CRITICAL NOTE

AN OBJECTION TO POSSIBLE-WORLD SEMANTICS FOR COUNTERFACTUAL LOGICS

In this note we present a fundamental difficulty for any attempt to validate a logic of counterfactuals via a possible-world semantics, where that semantics is conceived of as a piece of philosophical analysis. We will not be concerned with possible-world semantics viewed purely formally as a piece of model theory, but with such semantics taken philosophically seriously as providing truth-conditions for counterfactuals. D. Lewis [3] and R. Stalnaker [5] certainly take them this way.

The difficulty arises from the combination of two points. The first is that there is a theorem any adequate logic of counterfactuals must have, and one it must not have. The second is that any possible-world semantics worthy of the name validates the former only if it also validates the latter.

The theorem any adequate counterfactual logic must have is:

$$\text{(T)} \qquad [(A \vee B) \mathrel{\Box\!\!\rightarrow} C] \supset [(A \mathrel{\Box\!\!\rightarrow} C) \,\&\, (B \mathrel{\Box\!\!\rightarrow} C)].$$

The case for this theorem is as strong as the case for any theorem can be in this kind of area. It is strongly intuitive and there are no good counter-examples to it. For example, we take the counterfactual 'If New Zealand had either not sent a rugby team to South Africa or had withdrawn from the Montreal games, then Tanzania would have competed' to be true precisely because we take *both* 'If New Zealand had not sent a rugby team to South Africa, Tanzania would have competed' and 'If New Zealand had withdrawn from the Montreal games, Tanzania would have competed' to be true. Again, it cannot be true that if I held either ticket number one thousand or ticket number one thousand and one, I would have won the lottery, unless *both* tickets would have led to winning a prize.

In discussion it has been suggested that the examples we have given are not really instances of '$[(A \vee B) \mathrel{\Box\!\!\rightarrow} C]$'. But if they are not, *what are*? Further, it is excessively *ad hoc* for a denier of (T) to hold that English counterfactuals of the form '$(A$ or $B) \mathrel{\Box\!\!\rightarrow} C$' are not instances of '$(A \vee B) \mathrel{\Box\!\!\rightarrow} C$' but of, say, '$(A \mathrel{\Box\!\!\rightarrow} C) \,\&\, (B \mathrel{\Box\!\!\rightarrow} C)$' (that is, that 'or' in the antecedent of a counterfactual functions as a wide-scope 'and'). This is for

Journal of Philosophical Logic 6 (1977) 355–357.

two reasons. First, the linguistic evidence for such an apparently strained thesis can only be the admitted interchangeability in English of a statement of the form '$(A$ or $B) \mathrel{\Box\!\!\rightarrow} C$' with the corresponding one of the form '$(A \mathrel{\Box\!\!\rightarrow} C)$ *and* $(B \mathrel{\Box\!\!\rightarrow} C)$'. And one who denies (T) has no obvious *explanation* of this; but one who affirms (T) does, namely, that the two are logically equivalent, for it is granted by all that the implication goes in the reverse direction.

Secondly, 'or' in the antecedent of a counterfactual behaves normally when negated. For example, 'If it had *not* been the case that either Tom or Harry was elected, then it would have been the case that Dick was elected' is clearly equivalent to 'If both Tom and Harry had not been elected, then Dick would have been'. On the other hand, the most plausible account the wide-scope theorist can give of '(*It is not the case that either* A *or* B) $\mathrel{\Box\!\!\rightarrow} C$' is as '$\sim [(A \mathrel{\Box\!\!\rightarrow} C) \,\&\, (B \mathrel{\Box\!\!\rightarrow} C)]$' or '$\sim (A \mathrel{\Box\!\!\rightarrow} C) \,\&\, \sim (B \mathrel{\Box\!\!\rightarrow} C)$'. And it is easy to construct counter-examples to both.

Reflection on related questions involving quantifiers further strengthens the case for (T). If it is true that if *someone* had witnessed the accident, the culprit would have been brought to justice, then it is true that if *anyone* had witnessed the accident, the culprit would have been brought to justice. This fits in nicely with (T) because of the familiar connections in the case of finite domains between existential quantification and disjunction, and between universal quantification and conjunction. Just as both '$[(A \vee B) \supset C] \supset [(A \supset C) \,\&\, (B \supset C)]$' and '$(\exists x\, Fx \supset A) \supset \forall x(Fx \supset A)$' are valid, so are '$[(A \vee B) \mathrel{\Box\!\!\rightarrow} C] \supset [(A \mathrel{\Box\!\!\rightarrow} C) \,\&\, (B \mathrel{\Box\!\!\rightarrow} C)]$' and '$(\exists x\, Fx \mathrel{\Box\!\!\rightarrow} A) \supset \forall x(Fx \mathrel{\Box\!\!\rightarrow} A)$'.

It is true that Lewis and Stalnaker's preferred systems do not contain (T). But this has rightly been taken to be a major objection to them, for instance, by Donald Nute [4] and Christopher S. Hill and Lewis G. Creary [2].

The theorem which everyone allows a logic of counterfactuals must *not* have is (N) $[(A \,\&\, B) \mathrel{\Box\!\!\rightarrow} C] \supset (A \mathrel{\Box\!\!\rightarrow} C)$.

Now consider a counterfactual $(p \mathrel{\Box\!\!\rightarrow} q)$, and suppose its truth-conditions given in terms of possible worlds. The fine detail is not important here. All that matters is that its truth or falsity be determined by the holding or failure to hold of various relations between various possible-worlds. Now consider $[((p \,\&\, r) \vee (p \,\&\, \sim r)) \mathrel{\Box\!\!\rightarrow} q]$ and make the same supposition about its truth-conditions. Then it must have the same truth-value as $(p \mathrel{\Box\!\!\rightarrow} q)$, because every p-world is a $(p \,\&\, r) \vee (p \,\&\, \sim r)$-world and conversely. For a theorist who seriously intends to provide truth-conditions for counterfactuals in terms of possible-worlds, there can be no truth-relevant difference between $(p \;\Box\!\!\rightarrow q)$ and $[((p \,\&\, r) \vee (p \,\&\, \sim r)) \;\Box\!\!\rightarrow q]$. (Apart from this argument, the intuitive plausibility of a principle of substutivity of truth-functional equivalents is obvious.[1])

But, by (T), from $[((p \,\&\, r) \vee (p \,\&\, \sim r)) \;\Box\!\!\rightarrow q]$ we can derive $((p \,\&\, r) \;\Box\!\!\rightarrow q)$. Thus, we have shown that if $(p \;\Box\!\!\rightarrow q)$, then $((p \,\&\, r) \;\Box\!\!\rightarrow q)$. But no restrictions were placed on p, q, or r, hence we have shown that (T) plus seriously intended possible-world semantics gives, disastrously, (N).

La Trobe University, Melbourne

BRIAN ELLIS
FRANK JACKSON
ROBERT PARGETTER

FOOTNOTE

[1] Kit Fine in [1] uses such a principle to derive the result of the next paragraph, but considers the difficulty a 'minor' one.

REFERENCES

[1] Kit Fine, Critical notice of Lewis, D., *Counterfactuals*, *Mind* 84 (1975) 451–8.
[2] Christopher S. Hill and Lewis G. Creary, review of Lewis, D., *Counterfactuals*, *Philosophy of Science* 42 (1975) 351–4.
[3] David Lewis, *Counterfactuals*, Oxford, 1973.
[4] Donald Nute, 'Counterfactuals and the Similarity of Words [sic]', *Journal of Philosophy* 21 (1975) 773–8.
[5] Robert C. Stalnaker, 'A Theory of Conditionals', in Rescher, N., ed., *Studies in Logical Theory, American Philosophical Quarterly* supp. monographs, Oxford, 1968.