

RELATIVE-SAMENESS COUNTERPART THEORY

DELIA GRAFF FARA

Princeton University

Just as set theory can be divorced from Ernst Zermelo's original axiomatization of it, counterpart theory can be divorced from the eight postulates that were originally stipulated by David Lewis (1968, p. 114) to constitute it. These were postulates governing some of the properties and relations holding among possible worlds and their inhabitants. In particular, counterpart theory can be divorced from Lewis's postulate **P2**, the stipulation that individuals are 'world bound'—that none exists in more than one possible world:

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

Lewis's motivation for this postulate at the time seemed a pragmatic one—rather than admit entities without clear criteria of identity among them, he would admit a new entity whenever a clear criterion of identity was wanting. Since the criteria for identity among individuals in different worlds were considered unclear, the putatively identical individuals being so different from one another, the strategy was to avoid the difficult and unclear question of transworld identity by admitting distinct individuals for distinct possible worlds (1968, p. 114). World-boundedness would later be motivated by Lewis's belief in concrete, or radical, modal realism (Lewis, 1986).

Counterpart theory can also be divorced from Lewis's own views about the nature of the counterpart relation used in the analysis of *de-re* modality. Lewis's counterpart relation was a similarity relation: as long as y is sufficiently similar to x , x has y as a counterpart when y is more similar to x than anything else in y 's world is (1968, p. 114f). Lewis's counterpart relation therefore failed to be either symmetric or transitive.¹ Nevertheless, a theory of *de-re* modality could be a counterpart theory while involving a counterpart relation that, unlike Lewis's, was symmetric or transitive, or both, and could also involve a counterpart relation that allowed for bare identity switching—for the possibility that each of two people could be just the way the other in fact is. In particular, a counterpart theory could involve a counterpart relation that was not based on qualitative similarity at all.

Counterpart theory can also be divorced from the particular use Lewis put it to. A theory of *de-re temporal* predication could be a counterpart theory.²

What makes a theory of *de-re* predication a counterpart theory? Consider any theory according to which *de-re* predications in the scope of some modal or temporal operator \odot are given an interpretation in terms of a relation R that entails the following: $\odot \Phi$ is true just in case there is a world or time (meeting such-and-such conditions) in which there are objects satisfying Φ that each bear relation R to the values of the names and free variables

Received: June 29, 2008; in revised form: February 7, 2008

¹ See pp. 115–116 for Lewis's explanation of these failures.

² As in, e.g., Sider (1996).

in Φ . The theory will be a counterpart theory if relation R is any relation other than the identity relation.³

Robert Stalnaker (1986, pp. 113–117) has distinguished three views which might motivate a philosopher to accept some version of counterpart theory: (i) radical modal realism, the view that counterfactual possible worlds and their inhabitants are as real, with as concrete an existence, as the actual world and its inhabitants, which requires that inhabitants of distinct possible worlds be distinct from each other (wherever they differ in their properties), that is, that *identity* (typically) does not hold across distinct worlds; (ii) anti-haecceitism, a rejection of any cross-world identity relation that does not depend on purely qualitative features of the inhabitants of the worlds involved; and (iii) the view that some statements of contingent identity are true.⁴

I have been motivated to develop a version of counterpart theory as a result of feeling forced (albeit reluctantly) to accept view (iii), which is not—as Stalnaker emphasizes—tied to Lewis’s radical modal realism or to anti-haecceitism. (I believe neither of these doctrines.) I have felt pushed to accept that some statements of contingent, as well as temporary, identity are true since I think that material objects are (actually now) identical to the matter that (actually now) constitutes them. And I believe that this holds no matter which world and time are the actual and current ones. People are identical to their bodies, which are in turn identical to the flesh, skin, and bones, etc., that make up those bodies. I am a ‘materialist’ in Lewis’s sympathetic terminology, a ‘one-thinger’ in Karen Bennett’s (2004) neutral terminology, and a ‘monist’ in the unsympathetic terminology of Kit Fine (2003). I believe that distinct material objects and portions of stuff cannot have the same material parts. They cannot be materially colocated. I cannot cogently hold this view, however, without adopting some form of counterpart theory in order to accommodate temporary or at least contingent identity. The reason I must accommodate temporary and contingent identity is that I accept the following two theses. First, human bodies (e.g.) *will* and *can* tolerate changes in their composition, since humans eat, excrete, slough, and the like. Second, portions of matter *will not* and *cannot* tolerate changes in their composition. Identity of portion-of-matter parts is the criterion of identity for portions of matter. So if the body will be composed of different portions of matter in the future, but the portion of matter it is now composed of will not, then the current identity between them is temporary. Similarly, if the body could be composed of different portions of matter, but the portion of matter it is actually composed of could not, then the actual identity between them is contingent.

Unlike Lewis, and a number of other counterpart theorists,⁵ I prefer to formulate the semantics of my version of counterpart theory for quantified modal and temporal logic by providing a definition of truth in a model (at a world and time, relative to a context), rather than by providing rules for translating sentences of quantified modal logic into those of first-order predicate logic with identity.

A. The leading idea. The leading idea is that relations of sameness, relativized to a *sort* of thing, can hold between objects that are not identical and that these relations

³ There need not be a *single* relation R involved in the analysis in order for the theory to be a counterpart theory; it could be that there is a family of relations involved, at least one of which is not the identity relation.

⁴ We find a counterpart theory motivated by view (iii) in the work of Gibbard (1975) and Lewis (1971), as well as in Stalnaker (1986).

⁵ Prominently, Forbes (1982, 1990) and Ramachandran (1989, 1990a).

are nevertheless the ones that are operative in *de-re* temporal and modal predications. For example, I think that the person sitting here now typing is not *identical* to the person who will fall asleep on my side of my bed tonight; they are not identical *entities*, for identity is too ‘unyielding a relation’ (to borrow Stalnaker’s words) to hold between entities composed of different matter.⁶ But I do think that the person sitting here now typing is nonetheless the *same person* as the one who will fall asleep on my side of my bed tonight. Because I think that the relation of *being the same person*—as contrasted with *absolute identity*, which I am proposing comes apart from it—is the one that is relevant for claims about what will be, or has been, or can be for a *person*, I hold the following:

- The truth of the claim that a certain person will be sleeping later does not depend on whether there is a later time at which an entity that is *identical* to that person is sleeping, but rather on whether there is a later time at which *that person* (an entity that is the same person) is sleeping.
- Similarly, the truth of the claim that a certain person could have six fingers does not depend on whether there is a possible world in which an entity that is identical to that person has six fingers, but rather on whether there is a possible world in which *that person* has six fingers.

Which relativized-sameness relation is operative in a given *de-re* temporal or modal predication is determined by context.⁷ It has been a doctrinaire assumption in recent philosophy, an assumption which I here reject, that if there is a relation of *absolute identity* at all that can hold over time, or across worlds, then *its* holding is a necessary condition of any sort-relativized sameness relation’s holding over time, or across worlds: for example, that personal ‘identity’ requires absolute identity.

There is only one identity relation: absolute identity. There are many sameness relations. They fall into at least two categories: those that are relativized to a quality (e.g., *color* or *height*) and those that are relativized to a sort of thing (e.g., *person* or *statue*). All sameness relations are weak equivalence relations.⁸ By a ‘weak equivalence relation’ I mean one that is symmetric, transitive, and weakly reflexive.⁹ The reflexivity should be weak since only a *person* can be the same person as something. If you do not have a daughter, then you cannot have the same daughter as anyone.

Let me now confine my attention to those relative-sameness relations that are relativized to a sort. Unless otherwise noted, ‘relative-sameness relation’ will henceforth include only sort-relativized sameness relations. Because all relative-sameness relations are weak equivalence relations, and because it cannot be at any single time (I hereby assume) that absolutely distinct entities bear one of these relative-sameness relations to one another, no entity will bear a relative-sameness relation to more than one entity at a single time in any world.¹⁰

⁶ I believe this brings me very close to the view that Timothy Williamson has called ‘crude reductionism’ (Williamson, 1998, p. 258).

⁷ This contextual variation in the relevant relation is as in Lewis’s revised (1971) version of counterpart theory.

⁸ Many doubt that the sameness relation *looking the same as* is transitive. I have argued that it is (Fara, 2001).

⁹ A relation is *weakly reflexive* when everything that bears the relation to anything at all bears the relation to itself.

¹⁰ This is codified in requirement D3(b) below.

Substituting a weak equivalence relation for Lewis's similarity-based counterpart relation allows us to avoid many of the objections that have been raised against Lewis's theory, which, as a haecceitist and merely moderate modal realist,¹¹ I want no part of.¹² But because *different* relative-sameness relations may be contextually associated with different variables, I am nonetheless able to achieve, as desired, the truth of certain statements of temporary and contingent identity. The theory I propose will be a counterpart theory, but because the relation I use will be so different from Lewis's counterpart relation, I will call it instead the '*rolemate* relation'.

In what follows, I first present the formal language for which I will subsequently be giving a relative-sameness (context-dependent) model theory. It will be a standard first-order language with identity and with tense as well as modal operators, including the indexical operators symbolizing 'actually' and 'now'. Comments of a philosophical and/or technical nature will be interspersed throughout, followed by some examples illustrating the mechanics of the truth definition, then some remarks concerning the consequences of the theory and in particular the lack of most of the bad consequences that have plagued Lewis's theory.

B. The language. Our language consists of the following primitive symbols, with standard formation rules and definitions for the usual nonprimitive symbols: (i) infinitely many variables, x_1, x_2, x_3, \dots ; (ii) for each n , infinitely many n -place predicates, $F_1^n, F_2^n, F_3^n \dots$;¹³ (iii) the symbol $=$, for (absolute) identity; (iv) symbols for conjunction and negation, \wedge and \neg ; (v) delimiters, (and); (vi) the universal quantifier, \forall ; (vii) the necessity operator, \Box ; (viii) the past and future tense operators, **P** and **F**; and (ix) the indexical modal and tense operators, **A** and **N**, for 'actually' and 'now'.

C. Definition of a model. A *model* \mathfrak{M} is a quadruple $\langle D, W, T, V \rangle$, where we have the following:

1. D is the (nonempty) domain of individuals (domains will not vary from world to world).
2. W is a nonempty set, understood to be representing a set of possible worlds.
3. T is any nonempty set of real numbers, which are understood to represent times.¹⁴
4. V is the interpretation function that assigns, at each world–time pair ('point') wt from $W \times T$, an extension of n -tuples from D^n to each n -place predicate F^n , denoted $V_{wt}(F^n)$.
5. There is no accessibility relation in the model. The semantics will yield **S5** as the modal logic for the propositional component.

¹¹ I am a haecceitist in Lewis's (1986) sense; whether that requires a belief in 'primitive thisness', I do not know. According to this view, there can be distinct possible worlds that are qualitatively just alike, yet which represent different *de-re* possibilities for an individual. For example, this view might allow for the possibility that I be just the way that my sister in fact is, with her being just the way that I in fact am.

¹² The objections I have in mind were all raised by either Hazen (1979) or Fara & Williamson (2005).

¹³ x, y, z , etc., may be occasionally used as variables and F, G, R , etc., as predicates so as to avoid unnecessary use of subscripts and superscripts.

¹⁴ The standard ordering relations on the numbers will be used to represent the standard temporal-ordering relations on times.

D. Definition of a context. A context c for a model $\mathfrak{M} = \langle D, W, T, V \rangle$ consists of the following three elements:

1. w_c , the (actual) world of the context, which is a member of W ;
2. t_c , the (current) time of the context, which is a member of T ; and
3. an assignment to each variable x of a relative-sameness relation, denoted R_c^x , where each of these relations is a relative-sameness relation (intuitively, one relativized to this or that *sort* of thing), where each of these relations is:

(a) a weak equivalence relation on $D \times W \times T$ (i.e., one that is symmetric, transitive, and weakly reflexive) and also

(b) a relation that is functional on its first right argument in the following sense: if $\langle d, w, t \rangle R_c^x \langle d', w', t' \rangle$ and $\langle d, w, t \rangle R_c^x \langle d'', w', t' \rangle$, then $d' = d''$.

Symmetry then guarantees functionality on the first argument in both directions: if $\langle d', w', t' \rangle R_c^x \langle d, w, t \rangle$ and $\langle d'', w', t' \rangle R_c^x \langle d, w, t \rangle$, then $d' = d''$.

The reason for requiring that each sameness relation be a relation on thing–world–time triples rather than a relation just on things is that whether one thing is the same F as another depends on what the two things are like and that in turn is something that can vary over time and across worlds. Portions of matter do not have all of their properties essentially. One may be human shaped at one time, scattered-dust shaped at a later time, and dolphin shaped in another possible world.

The reason for requiring weak equivalence, recall, is that weak equivalence is a logical property of relative sameness, whether that sameness be relativized to a quality (*the same color as*) or a sort of thing (*the same ship as*).

The reason for requirement (b), recall, is that we do not want distinct entities to be, for example, the same person as each other at any single world and time, since this would allow a single person to have incompatible properties. Weak equivalence alone, without requirement (b), would allow for the possibility of Ernie and Bert being the same person as each other, at a given world and time, even though they are in different places in that world at that time.

Definition f_c^x . Each variable x is thus correlated by the context with what I will call a *rolemate function* f_c^x defined as follows:

$$f_c^x(d_{wt}, w', t') = \begin{cases} d' & \text{if } \langle d, w, t \rangle R_c^x \langle d', w', t' \rangle, \\ D & \text{if there is no such } d'. \end{cases}$$

We say that f_c^x is an x -*rolemate function* and that it is *based on* R_c^x . Given an entity d at point wt , a world w' , and a time t' , each rolemate function picks out the entity's rolemate at $w't'$ if it has one.¹⁵

Each relative-sameness relation and each rolemate function based on it are thus relativized in a number of ways. First, they are both relativized to a variable—the one to which they are assigned. Second, they are relativized to a context—the one that assigns them to that variable. They are also relativized to a point wt , since the identity of d 's rolemate at $w't'$ will depend on how d is at world w and time t .

¹⁵ Clause D(3b) ensures that there is at most one such d' .

For example, if d is a scattered portion of dust at wt , then it cannot be the same *person* as anything at $w't'$, given how it is at wt . It could be, however, if it were a human body at wt . We do not want the *same-person-as* relation to cover the entire domain since people are not people forever or necessarily (they can and will die and turn to dust), and so do not bear the *same-person-as* relation to something at every world and time.¹⁶ Nevertheless, it simplifies the statement of the semantical rules if we define the corresponding rolemate function to be totally defined. We do this by conventionally assigning as a rolemate something that is guaranteed not to be in the domain, since something outside of the domain will be guaranteed not to be in the extension of any primitive predicate. D itself is just such an object.¹⁷ (We assume that the domain is not among the individuals in the domain.) We will say that an object has no rolemate at any world in which the rolemate function assigns D to it. For brevity, relativizations on the rolemate functions may be left implicit when doing so should cause no confusion.

These rolemate functions will be context-dependently invoked in the analysis of *de-re* modal and temporal predication and in that sense will serve the same purpose in this theory as Lewis's counterpart relations do in his (1971) context-dependent counterpart theory. We will therefore call this version of counterpart theory 'rolemate theory'.

E. Assignment functions. An *assignment function* g for a model $\langle D, W, T, V \rangle$ is a (total) function from the variables of the language to $D \cup \{D\}$, a set that is guaranteed to contain the image of every rolemate function. We require additionally that $g(v) \in D$ for some variable v .¹⁸

D is to be considered a 'non-entity' in the sense that it is never in the extension of any primitive predicate and is just conventionally assigned as the value of an R_c^x -based rolemate function at points where, intuitively, we would want the object in question to lack an R_c^x -based rolemate (e.g., because nothing is the *same person* as it is at that point).

F. Notation. I use the following abbreviations and notation:

1. For any (open or closed) sentence Φ , $\models_{wt}^{\mathfrak{M}cg} \Phi$ abbreviates the claim that Φ is true at point wt according to model \mathfrak{M} , context c , and assignment g .¹⁹
2. We take *strong validity*, which we write simply as \models , to be truth at every world and time relative to every model, context, and assignment:

$$\models \Phi \quad \text{iff} \quad \models_{wt}^{\mathfrak{M}cg} \Phi \text{ for every } w, t, \mathfrak{M}, c, \text{ and } g.^{20}$$

¹⁶ When I say that a relation does not 'cover' the entire domain, I mean that not everything in the domain bears or is borne the relation to or by something else.

¹⁷ Because we need to allow for indefinite nestings of modal and tense operators, each f_c^x needs to be defined for $\langle d, w, t \rangle$ not only when $d \in D$ but also when $d = D$. According to the definitions in D3, the value of the function is D in such cases.

¹⁸ See remark J17 below (page 185) for discussion.

¹⁹ Whenever \mathfrak{M} , c , and g occur together as superscripts, with w and t as subscripts or superscripts, it is to be understood that c is a context for \mathfrak{M} , that g is an assignment for \mathfrak{M} , that $w \in W$, and that $t \in T$. It should also be assumed that D , W , T , and V are from \mathfrak{M} whenever it would make sense to do so.

²⁰ The strong-weak terminology is borrowed from Hodes (1984).

3. Similarly, there is *strong implication*:

$$\Gamma \models \Phi \quad \text{iff} \quad \models_{wt}^{\mathfrak{M}cg} \Phi \text{ for every } w, t, \mathfrak{M}, c, \text{ and } g \text{ such that } \models_{wt}^{\mathfrak{M}cg} \gamma \text{ for all } \gamma \in \Gamma.$$

4. We take *truth in a model \mathfrak{M} relative to a context c* , which we write $\models^{\mathfrak{M}c}$, to be truth in \mathfrak{M} relative to c and every assignment g at the world and time of the context—that is, for any model \mathfrak{M} and context c for \mathfrak{M} ,

$$\models^{\mathfrak{M}c} \Phi \quad \text{iff} \quad \models_{w_c t_c}^{\mathfrak{M}cg} \Phi \text{ for every assignment } g.$$

5. Φ is *weakly valid*, which we write $\models_{w_c t_c}$, just in case it is true in every model relative to any context:

$$\models_{w_c t_c} \Phi \quad \text{iff} \quad \models^{\mathfrak{M}c} \Phi \text{ for any } \mathfrak{M} \text{ and } c.$$

6. Similarly, there is *weak implication*:

$$\Gamma \models_{w_c t_c} \Phi \quad \text{iff} \quad \models^{\mathfrak{M}c} \Phi \text{ for every } \mathfrak{M}, c \text{ such that } \models^{\mathfrak{M}c} \gamma \text{ for all } \gamma \in \Gamma.$$

7. $g[d_1, \dots, d_n/v_1, \dots, v_n]$ is that assignment function which assigns the value d_i to the variable v_i , but is otherwise just like g (where $v_i \neq v_j$ iff $i \neq j$).

8. Where it should cause no confusion, I use the logical symbols \Rightarrow , \forall , and \exists to abbreviate the English expressions of the material conditional and of universal and existential generalization.

9. Φ^y/x is the result of replacing all free occurrences of x in Φ with y , uniformly relettering occurrences of y already in Φ as necessary to ensure that new occurrences of y are free in Φ .

G. Truth definition. Now we present the definition of truth in a model. All clauses except those for the modal and tense operators are classical.

1. Identity: $\models_{wt}^{\mathfrak{M}cg} x = y \quad \text{iff} \quad g(x) = g(y)$.²¹

Given this truth condition for identity statements, we are entitled to say that identity and distinctness are treated as neither contingent nor temporary in the sense that we never have a pair of objects in the extension of the identity relation at one world or time, but not at some other world or time. According to our truth clause, the identity symbol expresses the one relation of absolute identity.²² Nonetheless, our truth clauses will provide for the truth of some *statements* of contingent and temporary identity (and distinctness).

2. Predication. If P is an n -place predicate, then

$$\models_{wt}^{\mathfrak{M}cg} P v_1 \cdots v_n \quad \text{iff} \quad \langle g(v_1), \dots, g(v_n) \rangle \in V_{wt}(P).$$

3. Negation: $\models_{wt}^{\mathfrak{M}cg} \neg \Phi \quad \text{iff} \quad \not\models_{wt}^{\mathfrak{M}cg} \Phi$.

²¹ We could have here decided to require that d be a member of D . Some important effects of this decision will be pointed out below.

²² Cf. page 169.

4. Conjunction: $\models_{wt}^{\mathfrak{M}cg} (\Phi \wedge \Psi)$ iff $\models_{wt}^{\mathfrak{M}cg} \Phi$ and $\models_{wt}^{\mathfrak{M}cg} \Psi$.
 5. Quantifiers: $\models_{wt}^{\mathfrak{M}cg} \forall x \Phi$ iff for all d in D , $\models_{wt}^{\mathfrak{M}cg[d/x]} \Phi$.

\Rightarrow Note that the quantifiers range over every object in the domain at every point.

6. Modals: $\models_{wt}^{\mathfrak{M}cg} \Box \Phi$ iff for all $w' \in W$,
 $\models_{w't}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w',t}), \dots, f_c^{v_n}(g(v_n)_{wt,w',t})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the (distinct) free variables in Φ .

\Rightarrow For example, a necessitation $\Box \Phi(x)$ with one free variable is true at a point wt (according to a context) just in case x has an R -based rolemate at t in every possible world, and in each of those worlds, at that time, x 's rolemate there satisfies Φ , where R is the relation contextually assigned to x by the context, that is, where R is R_c^x .

7. Tenses:

- (a) $\models_{wt}^{\mathfrak{M}cg} \mathbf{F}\Phi$ iff for some $t' > t$,
 $\models_{wt'}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w,t'}), \dots, f_c^{v_n}(g(v_n)_{wt,w,t'})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ ;
 (b) $\models_{wt}^{\mathfrak{M}cg} \mathbf{P}\Phi$ iff for some $t' < t$,
 $\models_{wt'}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w,t'}), \dots, f_c^{v_n}(g(v_n)_{wt,w,t'})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ .

8. Indexicals:

- (a) $\models_{wt}^{\mathfrak{M}cg} \mathbf{A}\Phi$ iff $\models_{w_c t}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w_c,t}), \dots, f_c^{v_n}(g(v_n)_{wt,w_c,t})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ ;
 (b) $\models_{wt}^{\mathfrak{M}cg} \mathbf{N}\Phi$ iff $\models_{w_t c}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w,t_c}), \dots, f_c^{v_n}(g(v_n)_{wt,w,t_c})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ .

H. Defined symbols. Standard definitions for nonprimitive symbols yield expected truth conditions. For example:

- Given that $\exists x$ is defined to be $\neg \forall x \neg$,
 $\models_{wt}^{\mathfrak{M}cg} \exists x \Phi$ iff for some $d \in D$, $\models_{wt}^{\mathfrak{M}cg[d/x]} \Phi$;
- Given that Ex (' x exists') is defined to be $\exists y x = y$,
 $\models_{wt}^{\mathfrak{M}cg} Ex$ iff for some $d \in D$, $g(x) = d$;
- Given that \diamond is defined to be $\neg \Box \neg$,
 $\models_{wt}^{\mathfrak{M}cg} \diamond \Phi$ iff for some $w' \in W$,
 $\models_{w't}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w',t}), \dots, f_c^{v_n}(g(v_n)_{wt,w',t})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ ;
- Given that \mathbf{G} ('it will always be the case that') is defined to be $\neg \mathbf{F} \neg$,
 $\models_{wt}^{\mathfrak{M}cg} \mathbf{G}\Phi$ iff for all $t' \in T$ such that $t' > t$,
 $\models_{wt'}^{\mathfrak{M}cg[f_c^{v_1}(g(v_1)_{wt,w,t'}), \dots, f_c^{v_n}(g(v_n)_{wt,w,t'})/v_1, \dots, v_n]} \Phi$, where v_1, \dots, v_n are the free variables in Φ .

I. Examples. These examples will treat only the modal operators but are straightforwardly extended to the analogous temporal cases.

1. $\models_{wt}^{\mathfrak{M}cg} \mathbf{A}Fx$
iff $\models_{wct}^{\mathfrak{M}cg[f_c^x(g(x)_{wt}, w_c, t)/x]} Fx$ (G8)
iff $f_c^x(g(x)_{wt}, w_c, t) \in V_{wct}(F)$ (G2)
(*Gloss: x has an actual-world rolemate in the extension of F at the actual world.*)
2. $\models_{wt}^{\mathfrak{M}cg} \mathbf{A}\neg Fx$
iff $\models_{wct}^{\mathfrak{M}cg[f_c^x(g(x)_{wt}, w_c, t)/x]} \neg Fx$ (G8)
iff $f_c^x(g(x)_{wt}, w_c, t) \notin V_{wct}(F)$ (G3, G2)
(*Gloss: x lacks an actual-world rolemate in the extension of F at the actual world.*)
3. $\models_{wt}^{\mathfrak{M}cg} \Box \exists yx = y$
iff $\forall w' \in W, \models_{w't}^{\mathfrak{M}cg[f_c^x(g(x)_{wt}, w', t)/x]} \exists yx = y$ (G6)
iff $\forall w' \in W, \exists d \in D, \models_{w't}^{\mathfrak{M}cg[f_c^x(g(x)_{wt}, w', t)/x][d/y]} x = y$ (G3, G5)
iff $\forall w' \in W, \exists d \in D, f_c^x(g(x)_{wt}, w', t) = d$ (G1)
(*Gloss: at every world, x has an x -rolemate.*)
4. $\models_{wt}^{\mathfrak{M}cg} \forall y(\forall x Fx \rightarrow Fy)$
iff $\forall d \in D, \models_{wt}^{\mathfrak{M}cg[d/y]} \forall x Fx \rightarrow Fy$ (G5)
iff $\forall d \in D$, either $\not\models_{wt}^{\mathfrak{M}cg[d/y]} \forall x Fx$ or $\models_{wt}^{\mathfrak{M}cg[d/y]} Fy$ (G3, G4)
iff $\forall d \in D$, either $\exists d' \in D \not\models_{wt}^{\mathfrak{M}cg[d/y][d'/x]} Fx$ or $d \in V_{wt}(F)$ (G5, G2)
iff either $\exists d' \in D, d' \notin V_{wt}(F)$ or $\forall d \in D, d \in V_{wt}(F)$ (G2)
(*Gloss: the extension of F at w either does not include the domain or it does.*)
This is a tautologous truth condition, so the sentence is strongly valid.
5. $\models_{wt}^{\mathfrak{M}cg} \forall x Fx \rightarrow Fy$
iff $\models_{wt}^{\mathfrak{M}cg} \forall x Fx \Rightarrow \models_{wt}^{\mathfrak{M}cg} Fy$ (G3, G4)
iff $\forall d \in D, \models_{wt}^{\mathfrak{M}cg[d/x]} Fx \Rightarrow g(y) \in V_{wt}(F)$ (G5, G2)
iff $\forall d \in D, d \in V_{wt}(F) \Rightarrow g(y) \in V_{wt}(F)$ (G2)
(*Gloss: if the extension of F at w includes the domain, then it contains $g(y)$.)*
This may be false since when $g(y) = D, g(y) \notin D$.
6. $\models_{wt}^{\mathfrak{M}cg} \Box \forall x Fx$
iff $\forall w' \in W, \models_{w't}^{\mathfrak{M}cg} \forall x Fx$ (G6)
iff $\forall w' \in W, \forall d \in D, \models_{w't}^{\mathfrak{M}cg[d/x]} Fx$ (G5)
iff $\forall w' \in W, \forall d \in D, d \in V_{w't}(F)$ (G2)
(*Gloss: at every world, the extension of F includes the domain.*)
7. $\models_{wt}^{\mathfrak{M}cg} \forall x \Box Fx$
iff $\forall d \in D, \models_{wt}^{\mathfrak{M}cg[d/x]} \Box Fx$ (G5)
iff $\forall d \in D, \forall w' \in W, \models_{w't}^{\mathfrak{M}cg[d/x][f_c^x(d_{wt}, w', t)/x]} Fx$ (G6)
iff $\forall d \in D, \forall w' \in W, f_c^x(d_{wt}, w', t) \in V_{w't}(F)$ (G2)
(*Gloss: every object in the domain has, at every world, an x -rolemate in the extension of F at that world.*)

8. $\models_{wt}^{\mathfrak{M}cg} \Box \forall x \neg Fx$
iff $\forall w' \in W, \models_{w't}^{\mathfrak{M}cg} \forall x \neg Fx$ (G6)
iff $\forall w' \in W, \forall d \in D, \models_{w't}^{\mathfrak{M}cg[d/x]} \neg Fx$ (G5)
iff $\forall w' \in W, \forall d \in D, \not\models_{w't}^{\mathfrak{M}cg[d/x]} Fx$ (G3)
iff $\forall w' \in W, \forall d \in D, d \notin V_{w't}(F)$ (G2)
(*Gloss*: at every world, F has an empty extension.)
9. $\models_{wt}^{\mathfrak{M}cg} \forall x \Box \neg Fx$
iff $\forall d \in D, \models_{wt}^{\mathfrak{M}cg[d/x]} \Box \neg Fx$ (G5)
iff $\forall d \in D, \forall w' \in W, \models_{w't}^{\mathfrak{M}cg[d/x][f_c^x(d_{wt}, w', t)/x]} \neg Fx$ (G6)
iff $\forall d \in D, \forall w' \in W, \not\models_{w't}^{\mathfrak{M}cg[d/x][f_c^x(d_{wt}, w', t)/x]} Fx$ (G3)
iff $\forall d \in D, \forall w' \in W, f_c^x(d_{wt}, w', t) \notin V_{w't}(F)$ (G2)
(*Gloss*: every object in the domain lacks, at every world, an x -rolemate that is in the extension of F at that world.)
10. $\models_{wt}^{\mathfrak{M}cg} \Box x = y$
iff $\forall w' \in W, \models_{w't}^{\mathfrak{M}cg[f_c^x(g(x)_{wt}, w', t), f_c^y(g(y)_{wt}, w', t)/x, y]} x = y$ (G6)
iff $\forall w' \in W, f_c^x(g(x)_{wt}, w', t) = f_c^y(g(y)_{wt}, w', t)$ (G1)
(*Gloss*: at every world, $g(x)$'s x -rolemate is identical to $g(y)$'s y -rolemate.)

J. Remarks.

1. Hazen (1979) made a number of objections to Lewis's original (1968) version of counterpart theory, all of which turn on the fact that Lewis's *similarity*-based counterpart relation allows for an individual to have multiple counterparts in a single world. The version of counterpart theory proposed here is not guilty of any but one of Hazen's charges against Lewis. Due to the philosophically well-motivated constraints (set out in D3) that have been placed on each of the involved relative-sameness relations, the rolemate relations based on them are *functional*—each object in the domain has at most one R -based rolemate at any given world and time for any given relative-sameness relation R . As a result of this, the theory survives all of Hazen's trenchant objections. The only one of Hazen's charges that our relative-sameness version of counterpart theory is guilty of is his (p. 325) objection that some statements of contingent identity are satisfiable.²³ From my perspective (as it was from Lewis's), this is a feature rather than a bug. The others of Hazen's objections are dealt with in the remainder of this remark:

- (a) $\Box Rxy \models \Box \exists y Rxy$,
in the theory proposed here (but not in Lewis's).

Hazen complained about the fact that Lewis's theory does not validate the above inference. The complaint is well founded. If it is necessary that x bears a certain relation to y , then it must also be necessary that x bears that relation to something. The inference is valid according to our intuitive understanding of necessity and existential quantification.

²³ See remark J4 below.

The inference is valid on our theory—in every case where Rxy is an atomic predication other than identity—because (now ignoring relativization to times) the truth of $\Box Rxy$ at a world w requires it to be, in every possible world, that $g(x)$ has an x -rolemate given the way it is at w , that $g(y)$ has a y -rolemate given the way it is at w , and that these rolemates satisfy Rxy at that world. This also suffices for the truth of $\Box\exists yRxy$ at w . The reasoning is as follows. If $g(x)$ and $g(y)$ have rolemates that satisfy (atomic) Rxy at every possible world, then their rolemates must exist at every world. On our theory this means that at each world they have a rolemate that is *in D* (by the clause for atomic predication (G2)). This entails that $g(x)$ has, at each world, a rolemate that bears the relation R to some object from D , which is the condition for the truth of $\Box\exists yRxy$.

Nevertheless, the inference does not in general hold good. It is not the case for us that $\Box\Phi xy \models \Box\exists y\Phi(xy)$ no matter what formula $\Phi(xy)$ may be. For example, the inference does not hold when $\Phi(xy)$ is $x = y$. The nonexistent is self-identical in every possible world, but not identical to an existent in any possible world.

$$(b) \Box x = y \not\models \Box\exists yx = y.$$

This results from the fact that we do not require the assignment function to assign an object in the domain to each variable (see Section E), but we do require that the existential quantifier range only over objects that are in the domain. This combination of choices was made independently of the counterpart-theoretic analysis of modality and is detachable from it, although with some significant effects on the logic.

Another case for which the inference fails is when $\Phi(xy)$ is $\neg Rxy$. Adam might, of necessity, have no mother. But it needn't therefore be possible for there to be someone who is not his mother.

$$(c) \Box\neg Rxy \not\models \Box\exists y\neg Rxy.$$

This results from the fact that we allow a negation to be true even when a term in its scope does not denote an object in the domain. This choice was made in order to preserve bivalence and, again, was made independently of the motivation for counterpart theory.

Even for the case of atomic predications, the implication does not hold for Lewis, as Hazen points out (p. 326). Lewis's view would allow for the truth of the premise but not the conclusion in the following situation: (i) every world that has a counterpart of *both* x and y is one in which all of x 's counterparts bear relation R to all of y 's counterparts, but (ii) x has a counterpart in some world in which y has no counterpart. Condition (i) suffices for the truth of the premise in (1a), while condition (ii) suffices for the falsity of the conclusion.

Although both the Lewis models and the ones proposed here allow for individuals to lack rolemates at any given world, the relevant difference in the semantics is that our semantics for *de-re* necessity could be equivalently stated by saying that *being- Φ* is necessary *of* an individual just in case in every world, *the* individual's rolemate in that world is Φ , where this definite description is in turn understood as having Russellian truth conditions. It is precisely because Lewis allows for multiple intraworld rolemates that this condition would be too strict for him.

Hazen also objected that the following existence-qualified version of (1a) does not hold for Lewis. It does hold for us when Rxy is an atomic predication other than identity, and for us it holds in general for the same instances of $\Phi(xy)$ for which it holds for (1a).

- (d) $\Box(Ex \rightarrow Rxy) \models \Box(Ex \rightarrow \exists yRxy)$,
in the theory proposed here (but not in Lewis's).

That Lewis's semantics and ours differ in their verdict on (1d) is perhaps more relevant than their difference in verdict on (1a). The more realistic cases are instances of (1d)—for example, when the premise is x is *essentially* y 's son.²⁴ (This is contrasted with x is *necessarily* y 's son, since x , if someone's son, will not exist necessarily).

These first two objections of Hazen's are extremely serious, since the implications should clearly hold *even from Lewis's own philosophical perspective*. Whether the remaining objections are extremely serious in that sense is less clear, but they are serious nonetheless.

Although Lewis does not consider a language with an actuality operator, Hazen points out (p. 330) that on the two most plausible ways of augmenting Lewis's semantics to account for an extended language with an actuality operator, we get the implausible satisfiability of the negation of one or another of (1e–1f) below: of the negation of (1e), if we say that $\mathbf{A}Fx$ is true just in case *all* of x 's actual-world counterparts actually satisfy F ; of the negation of (1f), if we say that $\mathbf{A}Fx$ is true just in case *at least one* of x 's actual-world counterparts actually satisfies F . The problem here is that Lewis's theory allows for an object to have multiple actual-world counterparts. But not so for us.

- (e) $\models \Box\forall x(\mathbf{A}Ex \rightarrow (\mathbf{A}Fx \vee \mathbf{A}\neg Fx))$, on the theory proposed here.
(f) $\models \Box\forall x(\mathbf{A}Ex \rightarrow \neg(\mathbf{A}Fx \wedge \mathbf{A}\neg Fx))$, on the theory proposed here.

2. Fara & Williamson (2005, pp. 10–12, 15–17) have recently extended Hazen's work by arguing that not only do the two most plausible candidate truth clauses for the actuality operator not yield an acceptable result, no other one does either.²⁵

The complaint about (1e) and (1f) applied to Lewis because he allows for multiple counterparts in the actual world. It does not apply to us because we do not allow for multiple rolemates at any time in the actual world. Our theory is like Lewis's, however, in that it does allow for an object to have no rolemate at all in the actual world. So it behooves us to consider those objections of Fara's and Williamson's that apply to Lewis because of that aspect of his theory.

They argue (9ff) that the only plausible ways of extending Lewis's theory to accommodate a language with an actuality operator yield the objectionable satisfiability *at the actual world* of the following:

- (a) $Fx \wedge \neg\mathbf{A}Fx$.

²⁴ We take this to mean that if x exists then x is y 's son.

²⁵ In this work, they include extensive criticism of efforts by Forbes (1982, 1985, 1990) and Ramachandran (1989, 1990a,b) to *revise* Lewis's 'translation scheme' while also accommodating an actuality operator.

This is satisfiable at the actual world on our semantics as well, but not in any objectionable way since its negation is weakly valid given a certain plausible condition (2d, below) on contexts, which is all that should be required of it.

More generally, the left-hand formulas below are not weakly valid. (Their converses, of course, are.)

- (b) $\not\models_{w_c t_c} Fx \rightarrow \mathbf{A}Fx$; (b') $\models_{w_c t_c} \mathbf{A}Fx \rightarrow Fx$;
 (c) $\not\models_{w_c t_c} Fx \rightarrow \mathbf{N}Fx$; (c') $\models_{w_c t_c} \mathbf{N}Fx \rightarrow Fx$.

The conditionals (2b) and (2c) are weakly valid, however, when we restrict our attention to contexts meeting a certain condition corresponding to one that would undoubtedly be met in any ordinary conversational context:

- (d) Whenever a variable x occurs only as an argument of the predicate F (in the sentence being evaluated), x is assigned a relative-sameness relation R_F which is such that if $d \in V_{wt}(F)$, then $\langle d, w, t \rangle R_F \langle d, w, t \rangle, \forall d \in D, \forall w \in W, \forall t \in T$.²⁶

This says (roughly) that any d that is F -ish is the same F as itself.

The condition needs to be generalized in order to account for more complex cases, but as stated, it has the following effect on the evaluation of (2b) and (2c): assuming that the antecedent Fx is true, at $w_c t_c$, on an assignment of an arbitrary $d \in D$ to x , that is, that $d \in V_{w_c t_c}(F)$, the constraint ensures that d has an x -rolemate at $w_c t_c$ (given the way it is at $w_c t_c$), namely, d itself, and hence that d has an x -rolemate at $w_c t_c$ that is in the extension of F at $w_c t_c$, in which case the consequents of (2b) and (2c) are both true on that assignment of d to x .

To focus on an intuitive example, suppose that F is the predicate *woman*. Then (2d) requires that a relation R_{woman} that is entailed by the *same-woman-as* relation be assigned to the variable x by the context. One appropriate relation would be the *same-person-as* relation. So any entity that is (actually now) a woman would be construed as (actually now) being the same person as some entity, namely, herself. That is enough to secure for any woman a *person*-mate at the actual world and current time, which suffices on the present view for the truth of the claim, of any woman, that she's actually now a woman.

Fara and Williamson strengthen the objection by considering the following sentence:

- (b) $\diamond \exists x (\mathbf{A}Fx \leftrightarrow \mathbf{A}\neg Fx)$.

The sentence turns out to be satisfiable for Lewis, on the most natural ways of extending his theory to account for an actuality operator because he allows for an absence of actual-world counterparts (and also because he allows for a multiplicity of them).

Our theory bans multiplicity of actual-world counterparts but does allow for absence. (This is so not just for the actual world but also for every possible world.) Nonetheless, for us, the sentence is strongly unsatisfiable. This results primarily

²⁶ We would not want to strengthen (2d) to a biconditional since when F is interpreted as, for example, *woman*, then a satisfactory R_F would be the relation *is the same person as*. But it need not be that everyone who's the same *person* as themselves be a woman, since men are people too.

from our Russellian-type interpretation for the actuality operator, combined with our classical semantics for negation. For us,

$$(f) \models \neg \diamond \exists x (\mathbf{A}Fx \leftrightarrow \mathbf{A}\neg Fx).$$

3. In the same work, Fara & Williamson include a criticism of Kripke variable-domain semantics (Kripke, 1963). They note that the sentence

(a) It might have been that everyone who is in fact rich was poor,

is naturally formalized with an actuality operator as

$$(b) \diamond \forall x (\mathbf{A}Rx \rightarrow Px),$$

but that this formalization is not correct within Kripke's variable-domain semantics, since it will be true in any model containing a world whose domain is disjoint from that of the actual world.

This is not so on our semantics. Although our semantics does not validate the Converse Barcan Schema (see (11a) below), it does not have domains that vary from world to world, and for that reason does deem as correct the above formalization of sentence (3a). Our truth condition for the formalization is this: *In some possible world, everything in the domain with a rich rolemate in the actual world is poor.*

4. As desired, necessity and permanence of identity and distinctness are not validated (4a–d), even in their existence-qualified versions (4e–h):

$$(a) \not\models \forall x \forall y (x = y \rightarrow \Box x = y);$$

$$(b) \not\models \forall x \forall y (x = y \rightarrow \mathbf{G}x = y);$$

$$(c) \not\models \forall x \forall y (x \neq y \rightarrow \Box x \neq y);$$

$$(d) \not\models \forall x \forall y (x \neq y \rightarrow \mathbf{G}x \neq y).$$

The reason for this is that something (or two things) may have a single rolemate at one world–time pair, but two rolemates at some other world–time pair, *as long as two different rolemate relations are involved*. More precisely, it may be that $g(x)$'s x -rolemate is identical to $g(y)$'s y -rolemate at some point though not at another, whether $g(x)$ and $g(y)$ are identical or not. This woman and that hunk of matter may now be the same person, but they might not have been, since that hunk of matter might have constituted (hence been) a different person.

And this can be so even when $g(x)$ and $g(y)$ *have*, respectively, an x -rolemate and a y -rolemate:

$$(e) \not\models \forall x \forall y (x = y \rightarrow \Box (Ex \wedge Ey \rightarrow x = y));$$

$$(f) \not\models \forall x \forall y (x = y \rightarrow \mathbf{G}((Ex \wedge Ey) \rightarrow x = y));$$

$$(g) \not\models \forall x \forall y (x \neq y \rightarrow \Box ((Ex \wedge Ey) \rightarrow x \neq y));$$

$$(h) \not\models \forall x \forall y (x \neq y \rightarrow \neg \mathbf{G}((Ex \wedge Ey) \rightarrow x \neq y)).$$

5. The necessity and permanence of identity will, however, hold for things of the same sort. There is a familiar intuitive example: if x is a planet that appears to be the brightest star in the evening sky and y is a planet that appears to be the brightest star in the morning sky, then since $x = y$, it couldn't but be that x is identical to y . Since the stage-setting here was enough to associate with both variables the same relative-sameness relation (*being the same planet as*), the claims that they are necessarily and permanently identical are both true. In general, (4a–d) and (4e–h)

will be valid when we restrict our attention to contexts that assign to x and y the same relative-sameness relation. This demonstrates that the question whether an identity is necessary on the present view is still separable from the question whether it is knowable *a priori*, in concord with Kripke's (1972) view.

6. Reflexivity, symmetry, and transitivity of identity are all validated:²⁷

- (a) $\models x = x$;
- (b) $\models x = y \rightarrow y = x$;
- (c) $\models (x = y \wedge y = z) \rightarrow x = z$.

7. Universal generalization and the rule of necessitation are both validity preserving. Therefore, the universal closures and necessitations of the identity laws in 6(a–c) are also valid.

8. Leibniz's law, construed as the principle of substitutivity of identicals, is not in general valid, for example:

- (a) $\not\models x = y \rightarrow (\Box(Ex \rightarrow Fx) \rightarrow \Box(Ey \rightarrow Fy))$.

The invalidity of this instance of substitutivity of identicals is desired given our philosophical underpinning, since a person (e.g.) is essentially a person (since nothing could be the same person as her without being a person), while her constituting matter is not (since something could be the same hunk of matter as it without being a person), although the person and the matter are identical. Whether a modal or temporal predication applies to an object depends on how we are naming or describing it or conceiving of it (as a person, as a human body, as a portion of matter, etc.). I follow Gibbard (1975, p. 201) in saying that for this reason we do not attribute a *property* to an object with such 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*.

The principles of substitutivity of identicals is, however, valid for the special case of predications not involving modal or temporal operators.

- (b) $\models x = y \rightarrow (\Phi \rightarrow \Phi^y/x)$,
(whenever x is not free in Φ in the scope of a modal or temporal operator).

9. The claim that everything exists necessarily, which is valid on Lewis' counterpart theory, is not valid for us, strongly or weakly. For us:

- $\not\models \forall x \Box Ex$.

Its truth condition on our theory can be glossed as follows: everything has a rolemate in every possible world,²⁸ whereas for Lewis its truth condition is glossed, everything is such that all of its counterparts exist. Our truth condition will not in general hold since for most relative-sameness relations R that could be assigned to the variable x , something will lack an R -based rolemate in some world. For example,

²⁷ Only the universal closure of reflexivity would have been validated had we chosen to require that $x = y$ be true only in cases where the assignment function assigns to them values in D . Symmetry and transitivity in any form (open or closed, necessitated or not) would still have been validated.

²⁸ Cf. example I3 above.

there are possible worlds in which there's nothing that is the same person as George W. Bush; there might have been no such person as George W. Bush.

The formula *is* true on our current version of the theory, however, when the relative-sameness relation contextually assigned to x is one that covers the entire domain. Intuitively, this amounts to saying—given weak reflexivity—that any object is the same F as itself at any world and time, where *being the same F as* is the relative-sameness relation contextually assigned to x . A plausible candidate is the relation *is the same entity as* (given that only something that is F can be the same F as itself). This feature of the theory is independent of the satisfiability of some statements of contingent identity. It results not from the fact that our theory is a counterpart theory, but rather from the incorporation of invariant domains into the models.

10. Despite the fact of invariant domains, the semantics does not validate the Barcan Schema, $\forall x \Box \Phi \rightarrow \Box \forall x \Phi$. (See examples (I6) and (I7), above.) Here is a countermodel for $\Phi = Fx$:

Let \mathfrak{M} , c , g be such that $D = \mathbb{N}$, and for some point wt and every point $w't' \neq wt$, $f_c^x(n_{wt}, w', t') = 2n$ for all n , and $V_{wt}(F) = \mathbb{N}$ but $V_{w't'}(F) = \{2n : n \in \mathbb{N}\}$. Then,

$$(a) \not\models_{wt}^{\mathfrak{M}cg} \forall x \Box Fx \rightarrow \Box \forall x Fx.$$

The crucial feature of the model is that the image of the domain under the x -rolemate function is, at at least one point, a proper subset of the domain. In fact, any countermodel to this simple instance of the Barcan Schema must also have this property, from which it follows, given other features of our semantics, that any such countermodel must have an infinite domain.²⁹ The reasoning is this. If $\forall x \Box Fx$ is true at a world (ignoring times), then every object in the domain must have an x -rolemate in the domain at every world. The rolemate must be *in the domain* (as opposed to in $D \cup \{D\}$) since these rolemates must be in the extension of F at each world. (D itself is never in the extension of any predicate.) This in turn requires, given the weak equivalence of each relative-sameness relation, that the image of the rolemate function be the same size as the domain in each world.³⁰ So if the domain is to be a proper superset of its equinumerous image under a function, it must be infinite.

11. The Converse Barcan Schema is also not validated, which it is on Lewis's theory.

For example:

$$(a) \not\models \Box \forall x Fx \rightarrow \forall x \Box Fx.$$

The reason in this case is exactly what one should expect within our rolemate version of counterpart theory, namely, that $g(x)$ may lack an x -rolemate at any point. Necessity on Lewis's semantics for the box involves quantification only over those possible worlds at which the object in question *has* a counterpart, while ours involves quantification over every possible world. Thus, on our semantics the

²⁹ This is not, however, always true of more complex instances, e.g., $\forall x \Box \neg Fx \rightarrow \Box \forall x \neg Fx$.

³⁰ Reminder: relative-sameness relations are one-one on the domain (D(3b)) but need not cover the domain; rolemate functions are total on the domain but take all the entities not covered by their corresponding relative-sameness relation to the same 'non-entity', namely, the domain itself.

antecedent here will be true while the consequent is false in the following sort of situation: every object in the domain is in the extension of F at every point while some object lacks an x -rolemate at some point. Here is an intuitively true counterscenario: It could only be that everything exists, yet it is not the case that everything could only have been the same *stone* as something that exists.

Note that here, we have violated constraint J(2d) on contexts. That constraint restriction does guarantee the validity of instance (11a) of the Converse Barcan Schema.

12. Kripke (1963, pp. 88–89) showed us that Prior’s (1956) proof of the Converse Barcan Schema makes essential use of (12a) below, a version of universal instantiation that is not valid on the present semantics. Its universal generalization, (12b), is valid, but only given a certain condition—that x does not occur free in the scope of a modal or temporal operator in Φ . This qualification, and the corresponding one of substitutivity of identicals (8b), constitute the main departures of the logic from a more standard S4-system with quantification and identity. These qualifications are needed: the relettering of free variables within the scope of modal or temporal operators is not truth-preserving since different variables may be associated by a context with different relative-sameness relations, hence with different rolemate functions.

(a) $\not\models \forall x\Phi \rightarrow \Phi^y/x$, but

(b) $\models \forall y(\forall x\Phi \rightarrow \Phi^y/x)$

(whenever x is not free in Φ in the scope of a modal or temporal operator).

The truth condition for (12a)—instantiating Fx for Φ —can be glossed as follows: if the extension of F includes the domain, then it contains $g(y)$. (See example I5 above.) This may fail since $g(y)$ may not be in the domain. It may be non-existent in the sense that it satisfies no formulas except those that deny it positive attributes or assert its self-identity.

The truth condition for (12b)—again instantiating Fx for Φ —is, in contrast, a tautology: if the extension of F includes the domain, then it includes the domain. (See example I4 above.)

The fact that $\forall x\Box Fx \rightarrow \forall y\Box Fy$ is not valid should be no more troubling *per se*, given that the model theory treats all modal and temporal operators as context-dependent, than it would if it were rendered invalid by a theory that modeled the contextual variability of quantifier domains by allowing $\forall x$ and $\forall y$ to be contextually assigned different domains. Of course, this is not to insist that it is unreasonable to be antecedently troubled by the treatment of modal and temporal operators as context-dependent in the way allowed for here.

13. One interesting fact about the present semantics is that some ‘negative’ instances of the Converse Barcan Schema *are* valid, for example:

(a) $\models \Box\forall x\neg Fx \rightarrow \forall x\Box\neg Fx$.

This is first-order equivalent to the following ‘positive’ instance of the existential version of the Converse Barcan Schema given one meager modal principle (duality):

(b) $\models \exists x\Diamond Fx \rightarrow \Diamond\exists x Fx$.

This is valid on the proposed semantics since whenever something has an x -rolemate in some world that is in the extension of F in that world, it must be that F has

some object from the domain—namely, the x -rolemate in question—in its extension in that world. The crucial feature of the semantics that makes for the difference between the positive atomic instance (11a) and the negated atomic instance (13a/b) is that satisfaction in a world of a modalized atomic formula Fx requires an object to meet a conjunctive condition: that it have a rolemate in that world and that its rolemate be in the extension of F in that world. Correlatively, satisfaction in a world of the modalized negation of that atomic sentence requires an object to meet only a disjunctive condition: either that it lack a rolemate in that world or that it have a rolemate that is not in the extension of F in that world.

Although the Converse Barcan *Schema's* universal version is equivalent to its existential version,³¹ corresponding instances of them need not be.³² This should not be surprising, as there is often a felt difference between corresponding instances of the two. For example, $\Box\forall xFx \rightarrow \forall x\Box Fx$ seems intuitively invalid—straightaway, *however F may be interpreted*—to those who believe that actual things exist only contingently. One merely need to note that it may be that in every world, everything is F , even though there are actual things that are not F in every possible world, since counterfactual worlds need not include among their population *all* of the actual things.

The corresponding $\exists x\Diamond Fx \rightarrow \Diamond\exists xFx$, in contrast, does not seem intuitively invalid *straightaway*, however F may be interpreted. It seems intuitively *valid*—even to those for whom contingent existence is a philosophical given. The reasoning is this: if there is in fact something that could have been F , then there is a possible world in which it *is* F ; and of course then in that possible world *something* is F (namely, the thing that as a matter of fact could have been F). Only once special, non-atomic predications (such as non-existence) are substituted for Φ in the existential version of the Converse Barcan Schema does it no longer seem valid to those who believe in contingent existence. And, in fact we cannot interpret the *atomic* F in (13b) as nonexistence to yield a false interpretation for (13b), since on that interpretation, F would necessarily have an empty extension, rendering the antecedent of (13b) *false!* I see it as an advantage of the present view that it brings out this felt difference between corresponding instances of the universal and existential versions of the Converse Barcan Schema.³³

14. The Qualified Converse Barcan Schema is valid, as it should be.³⁴

$$\models \Box\forall x\Phi \rightarrow \forall x\Box(Ex \rightarrow \Phi).$$

15. The S5 principle $\Diamond\Phi \rightarrow \Box\Diamond\Phi$ is not valid. It will be instructive to compute the truth condition of a simple instance, then explain intuitively why it can fail to be

³¹ Where two schemas are equivalent just in case any instance of one is a logical consequence of some instances of the other.

³² Two instances *correspond* in the intended sense when Φ gets substituted in each schema with the same formula. A qualification is called for here: corresponding instances *will* be equivalent when the quantifiers in them do not bind a variable.

³³ It should be pointed out, however, that our semantics does validate the counterintuitive instance of the diamond version of the Converse Barcan Schema: $\exists x\Diamond\neg Ex \rightarrow \Diamond\exists x\neg Ex$. This is due to the simplifying assumption of invariant domains, which was independent of the contingent-identity motivation and which could be dropped or not according to taste.

³⁴ Cf. Stalnaker's (2003a, p. 151) discussion.

true. Again we drop relativization to times to simplify exposition. The function symbol f here abbreviates the x -rolemate function f_c^x . (See Def f_c^x , p. 171.)

$$\models_w^{\mathfrak{M}cg} \diamond Fx \rightarrow \square \diamond Fx$$

$$\text{iff } \models_w^{\mathfrak{M}cg} \diamond Fx \Rightarrow \models_w^{\mathfrak{M}cg} \square \diamond Fx \quad (\text{G3, G4})$$

$$\text{iff } \exists w' f(g(x)_w, w') \in V_{w'}(F) \Rightarrow \forall w' \models_{w'}^{\mathfrak{M}cg[f(g(x)_w, w')/x]} \diamond Fx \quad (\text{G3, G6, G2})$$

$$\text{iff } \exists w' f(g(x)_w, w') \in V_{w'}(F) \Rightarrow \\ \Rightarrow \forall w' \exists v f(f(g(x)_w, w')_{w'}, v) \in V(F) \quad (\text{G3, G6, G2})$$

(Gloss: There's some or other world at which $g(x)$ has an F -ish rolemate only if at *every* world, $g(x)$ has a rolemate that has an F -ish rolemate at some or other world.)

This may fail to be true, intuitively, since an entity may have a rolemate at some world without having a rolemate at *every* world.

16. An instance of $\diamond \mathbf{A}\Phi \rightarrow \mathbf{A}\Phi$ may fail for a similar reason. $\diamond \mathbf{A}\neg Fx \rightarrow \mathbf{A}\neg Fx$ is true (roughly) just in case there is at least one possible world at which x lacks a rolemate with an F -ish rolemate in the actual world *only if* x itself has no F -ish rolemate in the actual world.

This condition may fail since x may have a rolemate at the actual world without having a rolemate in *every* possible world. The following serves as a falsifying scenario. We set Φx to be $\neg x = \text{me}$. I could have been the same person that I in fact am even if Gore had been declared winner of the United States 2000 presidential election; but it is not the case that I could have been the same person that I in fact am no matter what had happened, for example, even if tens of thousands of years ago a super-volcano had wiped out all human life. But of course, either way, I *am* the same person that I in fact am.

17. When defining an assignment function g (Section E, page 172), we allowed $g(v_j)$ to take D as a value while requiring that $g(v_i) \in D$ for at least one variable v_i . Given our understanding of D as the nonexistent, we may put this by saying that some, though not all, variables may have the nonexistent as their value. For us, to be is not to be the value of a variable; rather, it is to be in the range of the quantifiers.

Different choices might have been made at that point. The two salient alternatives are (i) to have allowed g to take values in $D \cup \{D\}$ with no further requirement that it take some value in D and (ii) to have allowed g to take values only in D . Our middle-ground choice can be given a purely logical justification.

Had we chosen option (i), to allow that no variable take a value in D , our logic would have been a free logic in the sense that it would invalidate the inference from universal to existential generalization and deem invalid the claim that something exists. That is, with option (i) we have the following problems: $\forall x \Phi \not\models \exists x \Phi$ and $\not\models \exists x Ex$.

Had we instead chosen option (ii), to require that no variable take a value outside of D , our logic would no longer have been a normal modal logic since the rule of necessitation would no longer be validity preserving. In particular, we would have $\models \forall x Fx \rightarrow Fy$, but $\not\models \square(\forall x Fx \rightarrow Fy)$. The reason for this departure would be that variables free inside the scope of a modal operator would be able to take a value outside of D in the sense that in any possible world, the value of any variable might lack a rolemate and thus be assigned something outside of D there by the rolemate function.

K. Conclusion. Our relative-sameness (context-dependent) version of counterpart theory is, as I see it, pretty much a straightforward model-theoretic development of the sortal-relative context-dependent counterpart theories presented by Lewis in his first (1971) revision of his original counterpart theory and by Gibbard (1975) in his ‘brief for contingent identity.’³⁵ It shares what to my mind is the main benefit of these theories, namely, the satisfiability of some statements of contingent identity, so as to preserve the principle of one material object to a place (my true underlying motive), while shedding many of their disadvantages. In particular, while allowing for the satisfiability of some statements of contingent and temporary identity, we have blocked the remaining ones of Hazen’s objections and all the objections of Fara & Williamson (2005), while invalidating nearly all instances of the Barcan and Converse Barcan Schemas (see remarks (10–14) above), yet while achieving a logic that validates most standard theorems and rules of quantified S4 with identity and an actuality operator,³⁶ though with some exceptions deriving from the fact that relettering within the scope of intensional operators is not truth preserving. Our way of incorporating contextual variation as associated with variable variation renders these exceptions expected and completely appropriate.³⁷ We have in particular:

1. Universal instantiation:

- (a) (Partially) closed: $\models \forall y(\forall x\Phi \rightarrow \Phi^y/x)$,
- (b) Strict: $\models \forall x\Phi \rightarrow \Phi^y/x$,

where x does not occur free in Φ in the scope of a modal or temporal operator;

2. Existence of actual world: $\models \diamond(\diamond\mathbf{A}\Phi \rightarrow \Phi)$;

3. Actuality commutes with truth functions:

- (a) $\models \mathbf{A}\neg\Phi \leftrightarrow \neg\mathbf{A}\Phi$,
- (b) $\models \mathbf{A}(\Phi \wedge \Psi) \leftrightarrow (\mathbf{A}\Phi \wedge \mathbf{A}\Psi)$;

4. Identity:

- (a) $\models x = x$,
- (b) $\models x = y \rightarrow (\Phi \rightarrow \Phi^y/x)$,

where x is not free in the scope of a modal or temporal operator in Φ ;

5. UG: $\models \Phi \Rightarrow \models \forall x\Phi$;

6. RN: $\models \Phi \Rightarrow \models \Box\Phi$;

³⁵ Ramachandran (1998) has also developed a version of sortal-relative counterpart theory. The possibility of developing a relative-sameness counterpart theory is discussed briefly by Sider (1999).

³⁶ See Hodes (1984) for axioms for actuality combined with quantified modal logic and identity.

³⁷ Stronger and perhaps more purely logical and standard systems could be derived from the present one by restricting the context in ways further than that already suggested. For example, we could drop the restrictions on universal instantiation and substitutivity of identicals if we required that the context assign the same relative-sameness relation to every variable. This would, undesirably, bring back the necessity and permanence of identity and distinctness. We could further restrict the contexts by requiring that they assign to each variable what we would probably call the *same entity as* relation—that relative-sameness relation R on $D \times W \times T$ which is an equivalence relation, not merely a weak equivalence relation. This would bring back not only all the necessity and permanence of identity and distinctness but also—given our use of invariant domains—all instances of the Barcan Schema and its converse.

7. K: $\models \Box(\Phi \rightarrow \Psi) \rightarrow (\Box\Phi \rightarrow \Box\Psi)$;
8. T: $\models \Box\Phi \rightarrow \Phi$; and
9. 4: $\models \Box\Phi \rightarrow \Box\Box\Phi$ (not valid on Lewis' theory).

Clearly, the main source of these advantageous departures from Lewis' counterpart theory derives primarily from our philosophically well-motivated requirement that objects have at most one rolemate at any single time in any given world and that the rolemate relation be transitive, symmetric, and weakly reflexive. I do not argue for this on a case by case basis, for each of the different natural rolemate relations that might come into play in any context in any model, but base it instead on the general logical properties of *sameness*. The idea that sameness, in *any* respect, could fail to be transitive or Euclidean has always seemed to me bizarre. Surprisingly, the rejection of transitivity for this or that sameness relation has not been an entirely unpopular view among philosophers, even among those who, unlike me, think that sameness relativized to a sort *requires* identity.

The models offered here allow for some but not all haecceitistic differences. For example, they allow *a* (at *wt*) to be the same person as *c* rather than *d* (at a different point *w't'*) while *b* (at *wt*) is the same person as *d* rather than *c* (at the point *w't'*), even though the purely qualitative properties that *a* has at *wt* are exactly the ones that *d* has at *w't'*, and likewise for *b* (at *wt*) and *d* (at *w't'*). There is no violation here of weak equivalence.

The models do not, however, allow for another classic case of haecceitism. While they do allow for the possibility that you could have had a twin, even though you do not in fact have one, they arguably do not allow for the distinctness of the possibilities that you could have been *either* of the two twins. Formally, this would require person *d* (at *wt*) to be the same person as twin-*a* at *w't*, *not* the same person as twin-*b* at *w't*, but the same person as twin-*b* but not twin-*a* at *w''t*.

Given symmetry and transitivity, the situation requires that 'twin-*a*' at *w't* be the same person as 'twin-*b*' at *w''t*, and similar reasons apply to '*b*' at *w't* and '*a*' at *w''t*. There is no violation here of weak equivalence. But it cannot be nonetheless. The situation involves a violation of any reasonable constraint on transworld personal identity. Twins *a* and *b* could exchange their qualities, or even their matter, but could not exchange their *personal identities*, even if they do exchange their names. In this model as described, it is misleading to use a different designation for the twin you're the same person as at *w't* from the one used for the twin you're the same person as at *w''t*. You're in the same situation with respect to your twindom at *w't* as the one you're in at *w''t*. You might have had a twin, but there is no difference between the possibility of your being the one and that of your being the other.

This apparently anti-haecceitist claim is therefore independent of other haecceitist theses one might affirm, such as the possibility of there being a possible world in which you and your sister have switched all your purely qualitative properties—where you're *you*, but otherwise just like her, and vice versa. The possible twins could exchange all of *their* qualitative properties, including their names. But there is no distinction between the possibility of your being the one twin and that of your being the other. I have ruled this out not because I wanted to develop a rolemate theory that avoids some of the worst pitfalls of Lewis' theory just by ruling out by fiat the possibility of multiple intraworld rolemates. Rather I have ruled it out because it emerged as a natural consequence of taking sort-relative sameness relations to be the ones relevant for *de-re* modal and temporal predications, while conservatively sticking to the position that *sameness* in any respect is both symmetric and transitive, while proposing, due to my desire to preserve the principle of one object to a

place, that *sameness* no more requires identity when it is relativized to a sort than it does when it is relativized to a quality.

L. Acknowledgments. I would like to dedicate this paper to my PhD advisor, Robert Stalnaker, having been unable (to my great regret) to produce a paper in time for his festschrift (Thomson & Byrne, 2006), despite the editors' generous deadline extensions. Had the timing been different on either side, I would have liked to have contributed some version of this paper. I did not realize just how profound Bob's influence on me had been until I came to have these views—similar, as I see it, to ones he has proposed, though not *quite* endorsed—even though I would not too long ago have thought the idea of contingent identity crazy, if not reprehensible. It is a real pleasure to study his work on this topic now that I am so much more sympathetic to, and perhaps equally tentative about, the sort of philosophical picture that underlies the logic presented here. Thanks additionally to the audience at the Banff International Research Station 'Mathematical Methods in Philosophy' conference where these ideas were informally presented, especially to the organizers of that enjoyable event—Aldo Antonelli, Alasdair Urquhart, and Richard Zach—and also for discussion to Harold Hodes, Agustín Rayo, Theodore Sider, and, as always, Michael Fara and Timothy Williamson. I am also indebted to Allan Gibbard's 'Lump1 and Goliath' paper (Gibbard, 1975), which (for better or worse) first got me to see that there could be any merit to the idea of contingent identity and which also already contains whatever good ideas there may be in this paper, and to Cian Dorr, whose criticisms resulted in significant improvement.

BIBLIOGRAPHY

- Bennett, K. (2004). Spatio-temporal coincidence and the grounding problem. *Philosophical Studies*, **118**(3), 339–371.
- Fara, D. G. (2001). Phenomenal continua and the sorites. *Mind*, **110**(440), 905–935. Published under the name 'Delia Graff'.
- Fara, M., & Williamson, T. (2005). Counterparts and actuality. *Mind*, **114**, 1–30.
- Fine, K. (2003). The non-identity of a material thing and its matter. *Mind*, **114**, 195–234.
- Forbes, G. (1982). Canonical counterpart theory. *Analysis*, **42**(1), 33–37.
- Forbes, G. (1985). *The Metaphysics of Modality*. Oxford: Oxford University Press.
- Forbes, G. (1990). Counterparts, logic and metaphysics: reply to Ramachandran. *Analysis*, **50**, 167–173.
- Gibbard, A. (1975). Contingent identity. *Journal of Philosophical Logic*, **4**, 187–221.
- Hazen, A. (1979). Counterpart-theoretic semantics for modal logic. *Journal of Philosophy*, **76**(6), 319–338.
- Hodes, H. (1984). Axioms for actuality. *Journal of Philosophical Logic*, **13**(1), 27–34.
- Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, **16**, 83–94.
- Kripke, S. (1972). Naming and necessity. In Davidson, D., & Harman, G., editors, *Semantics of Natural Language*. Dordrecht: Reidel, pp. 253–355. Reprinted as Kripke [1980].
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, D. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy*, **65**(5), 113–126.
- Lewis, D. (1971). Counterparts of persons and their bodies. *Journal of Philosophy*, **68**(7), 203–211.

- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- Prior, A. N. (1956). Modality and quantification in S5. *Journal of Symbolic Logic*, **21**(1), 60–62.
- Ramachandran, M. (1989). An alternative translation scheme for counterpart theory. *Analysis*, **49**, 131–141.
- Ramachandran, M. (1990a). Contingent identity in counterpart theory. *Analysis*, **50**, 163–166.
- Ramachandran, M. (1990b). Unsuccessful revisions of CCT. *Analysis*, **50**, 173–177.
- Ramachandran, M. (1998). Sortal modal logic and counterpart theory. *Australasian Journal of Philosophy*, **76**, 553–565.
- Sider, T. (1996). All the World's a stage. *Australasian Journal of Philosophy*, **74**(3), 433–453.
- Sider, T. (1999). Critical study of Michael Jubien's *Ontology, Modality and the Fallacy of Reference*. *NOÛS*, **33**, 284–294.
- Stalnaker, R. (1986). Counterparts and identity. *Midwest Studies in Philosophy*, **11**, 121–140. Page references are to reprinted version Stalnaker [2003b].
- Stalnaker, R. (2003a). The interaction of modality with quantification and identity. In Sinnott-Armstrong, W., Raffman, D., & Asher, N., editors, *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press. pp. 12–28.
- Stalnaker, R. (2003b). *Ways a World Might Be*. Oxford: Oxford University Press.
- Thomson, J. & Byrne, A., editors. (2006). *Content and Modality: Themes from the Philosophy of Robert Stalnaker*. Oxford: Oxford University Press.
- Williamson, T. (1998). Bare possibilia. *Erkenntnis*, **48**, 257–273.
- Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, **65**, 261–281. English translation in Jean van Heijenoort, editor. (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, pp. 199–215.

PHILOSOPHY DEPT
212 1879 HALL
PRINCETON UNIVERSITY
PRINCETON, NJ 08544
USA
Email: graff@princeton.edu