Philosophical Papers

Volume I

DAVID LEWIS

that its characteristic i-relation will be 2-place. (Such an i-relation is often called an *accessibility* relation between worlds). And consider Becker's principle for this relative necessity (but with '∃' still defined in terms of necessity itself): $\Box^i\phi\dashv3\Box^i\Box^i\phi$; that is, $\Box(\Box^i\phi \supset \Box^i\Box^i\phi)$. It is often said that Becker's principle holds just in case accessibility is transitive, which is correct if ϕ is a closed sentence. But for open ϕ Becker's principle holds just in case

$$\forall x_1\forall y_1\forall x_2\forall y_2\forall x_3\forall y_3(Ix_1y_1 \ \& \ Ix_2y_2 \ \& \ Ix_3y_3 \ \& \ Cx_2x_1 \ \& \ Cx_3x_2$$
$$\& \ R^iy_2y_1 \ \& \ R^iy_3y_2 \ . \supset . \ Cx_3x_1 \ \& \ R^iy_3y_1)$$

even if neither accessibility nor the counterpart relation is transitive.

<div align="center">

Postscripts to

"Counterpart Theory
and
Quantified Modal Logic"[1]

</div>

A. BEING IN A WORLD

I took as primitive the notion of an individual being *in* a possible world. I would now wish to distinguish the relation I had foremost in mind from two others.

A world is a large possible individual; it has smaller possible individuals as parts. A galaxy, a planet, a man, an electron—these things inhabit their world simply by being parts of it. Just as the electron is part of the man, and the man in turn is part of his planet which is part of its galaxy, so the galaxy in turn is part of its world. And so are its parts, and their parts, . . . , since the relation of part to whole is transitive. Any possible individual is part of a world, and in that sense it is *in* a world. (As a special case, a world is an improper part of itself.) Worlds do not overlap; unlike Siamese twins, they have no shared parts. Thus I stand by Postulate 2, construed as meaning that no possible individual is part of two worlds.

However, the possible individuals are not all the individuals. I wish to impose no restrictions on mereological summation of individuals, hence I must grant that there are individuals consisting of parts from several worlds. But such a cross-world sum is not a *possible* individual. There is no way for the whole of it to be actual.

No matter which world is actual, at most a proper part of it actually exists. It is not in any world, in the sense just discussed, for it is not part of any world. But it is *partly in* each of many worlds, overlapping different worlds in virtue of different ones of its parts.

Finally, there are the non-individuals: that is, the sets. Provisionally,[2] my ontology consists of iterative set theory with individuals; the only unorthodox part is my view about what individuals there are. I take it that the part-whole relation applies to individuals, not sets. Then no set is in any world in the sense of being a part of it. Numbers, properties, propositions, events[3]—all these are sets, and not in any world. Numbers *et al.* are no more located in logical space than they are in ordinary time and space. Even a sequence of possible individuals all from the same world is not, strictly speaking, itself in that world.

When we evaluate the truth of a quantified sentence, we usually restrict the domain and quantify over less than all there is. If we evaluate a quantification at a world, we will normally omit many things not in that world, for instance the possible individuals that inhabit other worlds. But we will not omit the numbers, or some of the other sets. Let us say that an individual exists *from the standpoint of* a world iff it belongs to the least restricted domain that is normally—modal metaphysics being deemed abnormal—appropriate in evaluating the truth at that world of quantifications. I suppose that this domain will include all the indivduals in that world; none of the other individuals; and some, but not all, of the sets. There will be many sets that even exist from the standpoint of all worlds, for instance the numbers. Others may not; for instance the unit set of a possible individual might only exist from the standpoint of the world that the individual is in.

Thus we have three relations: being *in* a world, i.e. being part of a world; being *partly in* a world, i.e. having a part that is wholly in that world; and existing *from the standpoint of* a world. Postulate 2, the principle that nothing is in two worlds, applies only to the first of these.

The language of counterpart theory, and the modal language it replaces, had best be understood as quantifying only over possible individuals. Modifications are called for if we wish to quantify over more of what there is.

B. MODAL CONTINUANTS

I granted, none too enthusiastically, that there are individuals, not wholly in any world and hence incapable of being actualized in their entirety, that consist of parts from several worlds. Some of these cross-world individuals are unified by counterpart relations. For instance, there is the mereological sum of myself and all my

---

[2] I am not sure what to say about universals, as advocated in D. M. Armstrong, *Universals and Scientific Realism* (Cambridge University Press, 1978), except for this: they are not to be confused with the sets of individuals that I call properties. If there are universals, they differ in many ways from properties and they meet completely different theoretical needs.

[3] See my "Events" (in the sequel to this volume).

counterparts. Or (what is not the same) there is a cross-world individual which is a maximal counterpart-interrelated sum of possible individuals of whom I am one. In fact, there are many such.

More precisely, let us call X a *world-stage* of Y iff (1) X is a possible individual, entirely in one world, and is part of Y; and (2) X is not a proper part of any other individual of which the same is true. And let us call Y *counterpart-interrelated* iff any two world-stages of Y are counterparts. And let us call Y *maximal counterpart-interrelated*, or for short let us call it a *modal continuant*, iff (1) Y is counterpart-interrelated; and (2) Y is not a proper part of any other individual of which the same is true.

At this point it lies close to hand to suggest that ordinary things—stars, loco-motives, cats, ourselves—are modal continuants, just as they are temporal contin-uants having many temporal parts and not wholly located at any one time. Then we can have our counterpart theory and our identity across worlds as well, and everyone's intuitions will be satisfied.

I am not impressed. I have no objection to the theory of modal continuants. It is an equivalent reformulation of counterpart theory, so it is just as right as coun-terpart theory itself. It does not offend against Postulate 2, which merely prohibited individuals from being *wholly* in more than one world.

But I cannot see that the theory of modal continuants has any intuitive advan-tage. On the one hand, it makes a problem. Given the intransitivity of the coun-terpart relation, doubtless I—or rather, as we should now say, the actual world-stage of me—am part of ever so many different modal continuants. Which one is me? The puzzle is like that of Methuselah in "Survival and Identity" (in this vol-ume). Then let us borrow the solution. To be sure, I am (we are?) infinitely many different modal continuants. No worries: if we count by stage-sharing rather than identity, there is only one. —Will this really do? For temporal continuants, the pathological cases are fictional. It is only fitting and proper that such extraordinary goings on should force us to say something a bit counterintuitive. For modal con-tinuants, *every case is pathological.* Everything that can happen does. If we adopt the theory of modal continuants, we must count by stage-sharing not only when we want to treat the uncanny Methuselah as if he (they?) were an ordinary person, but also in the case of the ordinary people who are all, modally speaking, Methu-selahs.

And to what end? Presumably, to make Hubert Humphrey come out literally right when, after losing, he thinks: I myself might have won.[4] He is supposed to think: I am a modal continuant with a world-stage that wins. That is: the modal continuant of which one world-stage thinks this thought has a world-stage—a dif-ferent one—that wins. (It isn't "the" but let that pass.) For even if Humphrey is a modal continuant, it doesn't take the whole of him to do such things as winning or thinking the thought. The continuant does them by having a world-stage that

[4]See Saul Kripke, "Naming and Necessity," in Donald Davidson and Gilbert Harman, eds., *Semantics of Natural Language* (Dordrecht: Reidel, 1972), p. 344.

does them (in the senses appropriate respectively to continuants and to world-stages), just as a temporal continuant does this or that through its stages. But what good is that? If Humphrey yearns to think only of himself and nobody else, it is no use that he the thinker is part of the same mereological sum as some winner. That much is provided by the thisworldly sum of Humphrey and Nixon! No; what matters is that the modal continuant is counterpart-interrelated, so that the thinker of the thought has a winner for a counterpart. Then why not just say so, and leave the modal continuant out of it?

I think intuition is well enough satisfied if we take "myself" to modify "might have won." Humphrey thinks that he himself, and not someone else who resembles him, has the modal property expressed by "might have won." And that is true on anybody's theory. In counterpart theory, it is true because Humphrey himself, in virtue of his very own qualitative character, is such as to have some winners for counterparts.

If more is wanted, I do not think the theory of modal continuants can provide it.[5]

C. VAGUENESS AND VARIETY OF COUNTERPART RELATIONS

I am by no means offering a wholehearted defense of "Aristotelian essentialism." For the essences of things are settled only to the extent that the counterpart relation is, and the counterpart relation is not very settled at all. Like any relation of comparative overall similarity, it is subject to a great deal of indeterminacy (1) as to which respects of similarity and difference are to count at all, (2) as to the relative weights of the respects that do count, (3) as to the minimum standard of similarity that is required, and (4) as to the extent to which we eliminate candidates that are similar enough when they are beaten by competitors with stronger claims.

Further, as with vagueness generally, the vagueness of the counterpart relation—and hence of essence and *de re* modality generally—may be subject to pragmatic pressures, and differently resolved in different contexts.[6] The upshot is that it is hard to say anything false about essences. For any halfway reasonable statement will tend to create a context that (partially) resolves the vagueness of the counterpart relation in such a way as to make that statement true in that context. So almost anything goes. The true-hearted essentialist might well think me a false friend, a Quinean sceptic in essentialist's clothing.

---

[5] If it really must be Humphrey himself who does the winning, the very same Humphrey who thinks the modal thought, the best solution is to think not of genuine worlds but of world-stories—ersatz worlds—so that we think of Humphrey's actual losing and possible winning as two stories about him, one true and the other false. It is my impression that something like this, and not a theory of modal continuants, was Kripke's intention.

[6] See the Appendix to "General Semantics" and Example 5 of "Scorekeeping in a Language Game" (both in this volume) for discussions of vagueness and the effects of context on its resolution.

To take one extreme, a suitable context might deliver an antiessentialist counterpart relation—one on which anything is a counterpart of anything, and nothing has any essence worth mentioning. Or, as Hazen has suggested,[7] we might somehow partition things into kinds, and take a counterpart relation on which anything is a counterpart of anything of its kind. That would make the essence of a thing simply be its kind. Such unselective counterpart relations as these would violate my Postulate 5, since things would have counterparts in their own worlds besides themselves. But I needn't insist on Postulate 5 for any counterpart relation that might be appropriate in any context; it is enough to take it as a condition governing the more commonplace of them.

At the opposite extreme, a suitable context might deliver a hyperessentialist counterpart relation—one on which nothing has any counterpart except itself. Then by Postulate 2 nothing has otherworldly counterparts, and hence nothing has any of its attributes nonessentially. This counterpart relation, of course, is simply identity.

A common sort of counterpart relation, especially in the context of recent discussions of essentialism, is one that gives decisive weight to perfect match of origins. (I count this simply as one respect of similarity, *pace* Jaakko Hintikka who says that I ought to have considered match of origins *as well as* similarity.[8]) It is this sort of counterpart relation that we need to make sense of such possibilities as that of Hitler's leading a blameless life while someone else moves into his actual role and ends by living a wicked life just like the life that Hitler actually lived.[9]

Not only may the vagueness of the counterpart relation be very differently resolved in different contexts. Also we may need to play off one counterpart relation against another in a single context.[10] Or we may need to play off one relation of comparative similarity that governs the counterpart relation against another that governs explicit judgements of similarity. For as Fred Feldman has observed, one might truly say:

> I could have been more like what you in fact are like than like what in fact I am, and at the same time, you could have been more like what I in fact am like than what you in fact are.[11]

[7] *Op. cit.*, p. 332.

[8] See his *The Intentions of Intentionality and Other New Models for Modalities* (Dordrecht: Reidel, 1975), pp. 127–29 and 209.

[9] See my *Counterfactuals* (Oxford: Blackwell, 1973), p. 41. For the case of blameless Hitler, see Kripke, "Naming and Necessity," pp. 288–89.

[10] See Hintikka's discussion of "two kinds of cross-identification," in "On the Logic of Perception" in his *Models for Modalities* (Dordrecht: Reidel, 1969); my *Counterfactuals*, pp. 42–43; and "Counterparts of Persons and Their Bodies" (in this volume). The last shows how to use a harmless multiplicity of counterpart relations to replace an undesired multiplication of entities. Denis Robinson has applied the same method to many other cases, in a lecture at the Australasian Association of Philosophy Conference in Sydney in August 1980.

[11] "Counterparts," *Journal of Philosophy* 68 (1971): 406–9. A parallel problem about similarity and counterfactuals is discussed in "Counterfactual Dependence and Time's Arrow" (in the sequel to this volume).

D. PAIRS OF COUNTERPARTS AND COUNTERPARTS OF PAIRS

Consider the twin brothers Dee and Dum. Together they comprise a pair. In this easy case, we may take the pair simply as a mereological sum; then it is a possible individual, not a set, so counterpart theory applies to it without need for any modification.

In another world are two duplicate planets, far apart in time and space. On one planet there live a pair of twin brothers, Dee_1 and Dum_1; and together they comprise the pair Dee_1-cum-Dum_1. Likewise on the other planet there live Dee_2 and Dum_2, who comprise Dee_2-cum-Dum_2. Dee_1 and Dee_2 resemble Dee equally, and quite well, and better than anything else in their world does, and so they are both his counterparts. Likewise Dum_1 and Dum_2 are both counterparts of Dum.

Now we ask a doubly *de re* modal question about Dee and Dum: might they have been not twin brothers, but totally unrelated inhabitants of separated planets? Intuitively, it seems not. They are essentially related (or, at any rate, related if they both exist). But counterpart theory disagrees, and gives an answer that seems counterintuitive. For Dee and Dum satisfy the counterpart-theoretic translation of the formula

$\Diamond$x and y are unrelated

as values respectively of its two variables. That is because there is a world in which there are unrelated counterparts of Dee and Dum respectively: namely, Dee_1 and Dum_2 in the world of the duplicate planets. (Or, for good measure, Dee_2 and Dum_1.)

But whether counterpart theory gives a counterintuitive answer depends on exactly how we put the question. Instead of a doubly *de re* question about Dee and Dum, we might better have asked a singly *de re* question about their pair, Dee-cum-Dum. Might *it* have been not a pair of twin brothers, but rather a pair of totally unrelated inhabitants of separate planets? Again it seems not. This time, counterpart theory agrees. For, under a reasonable counterpart relation, the pair Dee-cum-Dum does not satisfy the translation of the formula

$\Diamond \exists y \exists z$(x is y-cum-z & y and z are unrelated).

That is so because, for instance, the only counterparts of Dee-cum-Dum in the world of the duplicate planets are Dee_1-cum-Dum_1 and Dee_2-cum-Dum_2, and each of these is a related pair. We needn't worry about the unrelated pairs, Dee_1-cum-Dum_2 and Dee_2-cum-Dum_1. Nothing requires us to count these pairs as counterparts of Dee-cum-Dum, and the important difference between related pairs and unrelated ones gives us an excellent reason not to.

Two morals. First, we must learn the right way to apply counterpart theory. It seems that we do best to avoid doubly *de re* formulations if we want to respect intuition. And that seems plausible enough: doubly *de re* formulations may seem natural enough in the artificial language of quantified modal logic, but it is not so clear that they are part of our ordinary modal thought. Second, we should not

accept any neat principle to the effect that a pair of counterparts is a counterpart of the pair, or any generalization thereof.[12]

E. DOES COUNTERPART THEORY CHANGE LOGIC?

If counterpart theory calls for the rejection of some popular modal principles, that needn't worry us. But if it forces us to reject principles of the logic of identity and quantification, that is more serious. Allen Hazen and Saul Kripke have said that it does.[13] I plead not guilty. I shall consider one case of the trouble they have in mind, but I think my defense carries over to other cases.

Consider schema (1) and sentence (2) below. (1) is a valid schema of the classical logic of identity and quantification. It says something very uncontroversial indeed, if by " $=$ " we really mean identity: if we have one and the same thing, what's true of it is the same as what's true of it. (2), on the other hand, is an invalid sentence of quantified modal logic, since its counterpart-theoretic translation is not (and should not be) a theorem of counterpart theory.

(1) $\forall x \forall y (x = y .\, —x—\mathrel{\overline{\overline{}}}—y—)$
(2) $\forall x \forall y (x = y \supset . \Diamond x \neq y \equiv \Diamond y \neq y)$

How can (2) fail? Would its denial mean that we have two different things that are contingently identical? Or perhaps one thing that is only contingently self-identical? No—the denial of (2) would mean no such nonsensical thing. To see what it would really mean, don't guess, but read the counterpart-theoretic translation. The translation of (2) turns out to say that nothing in the actual world has more than one counterpart in any other world. Then its denial says that something actual does have two counterparts in a single world. The case of Dee, Dee_1, and Dee_2 could serve as an illustration, if Dee is actual. Again, we are in trouble when a doubly *de re* formulation meets something with double counterparts.

So (1) is not to be challenged while its instance (2) may well be false! How is that? —I reply that (2) is not an instance of (1), so there is nothing at all wrong with accepting (1) and rejecting (2).

Compare another invalid sentence, (3). How can we accept (1) and reject (3)? — That's easy: to take (3) as an instance of (1) is to commit a fallacy of confusing bound variables.

(3) $\forall x \forall y (x = y \supset . \exists y\, y \neq x \equiv \exists y\, y \neq y)$

To make (3) an instance of (1), the final occurrence of "y" would have to be bound to the initial "$\forall y$". It isn't. It's bound to the nearby "$\exists y$".

---

[12]I have more or less followed Hazen, *op. cit.*, in stating and solving the problem. But with this minor technical difference: he builds the solution into a modified counterpart-theoretic semantics for multiply *de re* formulations, whereas I apply it in unmodified counterpart theory by steering clear of the multiply *de re*.

[13]Hazen, *ibid.* Kripke, "Naming and Necessity," footnote 13.

I say much the same about (2). To treat it as an instance of (1) is to confuse bound variables. The abbreviated notation of quantified modal logic conceals the true pattern of binding, but it is revealed when we examine the counterpart-theoretic translation. It is as follows.

$$(2) \quad \forall x \forall y (x = y \supset \Diamond\, x \neq y \equiv \Diamond\, y \neq y)$$

The diamonds conceal quantifiers that bind the occurrences of "x" and "y" that follow. (Indeed, in the translation, the variables that follow would not be "x" and "y" at all, but different variables altogether.) The diamonds also conceal occurrences of "x" and "y" that are bound by the initial "∀x" and "∀y". This pattern of binding is not right for an instance of (1), any more than the pattern in (3) is.

So counterpart theory is no threat to standard logic. It is only a threat to simplistic methods of keeping track of variable-binding and instancehood when we are dealing with the perversely abbreviated language of quantified modal logic.

F. NONEXTENSIONALITY TOLERATED

Whatever else may fairly be said against the language of quantified modal logic, I withdraw my complaint against its nonextensionality. See "'Tensions'" (in this volume), in which I argue that such violations of extensionality as it commits are of no deep significance. If we want to restore extensionality, we need only reanalyze the language in an otherwise pointless, but harmless, way.

G. ATTITUDINAL MODALITIES ABANDONED

In view of the arguments of "Attitudes *De Dicto* and *De Se*" (in this volume), I withdraw my statement that implicit knowledge, belief, etc. are relative modalities, expressible by quantification over restricted ranges of worlds. They are, however, something closely analogous to that: we need only put possible individuals in place of the worlds.

H. CONVENTIONALITY OF POSTULATE 2 DISOWNED

Footnote 2 has given some readers the impression that I regard Postulate 2 as a mere convention, and that we could just as truly say that some things are identical with their otherworldly counterparts after all. Not so. I was alluding to the possibility of a hybrid theory—a theory opposed to my own, a theory which I take to be false—according to which there are identities across worlds, but we use the counterpart relation anyway.