

Commentary
David Lewis
Counterfactuals
Oxford, Blackwell, 1973 *

Vittorio Morato[†]
vittorio.morato@unipd.it

Counterfactuals, published in 1973, was the culmination of work (both technical and philosophical) that Lewis had done in previous years (see Lewis (1971), Lewis (1973b)). The bulk of his analysis is that counterfactuals are some sort of *variably* strict conditionals, whose semantics can be given in terms of (ordered sets of) possible worlds.

After the publication of *Counterfactuals*, the only articles Lewis wrote directly about the semantics of counterfactuals were Lewis (1977), Lewis (1979), and Lewis (1981). The first is a defense of Lewis's approach for counterfactuals with disjunctive antecedents (more on this below); the second is a full development (after an embryonic presentation in the book) of Lewis's approach to *backtracking counterfactuals*, conditionals of the form "if A were to happen at time T_A , then B would happen at time T_B "; the third is a comparison (indeed, a proof of equivalence) of Lewis's approach with that developed by A. Kratzer in 1981, where the factual background against which a counterfactual is evaluated is a set of premises, rather than an ordering of worlds. The principal changes to the 1986 "revised printing" edition consisted just of some corrections in the completeness results of chapter 6, plus other minor corrections. The core semantic analysis presented in *Counterfactuals* thus remained quite stable from its first publication onward.

While squarely a book about the semantic analysis of counterfactuals (at the time, more often called "subjunctive conditionals"), the book frequently steps outside this area to discuss topics in metaphysics, philosophy of language and philosophy of science: there are in fact many interesting passages about the

* Based on the revised printed edition, 1986.

[†] Department of Philosophy, University of Padua, Italy.

metaphysics of possible worlds, an analysis of the notion of law of nature and discussions about the vagueness, contexts and applications of the semantic approach developed for counterfactuals to other notions like conditional obligation, temporal notions and contextually definite descriptions. The style of the book, however, is – as it is typical of Lewis – very concise, crisp and forthright (it is less than 150 pages).

The book is organized into six chapters. The first chapter (nearly a third of the whole book) is a presentation of the semantics, while chapter 2 presents some “alternative reformulations”. In chapter 3, Lewis compares his analysis with what he calls “metalinguistic theory” – a term he uses to cover the theories of Chisholm, Goodman and Mackie – and with Stalnaker’s approach. Chapter 4 is a discussion of the two fundamental conceptual blocks of the theory, namely possible worlds and comparative similarity. In chapter 5, Lewis studies the presence of variably strict conditionals in other areas of intensional logic, namely deontic logic (conditional obligation), temporal logics (constructions like “when next” and “when last”), and “egocentric logic” (a term due A. Prior (1968) – quite unusual nowadays – that basically covers pre-Kaplanian attempts to develop a logic for indexicals). Finally, chapter 6 presents a regimentation of the materials into formal systems, for which Lewis proves completeness and decidability results. In the revised printing edition, the book ends with an appendix that contains a bibliography of related work, annotated by Lewis himself.

With respect to the philosophical background at the time of publication and abstracting from the technicalities that will be discussed in detail below, Lewis’s approach is characterized by two theses:

- Counterfactuals have truth conditions;
- The truth-conditions of counterfactuals could be given in terms of possible worlds.

Surely, neither of these two features were taken for granted at the time of publication (nor, in some respects, are they even today);¹ especially the possible worlds analysis was quite novel. At that time, the landscape was still dominated by Goodman’s “cotenability” approach. According to the cotenability approach, a counterfactual conditional of the form $\phi \ [\] \rightarrow \psi$ is

¹ See (Barker, 2011) for a very recent attempt to criticize possible worlds semantics for counterfactuals.

true if and only if ψ can be derived (according to some laws) by ϕ and some other contingently true propositions. On this view, counterfactual conditionals are not really sentences – not entities to which truth conditions can be assigned – but rather elliptical presentations of arguments.

The possible worlds approach to counterfactuals, however, was not original with Lewis. W. Todd (1964) was probably the first author to lay the groundwork for such an analysis, as this quotation (p. 107) quite explicitly reveals:

When we allow for the possibility of the antecedent's being true in the case of a counterfactual, we are hypothetically substituting a different world for the actual one. It has to be supposed that this hypothetical world is as much like the actual one as possible so that we will have grounds for saying that the consequent would be realized in such a world.

This line of research was then fully developed by Stalnaker (1968) and, more formally, by Stalnaker and Thomason (1970). As Stalnaker himself acknowledged, his approach and Lewis's, though quite similar, were developed independently from one another.²

The basic idea of Lewis's analysis is well expressed in the very first paragraph of the book:

“If kangaroos had no tail, then they would topple over” seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.

This paragraph contains *in nuce* almost all essential elements of Lewis's analysis. Let us start then from here to develop, step by step, a complete and more precise formulation of the truth-conditions for counterfactuals.

The general idea is that a counterfactual of the form $\phi \rightarrow \psi$ is true, in a possible world i , if and only if, in any world where ϕ is true and that resembles i as much as the truth of ϕ permits it to, ψ is true. We will see in a moment what it means that a world resembles another *as much as the truth of a proposition ϕ permits it to*. The important thing to notice, for now, is a structural feature. The form of this preliminary formulation of the truth-conditions for counterfactuals is the following:

² Stalnaker's acknowledgment is reported by Bennett (1974, p. 308).

$\phi \Box \rightarrow \psi$ is true in i if and only if for any world w , such that _____, ψ is true in w :

where “_____” stands for a restrictive condition that we will analyze in a moment. The left-hand side of these truth-conditions is structurally similar to the left-hand side of the truth-conditions of another, quite familiar, formula, namely $\Box\psi$:

$\Box\psi$ is true in i if and only if, for any world w such that _____, ψ is true in w .

In the case of the truth-conditions for $\Box\psi$, the restrictive condition is filled up by the specification of an accessibility relation between i and the possible worlds universally quantified over.

In the case of the truth-conditions for $\phi \Box \rightarrow \psi$, the restrictive condition is to be filled up by (the formal translation of) a condition having to do with i and ϕ . This condition is “being a world that resembles i as much as the truth of ϕ permits it to”. It is then expected that such a condition determines a class of worlds. The counterfactual is then said to be true if and only if ψ is true in such a class of worlds.

Given that the condition that restricts the universal quantification in a counterfactual conditional is defined relative to the antecedent of the counterfactual (and to the world of evaluation, but this is true also for the accessibility relation used to restrict the quantification over worlds in the clause for \Box), a nice way, suggested by R. Stalnaker, to present Lewis’s approach is to say that in his analysis the antecedents of conditionals “act like necessity operators on their consequents”.³ We will see how this analogy will reveal itself to be very useful for settling semantic issues for counterfactuals.

The class of worlds determined by ϕ and i is such that in it there are no worlds where ϕ is true and ψ is false. We could then say that a counterfactual like $\phi \Box \rightarrow \psi$ is true in i if and only if the material conditional $\phi \rightarrow \psi$ is true in every member of the relevant class of worlds. We could slightly change our clause for $\Box \rightarrow$ in order to register this new information:

$\phi \Box \rightarrow \psi$ is true in i if and only if, for every possible world y such that resembles i as much as the truth of ϕ permits it to, $\phi \rightarrow \psi$ is true in y .

³ For this view, see Stalnaker (1978, p. 93).

This clause, however, has two problems: the first is that it still contains the too informal and unexplained phrase “such that resembles i as much as the truth of φ permits it to”; the second is that the left-hand side of the biconditional could be vacuously satisfied in the case where there are no possible worlds that resemble i as much as the truth of φ permits them to, typically in the case where φ is impossible. The first problem could be solved by transforming the phrase “such that resembles i as much as the truth of φ permits it to” into the more tractable “such that φ is true in y and y resembles i ”, where y is bound by the universal quantifier. The new clause is surely more tractable, even if less expressive than the original: if φ is true in a world x that resembles i , then x is a world that resembles i as much as the truth of φ permits it to. Being true in a world is no conventional matter, so if something is true in a world, it is “permitted” to be true in such a world. For Lewis, the role of the informal clause (and in particular the use of “permits”) was that of highlighting the fact that the relevant worlds to consider when evaluating in i a counterfactual like $\varphi \ [] \rightarrow \psi$ are not those where φ is true and *everything else* is as it is in i . The reason is that, for Lewis, there are no such worlds. Or better, these worlds would be surprisingly far different from the actual, so different as to become irrelevant for the evaluation of the counterfactual. For example, the worlds where kangaroos have no tails and everything else is as it actually is are worlds less similar to the actual world than are the worlds where there are series of further deviations from actuality that “accommodate” the absence of tails in kangaroos in such worlds (due to a difference in the genetic set-up of kangaroos, for example). Here is what Lewis writes with respect to the similarity and difference trade-off:

Respects of similarity and difference trade-off. If we try too hard for exact similarity to the actual world in one respect, we will get excessive differences in some other respect. (Lewis, 1973a, p. 9)

We should not expect, however, that, in translating an informal condition in quasi-formal terms, every aspect of the informal idea will be explicitly preserved. As far as the truth-conditions of counterfactuals are concerned, we can live with the bare-bones formulation given in terms of truth of φ and similarity to i and leave more sophisticated features to the informal interpretation of our primitives (in our case, the similarity relation between worlds). In light of this, we can now write again a new formulation of our truth-conditions:

$\phi \Box \rightarrow \psi$ is true in i if and only if, for every possible world y such that ϕ is true in i and y resembles i , ψ is true in y .

The second problem (i.e., the eventual vacuous truth of the left-hand side of the biconditional) is not a problem *per se*; a distinctive feature of Lewis's semantics for counterfactuals is that counterfactuals with impossible antecedents are true. Here is what Lewis writes:

Confronted by an antecedent that it is not really an entertainable supposition one might react with a shrug: If that were so, anything you like would be true. (Lewis, 1973a, p. 23)

Furthermore, Lewis claims, counterfactual conditionals with impossible antecedent are asserted, by way of *reductio*, in philosophical, logical or mathematical arguments and need therefore to be taken as true in those contexts.

Note that the case of vacuous truth for $\phi \Box \rightarrow \psi$ could again be seen in analogy with what happens in the case of $\Box \psi$. In a world i such that it is not in a relation of accessibility with any other world (nor with itself), every formula is necessary: given an arbitrary, $\Box \psi$ is true in i . Analogously, in a world i that is not in the relation of accessibility with any world where ϕ is true, for any ψ , every counterfactual of the form $\phi \Box \rightarrow \psi$ is true.

What we want our truth-conditions to reveal, however, is what happens in the “normal” cases, those where ϕ is an entertainable supposition, namely where ϕ is possible. But inserting explicitly the possibility of ϕ in the truth-conditions for counterfactuals is going to add some complications. In particular, we would have to deal with three variables for worlds (the variable for the world of evaluation i , the variable introduced by the restricted universal quantification and the new variable, existentially quantified, for the world where ϕ is true), but with a relation of similarity with only two places.

The problem could be solved by introducing a three-place relation of similarity among worlds, *comparative similarity* (see Lewis, 1973a, p. 48):

$x \leq_i y =_{\text{def.}}$ x is at least as similar to i as the world y is.

Before discussing some properties of \leq_i , let us finally give the final formulation of the truth-conditions for counterfactuals in terms of this new relation of comparative similarity:

$\phi \Box \rightarrow \psi$ is true in i if and only if either

- *Vacuous case*: ϕ is false in any world accessible from i (i.e., ϕ is impossible); or
- *Non-vacuous case*: there is a world x accessible from i such that ϕ is true in x and, for any world y , if $y \leq_i x$, then ψ is true in y .

A counterfactual is (non-vacuously) true in a world i if and only if, if there is at least an accessible world x where the antecedent is true, then the consequent is true in every world at least as close to i as it is x .

In order for the truth-conditions to be working, we have to assume, as usual, that the standard binary relation of accessibility R and the new two-place relation \leq_i are defined for any possible world i with respect to any possible world. The latter notion will then generate an ordering of all possible worlds with respect to their comparative similarity to i . In order to see what kind of ordering is generated by \leq_i , we need to know its properties:

- \leq_i is *transitive*: whenever $x \leq_i y$ and $y \leq_i k$, then $x \leq_i k$;
- \leq_i is *strongly connected*: for every x and y , either $x \leq_i y$ or $y \leq_i x$.

I assume that the role and meaning of transitivity is clear. The role of strong connectivity is that of assuring the possibility of comparisons (with respect to i) of any arbitrary pair of possible worlds. Given that strong connectivity entails reflexivity, this condition implies also that i is at least as close to itself than any other world is.⁴

To these conditions on \leq_i , at least these other two features should be added:

- Every possible world i is accessible to itself (i.e., R is a reflexive relation);
- Every possible world i is "strictly minimal" with respect to \leq_i , namely for any world x (different from i), $i <_i x$ (i is more similar to itself than any other world is).

The strict minimality condition is responsible, in Lewis's approach, for the fact that counterfactuals with true antecedents might be true. What happens is that counterfactuals with a true antecedent reduce to material conditionals (see (Lewis, 1973a, p. 26)). In order to see this, suppose that ϕ is true at i ; then

⁴ In passing, note that the relation of comparative similarity between worlds used to evaluate counterfactuals is slightly different from the counterpart relation of similarity used to evaluate de re modal claims: in particular, the counterpart relation is a non-transitive similarity relation and is also non-symmetric.

there is a world accessible from i (i itself, given that it is assumed that the relation of accessibility is reflexive) where ϕ is true. If ψ is true in i , then in every world as close to i as i is to itself, $\phi \rightarrow \psi$ is true. This fact is granted by the strict minimality condition that assures that there is no other world but i that is as similar to i as i . Therefore, where ϕ and ψ are true in i , $\phi \Box \rightarrow \psi$ is true in i .

Contrary to what appears to be the case, it is a very welcome feature of Lewis's approach that "counterfactuals" with true antecedents turn out true. The term "counterfactuals" is in effect quite misleading, and if used nonetheless, it should not be used in a sense that presupposes or implies the falsity of its antecedent. As T. Williamson (2008, p. 137) rightly claims, in some contexts, we can assert a sentence like "if Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show" to be abductive evidence (by inference to the best explanation) for its antecedent, "Jones took arsenic".

But the eventual later empirical discovery that in effect Jones took arsenic does not make the previous assertion of the counterfactual inappropriate, rather it would constitute evidence for it. The counterfactual is true because we discover its antecedent to be actually true. Counterfactuals are "counter"-factuals not because they imply the falsity of the antecedent, but because their evaluation requires comparisons of alternative possible situations with the actual ones.

It is also interesting to note what properties \leq_i does *not* have. In particular, \leq_i is not anti-symmetric. That \leq_i is not anti-symmetric means that from $x \leq_i y$ and $y \leq_i x$, it does not necessarily follow that $x = y$: there might be two distinct worlds, none of which is more similar to i than the other. The non-anti-symmetricity of the \leq_i relation is one of the two main features that distinguishes Lewis's approach from Stalnaker's.⁵ The effect of this choice is

⁵ The other is the falsity, in Lewis, of the "limit assumption", namely the thesis that, when evaluating the counterfactual $\phi \Box \rightarrow \psi$ in i , there is always a closest to i world where ϕ is true; the effect of such an assumption is that when we know that a formula ϕ is true in some world, we know also that it is true in some world that is the closest world to i . This feature is very useful when proving the validity or the invalidity of counterfactuals. The falsity of the limit assumption implies that, when we have infinitely many worlds, there is no the closest to i world, but only an infinite series of closest and closest worlds. Note, however, that if the worlds are finite in number, the limit assumption automatically holds, even in Lewis.

that in Lewis (and not in Stalnaker) the following two schemas (“conditional excluded middle” and “distribution”) are not valid:

$$\begin{aligned} & (\varphi \Box \rightarrow \psi) \vee (\varphi \Box \rightarrow \neg\psi) \\ & (\varphi \Box \rightarrow \psi \vee \chi) \rightarrow ((\varphi \Box \rightarrow \psi) \vee (\varphi \Box \rightarrow \chi)) \end{aligned}$$

A countermodel for the first schema is given by two worlds equally similar to i , say x and y , such that φ is true in both of them and ψ is true in x but false in y . Given that a counterfactual, to be true, has to be true in every world similar to i , none of the disjuncts of our formula is true in our model. A countermodel for the second schema is given by letting φ , ψ and $\neg\chi$ be true in x and φ , χ and $\neg\psi$ be true in y .⁶

The conditional excluded middle is taken by Lewis to be a very “plausible” principle, especially because it explains why, in natural language, we do not usually distinguish between external and internal negation of a conditional. The sentence “it is not the case that if you had walked on the ice, it would have broken” seems to us perfectly equivalent to “if you had walked on the ice, it would not have broken”. Given that the conditional excluded middle is equivalent to $\neg(\varphi \Box \rightarrow \psi) \rightarrow (\varphi \Box \rightarrow \neg\psi)$ and everyone agrees that the converse, namely $(\varphi \Box \rightarrow \neg\psi) \rightarrow \neg(\varphi \Box \rightarrow \psi)$, is independently plausible, we have $\neg(\varphi \Box \rightarrow \psi) \leftrightarrow (\varphi \Box \rightarrow \neg\psi)$. The problem, for Lewis, is that, given a plausible (for him) definition of the “might” counterfactual $\varphi \Diamond \rightarrow \psi$ as $\neg(\varphi \Box \rightarrow \neg\psi)$, we can prove, by conditional excluded middle (now in the form $\neg(\varphi \Box \rightarrow \neg\psi) \rightarrow (\varphi \Box \rightarrow \psi)$), that $\varphi \Diamond \rightarrow \psi$ entails $\varphi \Box \rightarrow \psi$; given that the other direction is obviously true, what we have in effect proved is the equivalence between $\Diamond \rightarrow$ and $\Box \rightarrow$. This is, of course, quite unwelcome, and it is basically the reason why Lewis gives up on this formula.

Failure of the distribution principle is quite understandable if we remind ourselves again of the similarities and between $\Box \rightarrow$ and \Box in Lewis’s approach. The analogue distribution principle for \Box ,

$$\Box(\varphi \vee \psi) \rightarrow \Box\varphi \vee \Box\psi$$

quite clearly fails (take ψ to be $\neg\varphi$, for example). Contrary informal intuitions for the validity of the distribution principle for counterfactuals could

⁶ A nice formalism for establishing invalidity in counterfactual logic is presented in Sider (2010, p. 208–216).

then be taken simply as symptoms of “scope muddles” typical of interactions between disjunction and intensional operators. Similar “muddles” are probably in play also in the case of positive informal judgments on the validity of the following inferences, called “Simplification of disjunctive antecedent”:

$$\varphi \vee \chi \boxed{\rightarrow} \psi \quad (\text{DISJ-1})$$

$$\varphi \boxed{\rightarrow} \psi$$

Or

$$\varphi \vee \chi \boxed{\rightarrow} \psi \quad (\text{DISJ-2})$$

$$\varphi \boxed{\rightarrow} \chi$$

The failure of Lewis’s (and Stalnaker’s) semantics to make such inferences valid is often presented as a drawback for both approaches and, in general, for any possible worlds approach to counterfactuals¹ (see (Nute, 1976); (Ellis, Jackson & Pargitter, 1977)). A countermodel to DISJ-1 in Lewis’s semantics is done by supposing the existence of a χ and $\neg \varphi$ world w such that, in every x , such that $x \leq_i w$, ψ is true in x ; a countermodel to DISJ-2 by supposing the existence of a φ and $\neg \chi$ world, such that in every x such that $x \leq_i w$, ψ is true in x .

The problem is that we seem to be normally disposed to infer from “if φ or χ would have been the case, then ψ would have been the case” the conclusion that “if χ would have been the case, then ψ would have been the case” or “if φ would have been the case, then ψ would have been the case”. For example, after a very boring evening spent at home watching TV, we could say something like “if we had gone to a cinema or to a theatre, it would have been definitely better”, and from this it is quite natural to conclude also that “if we had gone to a cinema, it would have been definitely better” or “if we had gone to a theatre, it would have been definitely better”. Accepting such schemas into the logic, however, would be particularly dramatic for Lewis. As pointed out in Fine (1975, p. 453), given the logical equivalence between φ and $(\varphi \wedge \chi) \vee (\varphi \wedge \neg \chi)$, from $\varphi \boxed{\rightarrow} \psi$ we can derive by substitution of logical equivalents

$(\phi \wedge \chi) \vee (\phi \wedge \neg\chi) [] \rightarrow \psi$, and then, by an application of DISJ-1, we can conclude $(\phi \wedge \chi) [] \rightarrow \psi$. But failure of this rule:

$$\frac{\phi [] \rightarrow \psi \quad (\text{STRENGTH})}{(\phi \wedge \chi) [] \rightarrow \psi}$$

is taken as a benchmark of Lewis's logic of counterfactuals. Failure of such a rule is one of the main arguments Lewis uses to prove that counterfactual conditionals are not to be understood as (and then not reduced to) special kinds of (non-variably) strict implications. Failure of STRENGTH is then essential for the identity of $[] \rightarrow$ as an autonomous type of conditional.⁷

One way in which the problem could be solved is by explaining away the evidence in favor of DISJ. In particular, what might be contested is the legitimacy of translating natural language counterfactuals with apparently disjunctive antecedents with a formula like $\phi \vee \chi [] \rightarrow \psi$. Interactions of other intensional operators with disjunctions in natural language might constitute a useful analogy. Take the case of permission. In natural language, from "it is permissible that ϕ or ψ " we seem quite naturally to be disposed to infer "it is permissible that ϕ and it is permissible that ψ ". But adding the following rule to deontic logic:

$$\frac{R(\phi \vee \psi) \quad (\text{P-DISJ})}{R\phi \wedge P\psi}$$

would have dramatic consequences, because it would imply the truth of $R\phi \rightarrow P\psi$; given the arbitrariness of ϕ and ψ , this formula would entail that, from the permission of doing something, anything is permitted (this problem is called "paradox of free choice permission"). One of the standard responses to such a case is to translate a sentence schema like "it is permissible that ϕ or ψ " not as its surface form would suggest, but rather like $R\phi \wedge P\psi$. Similarly, one can adopt the same strategy for counterfactuals by translating natural language

⁷ Actually, and more dramatically, if one agrees to contextually define $\Box\phi$ as $\neg\phi [] \rightarrow \phi$, then one can prove, by the essential use of DISJ and substitution of logical equivalents, that $\phi [] \rightarrow \psi$ entail $\phi \rightarrow \psi$; see Loewer (1976, p. 532).

counterfactuals like “if ϕ or χ would have been the case, then ψ would have been the case” by $\phi \Box \rightarrow \psi \wedge \chi \Box \rightarrow \psi$. The reason for doing so would be mainly pragmatic: we seem to utter counterfactuals with disjunctive antecedents *only* in case we are *already* willing to assert the two simplified counterfactuals; we naturally tend to infer the two simplified counterfactuals from the disjunctive one only because it would be inappropriate to assert the latter without already believing the former. But then, the relation between the two simplified counterfactuals and the counterfactual with a disjunctive antecedent is not one of entailment, but of presupposition, and therefore there is no reason to consider DISJ, in its original formulation, a logically valid rule of inference.⁸

The truth conditions given here in terms of a relation of comparative similarity are not those preferred by Lewis as given in the first chapter of *Counterfactuals*. In the book, the “official” truth-conditions are given in terms of a nested system of “spheres”, where a sphere is a set of possible worlds within a certain degree of similarity to a given world. In this formulation, a counterfactual $\Box \rightarrow$ is true in i if either no sphere around i contains a ϕ -world or some sphere contains a ϕ -world and no ϕ and $\neg\psi$ world.

The two formulations are, however, perfectly equivalent (Lewis himself proves this; see Lewis, 1973a, p. 49–50), but I think that the one in terms of comparative similarity is to be preferred. It is surely more fundamental. This is because spheres must be ultimately explained in terms of comparative similarity. A sphere is a set of worlds whose members are more (or less) as similar to a given world than the non-members. Any condition on spheres – their being nested, centered around a world, closed under unions, etc.⁹ – is thus justified on the basis of corresponding conditions on the comparative similarity relation.

Furthermore, as Fine (1975, p. 457) pointed out, the formulation in terms of \leq_i has the advantage of being first-order, while the formulation in terms of spheres presupposes an assignment to each world of a set of sets of worlds. Use of \leq_i would also allow the substitution of the intensional language of modal and counterfactuals operators with an extensional language containing just first-order quantifications over worlds. In the same way in which modal languages

⁸ On this see (Bennett, 2003, p. 168–171).

⁹ Cf. Lewis (1973a, p. 14).

containing \Box and \Diamond could be substituted by first-order non-modal languages containing quantifiers over worlds and a group of new interpreted predicates (see Lewis, 1968), counterfactual languages containing $[\]\rightarrow$ and $\Diamond\rightarrow$ could be substituted by first-order non-modal languages containing the interpreted predicates R (for "... is accessible from ...") and \leq_i (for "... is as similar to ... as ...").

There is, however, at least an expository advantage that spheres have over comparative similarity. By using spheres, the difference between counterfactual conditionals and strict conditionals, as I said, a benchmark of Lewis's analysis, could very vividly (almost graphically) be made.¹⁰ To a (constantly) strict conditional there is an assignment to each i of a single sphere of accessibility, while to a variably strict conditional (i.e., a counterfactual) there is an assignment to each i of a system of spheres. A counterfactual is non-vacuously true in case the corresponding material conditional is true in every world belonging to at least one sphere around i ; on the contrary, the strict conditional is true if the corresponding material conditional is true in every world accessible-to- i . To the counterfactuals is assigned a structured space of metaphysical accessibility, whose structure is given by the similarity relation dividing this space into spheres; the space of accessibility of strict conditionals is instead simply left unstructured. The counterfactual "if the bush had not been there, the rock would have ended in the lake" is true in the actual world, even if there is an accessible world where the bush is not there and the rock does not end in the lake; what is needed for the counterfactual to be true is rather a special sub-group of accessible-to- i worlds such that in every world where the bush is not there, the rock ends in the lake.

Of course, we can tweak accessibility relations to detect just those worlds where the corresponding material conditional is true. Restriction of the accessibility relation used to evaluate strict conditionals could be obtained in various ways. One way would be to make the accessibility relation contextually determined. This strategy is at the heart of quite recent attempts to analyze counterfactuals as kinds of highly context-sensitive (non-variably) strict conditionals. According to Fintel (2001), for example, in the case of counterfactuals, the determination of the accessibility relation useful to

¹⁰ See Lewis (1973a, p. 11).

evaluate the conditional, what he calls “the modal horizon”, is contextually determined and dynamically changes throughout a discourse. Relativization of the accessibility relation to dynamically determined modal horizons could be quite useful to explain certain linguistic phenomena involving counterfactuals, like sensitivity to order of utterance. If in a conversation I first utter the sentence “If I had struck the match, it would have lit” and then “If I had struck the match in outer space, it would have lit”, the first sentence seems to be true and the second false. But if, in another conversation, the order of utterance is reversed, the latter sentence still seems false, but now the former also seems false.¹¹ This is because the relevant modal horizon for the conversation is that of the first uttered counterfactual and the horizon of the second counterfactual is wider than the one of the first counterfactual. Being wider, it contains worlds (which remains accessible throughout a discourse) where I strike the match, but it does not light and where therefore the material conditional is false.

It is difficult to say whether these contemporary and sophisticated contextual approaches represent a real novelty or are simply notational variants of the “classic” approaches given by Lewis and Stalnaker. The author of this commentary suspects that they are. Already in the seventies, however, just envisaging the (at the time open) possibility of treating counterfactuals as contextually strict conditionals, Lewis defined such strategies “defeatist” because, he wrote: «It consigns to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket» (Lewis, 1973a, p. 13). This quotation reveals quite nicely and from a quite different angle, in what consists Lewis’s often celebrated “systematic philosophy”; Lewis’s systematicity consists not only in his capacity to clarify a great number of problems belonging to different philosophical areas, but also, and more importantly, in his capacity to tackle such problems by means of robust theorizing.

¹¹ This is an example taken from Sider (2010, p. 225).

REFERENCES

- Barker, S. (2011). Can Counterfactuals really be about Possible Worlds? *Nous*, 45(3), 557–576.
- Bennett, J. (1974). Counterfactuals and Possible Worlds. *Canadian Journal of Philosophy*, 4(2), 381–402.
- Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
- Ellis, B., Jackson, F., & Pargitter, R. J. (1977). An Objection to Possible Worlds Semantics for Counterfactuals Logic. *Journal of Philosophical Logic*, 6, 35–57.
- Fine, K. (1975). Critical Notice of Counterfactuals. *Mind*, 84 (335), 451–458.
- Fintel, K. von. (2001). Counterfactuals in a Dynamic Context. In M. Kenstowicz (Ed.), *Ken Hale: a Life in Language*. Cambridge, MA: MIT Press, 123–152.
- Kratzer, A. (1981). Partition and Revision: the Semantics of Counterfactuals. *Journal of Philosophical Logic*, 10, 201–216.
- Lewis, D. K. (1968). Counterpart Theory and Quantified Modal Logic. *Journal of Philosophy*, LXV(5), 113–126.
- Lewis, D. K. (1971). Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, 37, 74–85.
- Lewis, D. K. (1973a). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. K. (1973b). Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*, 2, 418–446.
- Lewis, D. K. (1977). Possible-Worlds Semantics for Counterfactuals: a Rejoinder. *Journal of Philosophical Logic*, 6, 359–363.
- Lewis, D. K. (1979). Counterfactual Dependence and Time's Arrow. *Nous*, 13(4), 455–476.
- Lewis, D. K. (1981). Ordering Semantics and Premise Semantics for counterfactuals. *Journal of Philosophical Logic*, 10, 217–234. (Reprinted as ch. 6 of Lewis (1998))

- Lewis, D. K. (1998). *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.
- Loewer, B. (1976). Counterfactuals with Disjunctive Antecedents. *Journal of Philosophy*, 73(16), 531–536.
- Nute, D. (1976). David Lewis and the Analysis of Counterfactuals. *Nous*, 10(3), 355–361.
- Prior, A. N. (1968). Egocentric Logic. *Nous*, 2, 191–207. (Reprinted as ch. XVII of Prior (2003))
- Prior, A. N. (2003). *Papers on Time and Tense* (new edition). Oxford: Oxford University Press.
- Sider, T. (2010). *Logic for Philosophy*. Oxford: Oxford University Press.
- Stalnaker, R. (1968). A Theory of Conditionals. In N. Rescher (Ed.), *Studies in Logical Theory*. Oxford: Blackwell, 98–112.
- Stalnaker, R. (1978). A Defense of Conditional Excluded Middle. In W. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–103). Dordrecht: Reidel.
- Stalnaker, R., & Thomason, R. (1970). A Semantic Analysis of Conditional Logic. *Theoria*, 36, 23–42.
- Todd, W. (1964). Counterfactual Conditionals and the Presuppositions of Induction. *Philosophy of Science*, 31(2), 101–110.
- Williamson, T. (2008). *The Philosophy of Philosophy*. Oxford: Blackwell.