento de mundos posibles que haría falsos los enunciados (2) y (3). Hay que analizar los mundos posibles más similares al real en que vale $\neg R$ (el antecedente de (2) y (3)). Esos mundos pueden clasificarse en cuatro grupos: grupo [2], de mundos en que valen P y B, grupo [4] de mundos donde valen P y $\neg B$, [6] donde valen $\neg P$ y B, y [8] donde valen $\neg P$ y $\neg B$. Un análisis muestra que en realidad los mundos (2) no son posibles ($\neg R$, P y B no son los tres compatibles) y los del grupo [8] no están entre los más similares al mundo real (registran 2 alejamientos del mundo real —donde P y B son verdaderos— en tanto que los grupos [4] y [6] muestran uno solo de ellos). Para la verdad de (2) y (3) interesan entonces los grupos [4] y [6]. Si los mundos de [4] son más similares al real que los de [6], (3) es verdadero; si los de [6] son más similares al real que los de [4], (2) es verdadero. Si ninguno de los dos grupos [4] y [6] contiene mundos más cercanos al real que cualquiera del otro grupo, se registra un "empate" como el antes descrito. En ese caso, (2) y (3) son falsos, la teoría de Lewis escapa al dilema y (6) es verdadero.

Pero la línea de solución basada en el "empate" presenta varios puntos débiles:

a) ¿Cómo determinar claramente que los mundos del grupo [4] más similares al real no son un poco más similares o un poco

menos similares que los del grupo [6]? Se requieren argumentos para justificar un ordenamiento de mundos con tal resultado.

b) Aun cuando encontremos una solución al problema (a), ¿por qué deberíamos estar seguros de que en otro ejemplo análogo vamos a encontrar también un ordenamiento que produzca un "empate"?

c) Si el ordenamiento de mundos de acuerdo con su similaridad depende del contexto de enunciación de cada contrafáctico, es probable que en caso de empate la situación sea inestable y un ligero cambio de contexto conduzca a (2) o (3), haciendo que la causalidad sea una relación dependiente del contexto.

Las objeciones de Dorothy Edgington

Se incluyen en el trabajo algunos comentarios sugeridos por un artículo de Dorothy Edgington no publicado todavía.

i) Edgington sugiere aceptar que "si el barómetro hubiera marcado otra presión, podría haber ocurrido que la presión fuera distinta o podría haber ocurrido que el barómetro estuviera funcionando mal". Esta oración puede interpretarse formalmente de distintos modos. Si se acepta la formalización

$$(\neg R \ \diamondsuit\!\!\rightarrow \neg P) \ \& \ (\neg R \ \diamondsuit\!\!\rightarrow \neg B)$$

la propuesta de Edgington proporciona una solución de la dificultad coincidente con la del presente trabajo, ya que la oración (6) propuesta más arriba se formaliza de idéntica manera.

ii) Edgington considera que su propuesta es plausible con esta condición de verdad de $\diamondsuit\!\!\rightarrow$:*

(7) $A \diamondsuit\!\!\rightarrow B$ es verdadero si hay mundos en los que se cumple $A\&B$ que no son demasiado disímiles del mundo real.

(7) permite la verdad de $A \diamondsuit\!\!\rightarrow B$ y $A \diamondsuit\!\!\rightarrow \neg B$ sin requerir condiciones tan estrictas como las de Lewis. Pero la propuesta trae algunas dificultades: (a) Dejan de ser aceptables las definiciones

* El autor aclara que, estrictamente hablando, Edgington propone una condición de *asertabilidad*.

(5a) y (5b); (b) la verdad de (6) no implica la falsedad de (2) y (3), y por consiguiente (6) no salva a Lewis de los problemas ya descritos; y (c) si se define $\Diamond\!\!\rightarrow$ como en (7) y $\Box\!\!\rightarrow$ como en (5b), se obtiene una semántica de contrafácticos que quizás transforma los contrafácticos necesarios en condicionales estrictos, en contra de las conclusiones de Lewis.

Conclusiones

La objeción de Flichman no es del todo concluyente, porque hay maneras de considerar falsos (2) y (3). Pero tal estrategia perjudica de todos modos la teoría de Lewis porque, o bien

(a) adoptamos esa estrategia sin la reinterpretación de $\Diamond\!\!\rightarrow$ propuesta por Edgington, en cuyo caso se obtiene una noción de causalidad dependiente del contexto y hay la posibilidad de que la estrategia no pueda defenderse en otros ejemplos, o bien

(b) seguimos las sugerencias de Edgington en cuyo caso la teoría de contrafácticos podría colapsar en una teoría de condicionales estrictos.

[*Raúl Orayen*]