

ROBERT C. STALNAKER

## A DEFENSE OF CONDITIONAL EXCLUDED MIDDLE\*

This paper is a polemic about a detail in the semantics for conditionals. It takes for granted what is common to semantic theories proposed by David Lewis,<sup>1</sup> John Pollock,<sup>2</sup> Brian Chellas,<sup>3</sup> and myself and Richmond Thomason<sup>4</sup> in order to focus on some small points of difference between the theory I favor and the others. I will sketch quickly and roughly the general ideas which lie behind all of these theories, and the common semantical framework in which these ideas are developed. Then I will describe the divergences between my theory and the others – I will focus on the difference between my theory and the one favored by Lewis – and argue that my theory gives a better account of the way conditionals work in natural language.

The differences between the theories I will be comparing may seem small and unimportant. Does it really matter very much whether we conclude that conditionals like *If Bizet and Verdi had been compatriots, Bizet would have been Italian* are false or (as I will suggest) neither true nor false? This judgment may not be important in itself, but it is not an isolated judgment. Conditionals interact with negation, quantifiers, modal auxiliaries like *may* and *might*, adverbs like *even*, *only* and *probably*. Small differences among analyses of conditionals may have consequences for many complex constructions involving conditionals. A small distortion in the analysis of the conditional may create spurious problems with the analysis of other concepts. So if the facts about usage favor one among a number of subtly different theories, it may be important to determine which one it is.

All of the semantic analyses of conditional logic that I have in mind are given within the possible worlds framework. All begin with the general idea that a counterfactual conditional is true in the actual world if and only if the consequent is true in some possibly different possible world or worlds. The world or set of worlds in which the consequent is said to be true is determined by the antecedent. These must be possible worlds in which the antecedent is true, and which are otherwise minimally different from the actual world. To make sense of this idea, the semantic theory needs a semantic determinant which selects the minimally different world or worlds, or which orders the possible worlds with respect to their comparative similarity

to the actual world. Truth conditions for conditionals are given relative to such a semantic determinant.

The semantic determinant for the account I prefer is a world selection function: a function  $f$  which takes a proposition and a possible world into a possible world.<sup>5</sup> A conditional, if  $A$  then  $B$ , is then said to be true in a world  $i$  if and only if  $B$  is true in  $f(A, i)$  – the possible world which is the value of the function for arguments  $A$  and  $i$ . Various constraints are placed on the selection function, constraints which are motivated by the intuitive idea that the nearest, or least different, world in which antecedent is true is the one that should be selected. For example, it is required that the world selected relative to proposition  $A$  be an  $A$ -world – a possible world in which  $A$  is true ( $f(A, i) \in A$ ). And if the actual world meets this condition, it is required that it be selected. (If  $i \in A$ , then  $f(A, i) = i$ ).

David Lewis's theory of conditionals is formulated in terms of a different semantic determinant. It states truth conditions for conditionals in terms of a three place comparative similarity relation instead of a selection function. Let  $C_i(j, k)$  mean that  $j$  is more similar to  $i$  than  $k$  is to  $i$ . For any fixed  $i$ , the relation is assumed to be transitive and connected, and so to determine a weak total ordering of all possible worlds with respect to each possible world. A counterfactual, if  $A$ , then  $B$ , is then said by Lewis's theory to be true if and only if there is an  $A$ -world  $j$  such that  $B$  is true in it, and in all  $A$ -worlds which are at least as similar to  $i$  as  $j$ .<sup>6</sup>

Part of the difference between the two theories of conditionals that I have sketched is superficial. The theory I favor could have been formulated in terms of a comparative similarity relation instead of a selection function, and a comparative similarity relation is definable in terms of the selection function as follows:  $C_i(j, k)$  if and only if for some proposition  $A$  such that both  $j$  and  $k$  are members of  $A$ ,  $f(A, i) = j$ . It can be shown, first, that this defined relation meets all the conditions Lewis imposes on his comparative similarity relation, and second, that Lewis's truth conditions, applied to the defined relation, coincide with the truth conditions given in my theory. The theories are not equivalent since the defined comparative similarity relation necessarily has properties beyond those imposed by Lewis's theory; specifically, the comparative similarity relation defined in terms of a selection function determines not just a weak total ordering, but a well ordering of all possible worlds with respect to each possible world. So my theory, formulated in terms of a comparative similarity relation, is a special case of Lewis's.

Lewis identifies two assumptions about the comparative similarity relation which my theory makes and this does not: he calls them the limit assumption

and the uniqueness assumption. The first is the assumption that for every possible world  $i$  and non-empty proposition  $A$ , there is at least one  $A$ -world minimally different from  $i$ . The second is the assumption that for every world  $i$  and proposition  $A$  there is at most one  $A$ -world minimally different from  $i$ . Lewis's theory, with the addition of these two assumptions, is essentially equivalent to mine.

Each of the two assumptions about the comparative similarity relation corresponds to an entailment principle in the semantics for conditionals. To accept the limit assumption is to accept the following consequence condition for conditionals: for any class of propositions  $\Gamma$  and propositions  $A$  and  $C$ , if  $\Gamma$  semantically entails  $C$ , then  $\{A > B : B \in \Gamma\}$  semantically entails  $A > C$ . To accept in addition the uniqueness assumption is to accept the validity of the principle of conditional excluded middle:

$$\Vdash (A > C) \vee (A > \sim C).$$

Lewis argues that it is not reasonable to make the two assumptions which distinguish his theory from mine. I will argue that one of the assumptions is reasonable to make, and that the other need not be made in application. I will also discuss a number of examples which I think tend to show that the analysis I have proposed gives a better account of the phenomena.

Let me look first at the uniqueness assumption. This is the assumption which rules out ties in similarity. It says that no distinct possible worlds are ever equally similar to any given possible world. That is, without a doubt, a grossly implausible assumption to make about the kind of similarity relation we use to interpret conditionals, and it is an assumption which the abstract semantic theory that I want to defend does make. But like many idealized assumptions made in abstract semantic theory, it may be relaxed in the application of the theory. In general, to apply a semantic theory to the interpretation of language as it is used, one need not assume that every semantic determinant is completely and precisely defined. In application, domains of individuals relative to which quantifiers are interpreted, sets of possible worlds relative to which modal auxiliaries are interpreted, propositional functions used to interpret predicates all may admit borderline cases even though the abstract semantic theory assumes well defined sets with sharp boundaries. To reconcile the determinacy of abstract semantic theory with the indeterminacy of realistic application, we need a general theory of vagueness. But given such a theory, we can reconcile the uniqueness assumption, as an assumption of the abstract semantics for conditionals, with the

fact that it is unrealistic to assume that our conceptual resources are capable of well ordering the possible worlds.

The theory of vagueness that I will recommend is the theory of supervaluations first developed by Bas Van Fraassen.<sup>7</sup> The main idea of this theory is this: any partially defined semantic interpretation will correspond to a class of completely defined interpretations – the class of all ways of arbitrarily completing it. For example, a partial ordering will correspond to a class of total orderings, and a domain with fuzzy borders will correspond to a class of domains with sharp borders. The theory of supervaluations defines the truth values assigned by partial interpretations in terms of the corresponding class of complete, two-valued, *classical valuations*. In this way, it explains the values under partial interpretations in terms of the kind of valuations assumed by idealized abstract semantic theories. A sentence is *true* according to a supervaluation if and only if it is true on *all* corresponding classical valuations, *false* if and only if it is false on *all* corresponding classical valuations and neither true nor false if it is true on some of the classical valuations and false on others.

Using the method of supervaluations, we may acknowledge, without modifying the abstract semantic theory of conditionals, that the selection functions that are actually used in making and interpreting counterfactual conditional statements correspond to orderings of possible worlds that admit ties and incomparabilities. In doing this, we are not resorting to an *ad hoc* device to save a theory, since the method of supervaluations, or some account of semantic indeterminacy, is necessary anyway to account for pervasive semantic underdetermination in natural language. Whatever theory of conditionals one favors, one must admit that vagueness is particularly prevalent in the use of conditional sentences.<sup>8</sup>

What effect does the recognition of indeterminacy by the introduction of supervaluations have on the *logic* of conditionals? None at all: it is one of the virtues of this method of treating semantic indeterminacy that it leaves classical two-valued logic virtually untouched.<sup>9</sup> Classical logical truths are true in *all* classical valuations, and so will be true in all classical valuations defined by any partial interpretation. Therefore, they will be true in all supervaluations. Also, since classical valuations are themselves special cases of supervaluations, any sentence true in all supervaluations will be true on all classical valuations. So, whatever the details of the particular classical semantic theory, the concept of logical truth defined by it will not be changed by the introduction of supervaluations.

For example, in the conditional logic C2 (the logic of the theory I am

defending), the principle of conditional excluded middle,  $(A > B) \vee (A > \sim B)$ , remains valid when supervaluations are added, even though there may be cases where neither  $(A > B)$  nor  $(A > \sim B)$  is true. It may be that neither disjunct is made true by every arbitrary extension of a given partial interpretation, but it will always be that each arbitrary extension makes true one disjunct or the other. The theory of supervaluations, applied to this logic of conditionals, gives the principle of conditional excluded middle the same status as it gives the simple principle of excluded middle.  $(B \vee \sim B)$  is logically true even though sometimes neither  $B$  nor  $\sim B$  is true.

My aim so far, in this defense of my analysis of conditionals against its close relatives, has been simply to neutralize one important objection to the analysis: that it makes an implausible assumption about our conceptual resources, the assumption that we need a well ordering of all possible worlds with respect to each possible world in order to interpret conditional statements. I have argued that in the context of a general recognition of semantic indeterminacy, the dispute over the uniqueness assumption should be regarded not as a dispute about how much and what kind of structure there is in the actual contextual parameter we use to interpret conditionals, but rather a dispute about what degree and kind of structure that parameter is aiming at: about what would count as a determinate complete interpretation. In practice, what the issue comes down to is a disagreement about whether certain counterfactual conditionals are false or neither true nor false, and about whether certain inferences involving conditionals are valid.

Before looking at some examples of inferences and judgments which I think support the analysis I have proposed, I should point out, as Lewis does, that the limit assumption cannot be neutralized by the introduction of supervaluations in the same way as the uniqueness assumption. In my defense of the principle of conditional excluded middle, I shall take for granted that this assumption is a reasonable assumption to make. Later, I will explain and defend this decision.

Let us look at some examples. I will begin with a familiar pair of counterfactual conditionals first discussed in 1950 by W. V. Quine:<sup>11</sup>

If Bizet and Verdi had been compatriots, Bizet would have been Italian.

If Bizet and Verdi had been compatriots, Verdi would have been French.

These examples have been taken, in the context of possible worlds analyses of

conditionals, to illustrate the possibility of virtual ties in closeness of counterfactual possible worlds to the actual world. Worlds in which Bizet and Verdi are both French or both Italian, it seems plausible to assume, are more like the actual world than worlds in which both are Argentinian or Japanese. But there is no apparent reason to favor a world in which both are French over one in which both are Italian, or vice versa. This seems right; it would be arbitrary to require a choice of one of the above counterfactuals over the other, but as we have seen, this is not at issue. What is at issue is what conclusion about the truth values of the counterfactuals should be drawn from the fact that such a choice would be arbitrary. On Lewis's and Pollock's analyses, both counterfactuals are false. On the analysis I am defending, both are indeterminate – neither true nor false. It seems to me that the latter conclusion is clearly the more natural one. I think most speakers would be as hesitant to deny as to affirm either of the conditionals, and it seems as clear that one cannot deny them both as it is that one cannot affirm them both. Lewis seems to agree that unreflective linguistic intuition favors this conclusion. He writes:

Given Conditional Excluded Middle, we cannot truly say such things as this:

*It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian . . .*

I want to say this, and think it is probably true; my own theory was designed to make it true. But offhand, I must admit, it does sound like a contradiction. Stalnaker's theory does, and mine does not, respect the opinion of any ordinary language speaker who cares to insist that it is a contradiction.<sup>12</sup>

Lewis goes on to say that the cost of respecting this ‘offhand opinion’ is too great, but as I have argued, the introduction of supervaluations avoid the need to pay the main cost that he has in mind.

Quine originally presented this example, not to defend one analysis of counterfactuals against another, but to create doubt about the possibility of any acceptable analysis. “It may be wondered, indeed,” he writes introducing the two Bizet–Verdi counterfactuals, “whether any really coherent theory of the counterfactual conditional of ordinary usage is possible at all, particularly when we imagine trying to adjudicate between such examples as these.”<sup>13</sup> There is a problem, Quine suggests, because we are required to *adjudicate* between the two. But why are we required to adjudicate? The argument is implicit, but I suspect that what Quine had in mind might be reconstructed as follows: “It is clear that if Bizet and Verdi had been compatriots, then

either Bizet would have been Italian, or Verdi French. But then one (and only one) of the two counterfactuals, *If Bizet and Verdi had been compatriots, Verdi would have been French*, or *If Bizet and Verdi had been compatriots, Bizet would have been Italian* must be true. How are we to adjudicate between them?" The crucial inference in this reconstructed argument relies on the distribution principle,  $(A > (B \vee C))$ , therefore  $(A > B) \vee (A > C)$ , a rule of inference that is equivalent, in the context of conditional logic, to the principle of conditional excluded middle. Quine takes for granted, by tacitly using this principle of inference, that a counterfactual antecedent purports to represent a unique, determinate counterfactual situation. It is because counterfactual antecedents *purport* to represent unique possible situations that examples which show that they may fail to do so are a problem. One should respond to the problem, I think, not by revising the truth conditions for conditionals so that it does not arise, but rather by recognizing what we must recognize anyway: that in application there is great potential for indeterminacy in the truth conditions for counterfactuals.

The failure of the distribution principle we have been discussing is a symptom of the fact that, on Lewis's analysis, the antecedents of conditionals act like necessity operators on their consequents. To assert *if A, then B* is to assert that *B* is true in every one of a set of possible worlds defined relative to *A*. Therefore, if this kind of analysis is correct, we should expect to find, when conditionals are combined with quantifiers, all the same scope distinctions as we find in quantified modal logic. In particular, corresponding to the distinction between  $(A > (B \vee C))$  and  $((A > B) \vee (A > C))$  is the quantifier scope distinction between  $(A > (\exists x)Fx)$  and  $(\exists x)(A > Fx)$ . On Lewis's account, even when the domain of the quantifier remains fixed across possible worlds, there is a semantically significant difference between these two formulas of conditional logic, and we should expect to find scope ambiguities in English sentences that might be formalized in either way.

Before seeing if such ambiguities are found in conditional statements, let us look at a case where the ambiguity is uncontroversial. The following dialogue illustrates a quantifier scope ambiguity in a necessity statement:

- X: President Carter has to appoint a woman to the Supreme Court.
- Y: Who do you think he has to appoint?
- X: He doesn't have to appoint any particular woman; he just has to appoint some woman or other.

Y, perversely, gives the quantified expression, *a woman*, wide scope in interpreting X's statement. X, in his response to Y's question, shows that he meant the quantifier to have narrow scope. The difference is, of course, not a matter of whether the speaker *knows* who the woman is. X might have meant the wide scope reading – the reading Y took it to have – and still not have known who the woman is. In that case, his response to Y's question would have been something like this:

X: I don't know; I just know it's a woman that he has to appoint.

In this alternative response, the appropriateness of the question is not challenged. X just confesses inability to answer it. This alternative reply is appropriate only if the speaker intended the wide scope reading.

Now compare a parallel dialogue beginning with a statement that is clearly unambiguous:

X: President Carter will appoint a woman to the Supreme Court.

Y: Who do you think he will appoint?

X: He won't appoint any particular woman; he just will appoint some woman or other.

X's response here is obviously nonsense.<sup>14</sup> There must be a particular person that he will appoint, although the speaker need not know who it is. If he does not know, the analogue of the alternative response is the one he will give:

X: I don't know; I just know it's a woman that he will appoint.

Now look at a corresponding example with a counterfactual conditional and consider which of the above examples it most resembles.

X: President Carter would have appointed a woman to the Supreme Court last year if there had been a vacancy.

Y: Who do you think he would have appointed?

X: He wouldn't have appointed any particular woman; he just would have appointed some woman or other.

Or, the alternative response:

X: I don't know; I just know it's a woman that he would have appointed.

If Lewis's analysis is correct, you should perceive a clear scope ambiguity in X's original statement. Y's question, and X's alternative response, should seem appropriate only when the strong, wide scope reading was the intended one. I do not see an ambiguity; X's first response seems as bad, or almost as bad, as the analogous response in the future tense case. And I do not think there is any interpretation for which Y's question shows a misreading of the statement.

There is still, on the analysis I am defending, a relevant difference between the future tense example and the counterfactual example – a way in which the latter is more like the necessity example than the former. In the future tense case, if X's initial statement is true, then it follows that Y's question has a correct answer, even if no one knows what it is.<sup>15</sup> But in the counterfactual case, this need not be true. There may be no particular woman of whom it is true to say, President Carter would have appointed *her* if a vacancy had occurred. This is possible because of the possibility of underdetermination, but it does not imply that there is any scope ambiguity in the original statement. The situation is analogous to familiar examples of underdetermination in fiction. The question, *exactly how many sisters and cousins did Sir Joseph Porter have?* may have no correct answer, but one who asks it in response to the statement that his sisters and cousins numbered in the dozens does not exhibit a misunderstanding of the semantic structure of the statement.<sup>16</sup> It is not surprising, from the point of view of the analysis I am defending, that the possible situations determined by the antecedents of counterfactual conditionals are like the imaginary worlds created by writers of fiction. In both cases, one purports to represent and describe a unique determinate possible world, even though one never really succeeds in doing so.

As we have seen, Lewis agrees that the analysis I am defending respects, as his does not, certain "offhand" opinions of ordinary language speakers. He argues that the cost of respecting these opinions is too high. But Lewis also recognizes – in fact emphasizes – that counterfactual conditionals are frequently vague, and he adopts the same account of vagueness that allows the analysis I am defending to avoid implausible assumptions about our conceptual resources. Why, then, does Lewis still reject this analysis? "Two major problems remain," he writes. "First, the revised version [C2, revised by the introduction of supervaluations] still depends for its success on the Limit Assumption . . . Second, the revised version still gives us no 'might' counterfactual."<sup>17</sup> I will conclude the defense of my analysis by responding to these two further problems. I will first argue that the limit assumption, unlike the

uniqueness assumption, is a plausible assumption to make about the orderings of possible worlds that are determined by our conceptual resources, and that the rejection of this assumption has some bizarre consequences. Second, I will say how I think *might* conditional should be understood, and argue that Lewis's analysis fails to give a satisfactory account of the relation between *might* and *would* conditionals.

When the uniqueness assumption fails to hold for a comparative similarity relation among possible worlds, then the selection function in terms of which conditionals are interpreted in C2 is left underdetermined by that relation. Many selection functions may be compatible with the comparative similarity relation, and it would be arbitrary to choose one over the others. But if the limit assumption were to fail, there would be too few candidates to be the selection function rather than too many. Any selection function would be forced to choose worlds which were less similar to the actual world than other eligible worlds. This is why the supervaluation method does not provide a way to avoid making the limit assumption.

The limit assumption implies that for any proposition *A* which is possibly true, there is a non-empty set of *closest* worlds in which *A* is true. Is this a plausible assumption to make about the orderings of possible worlds which are relevant to the interpretation of conditionals? If one were to begin with a concept of overall similarity among possible worlds which is understood independently of its application to the interpretation of conditionals, this clearly would be an arbitrary and unjustified assumption. Nothing that I can think of in the concept of similarity, or in the respects of similarity that are relevant, would motivate imposing this restrictive formal structure on the ordering determined by a similarity relation. But, on the other hand, if one begins with a selection function and thinks of the similarity orderings as induced by the selection function, the assumption will not be arbitrary or unmotivated: the fact that it holds will be explained by the way in which the orderings are determined. To the extent that an intuitive notion of similarity among possible worlds plays a role, it is a device used for the purpose of selecting possible worlds. Given this rule, it is not unreasonable to require that the way respects of similarity are weighed should be such as to make selection possible.

Even if we take the selection function as the basic primitive semantic determinant in the analysis of conditionals, we still must rely on some more or less independently understood notion of similarity or closeness of worlds to describe the intuitive basis on which the section is made. The intuitive idea is something like this: the function selects a possible world in which the

antecedent is true, but which otherwise is as much like the actual world, in relevant respects, as possible. So, one might argue, we still need to give some justification for the limit assumption. How can we be sure that it will be possible to select a world, or a set of worlds, on this basis? Consider one of Lewis's examples: suppose that this line, ——, were more than one inch long. (The line is actually a little less than one inch long.) Every possible world in which the line is more than one inch long is one in which it is longer than it needs to be in order to make the antecedent true. It appears that the intuitive rule to select a world that makes the minimal change in the actual world necessary to make the antecedent true is one that cannot be followed.

The qualification in the intuitive rule that is crucial for answering this objection is the phrase, 'in relevant respects'. The selection function may ignore respects of similarity which are not relevant to the context in which the conditional statement is made. Even if, in terms of some general notion of overall similarity, *i* is clearly more similar to the actual world than *j*, if the ways in which it is more similar are irrelevant, than *j* may be as good a candidate for selection as *i*.<sup>18</sup> In the example, it may be that what matters is that the line is more than one inch long, and still short enough to fit on the page. In this case, all lengths over one inch, but less than four or five inches will be equally good.

But what about a context in which every millimeter matters? If relative to the issue under discussion, every difference in length is important, then it is just inappropriate to use the antecedent, *if the line were more than an inch long*. This would, in such a context, be like using the definite description, *the shortest line longer than one inch*. The selection function will be undefined for such antecedents in such contexts.

To summarize: from a naive point of view, nothing seems more obvious than that a conditional antecedent asks one to imagine a possible situation in which the antecedent is true. To say *if pigs could fly* is to envision a situation, or a kind of situation, in which pigs can fly. This is the motivation for making a selection function the basic semantic determinant. But it is equally obvious that the basis for the selection is some notion of similarity or minimal difference between worlds. The situations in which pigs can fly that you are asked to envision are ones which are as much as possible like the actual situation. The problem is that it is theoretically possible for these two intuitions to clash. There could be similarity relations and antecedents relative to which selection would be impossible. Lewis's response to this problem is to generalize the analysis of conditionals so that selection is no longer essential. The

alternative response, which seems to me more natural, is to exclude as inappropriate antecedents and contexts in which the relevant similarity relation fails to make selection possible. Given that the appropriate similarity notion is one that may ignore irrelevant respects of similarity altogether, this exclusion should not be unreasonably restrictive.

I think a closer look at our example will support the conclusion that Lewis's response to the problem is intuitively less satisfactory than the one I am suggesting. On Lewis's analysis, every conditional of the following form is true: *If the line had been more than one inch long, it would not have been x inches long*, where  $x$  is any real number. This implies (given Lewis's analysis of *might* conditionals) that there is no length such that the line might have had that length if it had been more than one inch long. Yet, the line might have been more than an inch long, and if it had been, it would have had some length or other. The point is not just that there is no particular length that the line *would* have had. More than this, there is not even any length that it *might* have had. That conclusion seems, intuitively, to contradict the assumption that the line might have been more than one inch long, yet on Lewis's account, both the conclusion and the assumption may be true.<sup>19</sup>

The second problem that Lewis finds with the analysis I am defending is that it gives us no account of the *might* conditional. Lewis analyzes this kind of conditional in terms of his *would* conditional as follows: the *might* conditional, *if A, it might be that B*, is true if and only if the *would* conditional, *if A, it would be that not-B*, is false. In Lewis's notation,  $(A \diamondrightarrow B) =_{\text{df}} \sim(A \Box \rightarrow \sim B)$ . Ordinary counterfactuals express a kind of variable necessity on the consequent, according to Lewis. *Might* counterfactuals express the corresponding kind of possibility.

It is clear that this definition conflicts with the analysis of conditionals I am proposing, since the principle of excluded middle, together with Lewis's definition of *might* conditionals, implies that a *might* conditional is equivalent to the corresponding *would* conditional. This is obviously an unacceptable conclusion, so if Lewis's definition is supported by the facts, this counts against an analysis that validates the principle of conditional excluded middle. But I will argue that Lewis's definition has unacceptable consequences, and that a more satisfactory analysis, compatible with the principle of excluded middle, can be given.

Note that Lewis's definition treats the apparently complex construction, *if . . . might*, as an idiom instead of analyzing it in terms of the meanings of *if* and *might*. This is not a serious defect, but it would be methodologically preferable – less *ad hoc* – to explain the complex construction in terms of

its parts. So I will begin by looking at uses of *might* outside of conditional contexts, and then consider what the result would be of combining the account of *might* suggested by those uses with our analysis of *if*.

*Might*, of course, expresses possibility. *John might come to the party* and *John might have come to the party* each say that it is possible, in some sense, that John come, or have come, to the party. I think the most common kind of possibility which this word is used to express is epistemic possibility. Normally, a speaker using one of the above sentences will be saying that John's coming, or having come, to the party is compatible with the speaker's knowledge. But *might* sometimes expresses some kind of non-epistemic possibility. *John might have come to the party* could be used to say that it was within John's power to come, or that it was not inevitable that he not come. The fact that the sentence, *John might come to the party, although he won't*, is somewhat strange indicates that the epistemic sense is the dominant one for this example. There is no strangeness in *John could come to the party, although he won't*. The epistemic interpretation seems less dominant in the past tense example: *John might have come to the party, although he didn't* is not so strange.

What I want to suggest is that *might*, when it occurs in conditional contexts, has the same range of senses as it has outside of conditional contexts. Normally, but not always, it expresses epistemic possibility. The scope of the *might*, when it occurs in conditional contexts is normally the whole conditional, and not just the consequent. This claim may seem *ad hoc*, since the surface form of English sentences such as *If John had been invited, he might have come to the party* certainly suggests that the antecedent is outside the scope of the *might*. But there are parallel constructions where the wide scope analysis is uncontroversial. For example, *If he is a bachelor, he must be unmarried*. Also, the wide scope interpretation is supported by the fact that *might* conditionals can be paraphrased with the *might* preceding the antecedent: *It might be that if John had been invited, he would have come to the party*.

The main evidence that *might* conditionals are epistemic is that it is unacceptable to conjoin a *might* conditional with the denial of the corresponding *would* conditional. This fact is also strong evidence against Lewis's account, according to which such conjunctions should be perfectly normal. On Lewis's account, *might* conditionals stand to *would* conditionals as ordinary *might* stands to *must*. There is no oddity in denying the categorical claim, *John must come to the party*, while affirming that he might come. But it would sound strange to deny that he would have come if he had been invited, while affirming that he might have come.

Consider a variation on the Supreme Court appointment dialogues discussed above:

- X: Does President Carter have to appoint a woman to the Supreme Court?
- Y: No, certainly not, although he might appoint a woman.

This is perfectly okay. Now compare:

- X: Would President Carter have appointed a woman to the Supreme Court last year if a vacancy had occurred?
- Y: No, certainly not, although he might have appointed a woman.

On Lewis's analysis, one should expect Y's second response to be as acceptable as his first.

One should not conclude from the conflict between the denial of the *would* conditional and the affirmation of the *might* conditional that these two statements contradict each other. To draw that conclusion would be to confuse pragmatic with semantic anomaly. On the epistemic interpretation, what Y does is to represent himself as knowing something by asserting it, and then to deny that he knows it. The conflict is thus like Moore's paradox, rather than like a contradictory assertion.

My account predicts, while Lewis's does not, that the example given above should seem Moore-paradoxical. I think it is clear that the evidence supports this prediction. Rich Thomason has pointed out that there are also examples of the reverse: cases for which Lewis's account predicts a Moore's paradox, while mine does not. Here too, I think it is clear that the evidence supports my account. Consider any statement of the form *If A, it might be that not-B, although I believe that if A then it would be that B*. Lewis's definition implies that such a statement is equivalent to a statement of the form *Not-C, although I believe that C*, and so implies that such a statement should seem Moore-paradoxical. But there is nothing wrong with saying *John might not have come to the party if he had been invited, but I believe he would have come*. As my account predicts, this statement is as acceptable as the parallel statement with non-conditional *might*: *John might not come to the party, although I believe that he will*.<sup>20</sup>

Lewis considers and rejects a number of alternatives to his analysis of *might* counterfactuals, including an analysis which treats them as *would* counterfactuals prefixed by an epistemic possibility operator. Here is his

counterexample: Suppose there is in fact no penny in my pocket, although I do not know it since I did not look. “Then ‘*If I had looked, I might have found a penny*’ is plainly false.” But it is true that it might be, for all I know, that I would have found a penny if I had looked.<sup>21</sup>

I do not think that Lewis’s example is *plainly* false since the epistemic reading, according to which it is true, seems to be one perfectly reasonable interpretation of it. I can also see the non-epistemic sense that Lewis has in mind, but I think that this sense can also be captured by treating the *might* as a possibility operator on the conditional. Consider not what is, in fact, compatible with my knowledge, but what would be compatible with it if I knew all the relevant facts. This will yield a kind of quasi-epistemic possibility – possibility relative to an idealized state of knowledge. If there is some indeterminacy in the language, there will still remain some different possibilities, even after all the facts are in, and so this kind of possibility will not collapse into truth. Propositions that are neither true nor false because of the indeterminacy will still be possibly true in this sense. Because *if Bizet and Verdi had been compatriots, Verdi would have been French* is neither true nor false, *If Bizet and Verdi had been compatriots, Verdi might have been French* will be true in this sense of *might*.

Now this interpretation of *might* conditionals is very close to Lewis’s. It agrees with Lewis’s account that *If A, it might be that B* is true if and only if *If A it would be that not-B* is not true. But my explanation has the following three advantages over Lewis’s: First, it treats the *might* as a kind of possibility operator on the conditional – an operator that can also operate on other kinds of propositions – rather than treating *if . . . might* as a semantically unanalyzed unit. With Lewis’s analysis of the *would* conditional, this cannot be done. Second, it treats this particular kind of *might* as a special case of a more general analysis – one that includes the ordinary epistemic interpretation as another special case. Third, it explains, as Lewis’s analysis cannot, why it is anomalous to deny the *would* conditional while affirming the corresponding *might*.

It may seem strange that I have called the use of *might* which expresses semantic indeterminacy a quasi-epistemic use, but I think that there is a general tendency to use epistemic terminology to describe indeterminacy, and to think of indeterminacy as a limiting case of ignorance – the ignorance that remains after all the facts that are in. I will conclude with an example that illustrates this tendency, as well, I think, as the general point that we tend to think of counterfactual suppositions as determining a unique possible situation.

If President Kennedy had not been assassinated in 1963, would the United States have avoided the Vietnam debacle? It is a controversial question. We will probably never know for sure. If we could look back into the minds of President Kennedy and his advisors, if we could learn all there is to learn about their policy plans and priorities, their expectations and perceptions, then maybe we could settle the question. But on the other hand, it could be that the answer turns on possible actions and events which are not determined by facts about the actual situation. In that case, we *could* never know, no matter how much we learned. In that case, even an omniscient God wouldn't know.<sup>22</sup> If this is true, then our failure to answer the question is not really an epistemic limitation, but we still use the language of knowledge and ignorance to characterize it. Even when we recognize that such a question really has no answer, we continue to talk and think as if there were an answer that we cannot know. This is, I think, because we tend to think of the counterfactual situations determined by suppositions as being as complete and determinate as our own actual world.

*Cornell University*

#### N O T E S

\* This paper was completed while I was a National Endowment for the Humanities Fellow at the Center for Advanced Study in the Behavioral Sciences. I am grateful both to NEH and to the Center for providing time off from teaching and an idyllic setting in which to work. I am also grateful to David Lewis and Richmond Thomason for valuable conversation and correspondence over the years on the topic of the paper.

<sup>1</sup> [7].

<sup>2</sup> [11].

<sup>3</sup> [2].

<sup>4</sup> [13] and [14].

<sup>5</sup> Propositions, in the formal semantic theory, are identified with sets of possible worlds, or equivalently, with functions from possible worlds into truth values. If  $A$  is a proposition and  $i$  is a possible world, then ' $i \in A$ ' means that  $A$  is true at  $i$ .

In [13] and [14], the selection function was a function from sentences rather than propositions, but it was assumed that this function assigned the same values to sentences which expressed the same proposition.

<sup>6</sup> This is the truth condition only on the assumption that there are some  $A$ -worlds, that is, that the antecedent is not necessarily false. Both Lewis's theory and mine stipulate that a counterfactual is vacuously true when its antecedent is necessarily false.

<sup>7</sup> See [8]. This theory, and theories based on the same general idea, have been widely applied. See, for example, [3], [4], [6], [9], [15], and [16].

<sup>8</sup> Lewis emphasizes the vagueness and context dependence of counterfactuals in his discussion of resemblance. See [7], pp. 91–95.

<sup>9</sup> This statement needs qualification when the language has a truth operator or other resources capable of expressing the fact that a statement lacks a truth value. In such cases, there will be a divergence between the validity of an argument from  $A$  to  $B$  and the logical truth of a statement  $A \supset B$ . See [18].

<sup>10</sup> The relationship between the semantics for C2 with supervaluations and Lewis's semantics is explored by Bas Van Fraassen in [17]. A C2 model with supervaluations is defined as a family of determinate C2 models. Van Fraassen shows that Lewis models which satisfy the limit assumption are equivalent to what he calls *regular* families of C2 models – families meeting a certain restriction. I believe that if one drops the regularity restriction, then the same equivalence holds between families of C2 models and the models of the semantic theory of simple subjunctive conditionals favored by John Pollock. For an exposition of Pollock's theory, see [11].

<sup>11</sup> [12], p. 14.

<sup>12</sup> [7], p. 80.

<sup>13</sup> [12], p. 14.

<sup>14</sup> Philosophers who deny truth to statements about future contingents may disagree. They may want to say that X's reply makes sense, and might be true. Readers who are inclined to this view may substitute a past tense example.

<sup>15</sup> I am assuming that statements about future contingents may be true. For one who treats future contingent statements as truth-valueless, and uses supervaluations to interpret truth-value gaps, the contrast between future tense and counterfactual examples that I am pointing to will disappear. See [16].

<sup>16</sup> This example from Gilbert and Sullivan's *H.M.S. Pinafore* is borrowed from David Lewis, [10].

<sup>17</sup> [7], pp. 82–83.

<sup>18</sup> David Lewis, in [8], suggests that the similarity ordering relevant to interpreting counterfactuals may, in some cases, give zero weight to some respects of similarity.

<sup>19</sup> See [5] and [11] for arguments against the limit assumption.

<sup>20</sup> This argument was given to me by Rich Thomason, private communication.

<sup>21</sup> [7], p. 80.

<sup>22</sup> Cf. [1].

## BIBLIOGRAPHY

- [1] Adams, Robert, 'Middle Knowledge and the Problem of Evil', *American Philosophical Quarterly* 14 (1977), 109–117.
- [2] Chellas, Brian, 'Basic Conditional Logic', *Journal of Philosophical Logic* 4 (1975), 133–228.
- [3] Field, Hartry, 'Quine and the Correspondence Hypothesis', *Philosophical Review* 83 (1974), 200–228.
- [4] Fine, Kit, 'Vagueness, Truth and Logic', *Synthese* 30 (1975), 265–300.
- [5] Herzberger, Hans, 'Counterfactuals and Consistency', *Journal of Philosophy* 76 (1979), 83–88.
- [6] Kamp, Hans, 'Two Theories About Adjectives', in Edward L. Keenan (ed.), *Formal Semantics of Natural Language*, Cambridge U. Press, Cambridge, 1975, pp. 123–155.
- [7] Lewis, David, *Counterfactuals*, Basil Blackwell, Oxford, 1973.

- [8] Lewis, David, 'Counterfactual Dependence and Time's Arrow', *Noûs* 13 (1979), 455–476.
- [9] Lewis, David, 'General Semantics', in Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, D. Reidel, Dordrecht, Holland, 1972, pp. 169–218.
- [10] Lewis, David, 'Truth in Fiction', *American Philosophical Quarterly* 15 (1978), 37–46.
- [11] Pollock, John, *Subjunctive Reasoning*, D. Reidel, Dordrecht, Holland, 1976.
- [12] Quine, W. V., *Methods of Logic*, Holt, Rinehart, and Winston, New York, 1950.
- [13] Stalnaker, Robert, 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, Basil Blackwell, Oxford, 1968, pp. 98–112.
- [14] Stalnaker, Robert and Richmond Thomason, 'A Semantic Analysis of Conditional Logic', *Theoria* 36 (1970), 23–42.
- [15] Thomason, Richmond, 'A Semantic Theory of Sortal Incorrectness', *Journal of Philosophical Logic* 1 (1972), 209–258.
- [16] Thomason, Richmond, 'Indeterminist Time and Truth-value Gaps', *Theoria* 36 (1970), 246–281.
- [17] Van Fraassen, Bas, 'Hidden Variables in Conditional Logic', *Theoria* 40 (1974), 176–190.
- [18] Van Fraassen, Bas, 'Singular Terms, Truth-value Gaps, and Free Logic', *Journal of Philosophy* 63 (1966), 481–495.