

Possibility Relative to a Sortal

Draft, Version 1.3, 18 March 2008

Delia Graff Fara
Princeton University

I. Introduction.

When I first started in philosophy I had at least one philosophical view, and I thought that this was a view that doing philosophy would never cause me to reject. It was the view that a person is identical to his body (THE SENSIBLE VIEW). It hadn't occurred to me that there could be a further question about whether a person's body is identical to the matter that composes it. But if this question had occurred to me, I would have affirmed it just as wholeheartedly.

I had no idea that this view would be so difficult to defend, however, or that many had gone to great and counter-intuitive lengths in order to defend it. I still want to defend the view, but to do so without going to counter-intuitive lengths. That is my plan here.

The main arguments against The Sensible View, in its various applications to different sorts of material things, proceed by noticing that a name for one thing is not always intersubstitutable—*salva veritate*—with a name for a thing that The Sensible View would have it being identical to. For example, a name for a person's body is not always intersubstitutable, *salva veritate*, with a name for the totality of matter that makes it up.

(1) BODY will soon lack FINGER-MATTER as a part even while both still exist, but

(2) BODY-MATTER will never lack FINGER-MATTER as a part while both still exist.

One is supposed to conclude from the substitution failure that the body and its constituting matter are not identical since they have different futures. What will happen to the one is not what will happen to the other.

Analogous and equally convincing examples of substitution failure for person-names and body-names are not as easy to devise.

(3) MARIA will soon be somewhere that Maria-Body is not, but

(4) MARIA-BODY will never be anywhere that Maria-Body is not.

It doesn't seem at first that these could both be true since it seems that Maria's body will go wherever she goes. If Maria is to be somewhere without her body being there then Maria, before she gets there, would have to lack the body she has now. For (3) and (4) to both be true, Maria has to *abandon* her present body. But whether Maria could abandon her present body is extremely contentious. Body abandonment is arguably not an open future for Maria. So a surer argument against this motivating application of The Sensible View must proceed from a different pair of examples.

Perhaps the body could go somewhere without Maria going there? This might happen if Maria's present body were to abandon *her*.¹ This, arguably, is an open future for the body since, arguably, this is what happens in death. But it is also arguable that this is not an open future for the body. Even after Maria's death she, *Maria*, will be buried in the ground just as her body is, and she will decay just as her body does. Arguably. There is room for disagreement about whether a body might in the future abandon the person whose body it is as much as there is room for disagreement about whether a person might in the future abandon her current body.

The failures of intersubstitution that support the arguments against The Sensible View do not always trade on questions about which sorts of futures are open futures. They may also trade on questions about which sorts of counterfactual possibilities are open possibilities. Even if it is not open whether Maria or her body will abandon the other *in the future*, it could still be that either might not have been associated with the other had things been different. It could be that the association of a person with a body is permanent without being essential to either. Whether Maria could have had a different body had things been different is in principle separable from the question whether she might, in the future, as a matter of fact, abandon the body she actually has. And vice-versa. There could be failures of intersubstitutability in modal or in temporal contexts.

Rather than cast around for convincing modal versions of the argument against *this* application of The Sensible View, let us note that there are convincing modal substitution failures in cases of other appealing applications of it. It's not just the identity of persons with their bodies that I'm interested in defending, but also the identity of material objects in general with the matter that constitutes them.

(5) BOAT could only have been boat shaped, but

(6) BOAT-WOOD might not have been boat shaped.

¹I speak of *abandonment* rather than *separation* or *disassociation* because *abandonment* is straightforwardly asymmetric. If *A* abandons *B*, then *A* but not *B* must exist when the abandonment is over.

I actually think there is room for denying (5), but let me accept it for the sake of argument. In fact, I might as well accept (3–4) and the like for the sake of argument. Granted that there is at least one incontrovertible case of modal substitution failure, my defense of The Sensible View would be piecemeal—with a different fragment for each *sort* of thing that the view applied to—if my defence could not cover (3–4) and (5–6).

There are other examples of substitution failure that aren't obviously within either a modal or a temporal context:

- (7) Carla loves MARIA, but
- (8) Carla does not love MARIA-BODY; and
- (9) CARLA loves Maria, but
- (10) CARLA-BODY does not love Maria.

But we will leave such examples aside for another time. Here we will provide a SENSIBLE EXPLANATION of the substitution failures in contexts that are explicitly either modal or temporal.

II. The Leading Idea.

The LEADING IDEA rests on an intuitive distinction. The distinction is that between *identity* and *sameness*. There is just one *identity* relation, but there are many *sameness* relations. As many have said, *identity* is that relation which everything bears to itself but to no other thing. But something may be the same as something other from itself, and in a number of different respects. *Sameness* relations fall under at least two categories: those that are relativized to a quality (e.g., *same color* or *same height*) and those that are relativized to a *sort* of thing (e.g., *same person* or *same boat*).

Symmetry and transitivity are logical properties of both of these kinds of sameness relation. If I'm the same height as you and you're the same height as Maria, then you're the same height as me (symmetry), and I'm the same height as Maria (transitivity). If RUSTY and SCARAB and also SCARAB and LANCHA are the same boat as each other, then so are RUSTY and LANCHA. Relative-sameness relations are not quite equivalence relations, however, since they are not all reflexive. Since Barack Obama is not a boat, he is not the *same* boat as anything, even himself. Obama does not bear the *same-boat* relation to himself. The only way for you to be the same boat as something is if you yourself are a boat. Only a boat

can bear this relation to something. And, of course, every boat does bear this relation to itself. Since all and only boats bear the relation to anything, and all boats bear the relation to themselves, we may say that the *same-boat* relation is *weakly reflexive*. A relation is *weakly reflexive* when anything that bears the relation to anything at all bears the relation to itself.

WEAK REFLEXIVITY: $\forall x(\exists yxRy \rightarrow xRx)$.

We will say that a relation is a ‘weak-equivalence relation’ if it is symmetric, transitive, and weakly reflexive. Weak equivalence is a logical property of all *relative-sameness* relations. (From here onwards, I will forget about relations of sameness relativized to a *quality*. As shorthand I will use ‘*relative-sameness* relation’ just to mean a *relation of sameness relativized to a sort of thing*

Some have used the term relative ‘identity’ to denote what I am calling *relative sameness*. Proponents of ‘relative identity’ have denied that there is any relation of *absolute* identity, instead only *same-F-as* relations. It is for this reason worth emphasizing that in admitting the various *relative-sameness* relations, I in no way reject absolute identity.

I do, however, want to wedge a greater gap between *relative sameness* and *absolute identity* than philosophers will usually allow space for. For example, I think the boat that Michael embarked on is the same boat as the one he will row ashore, even though these boats are not *absolutely identical*. I am happy to dispense with the qualification ‘absolute’ since I do not think that there is any kind of *identity* other than absolute identity. The boats are the same though not identical,² for identity is ‘too unyielding a relation’ (to borrow Robert Stalnaker’s words) to hold of entities composed of different matter.

Granting this much, what should we say about the truth of the following temporal and modal claims? What conditions must be met in order for them to be true?

(11) JORDAN, Michael’s boat, will be rowed ashore;

(12) JORDAN might have been rowed ashore.

The unpopular answer, derived from David Lewis (1968, 1971, 1986), is that these sentences are true just in case there is a future time (11), or a possible world (12), at which

²When we say that the boats are the same, we mean (perhaps defeasibly) that the boats are the same *boat*. This is like the fact that when we say that a certain fly is small, we mean—defeasibly—that the fly is small for a *fly*.

something that's a *counterpart* of Jordan is rowed ashore. It doesn't matter to us yet what relation this *counterpart* relation is.³ The important point for us now is that it may hold between things that are not identical to each other.

The popular answer, taken to derive from Saul Kripke (1963) and (1972), is that these sentences are true just in case (A.) there is a future time, or a possible world, at which something absolutely identical to Jordan is rowed ashore. Do we accept the popular answer, so phrased? For Kripke, and those who follow him on this question, there is no distinction between this condition and the condition that (B.) there be a future time or possible world at which something that's the *same boat* as Jordan is rowed ashore. But for us, the two conditions come apart.

The leading idea for our answer is that (B.), the second phrasing of the popular view, is the correct one. The idea is that a future claim that *a* will be Φ , or a modal claim that *a* could have been Φ , is true when there is in the future, or in another possible world, something to which *a* bears a *sameness* relation that *is* Φ at that future time, or in that other possible world. Which *sameness* relation need be involved (*same person, same body, same boat, et cetera*) is determined by context.⁴ We will have to sharpen this up in a bit. But before we do, I want to point out that this leading idea about the truth conditions for *de-re predications* (temporal or modal) thus has something in common with both the popular view and the unpopular view.

It shares with the popular view the idea that future and counterfactual possibilities for *y* may suffice for the truth of claims about what *x* will be or could have been like only when *x* bears a *sameness* relation to *y*. Because the theory has this feature, I call it a *sameness* theory of *de-re* predication.

It shares with the unpopular view the idea that future and counterfactual possibilities for *y* may suffice for the truth of claims about what will or could happen to *x* even when *x* is not identical to *y*. Because the theory has this feature, I consider it to be a *counterpart* theory of *de-re* predication.

³I depart from Lewis in using the word 'counterpart' to denote the cross-time as well as the cross-world relation involved. [Fair enough?] Since I will continue to use the term in this way, I would have to give a disjunctive definition of the relation—one part for the trans-world component; another for the cross-temporal component—if I were to be thorough. But I won't be thorough. I will focus solely on the modal case. See Lewis (1976).

⁴This is as in Lewis (1971) and Allan Gibbard (1975), whose views on contingent identity inspired me to develop the view presented here. I differ with each of them on a number of fundamental issues, among them, the question whether objects are extended in time.

Although I call the theory a *counterpart* theory, I should emphasize that there are nonetheless many points of disagreement with the postulates that were originally stipulated by David Lewis to constitute counterpart theory. In particular, I deny Lewis’s postulate **P2**, the stipulation that individuals are ‘world-bound’—that none exists in more than one possible world:

$$\mathbf{P2} : \forall x \forall y \forall z ((x \text{ is in possible world } y \wedge x \text{ is in possible world } z) \rightarrow y = z).$$

III. Sharpening the Idea.

World-Time Relativization.

Whether or not x is the same *boat* as y —for example—will depend on what x is like and on what y is like. But things change over time (‘cross-temporally’), and they differ from how they might alternatively have been (‘cross-worldly’).⁵ So if the sameness of x and y is to depend on what they’re like, it must depend on what they’re like at a time and a world. Since we’re primarily interested in cross-temporal and cross-world sameness—whether x is the same boat as y given the way x is at the world–time ‘point’ wt and the way y is at a different point $w't'$ —we need to further relativize *relative-sameness* relations, to worlds and times. The *relative-sameness* relations appealed to in our theory of *de-re* predication will officially be understood as relations holding between thing–world–time triples. In other words, whether $\langle x, w, t \rangle$ bears a *relative-sameness* relation to $\langle y, w', t' \rangle$ depends on what x is like at wt and what y is like at $w't'$. If $\langle x, w, t \rangle$ is to be the same *statue* as $\langle y, w', t' \rangle$, then x ’s shape at wt must be similar to y ’s shape at $w't'$ —arguably. This last statement brings up three important further points.

First, the necessary conditions for cross-world or cross-temporal sameness, relativized to this or that sort of thing, will not *per se* constitute part of the theory here. We are interested only in the formal properties of relative sameness. The conditions for sameness of statues over time, or across worlds, is the province of aesthetics, not of our theory. The conditions for sameness of persons will be part of the theory of personal ‘identity’, not of our theory. The conditions for sameness of boats, of the theory of property rights (among others).

Second, if we were to maintain the aesthetic view that similarity of shape is necessary for cross-world or cross-temporal sameness of a statue while maintaining the logical view

⁵We extend all claims about a single *relative-sameness* relation (e.g., *same boat*) to other example *relative-sameness* relations (e.g., *same water*, *same chair*, *same person*), making changes as necessary.

that this relation is transitive, then we should admit that the relation is vague. We could then appeal to a theory of vagueness to evaluate the effects of this on our theory of *de-re* predication.

Third, if we were to give up the aesthetic view that similarity of shape is necessary for the cross-temporal sameness of a statue and hold instead that change in shape is permitted as long as it is sufficiently slow and gradual, then we would have to allow that whether $\langle x, w, t \rangle$ is the same statue as $\langle y, w, t' \rangle$ depends not only on what x is like at w, t and what y is like at w, t' , but also on what some potentially distinct things are like at intervening times.

Sameness and Identity.

We have relativized the (sort-relativized) sameness of objects to worlds and times. This is not to admit funny objects— things-at-a-world-at-a-time, that are somehow world–time ‘slices’ of (somehow) modally-extended and (somehow) temporally-extended objects. We are genuinely regarding these relations as relations on triples. The first member of each triple is a normal object. Normal objects do not have some of their parts in one possible world, some other parts in some other worlds. And they do not have some but not all of their parts at different times. These relations are symmetric, transitive and weakly reflexive *as* relations on triples.

SYMMETRY: $\langle d, w, t \rangle R \langle d', w', t' \rangle \rightarrow \langle d', w', t' \rangle R \langle d, w, t \rangle$.

TRANSITIVITY: $(\langle d, w, t \rangle R \langle d', w', t' \rangle \wedge \langle d', w', t' \rangle R \langle d'', w'', t'' \rangle) \rightarrow \langle d, w, t \rangle R \langle d'', w'', t'' \rangle$.

WEAK REFLEXIVITY: $\exists d' \exists w' \exists t' \langle d, w, t \rangle R \langle d', w', t' \rangle \rightarrow \langle d, w, t \rangle R \langle d, w, t \rangle$.

Like all relations, *relative-sameness* relations hold *at* worlds and at times, even when construed as relations on thing–world–time triples. This is analogous to one thing–world–time triple being preferred by me to another thing–world–time triple. The *preferred-by-me* relation is one that holds at a world and a time, even though it itself may hold among thing–world–time triples. I now prefer the triple $\langle \text{Clarissa}, @, 12 \text{ May } 2006 \text{ } 3 : 45 \rangle$ to $\langle \text{Clarissa}, @, 12 \text{ May } 2006 \text{ } 3 : 42 \rangle$ because I now prefer triples with later times to those with earlier times, all else being equal. But that could change, and it might not have been so. Which triples I prefer to which is something that may vary cross-temporally as well as cross-worldly. The relation is not a permanent or essential one. I suspect that *relative-sameness* relations on thing–world–time triples do, in contrast, hold permanently and essentially when they hold at all. But I officially leave this question open. Nothing here explicitly depends on it.

We do not deny, however, that there are normal *relative-sameness* relations that hold among things *simpliciter* rather than thing–world–time triples, i.e., without being further relativized to world and times. I am the same person as myself, the same mother as myself, the same philosopher as myself. . . , *simpliciter*. Nothing distinct from me could be the same woman as myself *simpliciter*, since one woman cannot have properties that are incompatible with others of her properties. A woman cannot be in two places at once, for example. We can put this point formally as follows. *Relative-sameness* relations on things *simpliciter* are not only weak equivalence relations, they are subsets of the *identity* relation on things. They will be *proper* subsets of the identity relation on things whenever they are relativized to a sort of thing that is non-vacuous, that is, to a sort of thing that some things do not fall under. For example, I am not the same *boat* or *statue* as myself since I am neither a boat nor a statue. The normal same-boat-as relation is that subset of the identity relation that contains those pairs $\langle d, d \rangle$ such that d is a boat. Vacuous sorts, if we admit them (as some do not) would include *thing* and *entity*, and perhaps *object*.

The fact that *relative-sameness* relations on things (*simpliciter*) are subsets of the identity relation can be represented formally as a kind of functionality condition on *relative-sameness* relations as they hold among thing–world–time triples:

$$\text{THING FUNCTIONALITY: } \langle d, w, t \rangle R \langle d', w, t \rangle \rightarrow d = d'.$$

It is because this condition holds that we may truly say of ourselves (when in the point w, t) that we are not the same person (philosopher, *et cetera*) as anything other than ourselves. So sameness really is a kind of identity after all. *Relative sameness* does on our view, as on the standard view, require (i.e., entail) *identity*. We differ from the standard view in holding that cross-world and cross-temporal *sameness* do not require cross-world or cross-temporal *identity*. In saying this, we do not thereby reject the notion of cross-world or cross-temporal identity. I am happy to admit these relations, and without them could not state the difference between my theory of *de-re* predication and the popular view.

Thought of in these terms, the representative difference between our view and the standard view of the relation between *identity* and *relative sameness* is that $\langle \text{Goliath}, @, \text{NOW} \rangle$ is the same_{DWT} lump of clay as $\langle \text{Lump1}, @, \text{NOW} \rangle$ just in case Goliath is the same_D lump of clay as Lump1, which holds, as all parties agree, only if Goliath is identical to Lump1. (Here I use the subscripts to indicate the type of the arguments of the *relative-sameness* relation—thing–world–time triples, or just things.)

An irrelevant but curious point is that we may construe cross-world and cross-temporal *sameness* relations, the ones on thing–world–time triples, as special cases of regular *sameness* relations, the ones on things. This is because thing–world–time triples are a special case of a thing.

There are further logical questions here which we could explore but won't since they are not difficult and not especially pertinent. For examples, how do we give a general definition of the correspondence between *relative-sameness* relations on triples and *relative-sameness* relations on things that holds above between the *same_{DWT}-lump-of-clay* and the *same_D-lump-of-clay* relations? And what set-theoretic relations hold between corresponding *relative-sameness* relations?

IV. Using the Idea.

We initially framed the problem for The Sensible View as a problem of substitution failure. The problem for The Sensible View can also be framed as the problem of having contingent and temporary identity as a commitment. If a wooden boat is identical to the wood that materially constitutes it, then we must say that the identity is a temporary one. The boat and the wood will not be identical later because later the boat will be constituted by different wood. The boat may have some planks replaced, lose some splinters, etc. But the wood, *that wood*, cannot later be constituted by different wood. Without those splinters, we do not have the same wood. Sameness of portion-of-matter parts is *the* criterion of sameness for portions of matter. Similarly, the identity between the boat and the wood must be a contingent one. The boat could have been constituted by different wood. But the wood itself cannot have been constituted by different wood.

The idea we've been developing is the idea that the boat here now, the one just departed from this shore, can be the same boat as that one there later, the one that will arrive at that shore, even though this boat here is not identical to that boat there. The point in developing this idea is to defend The Sensible View that wooden boats are identical to the wood that constitutes them.

The defense consists of an explanation and justification of the truth of statements of temporary and contingent identity. We use the leading idea to do this.

TEMPORARY IDENTITY: $b = w \wedge \text{WILL} : b \neq w$.

Conjunct one is true because the boat and the wood now consist of exactly the same matter. The second conjunct is true because at some later time, something that's the same boat as this one will be distinct from something that's the same wood as this wood. The criteria for later sameness of boat are not the same as the criteria for later sameness of wood.

CONTINGENT IDENTITY: $b = w \wedge \diamond : b \neq w$.

Similarly, for modal claims. What matters for the truth of the second conjunct is not cross-world identity to distinct objects, but rather cross-world sameness to distinct objects, which is made possible by relativization of the samenesses to two different sortals.

The idea is that we have truth conditions indicated by the following statements.⁶

(13) $\text{WILL}:\Phi a$ is true just in case at some future time, there's something that's the same F as a that satisfies Φ .

(14) $\Box\Phi a$ is true just in case in every possible world there's something that's the same F as a that satisfies Φ .

In each case, the sortal F is determined by context, where this may include words occurring elsewhere in the sentence. Different sortals may be associated with different names or variables as they occur in the scope of modal or tense operators, so that we have the following, for example.

(15) $\diamond\Phi xy$ is true just in case in some possible world there's something that's the same F as x that bears the relation Φ to something that's the same G as y .

It should be clear how to generalize this to cases involving more free variables. The pertinent detail is that different variables may be associated with different sortals and hence different relative-sameness relations. In particular, we have:

(16) $\diamond x \neq y$ is true just in case in some possible world there's something that's the same F as x that's distinct from something that's the same G as y .

Let us introduce the following terminology. x and y are ' F -mates' just in case x is the same F as y . The we have for instance that:

⁶For my explicit and elaborate statement of truth clauses for modal and temporal claims, see Fara (2008).

- (17) $\diamond b \neq w$ is true just in case there's a possible world in which b has a BOAT-mate that's distinct from w 's WOOD-mate.

V. Defending the Idea.

Substitution Failures? No Problem.

One could complain, against the theory, that it does not validate Leibniz's Law. Once the complaint is accurately stated, however, it should not worry us too much. There are two different versions of the Leibniz's-Law conditional, (I) the sentence-schema version, and (II) the property-version.

LEIBNIZ LAW I. $\forall x \forall y (x = y \rightarrow (\Phi x \rightarrow \Phi y))$,⁷

LEIBNIZ LAW II. $\forall x \forall y (x = y \rightarrow \forall z (z \text{ is a property} \rightarrow (x \text{ has } z \rightarrow y \text{ has } z)))$.

The first version of Leibniz's Law is properly called the principle of SUBSTITUTIVITY OF IDENTICALS rather than Leibniz's Law. The two may come apart since it may be that x doesn't occur referentially in Φx . In saying that x is Φ , we may not be attributing a genuine property to x . Substitutivity of Identicals simply is not a valid principle. A classic counterexample is Quine's: 'Giorgione was so-called because of his size'. Whether Giorgione satisfies the predicate '_____ was so-called because of his size' depends on how we are referring to him. That relative-sameness counterpart theory invalidates the substitution principle is not *per se* a problem with it. Moreover, we have an explanation for why it fails in cases of attributions using modal and temporal predicates. Our case is '_____ will be/might have been constituted by different wood'. Whether BOAT satisfies this predicate depends on whether we are thinking of it or describing it as a boat or as a portion of wood-matter. I follow Gibbard (1975, 201) in saying that for this reason we do not attribute a *property* to an object with such temporal and modal predications, and hence do not have a counterexample to Leibniz's Law properly understood as the principle that only distinct objects can diverge in their *properties*.

If the opponent of our version of counterpart theory grants us our view that the truth of attributions using modal and temporal predicates depends on how we are thinking, describing, or conceiving of the object that the predicate is being attributed to, then she must grant

⁷More pedantically: $\forall x \forall y (x = y \rightarrow (\Phi \rightarrow \Phi^y/x))$, where Φ^y/x is the result of replacing all free occurrences of x in Φ with y , uniformly relettering as required, so as to avoid capturing new occurrences of y with a quantifier in Φ .

us that the concomitant failures of substitutivity are desirable. The substitutivity-failure objection is no objection to the view by itself. It constitutes an objection only together with a flat-out denial of our view—only together, for example, with the claim that the occurrence of the names GOLIATH and LUMP1 are purely referential as they occur in the context of modal and temporal predications.

The technical explanation of the invalidity of the Principle of Substitutivity is that different variables may be associated with different relativized-sameness relations (BOAT-mate WOOD-mate, *et cetera*). The principle of Substitutivity of Identicals is, however, valid for the special case of predications not involving modal or temporal operators.

Similarity-Based Counterparts? Problems.

My further defense of the relative-sameness—crucially context-dependent—version of counterpart theory proceeds by comparing it to David Lewis’s (1971) amended, context-dependent version of counterpart theory. For Lewis, as for us, different counterpart relations may be associated with different names or variables.⁸ The crucial differences between the two views are (i) the difference between the formal properties of our relative-sameness counterpart relations and his similarity-based counterpart relations and (ii) the difference in intuitive attractiveness of the appeal to sameness rather than similarity as the requirement for truthful attributions of necessity and possibility.

Lewis’s counterpart relations were each a kind of *comparative*-similarity relation.⁹ These relations differed from each other by differing in which respects of similarity were given most weight. So there were *person*-counterparts, *body*-counterparts, *statue*-counterparts, *et cetera*. *X* has *Y* as an *F*-counterpart, on Lewis’s view, just in case *Y* is sufficiently *F*-similar to *X* (i.e., similar in respects important for *F*-hood), and *Y* is no less *F*-similar to *X* than anything else is in the world that *Y* occupies. Lewis’s counterpart relations therefore failed to be either symmetric or transitive.¹⁰

In fact, each of Lewis’s *F*-counterpart relations allowed for an individual to have multiple *F*-counterparts in another possible world. When *Y* and *Z*, occupying the same world, are

⁸Lewis does not present a formal theory once he adds context dependence to his (1968) theory. But in describing his amended theory, I formalize it where convenient in the way that seems to me most natural or straightforward.

⁹From now on I ignore temporal predication and discuss only modal predication. For me, the two cases are exactly analogous, since I am a ‘three-dimensionalist’. But for Lewis, a ‘four-dimensionalist’, the counterpart relation involved in temporal predication are very different from the ones he uses for modal predication. The relation for temporal predication is that of being temporal parts of the same temporally-extended objects. Lewis wouldn’t even call these counterpart relations. I do, however, since they are not the identity relation.

¹⁰See pp115–16 for Lewis’s explanation of these failures.

equally F -similar to X and no less so than anything else in their world, then both Y and Z are F -counterparts of X .¹¹

The admission of multiple counterparts raises a question about how to interpret claims of necessity and possibility. If necessity is truth at every possible world, then in order for necessarily- Φ to be true of you, do you have to have some or other counterpart that's Φ in every possible world? And in worlds where you have multiple counterparts, do all of them have to be Φ in that world? Lewis's answers were no and yes, respectively. he proposed that in order for necessarily- Φ to be true of you, in any world in which you have counterparts, each of those counterparts must be Φ . So for necessarily- Φ to be true of you, it needn't be that something satisfies Φ in each and every possible world. In worlds where you have no counterpart, it might be that nothing at all does.

Analogously, if possibility is truth at some possible world, then in order for possibly- Φ to be true of you, does there have to be at least one world in which all of your counterparts are Φ , or just in which at least one of your counterparts is Φ ? For Lewis, it was the latter.

For us, in contrast, multiple counterparts are banned, due to the weak reflexivity and thing functionality of relative-sameness relations. Let's see why. If $\langle d, w \rangle$ and $\langle d', w' \rangle$ are F -mates, and $\langle d, w \rangle$ and $\langle d', w' \rangle$ are F -mates, then by symmetry and transitivity, $\langle d', w' \rangle$ and $\langle d', w' \rangle$ are F -mates, which by thing functionality requires that $d' = d'$. (We've now suppressed relativization to times in order to improve readability.)

Given the ban on multiple counterparts, our truth conditions for claims of necessity and possibility can be stated in a way that allows for easy comparison with Lewis. For us, necessarily- Φ is true of you when in every possible world, *the* F -mate of you in that world is Φ . Analogously, possibly- Φ is true of you when there's some possible world in which *the* F -mate of you in that world is Φ . The definite descriptions here are understood as having Russellian (1905) truth conditions. If you have no F -mate in a possible world then it is *false* to say of you that *the* F -mate of you in that world is Φ .¹²

Unlike Lewis, we recognize the standard distinction between your being necessarily- Φ and your being *essentially*- Φ . On our theory, for you to be necessarily- Φ it must be that in each possible world your F -mate in that world is Φ . In order for you to be essentially- Φ , it must be that in every possible world in which you *have* an F -mate, your F -mate in that world is Φ . The distinction is desirable since even if you're a person, there will be many possible

¹¹From here onwards, when we refer to one of Lewis's counterpart relations, in the singular, we make general claims about any one of his counterpart relations.

¹²Here we ignore complications raised by predicates such as 'doesn't exist'.

worlds in which nothing at all is the same person as you. Many possible worlds contain only barren deserts, unpopulated by life of any kind. So it's not at all necessary that you're female. But if you are female, then you are *essentially* so. Nothing could be *you* without being female.¹³

This important difference between the two theories comes into sharp relief when we consider the question of necessary versus essential existence. On our theory, a person exists necessarily only if in every possible world, something existing there is the same person as him. A person exists essentially if in every possible world in which that person has a person-mate, his person-mate exists there. The essential-existence condition is trivially met since it's not possible to be a person without existing. All people, like most—arguably all—sorts of things, exist essentially. Existence is an essential feature of their existence. The necessary-existence condition is very difficult to meet, however. As I said before, many possible worlds are barren deserts, unpopulated by life of any kind.

On Lewis's theory, a person exists necessarily as long every possible world in which she has a person-counterpart is one in which all of her person-counterparts exist. This condition is trivially met, not just by people but by every sort of existing thing whatsoever. On Lewis's theory, though not on ours, everything exists necessarily.

More Problems for Similarity-Based Counterpart Theory.

I would like to cover two further important problems for similarity-based counterpart theory that aren't problems for relative-sameness counterpart theory. These are developments of problems which to my knowledge were first pointed out by Allen Hazen (1979).

THE FIRST.

(18) Necessarily, Adam is Abel's father,

$\Box Rab.$

We have not yet been specific about Lewis's truth conditions for modal sentences with multiple free variables. If we have a sentence $\Phi\vec{x}$ containing the free variables x_1, \dots, x_n then $\Box\Phi\vec{x}$ is true just in case every world in which all of the x s have a counterpart is one in which the sequence of counterparts \vec{x}_c satisfies Φ . As for us, the counterpart relation may differ for distinct variables. We can think of each x_i as being associated with the F_i -counterpart relation, allowing that the various F_i s may be different from one another.

¹³When I say here that nothing could be you without being female, I mean that nothing could be *the same person as you* without being female. I take it that this can be left implicit, since context sufficed to associate the *person-mate* relation with 'you'.

For possibility: $\diamond\Phi\vec{x}$ is true just in case there's *some* world in which all of the x s have a counterpart and there's some sequence of counterparts \vec{x}_c there that satisfies Φ . Again, we can think of each x_i as being associated with potentially different F_i -counterpart relations.

There's now a respect in which the truth conditions of (18) are too weak and a respect in which they are too strong.

They are too weak because they render the sentence compatible with the possibility of Adam's having had no children at all. For they require merely that every possible world in which *both* Adam and Abel have a counterpart is one in which Adam is father of Abel. The existence of a world in which Adam has a counterpart, but Abel doesn't does not suffice to render the sentence false. Here's another way of making the same point: the sentence is entailed by it's merely being essential to *Abel* that he have Adam for a father.

The truth conditions for (18) are, on the other hand, too strong because they render the sentence incompatible with both men's having had identical twins, each with a life as similar as his twin's to that of their common actual-world counterpart. For then both Adam₁ and Adam₂ would be counterparts of actual Adam, and likewise for Abel₁ and Abel₂. Neither son in this counterfactual world would be son to *both* fathers. So as a matter of fact it's not necessary that Adam is Abel's father, since we here have a world in which it's not the case that each of Adam's counterparts is father of each of Abel's counterparts. In fact, it is not even the case, given the possibility we're considering, that it's *essential* to Abel that he have Adam for a father, since we here have a world in which Abel (as well as Adam) does have a counterpart, but in which not every counterpart of Adam is that counterpart's father.

On our theory, in contrast, the sentence is false for the right reasons. It is false because there are possible worlds in which Adam has no counterpart that had children at all.

And unlike Lewis's theory, our theory gets the truth conditions right for the corresponding essentiality claims about Adam and Abel.

(19) It is essential to Adam that he have Abel for a son,

$$\Box(Ea \rightarrow Rab).$$

(20) It is essential to Abel that he have Adam for a father,

$$\Box(Eb \rightarrow Rab).^{14}$$

¹⁴ Eb is an abbreviation for $\exists yb = y$

We render (19) true just in case every possible world in which there's something that's the same person as Adam (or, in which Adam has a counterpart) is one in which that person has as progeny something that's the same person as Abel.

We render (20) true just in case every possible world in which there's something that's the same person as Abel (or, in which Abel has a counterpart) is one in which that person has as a father something that's the same person as Adam. Whether (20) is true is an independent question concerning the 'essentiality of origins'.

THE SECOND FURTHER PROBLEM.

Here is a problem that arises once we try to accommodate the addition of an 'actuality operator' to our language.¹⁵ With such an addition the question arises, for Lewis's theory, whether it is to have a universal semantics, like that given for the necessity operator, or will it have an existential semantics, like that given for the possibility operator. Is $\mathbf{ACT}\Phi x$ to be true if *all* of x 's counterparts in the actual world satisfy Φ , or just if at least one of them does. As Hazen (p330) points out, either choice leads to a problem. The universal semantics allows for the truth of (21) while the existential semantics allows for the truth of (22).

(21) There could have been an object that actually exists without being either actually F or actually not F ,

$$\diamond \exists x(\mathbf{ACT}Ex \wedge (\neg \mathbf{ACT}Fx \wedge \neg \mathbf{ACT}\neg Fx));$$

(22) There could have been an object that actually exists while being *both* actually F and actually not F ,

$$\diamond \exists x(\mathbf{ACT}Ex \wedge (\mathbf{ACT}Fx \wedge \mathbf{ACT}\neg Fx)).$$

For the truth of (21) it suffices, given the universal semantics, that the x in question have an actual counterpart that's F and one that's not F . For the truth of (22) first it suffices, given the existential semantics, that the x in question have an actual counterpart that's F and one that's not F . In the first case this possibly actually existing thing lacks too many properties; in the second case it has too many. [Delete those two sentences?]

For us, since no object has more than one actual-world counterpart, for any object with an actual-world counterpart (i.e., that satisfies $\mathbf{ACT}Fx$), its actual-world counterpart will satisfy exactly one of $\mathbf{ACT}Fx$ and $\mathbf{ACT}\neg Fx$.

¹⁵The problem is explored more deeply by Michael Fara and Timothy Williamson (2005). See also Fara (2008, Remark K2) for more extended discussion of a possible response to Fara & Williamson on behalf of relative-sameness counterpart theory.

VI. Conclusion

The principle difference between similarity-based counterpart theory and relative-sameness counterpart theory, which causes their difference in handling of the problems just discussed is that latter allows for multiple counterparts within a world while the former does not. The difference, moreover, is not merely stipulated by either party, but is rather derivable from logical properties of comparative similarity as versus sameness.

The problems raised for Lewis's theory stemmed in the first case from our consideration of the universal truth condition that Lewis in fact proposed for the necessity operator:

LEWIS $_{\square}$: $\square\Phi\vec{x}$ \equiv In any world in which \vec{x} has a counterpart \vec{x}_c , every counterpart \vec{x}_c of the series \vec{x} satisfies Φ .

We did not consider alternative proposals that Lewis might have made. One might wonder whether Lewis's theory could be rescued from the difficulties we considered by revising the truth condition for the Box. Due to the weakness of this truth condition—its verification of $\Phi\vec{x}$ even if some x_i lacks a counterpart in some possible world (i.e., exists only contingently)—one might consider strengthening the condition to require that $\square\Phi\vec{x}$ be true only if each x_i exists necessarily (i.e., has a counterpart in each possible world). This strengthening could be provided for by:

KAPLAN $_{\diamond}$: $\square\Phi\vec{x}$ \equiv In every world, the series \vec{x} has some counterpart series \vec{x}_c that satisfies Φ .¹⁶

This strengthens the truth condition for the Box in the desired way, but Lewis notes that the condition creates a dilemma. He regards acceptance of either horn of the dilemma as worse than acceptance of the weakness of the truth condition displayed in LEWIS $_{\square}$. The dilemma is created because the change from LEWIS $_{\square}$ to KAPLAN $_{\square}$ raises a question about whether or how to revise the truth clause for LEWIS $_{\diamond}$:

LEWIS $_{\diamond}$: $\diamond\Phi\vec{x}$ \equiv In some world, the series \vec{x} has some counterpart series \vec{x}_c that satisfies Φ .

The revision of LEWIS $_{\diamond}$ that would maintain the duality of \square and \diamond (meaning that $\square A \equiv \neg\diamond\neg A$ for any formula A) would render the following satisfiable:

¹⁶Lewis (p119) attributes the suggestion to David Kaplan.

- (23) Something exists that could be distinct from itself;
 $\exists x \diamond x \neq x$.

One way for Lewis to avoid the dilemma of having to accept the possibility of self distinctness or to reject the duality of \square and \diamond would be to reject multiple intra-world counterparts. Lewis cannot reject multiple counterparts out of hand, however, for this would be incompatible with the comparative-similarity-basis of his counterpart relation. Two things can always be equally similar to a third thing. For him to give up the comparative-similarity basis of the counterpart relation would be to admit anathema-to-him haecceities.

Our relative-sameness based counterpart relation allows for the following sensible and attractive trio of views:

1. Being in a state is open as a future or counterfactual possibility for a person (boat, statue, *et cetera*) just in case *that very person*—i.e., an entity that's the same person as it—is in that state at some future time or in some possible world.
2. Persons are identical to their bodies, which are identical to the matter that constitutes them.
3. The view is not subject to the problems that allowance of multiple counterparts creates for Lewis's theory.

References

- Fara, Delia Graff (2008), 'Relative-Sameness Counterpart Theory', *Review of Symbolic Logic*. Forthcoming.
- Fara, Michael & Williamson, Timothy (2005), 'Counterparts and Actuality', *Mind* **114**: 1–30.
- Gibbard, Allan (1975), 'Contingent Identity', *Journal of Philosophical Logic* **4**: 187–221.
- Hazen, Allen (1979), 'Counterpart-Theoretic Semantics for Modal Logic', *Journal of Philosophy* **76**(6): 319–338.

- Kripke, Saul (1963), 'Semantical Considerations on Modal Logic', *Acta Philosophica Fennica* **16**: 83–94.
- Kripke, Saul (1972), 'Naming and Necessity', in D. Davidson & G. Harman, eds., *Semantics of Natural Language*, Reidel, Dordrecht, pp. 253–355. Reprinted as Kripke (1980).
- Kripke, Saul (1980), *Naming and Necessity*, Harvard University Press, Cambridge, MA.
- Lewis, David (1968), 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* **65**(5): 113–126.
- Lewis, David (1971), 'Counterparts of Persons and Their Bodies', *Journal of Philosophy* **68**(7): 203–211.
- Lewis, David (1976), 'Survival and Identity', in A. Rorty, ed., *The Identities of Persons*, University of Chicago Press, pp. 17–40.
- Lewis, David (1986), *On the Plurality of Worlds*, Blackwell.
- Russell, Bertrand (1905), 'On Denoting', *Mind* **14**: 479–493.
- Stalnaker, Robert (1986), 'Counterparts and Identity', *Midwest Studies in Philosophy* **11**: 121–140. Page references are to reprinted version (Stalnaker 2003).
- Stalnaker, Robert (2003), *Ways a World Might Be*, Oxford University Press, Oxford.