

## CRITICAL NOTE

### POSSIBLE-WORLD SEMANTICS FOR COUNTERFACTUAL LOGICS: A REJOINDER

Ellis, Jackson, and Pargetter [1] claim that a certain fact about counterfactual inference gives rise to a fundamental difficulty for any attempt to provide a possible-world semantics “worthy of the name” for counterfactuals. I disagree.

The problem arises as follows. Consider what seems to be a counterfactual with a disjunctive antecedent:

- (1) If either Oswald had not fired or Kennedy had been in a bullet-proof car, then Kennedy would be alive today.

It clearly implies two simpler counterfactuals:

- (2) If Oswald had not fired, then Kennedy would be alive today.  
(3) If Kennedy had been in a bullet-proof car, then Kennedy would be alive today.

Indeed, it seems equivalent to their conjunction. The same goes for parallel cases. Hence we might think that Simplification of Disjunctive Antecedents, the inference pattern

$$\frac{(A \vee B) \Box \rightarrow C}{\therefore A \Box \rightarrow C} \quad \text{or} \quad \frac{(A \vee B) \Box \rightarrow C}{\therefore B \Box \rightarrow C},$$

is a generally valid rule of counterfactual logic. We might also think that Interchange of Logical Equivalents is a generally valid rule. Unfortunately, as is shown in [1], Simplification and Interchange combine to give clearly unacceptable inferences.

Why is this especially a problem for the possible-world semanticist? It is everyone's problem. Whatever sort of semantics or non-semantics you favor to validate the counterfactual logic of your choice, you dare not combine Simplification with Interchange. Either you must explain away the evidence in favor of Simplification or you must explain some failures of Interchange.

Our problem was first discussed in print, so far as I know, by Fine [3].

He gave three solutions, stating no preference among them. *Each one is within possible-world semantics*. Moreover, each seems to be “taken philosophically seriously as providing truth-conditions”, not just “viewed purely formally as a piece of model theory”. Ellis *et al.* are aware of Fine’s discussion. Why, then, do they think the possible-world semanticist has any difficulty worse than an *embarras de richesse*?

Fine’s first solution is to stick with Interchange and explain away the evidence in favor of Simplification. (This is my choice. But I need not defend it at length; that has been well done by Loewer [6] and McKay and van Inwagen [4].) The strategy is to question that sentences like (1) really have the logical form  $(A \vee B) \Box \rightarrow C$ . Perhaps the ‘either . . . or’ in (1) is something other than standard disjunction. Perhaps it is conjunction, with wide scope; then the logical form of (1) would be the conjunction of the logical forms of (2) and (3). Or perhaps ‘if either . . . or . . . then . . .’ is a three-place operator. We could symbolize and define it as follows:

$$\left. \begin{array}{l} A \\ B \end{array} \right\} \Box \rightarrow C =^{\text{df}} (A \Box \rightarrow C) \& (B \Box \rightarrow C),$$

thereby providing for the desired equivalence between (1) and the conjunction of (2) and (3). Either way, Simplification of Disjunctive Antecedents is not needed to explain the evidence. The seeming disjunctive antecedent is an illusion of surface structure.

Ellis *et al.* do note that this solution has been proposed, but they have only two things to say about it. First, they rightly say that the possible-extra ‘if’ is not decisive evidence in favor. Second, they ask: if such sentences as (1) are not instances of the form  $(A \vee B) \Box \rightarrow C$ , what *are*? Answer: perhaps nothing – why must every logical form find an expression in ordinary English? Or perhaps such clumsy affairs as this:

- (1') Kennedy would be alive today if the following had been the case: either Oswald did not fire or Kennedy was in a bullet-proof car.

My linguistic intuitions don’t settle whether (1') is the desired instance of  $(A \vee B) \Box \rightarrow C$  or whether it is merely a stylistic variant of (1). But nothing depends on the answer.

Ellis *et al.* could have given a better reason to dislike the first solution. Isn’t it badly *ad hoc* to solve a problem in counterfactual logic by

complicating our treatment of 'or'? When we have a simple, familiar, unified treatment (marred only by the irrelevant question of exclusivity) wouldn't it be more sensible to cherish it? I reply that if I considered our present problem in isolation, I would share these misgivings. But parallel problems arise from other constructions, so our nice uncomplicated treatment of 'or' is done for in any case. Consider:

- (4) I can lick any man in the house, or drink the lot of you under the table.
- (5) It is legal for you to report this as taxable income or for me to claim you as a dependent.
- (6) Holmes now knows whether the butler did it or the gardener did.

Take the standard treatment of 'or'. Try wide or narrow scope; try inclusive or exclusive. (4)–(6) will prove as bad as (1).

Fine's second solution is to stick with Simplification, thus explaining the evidence in its favor in the least devious way possible, and to restrict Interchange enough to keep out of trouble. Such a solution has been advocated by Nute [7]. Ellis *et al.* give the impression that this strategy could not be joined with possible-world semantics "worthy of the name", but why not?

The solution could work like this. First you give truth conditions, using your favorite version of 'serious' possible-world semantics, for all counterfactuals except those with disjunctive antecedents. (You might reserve other cases also for special treatment, but let us ignore that.) Then you cover the special case by specifying that a counterfactual of the form  $(A \vee B) \square \rightarrow C$  is true at a world iff both its simplifications,  $A \square \rightarrow C$  and  $B \square \rightarrow C$ , are true there. Thereby you validate Simplification; but you explain the failure of Interchange between a disjunction and an equivalent non-disjunction in the antecedent position.

The only clue Ellis *et al.* give about what they find wrong with this solution is an argument near the end of the paper. They consider a special case, but their argument could be generalized as follows. Premise: according to possible-world semantics, the semantic character of a sentence is fully determined by the set of worlds where that sentence is true. Premise: equivalent sentences, even if one is a disjunction and the other is not, have the same truth set. So they have the same semantic character. So they can be

interchanged without changing the semantic character of compound sentences of which they are constituents.

The first premise is false, at least if possible-world semantics is that which is so-called by its practitioners. (For one example among many, see the discussion of 'meanings' by Lewis [5].) But let that pass. The second premise is right. However, to reach the conclusion we need another premise, which Ellis *et al.* do not mention. Call it the Principle of Immediacy:

*The semantic character of any compound sentence depends entirely on the semantic characters of its immediate constituents.*

Applied to the case at hand, Immediacy demands that the semantic character of a sentence of the form  $(A \vee B) \Box \rightarrow C$  must depend entirely on the characters of its immediate constituents  $(A \vee B)$  and  $C$ , and not at all on the characters of its non-immediate constituents  $A$  and  $B$ . The argument shows only what should have been plain from the start: that the second solution violates Immediacy. What's bad about that?

Immediacy is not sacred. It is customary; it is convenient; it is somehow associated with Frege. But it is not required as a condition of coherence or rigor in semantics. The example at hand is enough to show that.

Ellis is elsewhere commendably open-minded about Immediacy. He prefers to analyze the semantic character of sentences in terms of constraints on rational belief, not truth at possible worlds; but the question of Immediacy cuts across disagreements about the character of semantic character. And in [2] Ellis has proposed a treatment of counterfactuals with disjunctive antecedents that violates Immediacy in just the way Fine's second solution does: such counterfactuals are made an exception to the general rule, and treated as if they were conjunctions of their simplifications. The semantic character of the immediate constituent  $(A \vee B)$  plays no part in determining the semantic character of  $(A \vee B) \Box \rightarrow C$ . Ellis must think that possible-world semantics is stuck with Immediacy in a way that rival approaches are not, but he gives no reason for this discriminatory attitude.

I know of only one reason to insist on Immediacy. (Saul Kripke brought it to my attention, but I am not sure how compelling he takes it to be.) The enterprise of parsing – of assigning constituent structure and logical form – seems badly underconstrained. Given a parsing that forces a violation of Immediacy (under our chosen approach to semantics) perhaps we can always

replace it by an acceptable alternative that doesn't. If so, then it might make sense as a methodological convention to standardize on parsings that respect Immediacy. In the case at hand, that convention tells us to reject the second solution because the first is better. That cannot help Ellis *et al.*, who claim that no solution is satisfactory.

Fine's third solution is to keep  $(A \vee B) \Box \rightarrow C$  as the logical form of (1) and the like; to keep Interchange; and to give a possible-world semantics that validates the instances of Simplification that are supported by the evidence but not the ones that make trouble. The natural objection is that the ones that make trouble are among the ones supported by linguistic evidence. I am not sure; perhaps the objection is right, but linguistic intuition falters. Be that as it may, Ellis *et al.* raise neither this objection nor any other against Fine's third solution.

*Princeton University*

DAVID LEWIS

#### POSTSCRIPT

Fine has informed me that, although his statement of his second solution may bear the interpretation I gave it, he intended something different: a solution within fact-theoretic semantics of the sort presented in B. van Fraassen, 'Facts and Tautological Entailments', *Journal of Philosophy* 66 (1969) 477–487. It remains true that the solution I ascribed to Fine is within possible-world semantics, and that Ellis *et al.* have given no good reason to reject it.

#### REFERENCES

- [1] B. D. Ellis, F. C. Jackson and R. J. Pargetter, 'An Objection to Possible-World Semantics for Counterfactual Logics', this issue, pp. 359–363.
- [2] B. D. Ellis, 'Epistemology, Validity and Truth', presented at the 1976 annual conference of the Australasian Association of Philosophy, Melbourne.
- [3] K. Fine, critical notice of Lewis, D., *Counterfactuals*, *Mind* 84 (1975) 451–8.
- [4] T. McKay and P. van Inwagen, 'Counterfactuals with Disjunctive Antecedents', *Philosophical Studies* (forthcoming).
- [5] D. Lewis, 'General Semantics', *Synthese* 22 (1970) 18–67.
- [6] B. Loewer, 'Counterfactuals with Disjunctive Antecedents', *Journal of Philosophy* 73 (1976) 531–7.
- [7] D. Nute, 'Counterfactuals and the Similarity of Worlds', *Journal of Philosophy* 72 (1975) 773–8.