August 11, 1969

Professor David Lewis
Department of Philosophy
University of California
Los Angeles, California

Dear Dave,

I have just been reading and thinking about your paper on counterpart theory, and I wanted to send you some comments.

My own inclinations as to a way of speaking don't seem to be very much like yours, since I think it is literally Hubert himself and not a mere counterpart who would have defeated Nixon under certain circumstances. In the same way I think that I am literally the same person today who did something yesterday, contrary to what you say on the top of page 115. Maybe this isn't too important, since I can interpret you as speaking in the latter case of time-slices and in the former case, in some sense, of "world-slices". Maybe you gain some generality over the other way of speaking, since, e.g., where in some cases I might have to say that the identity criteria over time for people and material objects are somewhat vague you can say that the counterpart relation is not an equivalence relation; so your method may have some advantages if one wishes to avoid vagueness. (For the modal case, however, see what I have to say in the formal comments below.) More important, perhaps, I don't think that similarity, in any significant sense, has much to do with individuation over possible worlds. It seems to me intuitively clear that I might have been a person radically different than I am now and that someone else at the same time might have resembled me in almost all the respects people would find important and that in that case we would not say that the other person would have been I. Anyway I won't elaborate on such philosophic questions, on which there might be disagreement, since I have some formal comments to make. The first of these is a quite conclusive objection to your theory the way it is set up now.

The objection I think you will accept as conclusive is this: Let a be an actual individual with two distinct counterparts in some possible world. Then, under your translation of quantified modal logic, if the object a is assigned as value to the variable y, $(x)\ \ (x \neq y)$ comes out true, while $\Diamond (y \neq y)$ is of course false. So the schema $(y)\ ((x)Fx \supset Fy)$ is not valid under your interpretation! The same counter example works for Kaplan's interpretation, page 119, provided that a has a counterpart in every world.

Has anyone pointed this out to you? I think the best way of fixing it would be simply to assume that an individual can have at most one counterpart in every world. An alternative might be to adopt the following more complicated translation scheme:

Given ϕ, define ϕ^{β} by induction or the number of quantifiers and connectives in ϕ. All clauses except those for $(\Box \phi)^{\beta}$ and $(\Diamond \phi)^{\beta}$ are the same as yours. For $(\Box \phi\}^{\beta}$, define it by induction on the number of free variables in ϕ. If ϕ has no or one free variable, the clause is the same as yours. If ϕ has two distinct free variables free, x and y, $(\Box \phi xy)^{\beta}$ is $x = y\ \&\ (\Box \phi xx)^{\beta}$ $:v:\ x \neq y\ \&\ (\beta_1)(u)(v)(W\beta_1\ \&\ Iu\beta_1\ \&\ Iv\beta_1\ \&\ Cux\ \&\ Cvy\ \beta_1 \supset$ $\phi^{\beta_1} uv)$

Here $(\Box \phi xx)^{\beta}$ has already been defined, since it contains only one free variable (it is assumed that if necessary the bound variables in ϕ have been changed so as to avoid confusion of bound variables in ϕxx). Also note that the second conjunct of the second disjunct is the way you would have translated $(\Box \phi xy)^{\beta}$, given ϕ^{β}; the pattern persists. For these variables, translate $(\Box \phi xyz)^{\beta}$ as

$$x = y\ \&\ (\Box \phi xyz)^{\beta}\ .v.\ x = z\ \&\ (\Box \phi xyx)^{\beta}\ .v.\ y = z\ \&\ (\Box \phi xyy)^{\beta}.\ v.\ x \neq y\ \&\ y \neq z\ \&\ x \neq z\ \&\ \psi,$$

where ψ is the way you would have translated $(\Box \phi xyz)^{\beta}$, given $(\phi xyz)^{\beta}$. And so on for more variables $(\Diamond \phi)^{\beta}$ is dual. Note that if each object in fact has at most one counterpart in each world, the translation collapes into the old one. Not only is the new translation much more complex than the old one, it seems to me despensible. Where you say that an individual a can have two counterparts, b and c in a single possible world, why not say that there are two distinct possible worlds, one in which b is "the" counterpart of a and one in which c is "the" counterpart. In the obvious sense, the new translation allows counterparts to be used only one at a time; in that case allowing an individual to have two counterparts in a single world seems formally pointless. I think the following technical claim is true. (I haven't checked it in detail as is also the case with some other technical claims made below): The valid schemata in quantified modal logic under the new interpretation are just the same as those under the old interpretation with the requirement that each object has at most one counterpart in a world. (Perhaps the difficulties you got into by allowing an object to have multiple counterparts in a world constitute a plausibility argument against your definition of counterpart in terms of similarity, since given this defintion the multiple counterparts do indeed seem plausible.)

Shouldn't you have mentioned that in addition to the controversial laws of modal logic like Brouwer's and Becker's principles that you interpretation rejects, it also rejects such ordinarily uncontroversial laws as $\Box(\phi \supset \psi) \supset . \Box\phi \supset \Box\psi$ and $\Box(\phi \,\&\, \psi) \supset \Box\phi$? This <u>isn't</u> meant to be a damning objection; what it means is that your $\Box$ operator behaves like the NMN of priors Q in <u>Time and Modality</u>. Probably you are already aware of this.

The following seems to mne perhaps more serious. If a language contains a name "a" of an individual no doubt you would treat it semantically the way you treat a free variable assigned to value a. Suppose the name is eliminated in favor of a predicate Ax (Quine's Ciceroizes). It would be natural to interpret the introduced predicate as holding in each world of precisely the counterparts of the object a. Then if ϕ_a is replaced by $(\exists x)(\phi x \,\&\, Ax)$ for atomic ϕ, $\Box\phi a$ (or, alternatively, ϕx where the free variable x is assigned the object a as value) need not have the same truth value as $\Box(\exists x)(\phi x \,\&\, Ax)$. If we translated ϕa as $(x)(\phi x \supset Ax)$, the difficulty would arise for $\Box \sim \phi a$. This is really another form of the problem on page 119, second paragraph; but the following is perhaps more serious. Suppose b is a counterpart of a, c is a counterpart of b, c is not a counterpart of a, and ϕx holds when x is assigned any counterpart of a in any world but fails when x is assigned C then $\Box\Box\phi x$ is false when x is assigned a as value, but $\Box\ \Box\ (\exists x)(\phi x \ \& \ Ax)$ is true (provided a has a counterpart in each world).

The difficulty of the preceeding paragraph can be put another way. Your interpretation allows some laws to be valid for atomic formulae containing just one free variable, eventhough some substitution instances of them are not valid when the atomic formulae are replaced by formulae with extra free variables as paramaters. How can you be sure that the modaic predicates in your formulae are not interpreted so as to refer to certain individuals not explicitly mentioned? (This happens with "Ciceroizes" and perhaps with more natural predicates as well.) You assume that your atomic predicates are, in some sense you don't define or explicate, "purely qualitative". I have some feeling for this, but am a bit queasy about the notion; anyway, this assumption certainly doesn't hold for all languages.

Finially, I am not sure that your counterpart relation really gives any additional generality over those who speak of identity across possible worlds or require that the counterpart relation be an equivalence relation. To consider one example: suppose we have objects a_1, a_2, a_3, in worlds H_1, H_2, H_3, where each individual is the counterpart of its predecessor, but a_3 is not the counterpart of a_1. Instead of speaking of an intransitivity of the counterpart relation, I think one could consider this a case of relative modality where the relation R between worlds is not transitive. That is to say, we could have H_1RH_2, H_2RH_3, but not H_1RH_3. The problem is, of course, that in the old model maybe H_1 was related to H_3. (especially in the non-relative case you consider where all worlds are related to each other). We can take of this by postulating an additional world H_3', qualitatively like H_3, and such that H_1RH_3', but a_3 (or the individual corresponding to it in H_3'), is not the counterpart of a_1 in H_3' (if in the old terminology H was the counterpart of a in H_3). Now we have four worlds instead of three and the process may go on creating an infinity of additional worlds, but the counterpart relation can thus be made into an equivalence relation.

In other words, where you speak of an intransitive counterpart relation and a transitive excessibility relation, by posting extra worlds one can get the same effect by assuming the reverse. I hope this isn't too vague; anyway, it seems to me that the valid schemata of quantified modal logic are just the same if you allow an intransitive counterpart relation as they would be if you allow instead an intransitive accessibility relation. Of course there is a real formal difference in the fact that if the accessibility relation is transitive but the counterpart relation is not, ϕ ϕ is valid for closed ϕ, whereas under the alternative this would not be so. Neither alternative would Becker's principle hold if ϕ contains free variables. As I have said above, though, the distinctions you approach makes between closed and open formulae seem to me to be dubious.

I had no idea that this letter was going to be so long. Maybe some of my remarks are still too sketchy. I hope the first technical difficulty, about (x)Fx Fy, which seems to me the most serious, is clear. Anyway, I want to ask you a favor. I hear fascinating rumors about your work on the counterfactual, which sounds very interesting. Could you tell me something about it? Regards to Steffi and to the people at U.C.L.A.

Best,

Saul Kripke

SK/mf