COUNTERPARTS OF PERSONS AND THEIR BODIES *

MATERIALISTS like myself hold that persons and their bodies are identical. But there is a simple argument to show that this identity thesis is refuted by the mere possibility that a person might switch bodies. To defeat the argument it seems necessary to revise my counterpart theory by providing for a multiplicity of counterpart relations. This revision has an odd result. Modal predictions may be *de re*, yet not referentially transparent.

The thesis I wish to defend here may be stated more precisely, as follows:

(T) Necessarily, a person occupies a body at a time if and only if that person is identical with that body at that time.

Note that the thesis (T) is formulated not in terms of identity itself, a two-place relation, but in terms of a derivative three-place relation of identity at a time. I wish to regard enduring things such as persons and bodies as aggregates—sets, mereological sums, or something similar—of momentary stages. Enduring things X and Y are *identical at* a time t if and only if they both have stages at t—that is, *exist at* t—and their stages at t are identical. Therefore X and Y are identical *simpliciter* if and only if they are identical whenever either one exists. Note that (T) does not say that persons and bodies must be identical *simpliciter*. It does imply that *if* a certain person occupies a certain body whenever either the person or the body exists, *then* the person and the body are identical. In such a case, all and only those stages which are stages of the person are stages of the body he occupies. But (T) also permits other cases: for instance, a body consisting of the stages of a certain person together with some final dead stages that are not stages of any person (and some initial prenatal stages that perhaps are not stages of any person); or a person consisting of stages of a certain body together with some initial or final ghostly stages that are not stages of any body; or even a body-switching person consisting partly of stages of one body and partly of stages of another body. A person consists of stages related pairwise by a certain relation we may call the relation of *personal unity*; a body consists of stages related pairwise by another relation we may call the relation of *bodily unity*.[1] Since for the most part

* I am indebted to M. J. Cresswell, David Kaplan, and John Perry for illuminating discussion of these matters.

[1] The so-called "problem of personal identity" is the problem of explicating the relation of personal unity between stages. This view is expounded more fully by John Perry in "Can the Self Divide?", presented at the APA Pacific Division meetings, March 1970.

persons occupy bodies and bodies are occupied by persons, it follows according to (T) that the two relations of unity are relations on almost the same set of stages. The exceptions are dead stages, perhaps prenatal stages, and perhaps ghostly stages. Moreover, if we leave out the dead or ghostly or perhaps prenatal stages, then at least for the most part the two relations of unity are coextensive. The exceptions would be body-switchers and perhaps split personalities. Nevertheless, the two relations of unity are different relations-in-intension; so they are coextensive only contingently if at all.

Now I shall present an argument against (T). I regard it as a simplified descendant of an argument put forth by Jerome Shaffer,[2] but I have changed it so much that he might not wish to acknowledge it as his own.

Body-switching is logically possible. Because I might have switched out of my present body yesterday, though in fact I did not, I and my actual present body are such that the former might not have occupied the latter today. Whether or not persons are identical with bodies they occupy, certainly persons are never identical with bodies they do not occupy. So we have:

(1) I and my body are such that they might not have been identical today.

Suppose that, as is surely at least possible, I occupy the same body from the time when it and I began until the time when it and I will end. Then, by (T), it and I are identical whenever it or I exist. Hence my body and I, enduring things, are identical *simpliciter*. By this identity and Leibniz's law, (1) yields (2):

(2) My body and my body are such that they might not have been identical today.

Since (2) is self-contradictory, (T) has apparently been refuted.

To rescue (T) without denying the possibility of body-switching and without denying that I might occupy the same body throughout the time that it or I exist, I plan to show that the step from (1) to (2) by Leibniz's law is illegitimate—in other words, that (1) is not referentially transparent with respect to the term 'I'.

I have used the familiar "are such that" construction to indicate that (1) and (2) are to be taken as *de re* rather than *de dicto* modal predications. That is, we are to consider what happens in other possible worlds to the things denoted *here* in our actual world by the terms 'I' and 'my body', not what happens in other worlds to the

[2] "Persons and Their Bodies," *Philosophical Review*, LXXV, 1 (January 1966): 59–77.

things denoted *there* by those terms. Suppose (1) were taken *de dicto*, as if it were this:

> It might have been the case that I and my body were not identical today.

There is no problem explaining why this is not referentially transparent. Its truth conditions involve the denotations in other worlds of the terms 'I' and 'my body'; even if these denote the same thing here in our world, they denote different things in some other worlds, for instance in a world in which I switched bodies yesterday. Hence they are not interchangeable. But it would be wrong to take (1) as *de dicto*, for the argument leading up to (1) would then be incoherent and question-begging. As we understand (1) in the argument, it seems true even given (T) because of the fact that I might have switched bodies yesterday. But (1) taken *de dicto* is a straightforward denial of an instance of (T), and the possibility of body-switching is irrelevant to its truth or falsity. For (1) taken *de dicto* is true today if and only if there is some world such that the things denoted in that world today by 'I' and 'my body' are not identical today. But in any world, 'my body' today denotes whichever body is today occupied by the person who is today denoted by 'I', regardless of whether that person occupied that body yesterday. If (T) is true, then in any world the person today denoted there by 'I' and the body today denoted there by 'my body' are identical today; so (1) taken *de dicto* is false. (It should be understood that when I speak of the denotation in another world of 'I' and 'my body', I am not concerned with any utterance of these terms by some inhabitant of the other world, but rather with the denotation of these terms in the other world on the occasion of their utterance by me here in our world.)

I have suggested elsewhere[3] that *de re* modal predications may best be understood by the method of counterparts. To say that something here in our actual world is such that it might have done so-and-so is not to say that there is a possible world in which that thing *itself* does so-and-so, but that there is a world in which a *counterpart* of that thing does so-and-so. To say that I am such that I might have been a Republican, but I am not such that I might have been a cockatrice, is to say that in some world I have a counterpart who is a Republican, but in no world do I have a counterpart who is a cockatrice. That is plausible enough, for the counterpart relation is a relation of similarity. X's counterparts in other worlds are all and

[3] "Counterpart Theory and Quantified Modal Logic," this JOURNAL, LXV, 5 (March 7, 1968): 113–126.

only those things which resemble X closely enough in important respects, and more closely than do the other things in their worlds. It is easier for a Republican than for a cockatrice to resemble me enough to be my counterpart.

The counterpart relation serves as a substitute for identity between things in different worlds. The principal advantage of the method of counterparts over the method of interworld identities is that if we adopted the latter in its most plausible form, we would say that things were identical with all and only those things which we would otherwise call their counterparts. But that could not be correct: first, because the counterpart relation is not transitive or symmetric, as identity is; and second, because the counterpart relation depends on the relative importances we attach to various different respects of similarity and dissimilarity, as identity does not.

To recapitulate: in each possible world there is a set of momentary stages and a set of enduring things composed of stages related pairwise by various relations of unity. An enduring thing and its stages exist only in one world, but may have counterparts in other worlds. We shall be concerned here only with counterparts of enduring things, though we can allow that stages also have their counterparts.

Applying the method of counterparts to the problem at hand, we immediately encounter a bothersome distraction. The translation of (2), which seemed self-contradictory, is this:

> There are a world W, a counterpart X in W of my body, and a counterpart Y in W of my body, such that X and Y are not identical today.

Unfortunately, this translation comes out true, but for an irrelevant reason. I, and also my body whether or not I am identical with it, might have been twins. My body therefore does have two different counterparts in certain worlds. Not only is the translation true; it seems to me to show that (2) itself is true. But the argument against (T) can easily be repaired. Replace (1) and (2) by:

(1′) I and my body are such that (without any duplication of either) they might not have been identical today.

(2′) My body and my body are such that (without any duplication of either) they might not have been identical today.

The argument works as well with (1′) and (2′) as it did with (1) and (2). Indeed, (1′) and (2′) correspond to (1) and (2) as we would have understood them if we had forgotten that I might have been twins.[4]

[4] I do not know how or whether (1′) and (2′) can be expressed in the language of quantified modal logic. That does not bother me. I know how to express them in English and in counterpart theory.

Applying counterpart theory to the repaired argument, we obtain these translations of (1′) and (2′):

(1*) There are a world W, a unique counterpart X in W of me, and a unique counterpart Y in W of my body, such that X and Y are not identical today.

(2*) There are a world W, a unique counterpart X in W of my body, and a unique counterpart Y in W of my body, such that X and Y are not identical today.

The argument against (T) seems to go through, using (1*) and (2*): (1*) seems true because I might have switched bodies yesterday; (2*) is self-contradictory; yet (1*) implies (2*) by Leibniz's law, given (T) and the supposition that I occupy the same body whenever I or it exist.

In defense of (T), however, I claim that (1*) is *false*, despite the fact that I might have switched bodies yesterday. What is true because I might have switched bodies is not (1*) but rather (1**):

(1**) There are a world W, a unique *personal counterpart* X in W of me, and a unique *bodily counterpart* Y in W of my body, such that X and Y are not identical today.

I now propose a revision of counterpart theory to the effect that, at least in the present context, (1**) rather than (1*) is the correct translation of (1′). What follows from (1**) by Leibniz's law, given (T) and the supposition that I occupy the same body whenever I exist, is not the self-contradiction (2*) but rather the truth:

> There are a world W, a unique personal counterpart X in W of my body, and a unique bodily counterpart Y in W of my body, such that X and Y are not identical today.

Two other truths follow from (1**) in the same way:

> There are a world W, a unique personal counterpart X in W of me, and a unique bodily counterpart Y in W of me, such that X and Y are not identical today.

> There are a world W, a unique personal counterpart X in W of my body, and a unique bodily counterpart Y in W of me, such that X and Y are not identical today.

However, the translation of (2′) is none of these. If the translation of (1′) is (1**), the translation of (2′) should be (2**):

(2**) There are a world W, a unique bodily counterpart X in W of my body, and a unique bodily counterpart Y in W of my body, such that X and Y are not identical today.

Though (2**) is not (2*), it is still a self-contradiction.

As we already noted, counterpart relations are a matter of over-all resemblance in a variety of respects. If we vary the relative importances of different respects of similarity and dissimilarity, we will get different counterpart relations. Two respects of similarity or dissimilarity among enduring things are, first, personhood and personal traits, and, second, bodyhood and bodily traits. If we assign great weight to the former, we get the *personal counterpart* relation. Only a person, or something very like a person, can resemble a person in respect of personhood and personal traits enough to be his personal counterpart. But if we assign great weight to the latter, we get the *bodily counterpart* relation. Only a body, or something very like a body, can resemble a body in respect of bodyhood and bodily traits enough to be its bodily counterpart.

If I am my body, then in many worlds there are things that are both personal and bodily counterparts of me and *ipso facto* of my body. These things, like me, are both persons and bodies. But in other worlds I (and my body) have neither personal counterparts nor bodily counterparts; or personal counterparts that are not bodily counterparts; or bodily counterparts that are not personal counterparts; or personal and bodily counterparts that are not identical. A world in which I switched out of my body—that is, my personal counterpart switched out of my bodily counterpart—yesterday is of this last sort. I and my body have there a personal counterpart that is a person but not a body and also a bodily counterpart that is a body but not a person. These are not identical today, and not identical *simpliciter*, though they were identical at times before yesterday since they shared their earlier stages. However, my personal counterpart is identical today with a different body. My bodily counterpart is identical today with a different person (if the body-switching was a trade) or with none.

We may draw an analogy between the relations of personal and bodily unity among stages of persons and bodies and the personal and bodily counterpart relations among enduring persons and bodies. If I ask of something that is both a stage of a body and a stage of a body-switching person "Was this ever in Borneo?" you should ask whether I mean this person or this body. If the former, I am asking whether the given stage is linked by personal unity to an earlier stage located in Borneo. If the latter, I am asking whether it is linked by bodily unity to an earlier stage located in Borneo. Similarly, if I ask of something that is both an enduring person and an enduring body "Might this have been an orangutan?" you should again ask whether I mean this person or this body. If the former, I

am asking whether it has an orangutan for a personal counterpart; if the latter, whether it has an orangutan for a bodily counterpart.

But the analogy is imperfect. The two relations of unity are equivalence relations, at least for the most part and as a matter of contingent fact. Therefore it is easy and natural to form the concept of an enduring person or body, consisting of stages linked together pairwise by a relation of personal or bodily unity. It is tempting to do the same with the counterpart relations, forming the concept of a superperson or superbody consisting of persons or bodies in different worlds, linked together by a personal or bodily counterpart relation. But this cannot be done in any straightforward way because counterpart relations are not equivalence relations. Like all similarity relations on a sufficiently variegated domain, they fail to be transitive because chains of little differences add up to big differences.

Why should I think it plausible to employ multiple counterpart relations to translate (1′) as (1**) rather than (1*)? Precisely because by doing so I escape the refutation of (T), and I am convinced of (T). I am offered a trade: instead of a multiplicity of kinds of thing I can have a multiplicity of counterpart relations. A *reductio* refutes the whole combination of assumptions that led to contradiction; if all but one of those assumptions are highly plausible, whichever remains is the refuted one. And in addition, if I contemplate the propositions I express by means of (1′) and (1**), it seems to me that they are the same.

I would like to present the translation of (1′) by (1**) and (2′) by (2**) as instances of a general scheme for translating English modal predications into sentences of counterpart theory with multiple counterpart relations. I do not know how to do this. Roughly, the idea is that the sense of a term somehow selects the counterpart relation that is to be used to find the counterparts of the thing denoted by that term. The terms 'I', 'you', 'that person', 'the lady I saw you with last night', 'George', all select the personal counterpart relation. 'This thing' (pointing at myself), 'this body', 'my body', 'that which will be my corpse after I die', all select the bodily counterpart relation. Similarly for indefinite terms (phrases of restricted quantification): 'everybody' selects the personal counterpart relation, whereas 'every body' selects the bodily counterpart relation. Even if everybody is his body, and conversely, "Everybody is such that he might have been a disembodied spirit" is true, whereas "Every body is such that it might have been a disembodied spirit" is false. The former means that each of those things which are both persons and bodies

has a disembodied spirit as personal counterpart, whereas the latter means that each of those same things has a disembodied spirit as bodily counterpart.

In certain modal predications, the appropriate counterpart relation is selected not by the subject term but by a special clause. To say that something, regarded as a such-and-such, is such that it might have done so-and-so is to say that in some world it has a such-and-such-counterpart that does so-and-so. With these "regarded as" clauses in mind, I might say that I translate (1′) as (1**) because I take it to be synonymous with (1″):

(1″) I, regarded as a person, and my body, regarded as a body, are such that (without any duplication of either) they might not have been identical today.

Likewise I translated (2′) as if it had contained two "regarded as a body" clauses.

If we are to have multiple counterpart relations, we may well wonder how many to have. One for every sortal? One for every natural kind? One for any common noun phrase whatever that can grammatically be inserted into 'regarded as a ___', even the phrase 'yellow pig or prime number'? One for any kind of entity, even kinds that cannot be specified in our language?[5] I do not know. Nor do I know whether one of the counterpart relations, corresponding perhaps to the clause 'regarded as an entity', can be identified with the single counterpart relation of my original counterpart theory.

It is customary to distinguish real essences of things from their nominal essences under descriptions. Now, however, we have a third, intermediate, kind of essence. My real essence consists of the properties common to all my counterparts. (Here I use the original single counterpart relation.) My nominal essence under the description 'person' consists of the properties common to all possible persons. My intermediate essence under the description 'person' consists of the properties common to all my personal counterparts. I have no reason to think that any two of these sets of properties are the same. It may even be that none of the three is properly included in any other, if my personal counterparts include some entities (robots, say) which are almost persons but not quite. Counterpart relations are vague, being dependent on the relative weights assigned to respects of similarity or dissimilarity. Hence real

[5] We could also put this question another way: given a three-place relation "___is a ___al counterpart of___", which kinds are appropriate middle arguments?

essences are vague in a way nominal essences are not. Intermediate essences under descriptions share this vagueness, for the new multiple counterpart relations are no less vague than the original counterpart relation.

In my original counterpart theory, any *de re* modal predication is referentially transparent. Something has the same counterparts however we may choose to refer to it. Given a *de re* modal predication, we find the thing denoted by the subject term in the actual world; then we consider what befalls that thing—or rather, its counterparts—in other worlds. Only the denotation of the subject term matters. We can substitute another subject term with the same denotation but different sense, and the truth value of the modal predication will not change.

But in the present revision of counterpart theory, *de re* modal predications are not in general transparent. Not only the denotation of the subject term matters, but also the counterpart relation it selects. If we substitute another subject term with the same denotation but different sense, it may change the truth value of the modal predication by selecting a different counterpart relation. Then even though the denoted thing here in our world remains the same, we have a different way of following the fortunes of that thing in other worlds.

Nevertheless, these modal predications are still *de re*, not *de dicto*. We still find the denoted thing in our actual world and then find counterparts of that thing elsewhere. We do not at all consider the things denoted by the subject term in other worlds, as we would in the case of a *de dicto* modal predication.

Transparency of modal predications can fail whenever the sense of the subject term is used to do anything beyond determining the actual denotation of the subject term. One further thing it might do is determine the denotation of the subject term in other worlds; that is the *de dicto* case. Another, and altogether different, further thing it might do is select a counterpart relation. (These two are not the only alternatives.) It is the latter, I suggest, that happens in the argument we are considering. Therefore we can accept (1′) as a consequence of the possibility that I might have switched bodies, reject (2′) as self-contradictory, and yet accept (T) and its consequence that if I occupy the same body whenever I or it exist then I am my body.

DAVID LEWIS

St. Catherine's College, Oxford