PAVEL TICHÝ

A COUNTEREXAMPLE TO THE STALNAKER-LEWIS ANALYSIS OF COUNTERFACTUALS

(Received 21 February, 1975)

Stalnaker's and Lewis's recent analyses of counterfactuals (see [1], [2], and [3]) are based on the idea that given a possible world W, other worlds can in principle be judged as to how similar they are to W. The authors assume, in other words, that possible worlds can be looked upon as approximations to W, and that of two such approximations one may be better than the other.

It will be convenient to adopt the following terminological convention: instead of saying that a sentence X is true in an approximation W' to W, we shall simply say that W' is an *X-approximation* to W.

The authors tell us little about what exactly it takes for one approximation to W to be better than another. They take this to be a virtue of their theory; since on their view, the vagueness of the notion of relative similarity among worlds affords an explanation of the (alleged) vagueness of counterfactual judgements. But, as David Lewis says, not anything goes.

One of the more obvious requirements which must be satisfied if the notion of world similarity is to make sense at all seems to be this:

(R) Let a non-causal sentence B be logically and causally independent from A in W. Moreover, let B be true in W. Then B is true in the best A-approximations to W.

For surely good A-approximations to W will be worlds where the causal dependences among propositions are the same as in W. But B is clearly true in some worlds of this sort; and since B is true in W, such worlds approximate W better than the others.

By way of a simple illustration of (R), consider a man – call him Jones – who is possessed of the following dispositions as regards wearing his hat. Bad weather invariably induces him to wear his hat. Fine weather, on the other hand, affects him neither way: on fine days he puts his hat on or leaves it on the peg, completely at random. Suppose, moreover, that actually the weather is bad, so Jones *is* wearing his hat. Writing A for

Philosophical Studies **29** (1976) 271–273.

'the weather is fine', (R) enables us to draw the natural conclusion that in the best A-approximations to the actual world, Jones is wearing his hat.

Now the analyses under consideration of counterfactuals of the form

(1) If A were the case then B would be the case

are as follows:

Stalnaker has submitted that

(S) (1) is true in world W if B is true in the best A-approximation to W.

From a purely formal point of view, (S) leaves something to be desired. The point is that there seems no guarantee that there will be a unique best A-approximation to W whenever (1) is true in W. It might very well happen that two A-approximations to W are better than all the remaining ones without one of them being better than the other. Or, there may be an infinite sequence of ever-better A-approximations to W. In neither of these two eventualities would (S) yield a truth-value for (1).

David Lewis has therefore proposed the following refinement of (S):

(L) (1) is true in world W if some AB-approximation to W is better than any $A\bar{B}$-approximation to W.

It is easy to see that (L) is but a marginal modification of (S). Both proposals codify the basic idea that if one wants to know whether (1) is true in a world W one looks and sees whether W can be better A-approximated by worlds where B holds than by worlds where B fails.

Now to see the inadequacy of the proposals, let us consider the sentences

(2) If the weather were fine, Jones might not be wearing his hat.

(3) If the weather were fine, Jones might (still) be wearing his hat.

Counterfactuals may be vague in general, but on the terms of the above illustrative story, the particular counterfactual statements (2) and (3) are undeniably true. It seems to me that any theory of counterfactuals which fails to yield this result, is bound to be inadequate.

But (S) and (L), *do* fail to yield the result. To see this, consider the sentence

(4) If the weather were fine, Jones would be wearing his hat.

It is readily seen that where A, B, and W are, respectively, 'The weather is fine', 'Jones is wearing his hat', and the actual world, the hypotheses of (R) are true. Hence by (R), B is true in the best A-approximations to W. But then, by (S) and/or (L), (4) is true in W, i.e. true *simpliciter*. (4) and (2), however, are contradictories. Thus on (S) and/or (L) – given that (R) is correct – (2) is false rather than true.

The proponents of (S) and (L) might, of course, reject (R) and insist (by my lights implausibly) that B is *false* in the best A-approximations to W, despite being true in W. But then, on (S) and/or (L) the sentence

(5) If the weather were fine, Jones would not be wearing his hat

would be true. (5) and (3), however, are contradictories. Thus on (S) and/or (L), (3) would be false rather than true.

There remains the possibility that the theories of Stalnaker and Lewis leave it open whether B is true or false in the best A-approximations to W. But if this were the case, the explications they offer for counterfactual statements would turn out to be substantially vaguer than the explicandum.

REFERENCES

[1] Robert C. Stalnaker, 'A Theory of Conditionals', in: Nicholas Rescher (ed.), *Studies in Logical Theory*, Oxford 1968.

[2] David Lewis, 'Counterfactuals and Comparative Possibility', *Journal of Philosophical Logic* **2** (1973), pp. 418–446.

[3] David Lewis, *Counterfactuals*, Oxford, Basil Blackwell, 1973.