I probably won't revise or publish this paper; feel free to cite.

# Beyond the Humphrey Objection*

Theodore Sider Draft of 2006

Counterpart theory has come a long way since the seventies. Its virtues are now generally appreciated. It has been extended to temporal discourse.[1] And it is less often dismissed out of hand, now that Saul Kripke's scornful words are no longer regarded as the last on the subject. But new critics have appeared, equally formidable if less dismissive. Counterpart theorists, both modal and temporal, owe them answers.

## 1. Counterpart theory: the current state of play

David Lewis's (1968) modal counterpart theory identifies *possibly being F* with having a counterpart—an appropriately similar object in another possible world—that is $F$. Kripke's complaint in *Naming and Necessity* was that while Hubert Humphrey cares very much that he might have won the 1968 U.S. presidential election, he "could not care less whether someone *else*, no matter how much resembling him, would have been victorious in another possible world." (1972, p. 45)

While certainly worthy of serious discussion, the argument now feels less compelling than it must have seemed then, at the height of the shock and awe immediately after *Naming and Necessity*. The argument can be taken various ways, each answerable. i) "Counterpart theory does not allow Humphrey himself to have the modal property of *possibly winning the election*, since only the counterpart wins." Reply: according to counterpart theory, the property of possibly winning *is* the property of having a counterpart who wins. Humphrey has a counterpart who wins, and so Humphrey himself (pound, stamp!) might have won.[2] ii) "If you ask Humphrey whether he cares that he might have won the election, he will say yes. If you ask him whether he cares that he has a counterpart who wins, he will say no. So, Humphrey takes different attitudes toward the properties *possibly winning* and *having a counterpart that wins*. So those are distinct properties." iii) "Look, it is just *obvious* that possibly winning is

*Thanks to Karen Bennett, David Chalmers, Andy Egan, Karson Kovakovich, Trenton Merricks, Sarah Moss, Laurie Paul, Josh Parsons, Adam Sennet, and especially John Hawthorne for helpful comments.

[1] See Hawley (2001); Sider (2001*a*, 1996*a*, chapter 5, section 8), and section 3 below.

[2] See Hazen (1979, pp. 321–322), Lewis (1986, p. 196), and Mondadori (1983).

not the same as having a counterpart who wins." Reply to ii) and iii): this is just the paradox of analysis. A reasonable person can care about a property under one description ("possibly winning") while not caring about the same property under another description ("having a counterpart who wins"), provided it is not obvious that the descriptions pick out the same property. Correct analyses need not be obvious to competent language users. Obviousness may count for something, but theoretical virtues are important as well in determining which analyses we ought to accept.[3]

And counterpart theory is indeed theoretically virtuous; that is the best argument for it. i) Unlike identity, the counterpart relation need not be an equivalence relation. This flexibility is welcome when dealing with various modal paradoxes (Lewis, 1968). ii) Bare-bones counterpart theory may be augmented by the claim that different counterpart relations, stressing different dimensions of similarity, count in different contexts. This context-sensitivity of *de re* modal predication matches our shifting *de re* modal intuitions, and also avoids certain other modal paradoxes (Lewis 1971; 1986, section 4.5). iii) Counterpart theory is consistent with a *qualitative metaphysics of modality*, according to which modality *de dicto* is more basic than modality *de re*—the most fundamental modal facts are purely qualitative, descriptive, general. For counterpart theory reduces *de re* modality to similarity and "intra-world" talk of possibilia (i.e., talk of possibilia and their features that is silent on the sorts of inter-world relations that would ground *de re* modality. Such talk may be taken in Lewisian fashion at face value, or reduced in the manner of the next section of this paper.) Anyone full of the reductionist spirit should welcome this feature of counterpart theory. iv) Counterpart theory can be applied to *de re* temporal predication, yielding gains parallel to i)–iii), plus others, in the philosophy of persistence and time.

While these theoretical virtues make counterpart theory attractive, a new wave of critics has arrived. The combination of modal counterpart theory with reductionism about possible worlds and objects has been challenged. Temporal counterpart theory faces special obstacles. And both temporal and modal counterpart theory face purely logical objections. The following sections present and then answer these objections.

[3] See Forbes (1987, p. 143), Hazen (1979, pp. 320–324), Sider (2001*a*, pp. 196–196).

## 2. Ersatz counterparts

Part of the seventies' recoil from counterpart theory may have been due to a failure to distinguish counterpart theory from other, independent, views of its most distinguished advocate. In particular, Lewis's (1986) infamous modal realism—his anti-reductionism about possibilia, which accords the same ontological status to possible talking donkeys and golden mountains as to actual donkeys and mountains—is not obligatory for counterpart theorists. Prima facie, one can combine counterpart theory with any of the strategies for reducing possibilia available on the current market, provided the strategy applies to quantification over possible individuals inhabiting possible worlds in addition to quantification over the worlds themselves.[4]

We nonLewisians do not have exactly Lewis's motive for accepting counterpart theory. Lewis regarded himself as being *forced* by his modal realism to accept counterpart theory. For a modal realist, the alternative to counterpart theory is a problematic sort of transworld identity: the presence of one and the same object as a common part of multiple worlds. My right hand has five fingers, but it might have had six. Given transworld identity, my hand itself must somehow have five fingers with respect to the actual world while having six fingers with respect to some other world. Given Lewis's insistence that having five fingers is a monadic property, this is hard to make sense of in a modal realist framework.[5] But, as many have observed, if possible worlds are conceived of as abstract entities rather than Lewisian concreta, "transworld identity" is unproblematic. A world with respect to which my hand has six fingers is simply a false story (or proposition, or whatever) about my hand. So far as the metaphysics of possibilia is concerned, transworld identity is unproblematic and counterpart theory is optional. NonLewisians nevertheless drawn to counterpart theory are drawn instead by its theoretical benefits: logical flexibility, solution to modal paradoxes, and so on.

The usual strategies for reducing talk of possibilia appeal to modal notions. Possible worlds, for instance, are defined as maximal consistent sets of sentences, or propositions, or states of affairs, where 'consistent' is taken modally. It would be nice if this notion of consistency could be reduced in some way, but that reduction is no easy task; the usual approach is to take modal consistency

[4]See Adams (1974); Heller (1998); Plantinga (1976); Rosen (1990); Stalnaker (1976). I defend my own reduction of possibilia in Sider (2002), but will couch my defense as a defense of a more familiar variety.

[5]See Lewis (1986, section 4.2).

as a primitive. Now, if this primitive notion of consistency applied to "singular" propositions—propositions about particular individuals—or if it were expressed by a sentential operator open to quantification-in, then counterpart theory would be superfluous; the primitive notion of consistency—and its dual, necessity—would already constitute *de re* modality. Thus: the counterpart theorist who wants to be a primitivist about necessity will take only a notion of *de dicto* necessity as primitive. This notion will apply only to purely general, or qualitative, propositions and sentences. The role of counterpart theory is then to reduce modality *de re*.

But combining counterpart theory with reductionism about possibilia is not as straightforward as it initially seems. One theme of a cluster of interrelated objections, put forward by Trenton Merricks (2003), is that no one reduction seems uniquely and intrinsically suitable. The reduction Lewis called "linguistic ersatzism" (1986, section 3.2) identifies possible worlds and individuals with linguistic entities: worlds with sets of closed sentences, individuals with sets of open sentences. (Lewis called these linguistic entities "ersatz" possible worlds and individuals.) Thus, a possible talking donkey would be identified with a set containing such members as '$x$ is a donkey' and '$x$ talks'. But this is not the only set one might reasonably produce. What of a set in which the free variable is $y$ rather than $x$? Or one containing French sentences? And there are multiple ways to construe the ontology of "sentences". Sentences can be regarded as set-theoretic sequences of their words, but there are different ways of constructing sequences from set theory, since there are multiple, equally good, ways to construct ordered pairs from unordered sets. One might try cutting down on the potential candidates by identifying the possible donkey with an equivalence class of candidates, but the equivalence class would be just another candidate alongside the rest.

In addition to being many in number and equally qualified, each candidate has an intrinsic shortcoming. The set described above does not seem particularly intrinsically suited to constitute *de re* modal facts. The linguistic ersatzer's definitions of "possible worlds" and "possible individuals" as these abstract entities feels more like a stipulation than a *discovery*. Sets of sentences are not born to be ersatz worlds and individuals. As Merricks puts it, "A set just sits there." (2003, p. 535).

Thus, we have two observations:[6]

[6]The observations are not unique to ersatz counterpart theory. Many ontological or logical constructions of non-fundamental entities lack uniqueness and overwhelming intrinsic merit.

O1: There are many equally suited candidate objects for the linguistic ersatzer's reduction

O2: These candidate objects are in some sense not particularly intrinsically suited for the role they are to play in the reduction.

I accept the observations, and distinguish two arguments based on them.[7]

The first focuses on O2. A critic might concede that counterpart theorists could introduce a new modal language with a stipulated ersatz counterpart-theoretic semantics. The critic might even grant, for the sake of argument, that this new language could serve many of the purposes served by ordinary English modal talk, perhaps better than any rival semantics. (Compare Carnap's

[7]These arguments are not precisely Merricks's, since my brand of linguistic ersatzism does not precisely match his intended targets. They are, I hope, the strongest Merricksian arguments that apply to me. Because of the differences between me and his targets, there are arguments in his paper that do not apply to me. For instance, he describes one of his targets thus: "first, *de re*…modality is reduced to what abstract structures represent. Second, abstract structures represent in virtue of what we do." (p. 530) He then objects that this view makes the nature of (p. 528), and the proper analysis of (p. 529), modal properties a function of what we do. These objections are effective against Merricks's target, but not against the view I want to defend. Let $C$ be any candidate counterpart-theoretic semantics (see p. 6 below), let $F$ be any one-place predicate, and let $P^F_C$ be the property *bearing the counterpart*$_C$ *relation to some possible individual*$_C$ *that is*$_C$ $F$. Pretend for a moment that there is a distinguished candidate semantics, $\mathscr{C}$. The analysis I accept is then the claim that the property *possibly being* $F$ just is the property $P^F_{\mathscr{C}}$. Whether something has this property has nothing to do with what we do, nor does the analysis of (i.e., identification with) *possibly being* $F$ as $P^F_{\mathscr{C}}$ have anything to do with what we do. My account does not invoke "representation" in the way that Merricks's target does. There is, of course, the relation of representation that holds between the *words* 'possibly being $F$' and the property $P^F_{\mathscr{C}}$, whose holding is indeed partly a matter of what we do, but that is irrelevant. There is also what I call "relative representation" (see p. 8 below): within any candidate semantics $C$, the notions of possible individuals, counterparthood, and property instantiation by possible individuals must be well-defined. But whether an entity (abstract object, linguistic item, or whatever) counts as a possible individual$_{\mathscr{C}}$, or as being$_{\mathscr{C}}$ $F$, or as a counterpart$_{\mathscr{C}}$ of another entity, has nothing to do with us; the definitions of these notions within a candidate semantics do not appeal to facts about us. (What if the entities are words, taken as entities that depend on us for their semantic properties? Then if $\mathscr{C}$'s definitions of the counterpart-theoretic notions involve semantic properties, we will get unwanted dependence on us. Answer: don't take "words" that way. Take words instead as abstract entities that have their semantic features intrinsically. Cf. Lewis (1962) on languages versus language.)

Now drop the pretense of a unique candidate semantics. The claim, then, is that there is no unique property of *possibly being* $F$. The expression 'possibly being $F$' is semantically indeterminate among many candidate meanings, each of which is analyzed as $\mathrm{P}^F_C$ for some candidate semantics $C$.

(1947, section 2) *explications*, which replace imperfect pre-existing concepts with better, more scientific, ones.) The critic would be careful to grant that the propositions statable in this new language were true before its introduction, and do not concern the counterpart-theorist's acts of stipulation, just as propositions about dinosaurs statable in English do not concern the conventions of English, and were true long before the introduction of the English language. But he would nevertheless complain that such an explication fails to yield an account of what *ordinary* speakers, all along, have meant by modal language. Ordinary speakers have made no stipulations like those just imagined. Ersatz counterparts are not particularly intrinsically suited to constitute modal facts, so how can the semantics of English be counterpart-theoretic if those in charge of its semantics—ordinary speakers—are so ignorant of counterparts?

Ersatz counterparts do not fit speakers' use of modal language because of stipulations. Rather, the fit is structural: the theoretical virtues cited in the previous section are facets of our use of modal language that fit the "shape" of counterpart-theoretic semantics. The context-sensitivity of our *de re* modal judgments matches the plurality of dimensions of similarity that ground counterpart relations, our patterns of modal inference match the intransitivity of the counterpart relation, and so on. The ability of counterpart-theoretic candidate semantic values to vindicate *de re* modal judgments and inferences constitutes their fit with use. Despite lacking the concept of a counterpart, ordinary speakers speak as if they are talking about counterparts; that is what makes it the case that they are in fact talking about counterparts.

This answer faces a challenge. A candidate counterpart-theoretic semantics, $C$, for a modal language $L$, consists of:

i) a choice of objects to count as the possible worlds and their inhabitants

ii) definitions, relative to the objects chosen in i), of the characteristic locutions needed to do counterpart-theoretic semantics for $L$ (e.g., 'it is true at world $w$ that snow is white', and 'object $x$ is a possible talking donkey')

iii) a choice of a relation or relations $R$ over the objects chosen in i) to count as counterpart relations.

For linguistic ersatzism, the objects in i) are (sets of) linguistic items, and the definitions in ii) exploit the fact that the objects chosen in i) contain linguistic items that are also present in the sentences of $L$. For instance, since a world is a set of sentences, we can say that a (non-modal) sentence of $L$ is true at $w$ iff it is a member of $w$. Relative to any such candidate $C$, one can do counterpart-theoretic semantics for $L$. Say that a candidate, $C$, is *charitable* if it assigns the "right" truth values to sentences of $L$, vindicates the "right" inferences in $L$, and so on, where the "right" truth values and inferences are those favored by ordinary speakers' use of $L$. The claim of the previous paragraph was that the semantics of English is counterpart-theoretic because counterpart-theoretic candidates are charitable. But this sort of charity is cheap, according to the objection. To obtain charitable candidates, we need look no further than set theory. Provided our modal language is consistent, and its logic is appropriate to counterpart theory, there *automatically* exist charitable counterpart-theoretic candidates within set theory; there is no need to construct candidates out of linguistic items. But these candidates need have nothing to do with modality, intuitively. So, the challenger concludes, charity is not enough. More is required of a candidate modal semantics than match with English usage.

More than charity is indeed required—this is the moral of Hilary Putnam's model-theoretic argument against realism.[8] Any consistent theory in a standard first-order (non-modal) language has a model in set-theory. So more than charity is required for an interpretation to be the intended interpretation of such a language; otherwise, any consistent theory, no matter how intuitively mistaken, will have an intended interpretation on which it is true, provided sets (or enough non-sets) exist. Likewise, more than charity is required of a candidate semantics for English modal talk. The challenge, then, is to say what makes the linguistic ersatzer's candidates, $C$, better than the cheap set-theoretic candidates.[9]

Merricks conceives of the challenge differently. He speaks of counterparts as "representing" possibilities, and raises problems for the ersatzer if counterparts do not "represent possibilities essentially and intrinsically, in and of themselves, not by way of interpretation" (2003, p. 535). Sets just sit there, as he says.

We must distinguish the kind of "intrinsic representation" that Merricks has in mind from a weaker notion. There is a weak sense in which the ersatzer's candidates must indeed represent worlds and individuals. Despite not really being talking donkeys, ersatz "talking donkeys" must play the theoretical role of Lewis's genuine talking donkeys within counterpart theory. But all this demands

[8]See Putnam (1978, part IV; 1980; 1981, chapter 2) for the argument, and Lewis (1984) for discussion.

[9]Even a modal realist like Lewis faces this question, though to a lesser degree. It's not as if his possibilia wear "Semantic value for QML" badges. Suitability for reference and meaning is a matter of degree.

is what we may call "relative" representation. Recall component ii) of each candidate counterpart-theoretic semantics, $C$. To do counterpart-theoretic semantics, relative to $C$, we need to make sense of locutions like "it is true at world $w$ that snow is white" and "individual $x$ is a talking donkey". To achieve this purpose, the ersatz worlds $w$ and individuals $x$ need not represent white snow and talking donkeys in any intrinsic sense; all that is required is that the locutions be well-defined by $C$, and thus that $x$ and $w$ "represent" snow and donkeys *relative to* $C$.

It really is incumbent on ersatzers to show why their candidates are better than the cheap set-theoretic candidates. If their candidates were "intrinsically representational", I suppose that would clearly establish their superiority. Perhaps candidates constructed from *sui generis* propositions, properties, states of affairs, and the like, would be intrinsically representational in the relevant sense. But ersatzers do not need to go this route. All they must do is show that their candidates—which allow *relative* representation—are better than the cheap ones.

And it is clear that they *are*. The relationship between those candidates and English is in no way arbitrary, precisely because of the natural manner in which the *relative* representation is defined. Linguistic ersatz worlds are like stories told in English, and represent (relatively) what their constituent sentences *say*. Likewise, ersatz individuals are like stories about individuals told in English. This superiority over the cheap set-theoretic candidates derives in large part from the conventional meaning of the words of English that they contain, but once it is clear that intrinsic representation is unnecessary, this source of superiority is unproblematic.[10]

A further sort of superiority is the systematic method used in the construction of ersatz candidates. In contrast to the cheap set-theoretic candidates, ersatz candidates employ a uniform method for defining relative representation; e.g., for *any* sentence $S$ and world $w$, $S$ is true at $w$ iff $S$ is a member of $w$. A simple, natural, rule, applies to all sentences. Notice that this superiority emerges only when considering *entire* candidates $C$, in the sense defined above (candidates consisting of choices of i) ersatz possibilia, ii) method of relative representation, and iii) ersatz counterpart relations). If we only look at an individual ersatz counterpart in isolation (e.g., an individual set of sentences) and ask what makes *it* suited to be a semantic value, we miss global dimensions

[10] One of Merricks's targets, Mark Heller, constructs ersatz worlds and individuals within pure mathematics, and so could not make use of this component of my reply.

of merit pertaining to entire candidates.

So far, we have focused on O2, ersatz candidates' apparent lack of intrinsic merit to be semantic values. The second argument against ersatz counterpart theory that I want to consider is based on O1, the observation that there are many ersatz candidates of equal merit. The argument is a version of Paul Benacerraf's (1965) puzzle about identifying numbers with sets. Linguistic ersatz counterpart theory identifies the fact that Humphrey might have won with a fact about ersatz counterparts. Corresponding to the many candidates for executing the ersatzer's construction, there are distinct candidate facts, $F_1$, $F_2$, .... Which of these can the ersatzer identify with the fact that Humphrey might have won? She cannot identify *more* than one with the fact that Humphrey might have won, for the $F_i$s are distinct from each other. Clearly, there is nothing in the convention-determining behavior of English speakers that chooses a *single* one.[11] So none of the candidates can be identified with Humphrey's possibly winning. So, ersatz counterpart theory is false.

The response is that English modal sentences are semantically indeterminate over the candidates. Anticipating this response, Merricks complains that the relevant semantic indeterminacy would be unlike both vagueness and "Quinean Indeterminacy" (2003, p. 534, n. 21). But so what? Indeterminacy comes in many different varieties. Merricks also worries that on some conceptions of linguistic ersatzism, there will be a great many candidates.[12] But again, so what? Sheer numbers is no obstacle (provided, of course, that each candidate has intrinsic merit to be a semantic value—i.e., provided that, for each, the objection from cheap structures within pure set theory can be answered).

(The multiplicity problem in metaphysics is not exactly parallel to the corresponding "problem" in mathematics. One can construct numbers, ordered pairs, groups, rings, vector spaces, etc., from sets in any of a number of ways. Faced with the question of which entities the numbers, pairs, etc., *really* are, a mathematician might arbitrarily *stipulate* a particular construction, since she is not trying to capture "what we were talking about all along". But for metaphysicians after the latter, indeterminacy is inevitable.)

Beyond their role in Merricks's arguments, O1 and O2 have further significance that I want to explore: they preclude the postulation of ontologically basic properties or relations of possibilia, for it is always a bad idea to apply ontologi-

[11] Epistemicists about vagueness could disagree here. See Sorensen (1988); Williamson (1994).

[12] Merricks (2003, pp. 534–535). He is there focusing on Heller's (1998) account, but claims later (p. 539) that his objections carry over to other accounts.

cally basic notions to arbitrarily constructible entities.[13] Imagine a counterpart theorist who wanted to take one of the following relations as ontologically basic:

- A relation of nomic accessibility between possible worlds (to account for laws of nature)
- A relation of nearness between possible worlds (to analyze counterfactuals)
- A counterpart relation

The problem comes from O1.[14] The posited relation $R$ is to apply to possibilia. According to ersatzism, possibilia are sets of sentences. So $R$ is a relation over sets of sentences. But which sets of sentences? The sets that *one* particular candidate $C$ identifies with possibilia? One needn't be overly hostile to metaphysics to recoil from the suggestion, for it elevates the choice of a method of construction, which seems wholly arbitrary, into a substantive question of metaphysics. Surely no one method of construction is privileged. And yet, the suggestion that $R$ is an ontologically basic relation pressures us to choose a single method. When doing the semantics for modal sentences in English, we will face the question of which possibilia are $R$-related. We must answer the question based on information about $R$'s holding over sets. But it might turn out that one candidate, $C$, counts sets $S_1$ and $S_2$ as possibilia that ought to turn out as $R$-related, whereas another candidate $C^*$ counts those very same sets as possibilia that ought to turn out as not being $R$-related. How, then, will we answer the question?[15]

Ersatzism limits the counterpart theorist's options: she cannot apply ontologically basic notions to possibilia. But how much of a limitation is this? Few modal counterpart theorists regard accessibility, nearness, or counterpart relations as ontologically basic. Counterparthood and nearness are usually

[13] The argument that follows parallels that of my 1996*b*.

[14] There also seems to be a problem in the vicinity of O2: sets of sentences seem insufficiently weighty to be the *relata* of fundamental properties or relations (other than properties and relations of pure set theory).

[15] My 1996*b* explores this more carefully. One might think to relativize the holding of $R$ to semantic candidates. The idea would be to turn $R$ into a three-place relation, relating sets $S_1$ and $S_2$ and candidate $C$ iff $S_1$ and $S_2$ represent $R$-related possibilia under $C$. But "candidates" are themselves arbitrarily constructible, which reintroduces the difficulty. See my 1996*b*, section 2.5.

taken to be similarity relations, which are ontologically underwritten by the properties of their *relata*, and most theorists analyze nomic accessibility in terms of laws of nature, not the other way around. Still, it is worth noting the closed doors.

It is also worth noting the doors closed to ersatzism about *temporal* counterparts. Temporal counterparts are what are usually called *temporal parts* of persisting entities, both past, present and future.[16] Unlike realism about possibilia, realism about past and future entities is a fairly common view, and is overwhelmingly favored by those who believe in temporal parts. Nevertheless, one can consistently combine anti-realism about past and future entities—*presentism*—with the doctrine of temporal parts, by claiming that temporal parts always exist at whatever moment is present. I myself reject presentism, but it seems to me that a presentist ought to appreciate the virtues of temporal counterpart theory as much as the next person. The presentist should therefore follow the lead of the anti-realist about possibilia, construct ersatz temporal counterparts from abstract entities, and apply the temporal counterpart relation to those entities. But here is where the closed doors come in: it is far more common to posit ontologically basic relations over *temporal* entities than over modal entities. Some want to posit ontologically basic causal relations, for instance. Such posits are ruled out by the multiplicity of candidate methods for constructing ersatz temporal counterparts.[17]

While modal counterpart theorists rarely posit *ontologically* basic properties or relations of possibilia, they often speak of *conceptually* basic properties and relations of possibilia—that is, they often use predicates of possibilia without reductively defining them. Lewis, for instance, says that the nearness relation for interpreting counterfactuals is a similarity relation that supervenes on the properties and relations instantiated within possible worlds, but probably cannot be *defined* by humans in terms of those properties and relations (1973, p. 95). Neither does he attempt a reductive definition of the counterpart relation in his writings on counterpart theory. Instead, he *explains* the (conceptually) *primitive* notion of a counterpart in various ways: he says it is a similarity relation, and makes various remarks about the relevant dimensions of similarity.

[16]That is the most natural view of temporal counterparts, at any rate; but see the next section.

[17]See my 2001*a* for discussions of temporal parts, the ontology of past and future entities (chapter 2), and the combination of presentism with the doctrine of temporal parts (chapter 3, section 4). While he does not defend temporal counterpart theory, Thomas Crisp (2007) is a presentist who applies ontologically fundamental relations to ersatz past and future entities, and thus faces the problem sketched here.

Does the multiplicity of methods for constructing counterparts cause trouble for this practice?

There is something a bit fishy about an ersatzer simply parroting Lewis here.[18] To simplify, suppose our ersatzer has chosen one particular method of constructing ersatz possibilia, and now wants to introduce a counterpart relation. To that end, she takes as primitive the predicate 'is a counterpart of', applied to ersatz possible individuals.[19] Though primitive, this predicate ought to be explained. But the ersatzer cannot say that it expresses a similarity relation, at any rate not in the most straightforward sense. For the similarities and dissimilarities between what it relates—abstract entities—are not of the relevant sort. Intuitively, the relevant similarities are between what those abstract entities *represent*. The ersatz possible individuals {$x$ is a talking donkey, ...} and {$x$ is a flying donkey, ...} should count as counterparts, not if those *sets* are (relevantly) similar, but rather, roughly, if a talking donkey ...would be relevantly similar to a flying donkey ....

This suggests taking a predicate of concrete objects, 'is relevantly similar to', as primitive, explaining it in the way Lewis does, but then *defining* the counterpart relation over ersatz possible individuals in terms of it.[20] Where ersatz possible individuals are understood to be sets of open sentences with only $x$ free, the obvious definition to try is:

> Ersatz individuals $a$ and $b$ are counterparts iff, necessarily, for any objects $o$ and $p$, IF: each member of $a$ is true when $o$ is assigned to the variable '$x$', and each member of $b$ is true when $p$ is assigned to the variable '$x$', THEN: $o$ and $p$ are (relevantly) similar

But this definition fails. The ersatz possible individuals $a$ and $b$ will specify *all* the features of the possible individuals they go proxy for, not just the intrinsic features. Each, in fact, completely describes the individual's entire possible world.[21] The descriptions contained in $a$ may well contradict the descriptions contained in $b$. If $a$ is proxy for a talking donkey, it will contain '$x$ is a talking donkey'; and if $b$ is proxy for a possible individual that happens to inhabit a world with no talking donkeys, then $b$ will contain '$x$ is such that there exist

[18] Compare Lewis (1986, p. 238).

[19] Must we speak of counterparts of actual concrete individuals as well as of ersatz individuals? No: the ersatz actual world contains ersatz actual individuals corresponding to the concrete actual individuals. Only these ersatz actual individuals have counterparts.

[20] I explore this in my 2002, pp. 302–304.

[21] See Lewis (1986, p. 149).

no talking donkeys'. Then the antecedent of the conditional in the definition will be vacuously satisfied, and the definition will automatically count $a$ and $b$ as counterparts, regardless of the other sentences contained in $a$ and $b$.

So one should instead state the definition as follows:

> Ersatz individuals $a$ and $b$ are counterparts iff, necessarily, for any $O$s and any $P$s, and any objects $o$ and $p$, IF: each member of $a$ is true when $o$ is assigned to the variable '$x$' **and all quantifiers in $a$ are restricted to the $O$s**, and each member of $b$ is true when $p$ is assigned to the variable '$x$' **and all quantifiers in $b$ are restricted to the $P$s**, THEN: $o$ and $p$ are (relevantly) similar

But even this succeeds only given a hefty metaphysical assumption: that for any two possible worlds $w$ and $v$ (the worlds of $a$ and $b$, respectively), there exists a third world containing copies of $w$ and $v$. Otherwise the antecedent of the definition will sometimes go vacuous.

It is a definition like this[22] that underlies the words "The ersatz possible individuals {$x$ is a talking donkey, ...} and {$x$ is a flying donkey, ...} are counterparts if a talking donkey ...would be relevantly similar to a flying donkey ...." If we accept the definition, well and good. But if not—if the hefty assumption daunts—then we are back to the drawing board in explaining the notion of a counterpart. The problem is not with applying undefined primitives to abstracta. It is with using such primitives without explaining them. We clearly *do* understand what *Lewis* means when he speaks of counterparts, and it is easy to slip into thinking that this understanding carries over to the ersatzer's predicate 'counterpart'. But that is an illusion. Lewis explained his predicate by appeal to similarity, whereas the ersatzer's predicate does not express a genuine similarity relation; and as we have seen, defining the ersatzer's 'counterpart' in terms of a genuine similarity predicate requires a substantive metaphysical assumption.

The ersatzer might try explaining 'counterpart' by citing examples of its holding. Suppose a certain possible talking donkey would indeed be relevantly similar to a certain possible flying donkey. Then the ersatzer could cite the following instance of the holding of the counterpart relation: "{$x$ is a talking donkey, ...} is a counterpart of {$x$ is a flying donkey, ...}". Other instances could be cited. And while such instances obviously do not uniquely *define* the counterpart relation, it might be alleged that they adequately explain it; we use the instances to latch on (well enough) to the intended relation over abstracta.

[22] Or a counterfactual one, which would raise similar issues.

Perhaps. But one worries that all we are really latching onto is the genuine similarity relation that would hold between any individuals represented by the ersatz individuals in the instances. Projecting from instances is easy if we're to project along some intuitive dimension of similarity, but since the ersatzer's counterpart relation is *not* a similarity relation, it's hard to see how we could project from the provided instances. At any rate, the ersatzer's choices here are: assume that any two possibilities can always be embedded in a single world; claim to understand 'counterpart', as applied to ersatz individuals, despite minimal and indirect explanation of that primitive; or find some further way to explain that primitive in terms of similarity.

## 3. Temporal counterpart theory

Temporal counterpart theory says that an object *will be* $F$ iff it has a future temporal counterpart that is (tenselessly) $F$, and that an object *was* $F$ iff it has a past temporal counterpart that is (tenselessly) $F$. Likewise for the metrical tense operators, which have no analog in modal logic: an object will be (was) $F$ in $n$ minutes iff it has a counterpart $n$ minutes in the future (past) that is $F$.

These truth conditions for tensed sentences ('$x$ will be $F$') are stated in a tenseless metalanguage: the 'is' in 'is $F$' is tenseless.[23] The tensed sentences I construe as sentences from standard tense logic, whose grammar is analogous to the grammar of quantified modal logic. While the ultimate goal is a semantics for English tensed sentences, temporal counterpart theorists should not claim that tense logic is a realistic model for the grammar of English tensed sentences (nor should modal counterpart theorists claim that QML perfectly represents the grammar of English modal sentences). Given the advances in linguistics proper since philosophy's linguistic turn, (mere) philosophers making semantic hypotheses can no longer pretend to be doing anything more than giving proto-theories. Counterpart theorists should claim only that something like counterpart theory for QML and tense logic will play a role in a more realistic syntax and semantics for natural language.

Temporal counterparts are entities from past and future times, just as modal

[23]Temporal counterpart theory is usually coupled with the B-theory of time, according to which this tenseless metalanguage is an ontologically fundamental language, but this is not inevitable. As noted in the last section, a presentist could accept temporal counterpart theory. That would involve giving an interpretation of the tenseless metalanguage in a yet more fundamental tensed language.

counterparts are entities from other possible worlds. Setting aside the anti-realism about past and future entities discussed in the previous section, the most natural view to take about these entities is *four-dimensionalism* (Sider, 2001*a*). On this view, the spatiotemporal world can be "sliced up" temporally as well as spatially; temporal counterparts are short-lived—perhaps instantaneous—slices of the material world, sometimes called temporal "stages".[24]

I have elsewhere discussed the question of which semantic hypothesis should be adjoined to four-dimensionalism: the "worm view" or the "stage view".[25] *Stages* we have met; (space-time) *worms* are aggregates of stages drawn from different times. The stage and the worm theories are answers to the question: which do ordinary language users name and quantify over, worms or stages? According to stage theorists, in the sentence 'I am currently typing', 'I' refers to my current stage. And in the sentence 'Every person in the room is typing', the quantifier ranges over current temporal stages. Likewise, it is current stages that satisfy predicates like 'is a person'. The worm theorist claims instead that we ordinary quantify over and name space-time worms. Whenever I use 'I', I refer to the aggregate of all of my temporal stages. 'Every person' ranges over all of the spacetime worms that are persons and are located at the present time.

Four-dimensionalists have traditionally preferred the worm view, but the stage view is also defensible. It does not imply that ordinary objects fail to persist over time, since persistence over time may be understood as requiring only that tensed sentences be true, construed counterpart-theoretically.

The stage and worm views thus differ over i) reference, and ii) the semantic analysis of tense. The stage theorist says that i) reference is to stages and ii) counterpart theory is the right theory of tense; the worm theorist says i) reference is to worms and ii) the right theory of tense involves "trans-time identity". These differences favor the stage view, in ways that parallel the benefits of modal counterpart theory. For instance, by claiming that different counterpart relations are invoked in different contexts, one can account for the inconstancy of our intuitions about tracing continuants over time. And one can block arguments for the the conclusion that distinct objects can share exactly the same parts at a time. Concerning a statue that tomorrow will be squashed,

[24] Can one combine temporal counterpart theory with "three-dimensionalism"? I take the three-dimensionalist's claim that a person, say, is "wholly present" at multiple times to imply that there are fundamental facts about which past and future times a person occupies. But then it seems likely that English tensed sentences concern these fundamental facts, rather than facts about counterparts. Compare my "trumping" argument (2001*a*, pp. 184–187).

[25] See my (1996*a*; 2001*a*), and also Hawley (2001); Haslanger (2003).

a stage theorist can vindicate our semantic intuition that each of the following is true: 'the statue = the lump' (since 'the statue' and 'the lump' refer to the same current stage), and 'the statue will not survive the squashing, whereas the lump will survive the squashing' (since the counterpart relations invoked by the two occurrences of 'will' are distinct.)[26]

The stage and worm views also differ over iii) the semantic values of predicates. As a result, they sometimes differ over the -adicy and intrinsicality of semantic values of predicates. Consider an ordinary present-tense sentence:

(T) Ted is sitting

If 'Ted' refers to a stage, then 'is sitting' expresses the monadic property, $S$, had by all and only the sitting stages. Whereas if 'Ted' refers to a space-time worm, the predicate 'is sitting' cannot express $S$. Instead, the worm theorist will say, 'is sitting' expresses the relation, $R$, that holds between space-time worm, $x$, and time, $t$, iff $x$'s stage at $t$ has property $S$. (T) is true, as uttered at $t$, iff Ted bears $R$ to $t$.

Or consider 'Lassie is a dog'. According to the stage view, 'is a dog' expresses a property, $D_S$, possessed by the referent of 'Lassie'—a stage. But whether something is a dog at a time is not merely a matter of what it is like at that time; it also matters what the thing was like in the past (and perhaps also in the future). Thus, $D_S$ is a highly relational property: a stage has $D_S$ only if it is counterpart-related to the right sorts of entities. In contrast, as Sally Haslanger (2003, section 7) has pointed out, a worm theorist might claim that 'is a dog' expresses a monadic property, $D_W$, had by the entire space-time worm 'Lassie'. This property is *not* relational in the way that $D_S$ is, since the dependence of whether $x$ is a dog on its future and past nature, for the worm theorist, amounts to dependence on the nature of $x$'s temporal *parts*, and properties concerning the (intrinsic) nature of one's parts are themselves intrinsic.[27]

[26]See Sider (1996*a*) and (Sider, 2001*a*, chapter 5, section 8). The shift in interpretation of 'will' cannot be assimilated to anything like ambiguity or ordinary indexicality, for one can rephrase the claim thus: 'The lump, but not the statue, will survive the squashing'. (Thanks to Sarah Moss here.) Such shifts in interpretation do not seem to occur with ambiguous or indexical phrases. Perhaps less overtly context-dependent expressions, for instance 'disgusting', or 'tall', or 'knows' provide better models. But the semantics of these expressions is not yet well-understood.

[27]The waters are actually muddier than this. Predicates like 'dog', 'horse' and 'chair' express *maximal* properties: something is disqualified on semantic grounds as being a dog, horse or chair if it is a large, undetached part of a dog, horse or chair (Sider, 2001*b*). Thus, Haslanger's

One might try to turn these differences over -adicy and intrinsicality of semantic values into an argument for one or the other view: for the stage view if one thinks that 'is sitting' expresses a monadic property, or for the worm view, if one thinks that 'is a dog' expresses a (relatively) intrinsic property.[28] But such arguments have a weakness. Let us distinguish between *metaphysical* and *semantic* intuitions about intrinsicality. (One could similarly distinguish metaphysical from semantic intuitions of -adicy.) Become a real metaphysician: forget about language and contemplate the world as it is in itself. Don't ask which predicates express intrinsic properties; ask simply whether various properties are intrinsic. Opinions about such matters are metaphysical intuitions of intrinsicality. Some of them, I think, are strong and highly justified.[29] What is less clear is whether we have strong semantic intuitions of intrinsicality, by which I mean intuitions of intrinsicality under linguistic descriptions—intuitions of the form "the property picked out by such-and-such predicate, whatever property that is, is intrinsic". Intrinsicality is a relatively theoretical notion, after all; why think we have *semantic* intuitions about it? The weakness in the arguments considered above is that they require semantic intuitions of intrinsicality and -adicy. For the worm and stage theories do not differ, with respect to any property, over whether it is intrinsic; they differ over which properties are expressed by our predicates.

(It is important to separate the two kinds of intuitions. Suppose you are initially inclined to intuit thus: "the property of being a dog is intrinsic". This might well be because you are simply *presupposing* the worm view, and thus presupposing that 'being a dog' refers to the property of *being a dog-worm*, and metaphysically intuiting that *this property* is intrinsic. It might well be that, once the question is raised of whether the worm-view or the stage-view is correct, your initial conviction disappears, and is replaced with agnosticism about whether "being a dog is intrinsic", despite your certainty that *being a*

claim that such predicates express intrinsic properties, given the worm view, is not strictly right. At best, these predicates express "locally intrinsic" properties: properties whose holding depends only on the intrinsic character of the immediate vicinity of their instances. But there are further sources of extrinsicality: perhaps something is a dog only if it is a member of a species that has evolved in the right sort of way. The more cautious point, then, is that the stage view makes these predicates much more relational than does the worm view.

[28] I argued in the first way in my (2000); Brian Weatherson argues (roughly) in the second way in his (MS) (using safer examples like 'grows while aging' rather than 'is a dog').

[29] We may even have modestly justified convictions that the most fundamental properties constituting shapes are intrinsic and monadic, regardless of their relation to English shape-predicates. To my mind, this highly metaphysical reading is the best one can make of Lewis's (1986, pp. 202–204) argument from temporary intrinsics.

*dog-worm* is intrinsic and *being a dog-stage* is extrinsic.)

Such is the case for the stage view and temporal counterpart theory. What obstacles do they face? The obstacles—and their solutions—are mostly parallel to those for modal counterpart theory. This is so for the Humphrey objection. It is also so for the problems of actuality considered in the next section. Presentists who want to be temporal counterpart theorists also face an analog of Merricks's objection (for nonpresentists the objection does not arise since temporal counterparts are then not ersatz.)[30] But there is a distinctively temporal problem, the problem of "timeless counting".

One reason to accept the stage view is its match with semantic intuitions when we *count* objects in the present tense. Case 1: If I have a single penny in my hand, 'How many ordinary coin-sized objects are in my hand' should be answered 'one'. This answer is predicted by the stage view, since there is just one coin-sized stage in my hand. But supposing that the coin will tomorrow be melted, preserving the copper but not the coin, the worm theory says instead that the answer is *two*, for in that case the coin-worm is numerically distinct from the hunk-of-copper worm (though they currently share a stage.) Case 2: if I will undergo fission tomorrow, the correct answer to 'how many persons are in the room now?' is one. This answer is predicted by the stage theory, but not by the worm theory, since the best worm-theoretic account of fission is Lewis's, which claims that cases of fission involve numerically distinct space-time worms that share stages before fission. Moral: in these cases, ordinary language counts stages, not worms.[31]

But for other sorts of counting, the stage-theoretic model seems incorrect. Suppose one takes the "timeless perspective", and asks "how many persons will there ever be?". The stage view incorrectly predicts that the answer is "infinitely many", assuming time is dense and there exists a stage at every moment within some continuous interval. Similarly for "how many persons have been sitting in my office during the last hour?"

In response to such objections, in earlier work I proposed a retreat from the pure stage view to a hybrid of the worm and stage views, by postulating indeterminacy between stage-theoretic and worm-theoretic semantics. While we usually quantify over and name stages rather than worms, sometimes we quantify over and name worms, in particular in the case of timeless counting

[30]See Merricks (2003, pp. 547–548, note 32).

[31]I argue against Lewis's (1976) method of counting by relations other than identity in Sider (1996*a*, section 3).

(Sider, 1996*a*, section VI). While I conceded at the time that this response was far from ideal, Sarah Moss has recently shown it to be unacceptable.[32]

Suppose I will undergo fission in a year. Thus, two space-time worms share a common stage at the present moment. How many people are sitting at my desk right now? The stage theory answers: one, for just one person stage sits at the desk. So far, so good. But now, Moss asks, how many people have been sitting at my desk during the last hour? The intuitive answer is still *one*, and yet, this answer is not forthcoming, whether we take 'people' to apply to worms or stages. Two person-worms, and infinitely many person-stages, are in the desk during that hour. And yet, *one* is the correct answer. We seem to be counting neither worms nor stages.

My original response to the problem of timeless counting was a bad idea. A better response both generates the correct counts in all the cases and eliminates the semantic indeterminacy.

The response employs the notion of a *perspective* from which a sentence is uttered, by which I mean an interval of time that, intuitively, the utterer thinks of as the temporal "topic" of the utterance. The perspective determines the range of (unembedded) quantifiers, referents of names, and what objects satisfy ordinary predicates. The general rule is this: the universe of discourse from a perspective consists of the restrictions of all space-time worms to that perspective. Quantifiers, from a perspective, range over the segments of spacetime worms confined to the perspective; and such segments are the extensions of predicates and the referents of names.

For ordinary present-tense predications of currently existing things, ordinary present-tense counting, and so on, the perspective is simply the present instant. From this perspective, we quantify over and name stages. For other utterances, the perspective is all of time, for instance an utterance of 'How many persons will there ever be?' From this perspective, it is entire space-time worms we quantify over and name. Thus, the account yields the pure stage and worm views as special cases, when the perspective becomes just the present moment, and all of time, respectively. But intermediate perspectives are possible as well. The perspective in Moss's example is naturally taken to be the last hour. Then, the things we name and quantify over are segments of space-time worms confined to the last hour. Since there is only one such segment of a person in my desk during the last hour—an hour-long segment of my space-time worm—we get the answer to Moss's question that we want:

[32] Personal communication. [Her argument has now appeared in print, in Moss (2012).]

"one". No ugly semantic shift has been posited, since quantifier domains and semantic values in every context are determined by the same, natural rule.

Once temporal perspectives are in the picture, it is natural to interpret tense operators in terms of them. 'It will be the case that $\phi$' means that $\phi$ is true in some future perspective, 'it was the case that $\phi$' means that $\phi$ is true in some past perspective, 'it is always going to be the case that $\phi$' means that $\phi$ is true in all future perspectives, and so on.[33] *De re* temporal claims of this sort must still be interpreted counterpart theoretically. The referents of names in a *de re* tensed claim are still determined by the speaker's perspective, but the perspectives quantified over by the tense operators determine the temporal lengths of the relevant counterparts. For instance, 'Fred' in 'Fred was standing', as uttered from a perspective $P$ consisting of the last hour, refers to Fred's segment $F$ from the last hour; and the sentence is true iff some perspective, $P$, is located before the last half hour and contains a temporal segment $P$ that is standing and is a counterpart of $F$. One more example: 'two people sat at my desk during the last hour, each of whom was previously standing', as uttered from the same perspective $P$, is true iff $P$ includes two worm-segments that are people, each of which has a standing counterpart drawn from some perspective before $P$. These remarks fall far short of being even a complete proto-theory, but they indicate a direction for future development.

# 4. The logic of 'actually'

## 4.1 Three problems

An operator for actuality should be added to the language of QML, to account for the full range of English modal claims, for instance 'it might have been that everyone actually rich was poor'.[34] Yet as Allen Hazen (1979) pointed out, there are purely logical obstacles to a counterpart-theoretic semantics for this operator.[35] Michael Fara and Timothy Williamson (2005) have recently strengthened Hazen's arguments and added some new ones.

Lewis's original counterpart theory (1968) consisted of rules for translating sentences from the language of QML into the language of counterpart theory.

[33]Humberstone (1979) develops a semantics for standard Priorian tense logic in which the points of evaluation are extended intervals of time.

[34]See Cresswell (1990) for an extensive discussion.

[35]Murali Ramachandran (1989) and Graeme Forbes (1982, 1985, 1990) offer alternate responses to Hazen, to which Fara and Williamson convincingly object.

The language of counterpart theory is a first-order language with variables $w, v, \dots$ ranging over possible worlds and variables $x, y, x_1, \dots$ ranging over possible individuals.[36] It has a primitive individual constant[37] @ for the actual world, a primitive predicate '$Ixw$' meaning that $x$ is *in* possible world $w$, and a primitive predicate '$Cxy$' meaning that object $x$ is a counterpart of object $y$ (let '$Cxyw$' abbreviate '$Ixw \wedge Cxy$'—"$x$ is a counterpart of $y$ in world $w$".) Call the world variables $w, v, \dots$ and the constant @, "world terms". Lewis first recursively defines a *translation* function. This function takes in a formula $\phi$ containing vocabulary of QML, and a world term, $v$, and returns a formula $\phi^v$, thought of as "the translation of $\phi$ relative to $v$". He then identifies the final translation of $\phi$—his analysis of $\phi$—with its translation $\phi^@$ relative to @. Here is the definition of Lewis's translation function:[38]

$$\phi^v \text{ is } \phi \text{ if } \phi \text{ is atomic} \tag{i}$$

$$(\sim\phi)^v \text{ is } \sim\phi^v \tag{ii}$$

$$(\phi \vee \psi)^v \text{ is } \phi^v \vee \psi^v \tag{iii}$$

$$(\forall \alpha \phi)^v \text{ is } \forall \alpha (I\alpha v \rightarrow \phi^v) \tag{iv}$$

$$(\Box \phi_{\alpha_1 \dots \alpha_n})^v \text{ is } \forall w \forall x_1 \dots \forall x_n [(Cx_1\alpha_1 w \wedge \dots \wedge Cx_n\alpha_n w) \rightarrow \phi^w_{x_1 \dots x_n}] \tag{v}$$

In clause (v), $\phi_{\alpha_1 \dots \alpha_n}$ is an arbitrary formula in which variables $\alpha_1, \dots, \alpha_n$ are free; $\phi^w_{x_1 \dots x_n}$ is the translation with respect to $w$ of the result of changing free $\alpha_i$s to $x_i$s in $\phi_{\alpha_1 \dots \alpha_n}$. Thus, clause v) translates $\Box\phi_{\alpha_1 \dots \alpha_n}$ as: "For any world $w$ and any counterparts $x_1 \dots x_n$ therein of $\alpha_1 \dots \alpha_n$, $\phi^w$ is true of $x_1 \dots x_n$". The derived clauses for the other connectives are what one would expect, including the following clause for the $\Diamond$ (defined in the usual way as $\sim\Box\sim$):

$$(\Diamond \phi_{\alpha_1 \dots \alpha_n})^v \text{ is } \exists w \exists x_1 \dots \exists x_n [(Cx_1\alpha_1 w \wedge \dots \wedge Cx_n\alpha_n w) \wedge \phi^w_{x_1 \dots x_n}]$$

("For some world $w$ and some counterparts $x_1 \dots x_n$ therein of $\alpha_1 \dots \alpha_n$, $\phi^w$ is true of $x_1 \dots x_n$".)

Given this, one natural way to introduce a sentential actuality operator ACT would be to define $(\mathsf{ACT}\,\phi_{\alpha_1 \dots \alpha_n})^v$ as follows:

$$\forall x_1 \dots \forall x_n [(Cx_1\alpha_1 @ \wedge \dots \wedge Cx_n\alpha_n @) \rightarrow \phi^@_{x_1 \dots x_n}]$$

[36] I follow Fara and Williamson (2005) in modifying Lewis's notation.

[37] I depart slightly from Lewis here.

[38] Here and henceforth, some convention must be introduced to insure that the variables introduced by translation rules do not become inadvertently bound to the wrong quantifiers.

("For any counterparts $x_1 \dots x_n$ of $\alpha_1 \dots \alpha_n$ in @, $\phi^{@}$ is true of $x_1 \dots x_n$".) This makes ACT a kind of universal quantifier over counterparts. But, as Hazen points out, this generates an intuitively unacceptable logic for ACT. Given the rest of the Lewisian definition of translation, the following intuitively contradictory formula turns out satisfiable:[39]

$$\Diamond \exists x(\mathsf{ACT}\, \exists y\, y{=}x \wedge \sim \mathsf{ACT}\, Fx \wedge \sim \mathsf{ACT} \sim Fx)$$

("There might have existed something that actually exists, but is neither actually $F$ nor actually not-$F$".) The reason is that a given non-actual entity may have two counterparts in the actual world. Counterpart theory generates the wrong logic for ACT, Fara and Williamson argue, and so is an incorrect analysis of modal discourse.[40]

This argument makes an assumption about the logic of ACT: that the displayed sentence is a logical falsehood. In fact, the sentence turns out to be a logical falsehood given the standard Kripke-semantics, which is based on transworld identity rather than counterpart theory. But this does not render the argument question-begging. The proponents of the argument do not assume that ACT obeys its intuitive logic *because* that is the logic generated by Kripke-semantics. Their (plausible) assumption about ACT's logic is intended to rest upon its own intuitive credentials.

In Lewis's translation scheme, the $\Box$ introduces universal quantifiers over counterparts. The $\Diamond$, on the other hand, introduces existential quantifiers. One might model ACT on the $\Diamond$ rather than the $\Box$, redefining $(\mathsf{ACT}\, \phi_{\alpha_1 \dots \alpha_n})^v$ as:

$$\exists x_1 \dots \exists x_n [C x_1 \alpha_1 @ \wedge \cdots \wedge C x_n \alpha_n @ \wedge \phi^{@}_{x_1 \dots x_n}]$$

("For some counterparts $x_1 \dots x_n$ of $\alpha_1 \dots \alpha_n$ in @, $\phi^{@}$ is true of $x_1 \dots x_n$".) But as Hazen notes, this existential reading of ACT renders satisfiable the following, equally intuitively contradictory, formula:

$$\Diamond \exists x(\mathsf{ACT}\, \exists y\, y{=}x \wedge \mathsf{ACT}\, Fx \wedge \mathsf{ACT} \sim Fx)$$

("There might have existed something that actually exists, and is actually $F$ but also actually not-$F$".) The reason again is that a non-actual individual can have multiple counterparts at the actual world.

[39]Let $F$ be perfectly precise, to avoid irrelevant worries about vagueness.

[40]Hazen himself is friendly to counterpart theory, and attempts to give a new version of counterpart theory that solves the problem. His proposal is interesting, but undermines one of the main benefits of counterpart theory: Lewis's defense of anti-haecceitism (discussed below).

A separate problem involving actuality, pressed by Fara and Williamson (2005), is that a non-actual individual might have *no* counterparts in the actual world. This results in the satisfiability of the following formula, whether we translate $\mathsf{ACT}$ in the first (universal) or second (existential) way:

$$\Diamond\exists x(\mathsf{ACT}\, Fx \leftrightarrow \mathsf{ACT}\sim Fx)$$

For if some non-actual $x$ has no actual-world counterpart, then for any $\phi x$ with just $x$ free, the universal translation schema makes $\mathsf{ACT}\ \phi x$ vacuously true, whereas the existential schema makes any such $\mathsf{ACT}\ \phi x$ vacuously false.

To the arguments of Hazen, and Fara and Williamson, I add one further, based on the fact that the counterpart relation need not be symmetric. Suppose there are two things in the actual world, $x$, and $y$, such that $x$ could have been exactly as $y$ in fact is. Then $x$ has a counterpart, $z$, that is exactly like $y$.[41] But now let us ask what things are counterparts of $z$ in the actual world. Counterparts of an object, Lewis tells us, are things that are sufficiently like that object, and more like that object than their worldmates. So it would seem that $y$ and only $y$ is a counterpart of $z$ in the actual world, for $y$ is exactly like $z$, and, moreover, is more like $z$ than $x$ is (and more like $z$ than anything else in the actual world, we may stipulate).[42] Let $F$ be some predicate satisfied by $x$ but not $y$. Then, whether our translation clause for $\mathsf{ACT}$ is existential or universal, the following statement comes out true:

$$\exists x(Fx \wedge \Diamond\, \mathsf{ACT} \sim Fx)$$

This statement is, intuitively, a logical falsehood.[43]

### 4.2 Solution to the problem of multiple counterparts in the actual world

Call relation $R$ a *thinning* of the counterpart relation iff it results from the counterpart relation by deleting just enough ordered pairs so that i) no object has two counterparts in any world, and ii) no two worldmates share a common counterpart. (That is: iff i) $R$ is a subset of the counterpart relation; ii) where

[41] $z$ could be $y$ itself, though it need not be.

[42] See Lewis (1968, pp. 114, 116).

[43] Analogs of all three problems confront a temporal counterpart-theoretic account of 'NOW'. Moreover, analogs of the first two problems arise for the metrical tense operators $\mathrm{WAS}_n$ ("it was the case precisely $n$ units ago that") and $\mathrm{WILL}_n$ ("it will be the case in precisely $n$ units that").

$x$ and $y$ are worldmates, $x = y$ if either $Rxz$ and $Ryz$, or $Rzx$ and $Rzy$; and iii) $R$ is not a proper subset of any other relation meeting conditions i) and ii).) There will of course be many thinnings of the counterpart relation, since there will be many ways of whittling the counterpart relation down to a "one counterpart per world" relation.[44] Where $w$ is a world and $R$ is a thinning, call $\langle w,R\rangle$ an *index*. I propose to treat $\Box$, $\Diamond$ and $\mathsf{ACT}$ as quantifiers over indices rather than possible worlds.

Consider a world of one-way infinite eternal recurrence, in which each age contains a single duplicate of me bouncing a basketball. This world can be considered as infinitely many possibilities. When the first duplicate is considered to be my counterpart, the world counts as a world in which I am the first of infinitely many basketball-bouncing duplicates. When the second duplicate is considered to be my counterpart, the world counts as a world in which I am the second of infinitely many basketball bouncers. And so on. Talk of a world when objects are "considered" as counterparts amounts to talk of an index: a world plus an arbitrary decision of which object is to count as my counterpart in that world. A "counterpart of $x$ in an index $\langle w,R\rangle$" is an object in world $w$ that bears $R$ to $x$.

Let us develop a modified counterpart-theoretic translation scheme in which indices play the roles of possible worlds. In what follows, I will use $i, j$, etc., as variables for indices. Where $i{=}\langle w,R\rangle$, I will refer to $w$ and $R$ as $w_i$ and $R_i$, respectively. Let $Cxyi$ abbreviate $Ixw_i \wedge R_i xy$ ("$x$ is $y$'s counterpart in index $i$"). The definition of $\phi^i$, the translation of $\phi$ with respect to $i$, is in essence Lewis's translation scheme with references to worlds replaced with references to indices:

$$\phi^i \text{ is } \phi \text{ if } \phi \text{ is atomic} \tag{i}$$

$$(\sim\phi)^i \text{ is } \sim\phi^i \tag{ii}$$

$$(\phi\vee\psi)^i \text{ is } \phi^i\vee\psi^i \tag{iii}$$

$$(\forall\alpha\phi)^i \text{ is } \forall\alpha(I\alpha w_i\rightarrow\phi^i) \tag{iv}$$

$$(\Box\phi_{\alpha_1\ldots\alpha_n})^i \text{ is } \forall j\forall x_1\ldots\forall x_n[(Cx_1\alpha_1 j\wedge\cdots\wedge Cx_n\alpha_n j)\rightarrow\phi^j_{x_1\ldots x_n}] \tag{v}$$

Next, we need the notion of the *actual* index, $i_@$, to play the role of the actual world @. What role is that? First, Lewis's final translation of a formula, $\phi$, in the language of QML, was defined as $\phi^@$—$\phi$'s translation in the actual

[44]Assuming the axiom of choice.

world. The analogous definition of the final translation of $\phi$ here would be $\phi^{i_@}$. Second, @ played a role in defining the translation of formulas containing $\mathsf{ACT}$; in the present setting we could define $(\mathsf{ACT}\,\phi_{\alpha_1\ldots\alpha_n})^i$ as:

$$\exists x_1 \ldots \exists x_n[Cx_1\alpha_1 i_@ \wedge \cdots \wedge Cx_n\alpha_n i_@ \wedge \phi^{i_@}_{x_1\ldots x_n}]$$

("$\phi$ is true of some counterparts of $\alpha_1\ldots\alpha_n$ in the actual index".)

Take "$i_@$" as short for "$\langle @, R_@\rangle$", where "$R_@$" is thought of as standing for the "actual thinning".[45] The problem is that there is no distinguished thinning to play the role of the actual thinning. It is sensible to require that the actual thinning be "@-reflexive"—that it assign each actual object as its own counterpart.[46] But this does not single out a unique thinning.

Two solutions are available.[47] One would be to define the final translation of $\phi$ as "$\exists R_@(R_@$ is @-reflexive $\wedge\, \phi^{\langle @,R_@\rangle})$", where '$R_@$ is @-reflexive' abbreviates '$\forall x(Ix@ \rightarrow R_@xx)$'. Alternatively, '$R_@$' can be treated in something like the way that Stalnaker (1981) treats the selection function in his semantics for counterfactual conditionals. Stalnaker's theory of counterfactuals contains a function symbol, 'selects', for a selection function, which assigns to any world $w$ and sentence $\phi$ one of the worlds that is most similar to $w$ in which $\phi$ is true. In fact, there is no single selection function that satisfies Stalnaker's constraints on 'selects'; he therefore regards 'selects' as semantically indeterminate over all the candidate functions that obey his constraints. When a counterfactual conditional is true relative to all such candidate functions, Stalnaker regards it as true simpliciter. When a counterfactual is false relative to all selection functions he regards it as false simpliciter. Otherwise he regards the sentence

[45] Expand the $w_i$ and $R_i$ notation in the obvious way, so that "$w_{i_@}$" and "$R_{i_@}$" abbreviate "@" and "$R_@$", respectively.

[46] This becomes important at the end of section 4.3.

[47] Which approach is better? Suppose closed sentence $P$ is a complete qualitative description of some world, $w$, which contains an individual, $x$, that is the only individual in that world satisfying predicate $F$. Suppose further that $x$ has two counterparts in the actual world, one of which satisfies $G$, the other of which does not. Now consider the sentences:

$$\Diamond[P \wedge \exists x(Fx \wedge \mathsf{ACT}\,Gx)]$$
$$\Diamond[P \wedge \exists x(Fx \wedge \mathsf{ACT}\sim Gx)]$$

On the first proposal, each turns out true. On the second proposal, each turns out indeterminate. One might argue thus: "if each of the quoted sentences is true then haecceitism is true. But haecceitism is not true. So the second proposal is preferable to the first." But this seems to presuppose that $\Diamond$ is a quantifier over worlds (see note 51).

as being neither true nor false. Likewise, "$R_@$" can be taken as semantically indeterminate over all the @-reflexive thinnings of the counterpart relation.[48]

This semantics solves Hazen's problem of actuality and multiple counterparts. The translations of:

$$\Diamond\exists x(\mathsf{ACT}\,\exists y\, y{=}x \wedge \sim\mathsf{ACT}\,Fx \wedge \sim\mathsf{ACT}\sim Fx)$$
$$\Diamond\exists x(\mathsf{ACT}\,\exists y\, y{=}x \wedge \mathsf{ACT}\,Fx \wedge \mathsf{ACT}\sim Fx)$$

are, respectively:

$$\exists i\exists x[Ixw_i \wedge \exists z(Czxi_@ \wedge \exists y(Iy@\wedge y{=}z))\wedge \\ \sim\exists z(Czxi_@\wedge Fz) \wedge \sim\exists z(Czxi_@\wedge\sim Fz))]$$

$$\exists i\exists x[Ixw_i \wedge \exists z(Czxi_@\wedge\exists y(Iy@\wedge y{=}z))\wedge \\ \exists z(Czxi_@\wedge Fz) \wedge \exists z(Czxi_@\wedge\sim Fz))]$$

Since no object has two counterparts in an index, neither of these sentences can ever turn out true.

Does the prohibition of multiple counterparts within an index undermine any of the benefits of counterpart theory? No.

Thinnings need not be equivalence relations, so the benefits that turn on this flexibility of the counterpart relation remain intact.

Nor does the prohibition threaten benefits that turn on multiple counterpart *relations*. Gibbard's (1975) Lumpl, a lump of clay, is synthesized in statue form, to form a statue, Goliath, and is subsequently vaporized, so that Lumpl and Goliath have exactly the same temporal career. We want to identify Lumpl with Goliath, even though we intuit that whereas Lumpl might have survived being squashed (though not being vaporized), it is not true that Goliath might have survived being squashed. Solution: the first modal claim, phrased using the name 'Lumpl', concerns Lumpl's *lump-counterparts*, whereas the second, phrased using the name 'Goliath', concerns Goliath's *statue-counterparts*. These are distinct counterpart relations, since they stress different dimensions of similarity. What determines which counterpart relation is used in the interpretation of a given term? This has not been worked out in much detail, but the idea is that

[48] One needn't follow Stalnaker in his supervaluational treatment of the semantically indeterminate sentences, however; one could approach the matter how epistemicists or nihilists (Braun and Sider, 2007) approach vagueness.

at least the sense of the term can be relevant: 'Lumpl' is introduced as a "lump-name", which triggers the lump counterpart relation; 'Goliath' is introduced as a "statue-name", which triggers the statue counterpart relation.[49] While this solution in a sense requires Lumpl/Goliath to have distinct counterparts at another world, it does not require two counterparts under any one counterpart relation.

The natural implementation of multiple counterpart relations in the index theory is to define an index as a sequence containing a possible world and a thinning for each counterpart relation. Which thinning is used in the interpretation of a given term will depend on which counterpart relation that term is associated with, just as in the original theory. Each thinning will assign exactly one counterpart to a given object in a given world, but the thinnings of the different counterpart relations may assign different objects within a single index. Thus,

$$\Diamond(\text{Lumpl survives being squashed and Goliath does not})$$

will turn out true, for there are indices in which the thinning of the lump counterpart relation assigns to Lumpl/Goliath something that survives being squashed, but in which the thinning of the statue counterpart relation assigns to Lumpl/Goliath a different entity, which does does not survive being squashed.

Unlike Lewis's original counterpart theory, the index theory renders the following sentence unsatisfiable:

$$\exists x \exists y (x{=}y \wedge \Diamond x \neq y) \tag{$\exists$LG}$$

But doesn't the multiple counterpart relations account of Lumpl and Goliath imply that Lumpl and Goliath are contingently identical? Yes and no. It is true that the defender of multiple counterpart relations will accept:

$$\text{Lumpl} = \text{Goliath} \wedge \Diamond\, \text{Lumpl} \neq \text{Goliath} \tag{LG}$$

For the natural interpretation of (LG), under the index-theoretic approach, is this: "Lumpl is identical to Goliath, and there is an index containing a lump-counterpart of Lumpl and a distinct statue-counterpart of Goliath". But we can explain why (LG) fails to imply its apparent existential generalization (∃LG).

[49] See Lewis (1971). The context of utterance can be relevant as well. 'It might have survived being squashed' ought to come out true when the antecedent of 'it' is 'Lumpl', but false when its antecedent is 'Goliath'.

According to the theory of multiple counterpart relations, names play a rich semantic role beyond simply having referents: they also trigger counterpart relations. 'Lumpl' and 'Goliath' in (LG) trigger distinct counterpart relations; but in moving to (∃LG), we have dropped the names in place of variables, which presumably cannot similarly trigger distinct counterpart relations.[50] The failure of existential generalization here is a bit like the failure of Quine's (1953) 'Giorgione was so-called because of his size' to imply 'Someone was so-called because of his size'.

Nor does the index-theoretic account threaten another benefit of counterpart theory that I have not previously mentioned: Lewis's defense of anti-haecceitism. Suppose the hypothesis of one-way eternal recurrence is *actually* true: in the actual world, there is an initial sequence of events with no predecessor, but subsequently this sequence is exactly repeated again and again without end. Thus, each "age" in this sequence contains a perfect doppleganger of me. Now, suppose that I am located in the first age, and let $P$ be a complete qualitative description of the world. It is therefore true that: $P$, and I inhabit the first age. Now, the following seems true:[51]

$$\text{Possibly: } P\text{, and I inhabit the second age} \qquad (\Diamond 2)$$

How can the counterpart theorist account for this? Under Lewis's original form of counterpart theory, we need a world in which $P$ is true that contains a counterpart of me in that world's second age. Since $P$ is a complete qualitative description of the actual world, such a world must qualitatively duplicate the actual world. *If* an object can have no more than one counterpart in a given world, then since I am a counterpart of myself, this world cannot be the actual world; it must be a distinct world, like the actual world in every qualitative respect, but in which my unique counterpart inhabits the second age. This would violate "anti-haecceitism", and thus the qualitative metaphysics of modality beloved of counterpart theorists (section 1), for the holding of the counterpart relation could not supervene upon the qualitative facts. Lewis's solution to this

[50] I suppose one could claim that (∃LG) has readings on which different counterpart relations are triggered by the different variables, but such readings sound strained.

[51] It *might* be reasonable to tap into anti-haecceitist intuitions and reject the alleged datum that ($\Diamond$2) is true. What would the difference *be*, one might ask, between the actual world and this alleged world in which I inhabit the second age? But notice that this presupposes thinking of the $\Diamond$ as a quantifier over worlds rather than indices. So if the index-theoretic approach is otherwise adequate, it seems preferable to accept the datum. At any rate, the index-theoretic approach has the flexibility to accommodate the alleged datum. Thanks to Karen Bennett.

problem was to claim that I have multiple counterparts in the actual world: each of my dopplegangers in each age is my counterpart (1986, p. 232). He thereby preserved the qualitative metaphysics of modality.

The index-theoretic account preserves this solution. Like Lewis's account, it is consistent with anti-haecceitism; it does not require the existence of duplicate possible worlds in which the facts of counterparthood differ. There is just one world in which $P$: the actual world. In that world, each of my dopplegangers is my counterpart. So for each doppleganger there is a thinning of the counterpart relation that assigns the doppleganger to me. And so, for each doppleganger there is an index in which the doppleganger is my one and only counterpart. (This does not require the theory to include any nonqualitative primitive notions: the counterpart relation is qualitative, and the notion of a thinning is defined quantificationally from the counterpart relation.) Therefore, ($\Diamond$2) turns out true.

### 4.3 Solution to the problem of no counterparts in the actual world

How is contingent existence usually handled in the *standard* (Kripkean) model theory of QML? There is no general consensus, but one natural approach is to declare an atomic sentence false, under an assignment of objects to its variables, at any world in which not all of those objects exist. Under this approach, Fara and Williamson's formula:

$$\Diamond\exists x(\mathsf{ACT}\,Fx \leftrightarrow \mathsf{ACT}{\sim}Fx)$$

is false in every model, for if $o$ does not exist at the actual world of a model, then when $o$ is assigned to $x$, $Fx$ is false at the actual world, so ${\sim}Fx$ is true at the actual world, and so $\mathsf{ACT}{\sim}Fx$ is true at every world of the model, whereas $\mathsf{ACT}\,Fx$ is false at every world. (In the standard model theory for $\mathsf{ACT}$, $\mathsf{ACT}\,\phi$ is true at any world iff $\phi$ is true in the actual world of the model.)

So why not just implement some analog of this approach within Lewis's translation scheme?—because nothing in that scheme corresponds to evaluating atomic formulas at worlds. In the translation of de re modal formulas, for instance $\forall x\Box Fx$, the quantifier over counterparts of $x$ is generated by the clause in the definition of translation for the $\Box$ (clause (v)), not by the clause for atomic formulas (clause (i)). That is in fact what generates the problem with $\mathsf{ACT}$ when an object fails to have a counterpart at the actual world. For when the $\mathsf{ACT}$ is interpreted like the other modal operators, introducing quantifiers over counterparts for the free variables in its scope, then if an object fails to have a counterpart, whole sentences beginning with ACT are rendered vacuously true or false (depending on whether the quantifiers introduced by ACT are universal or existential). Thus, in the case of Fara and Williamson's formula, in effect both ACT $Fx$ and ACT$\sim Fx$ get the same vacuous truth value.

So, pressing the question, why did Lewis choose to locate quantifiers over counterparts in the translation clauses for modal operators rather than in the clause for atomic sentences? Why didn't he instead proceed as follows? Simplify the clause for the $\Box$ thus:

$$(\Box\phi)^v \text{ is } \forall w\phi^w \tag{v$'$}$$

(The derived clause for the $\Diamond$ is then: $(\Diamond\phi)^v$ is $\exists w\phi^w$.) And complicate the clause for atomics thus:

$$(R\alpha_1\dots\alpha_n)^v \text{ is } \exists x_1\dots\exists x_n(Cx_1\alpha_1 v\wedge\cdots\wedge Cx_n\alpha_n v\wedge Rx_1\dots x_n) \tag{i$'$}$$

("$R$ is true of some counterparts of $\alpha_1\dots\alpha_n$ in $v$".) Thus, we treat an atomic sentence as false at a world if one of the values of its variables does not exist (lacks a counterpart) at the world.[52]

Doing things this way would immediately solve the problem of actuality and no counterparts in the actual world. For to this modified translation scheme we can add the obvious clause for ACT:

$$(\mathsf{ACT}\,\phi)^v \text{ is } \phi^@ \tag{vi}$$

Now Fara and Williamson's sentence, $\Diamond\exists x(\mathsf{ACT}\,Fx\longleftrightarrow\mathsf{ACT}\sim Fx)$, translates as a logical falsehood:

$$\exists w\exists x[Ixw\wedge(\exists y(Cyx@\wedge Fy)\longleftrightarrow\sim\exists y(Cyx@\wedge Fy))]$$

Which is what we want. So why didn't Lewis do things this way?

There are two reasons not to choose this route. The first is that $\Box Fx$ no longer says that all of $x$'s counterparts are $F$, for it says that every world contains *some* counterpart of $x$ that is $F$. It would not help to make the clause for translating atomics read: "$R$ is true of *all* counterparts of $\alpha_1\dots\alpha_n$ in $v$", or even "$\alpha_1\dots\alpha_n$ have counterparts in $v$, and $R$ is true of *all* counterparts of $\alpha_1\dots\alpha_n$

[52]Murali Ramachandran (1989) proposed something like this for somewhat independent reasons. He does not take up the problems I consider below. Fara and Williamson correctly point out that Ramachandran's theory fails to solve the problem of actuality and multiple counterparts, but neither they nor Ramachandran consider its value for solving the problem of actuality and no counterparts.

in $v$".[53] The first makes an atomic true in any world in which its variables do not denote, and so makes '$\exists x(x{=}\text{me} \wedge \Diamond\, x$ is a basketball)' true (since there are worlds in which I lack a counterpart.) And the second implies that for $\Diamond Fx$ to be true, it is not enough that some counterpart of $x$ be $F$; it must also be true that in some world, *all* of $x$'s counterparts are $F$. 'I could have inhabited the first age of a world of eternal recurrence' would turn out false.

This reason for not choosing this route depends on the fact that Lewis allows more than one counterpart per world. If we were prepared to disallow this, then the simple rule (i′) would suffice. In that case, $\Diamond Fx$ would be true iff some counterpart of $x$ is $F$, and $\Box Fx$ would be true iff i) $x$ has a counterpart in every world, and ii) all of $x$'s counterparts are $F$. $\Box Fx$ would thus be stronger than it is in Lewis's translation scheme, which only requires ii). But interpreting the $\Box$ so that $\Box Fx$ requires necessary existence of $x$ is not unnatural; moreover, the Lewisian condition ii) can always be expressed by $\Box(\exists y\, y{=}x \rightarrow Fx)$. A ban on multiple counterparts would not have been plausible for Lewis, but as we saw in the previous section, the ban can be implemented in an index-theoretic approach. So, solving the problem of no counterparts in the actual world with index-theoretic versions of (i′) and (v′) remains a live option.

But there is a second problem with (i′) and (v′), and thus a second reason favoring Lewis's original choice to place quantifiers over counterparts in the translation clauses for modal operators rather than atomics. Under (i′) and (v′), the translation of $\Diamond\Diamond Fx$ is $\exists w \exists v \exists y(Cyxv \wedge Fy)$, which is equivalent to $\exists v \exists y(Cyxv \wedge Fy)$—in effect, "some counterpart of $x$ is $F$". That's wrong; the translation ought to be: "*some counterpart of* some counterpart of $x$ is $F$". Since the counterpart relation need not be transitive, these are not equivalent. The present translation scheme equates $\Diamond\phi$ and $\Diamond\Diamond\phi$.

Lewis's original translation scheme does not equate $\Diamond\Diamond\phi$ and $\Diamond\phi$. Why not? In addition to generating the quantifier $\exists w$, Lewis's clause for the $\Diamond$ generates existential quantifiers for counterparts in $w$, one for each free variable in the contained formula. Thus, a pair of $\Diamond$s generates something of the form: "there exists a world $w$, and counterparts $x_i$ of $\alpha_i$ therein, and there exists a world $w'$, and counterparts $y_i$ of $x_i$ therein, …". Lewis's translation of a string of modal operators "touches down" in each world it generates, generating quantifiers over counterparts in that world, before proceeding to the next world. The new clause (v′) does not.

Unfortunately, this feature of Lewis's translation scheme—the fact that

[53] The latter is Ramachandran's (1989) proposal.

quantifiers over counterparts are generated by modal operators rather than atomics—is, as we saw, precisely the feature that created the problem of ACT and lack of counterparts in the actual world. What we need is a way to keep the quantifiers over counterparts in the clause for atomics, but nevertheless "touch down" when evaluating strings of modal operators. The final theory I will propose is a little complicated, so I will begin with a simplified first pass.

Call index variables and $i_@$, "index terms". In the index-theoretic framework, let us translate formulas relative to *sequences* of index terms rather than individual index terms. Understand the translation of $\phi$ relative to the null sequence as simply $\phi$. Define $\phi^{i_1 \dots i_m}$ thus:

$$(R\alpha_1 \dots \alpha_n)^{i_1 \dots i_m} \text{ is } \exists x_1 \dots \exists x_n (Cx_1\alpha_1 i_1 \wedge \dots \wedge Cx_n\alpha_n i_1 \wedge (Rx_1 \dots x_n)^{i_2 \dots i_m}) \quad \text{(i)}$$

$$(\sim\phi)^{i_1 \dots i_m} \text{ is } \sim\phi^{i_1 \dots i_m} \quad \text{(ii)}$$

$$(\phi \vee \psi)^{i_1 \dots i_m} \text{ is } \phi^{i_1 \dots i_m} \vee \psi^{i_1 \dots i_m} \quad \text{(iii)}$$

$$(\forall x\phi)^{i_1 \dots i_m} \text{ is } \forall x(Ixw_{i_m} \rightarrow \phi^{i_1 \dots i_m}) \quad \text{(iv)}$$

$$(\Box\phi)^{i_1 \dots i_m} \text{ is } \forall i\phi^{i_1 \dots i_m, i} \quad \text{(v)}$$

$$(\mathsf{ACT}\,\phi)^{i_1 \dots i_m} \text{ is } \phi^{i_@} \quad \text{(vi)}$$

As before, the final translation of a modal formula $\phi$ is its translation $\phi^{i_@}$ relative to '$i_@$'.

Here is the intuitive idea. To translate a formula, begin by translating at a single-membered sequence: the actual index. Each subsequent modal operator in the formula then adds a new index to the sequence. For the most part, the final index $i_m$ added to the sequence $i_1 \dots i_m$ plays the role of the single translation world in Lewis's theory (it restricts the quantifier in the clause for the $\forall$, and it is the index in which atomics are ultimately evaluated.) The earlier members of the sequence keep track of all the indices that have been visited in the course of translating the formula ("touching down"). They are inert until it is time to translate atomic formulas. Then, when translating $R\alpha_1 \dots \alpha_n$, what is in effect required for truth is that a chain of counterparts, beginning in $i_1$ and continuing through the intermediate indices to $i_m$, connect the values of $\alpha_1 \dots \alpha_n$ to objects in $i_m$; those latter objects must then satisfy $R$.

The first pass is not quite right. In the case of variables among $\alpha_1 \dots \alpha_n$ that were initially introduced deeper in the formula than the modal operator corresponding to $i_1$, the chain of counterparts should not extend all the way back to $i_1$. For instance, the translation of $\Diamond\exists xFx$ is currently:

$$\exists i\exists x(Ixw_i \wedge \exists y(Cyxi_@ \wedge \exists z(Czyi \wedge Fz)))$$

But this requires $x$ to have a counterpart at $i_@$. Informally speaking, what should happen instead is this: for each $\alpha_i$ in $(R\alpha_1\dots\alpha_n)^{i_1\dots i_m}$, the chain of counterparts leading to $i_m$ should begin, not at $i_1$, but rather at the index at which $\alpha_i$ was introduced.

To accomplish this, let the sequences for translating formulas include individual variables as well as index terms. Where $i$ stands for an arbitrary index term, $S$ stands for an arbitrary sequence of zero or more individual variables or index terms, and $V$ and $V_1$ stand for arbitrary sequences of zero or more individual variables only, the new definition of $\phi^S$ runs as follows:

$$(R\alpha_1\dots\alpha_n)^{V,i,V_1,S} \text{ is } \exists x_1\dots\exists x_p(Cx_1\alpha_{q_1}i\wedge\cdots\wedge Cx_p\alpha_{q_p}i\wedge(R')^{V',V_1,S}) \tag{ia}$$

$$(R\alpha_1\dots\alpha_n)^{V} \text{ is } R\alpha_1\dots\alpha_n \tag{ib}$$

$$(\sim\phi)^S \text{ is } \sim\phi^S \tag{ii}$$

$$(\phi\vee\psi)^S \text{ is } \phi^S\vee\psi^S \tag{iii}$$

$$(\forall x\phi)^{S,i,V} \text{ is } \forall x(Ixw_i\rightarrow\phi^{S^-,i,V,x}), \tag{iv}$$

$$(\Box\phi)^S \text{ is } \forall i\phi^{S,i} \tag{v}$$

$$(\mathsf{ACT}\,\phi)^S \text{ is } \phi^{VARS(S),i_@} \tag{vi}$$

where

- $\alpha_{q_1}\dots\alpha_{q_p}$ in (ia) are those variables (if there are none then let the translation be $(R\alpha_1\dots\alpha_n)^{V,V_1,S}$) among $\alpha_1\dots\alpha_n$ that occur in $V$, and $R'$ and $V'$ are the results of replacing $\alpha_{q_1}\dots\alpha_{q_p}$ in $R\alpha_1\dots\alpha_n$ and $V$ with $x_1\dots x_p$.

- $S^-$ in (iv) is the result of deleting all occurrences of $x$ from $S$. (There are such occurrences only in the special case where a quantifier occurs within the scope of another quantifier with the same variable.)

- $VARS(S)$ is the sequence of the individual variables in $S$ (i.e., the result of deleting the world variables in $S$)

This translation scheme does not equate $\Diamond\Diamond\phi$ and $\Diamond\phi$. Nor does it require, in order for $\Diamond\exists xFx$ to be true, that some possible $F$ have a counterpart in the actual world. Nor does it translate as true what seem to be contradictory sentences of QML containing ‘ACT’, in cases where nonactual objects have no counterparts, or more than one. Nor does it undermine any of the benefits of counterpart theory. Nor is it vulnerable to the third problem considered in section 4.1, which arose because the counterpart relation need not be symmetric. $\exists x(Fx\wedge\Diamond\,\mathsf{ACT}\sim Fx)$ translates as:

$$\exists x[Ixi_{@}\wedge Fx\wedge\exists w\sim\exists y(Cyxi_{@}\wedge Fy)]$$

which cannot be true since the actual thinning $R_{@}$ (the thinning in $i_{@}$) is @-reflexive.[54] This is the translation scheme we counterpart theorists should embrace.

Timothy Williamson raised the following serious problem for the version of counterpart theory defended in this section. I'm not sure how to respond. Whereas the following formula comes out invalid:

$$\exists x(Fx\wedge\Diamond\,\mathsf{ACT}\sim Fx)$$

the following, intuitively equivalent, formula does not come out invalid:

$$\exists x(Fx\wedge\Diamond\exists y(y{=}x\wedge\mathsf{ACT}\sim Fy))$$

This latter formula translates as follows:

$$\exists x(Ixi_{@}\wedge Fx\wedge\exists i\exists y(Iyw_i\wedge\exists z(Czxi\wedge y{=}z)\wedge\sim\exists z(Czyi_{@}\wedge Fz)))$$

The counterpart relation need not be symmetric, so $y$, which is $x$'s counterpart in $i$, may have a counterpart, $z$, in $i_{@}$, which is not $x$. Intuitively, here's what's going on. This sort of problem is blocked in the first formula because inside '$\mathsf{ACT}\,Fx$', we look to $x$ itself in the actual world, since the variable $x$ was introduced via a quantifier back in the front of the formula. But in the second formula, the variable $y$ is introduced, not at the beginning of the formula, but inside the scope of the $\Diamond$. The problem, in essence, is that my semantics treats these variables very differently.

[54]Note the importance of the chosen translation clause for ACT, rather than the alternate clause: $\phi^{S,i_{@}}$. Informally: when evaluating $\mathsf{ACT}\,\phi(\alpha_1\ldots\alpha_n)$, instead of *continuing* the chain of counterparts leading from $\alpha_1\ldots\alpha_n$ through the worlds visited so far along to objects in $i_{@}$ (which might lead to objects other than $\alpha_1\ldots\alpha_n$), we *directly* find counterparts of $\alpha_1\ldots\alpha_n$ themselves in $i_{@}$. In the case of variables among $\alpha_1\ldots\alpha_n$ that originated outside of the scope of any modal operators, these counterparts will be the values of those very variables, since $R_{@}$ is @-reflexive.

## References


Adams, Robert Merrihew (1974). "Theories of Actuality." *Noûs* 8: 211–31. Reprinted in Loux 1979: 190–209.

Benacerraf, Paul (1965). "What Numbers Could Not Be." *Philosophical Review* 74: 47–73.

Braun, David and Theodore Sider (2007). "Vague, So Untrue." *Noûs* 41: 133–56.

Carnap, Rudolf (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Cresswell, M. J. (1990). *Entities and Indices*. Dordrecht: Kluwer.

Crisp, Thomas (2007). "Presentism and the Grounding Objection." *Noûs* 41: 90–109.

Fara, Michael and Timothy Williamson (2005). "Counterparts and Actuality." *Mind* 114: 1–30.

Forbes, Graeme (1982). "Canonical Counterpart Theory." *Analysis* 42: 33–37.

— (1985). *The Metaphysics of Modality*. Oxford: Oxford University Press.

— (1987). "Is There a Problem about Persistence?" *Aristotelian Society, Supplementary Volume* 61: 137–55. Reprinted in Haslanger and Kurtz 2006: 267–84.

— (1990). "Counterparts, Logic, and Metaphysics: Reply to Ramachandran." *Analysis* 50: 167–73.

Gibbard, Allan (1975). "Contingent Identity." *Journal of Philosophical Logic* 4: 187–221. Reprinted in Rea 1997: 93–125.

Haslanger, Sally (2003). "Persistence through Time." In Michael J. Loux and Dean W. Zimmerman (eds.), *Oxford Handbook of Metaphysics*, 315–54. Oxford: Oxford University Press.

Haslanger, Sally and Roxanne Marie Kurtz (2006). *Persistence: Contemporary Readings*. Ed. Sally Haslanger and Roxanne Marie Kurtz. Cambridge, MA: MIT Press.

Hawley, Katherine (2001). *How Things Persist*. Clarendon Press.

Hazen, Allen (1979). "Counterpart-Theoretic Semantics for Modal Logic." *Journal of Philosophy* 76: 319–38.

Heller, Mark (1998). "Property Counterparts in Ersatz Worlds." *Journal of Philosophy* 95: 293–316.

Humberstone, I. L. (1979). "Interval Semantics for Tense Logic: Some Remarks." *Journal of Philosophical Logic* 8: 171–96.

Kripke, Saul (1972). "Naming and Necessity." In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 253–355, 763–9. Dordrecht: D. Reidel. Revised edition published in 1980 as *Naming and Necessity* (Harvard University Press, Cambridge, MA).

Lewis, David (1962). "Languages and Language." In Herbert Feigl and Grover Maxwell (eds.), *Scientific Explanation, Space, and Time*, volume 3 of *Minnesota Studies in the Philosophy of Science*, 3–35. Minneapolis: University of Minnesota Press.

— (1968). "Counterpart Theory and Quantified Modal Logic." *Journal of Philosophy* 65: 113–26. Reprinted in Lewis 1983: 26–46.

— (1971). "Counterparts of Persons and their Bodies." *Journal of Philosophy* 68: 203–11. Reprinted in Lewis 1983: 26–46.

— (1973). *Counterfactuals*. Oxford: Blackwell.

— (1976). "Survival and Identity." In Amelie O. Rorty (ed.), *The Identities of Persons*, 17–40. Berkeley: University of California Press. Reprinted in Lewis 1983: 55–77.

— (1983). *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.

— (1984). "Putnam's Paradox." *Australasian Journal of Philosophy* 62: 221–36. Reprinted in Lewis 1999: 56–77.

— (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Loux, Michael J. (ed.) (1979). *The Possible and the Actual*. Ithaca, NY: Cornell University Press.

Merricks, Trenton (2003). "The End of Counterpart Theory." *Journal of Philosophy* 100: 521–49.

Mondadori, Fabrizio (1983). "Counterpartese, Counterpartese*, Counterpartese$_D$." *Historie, Epistemology, Language* 5: 69–94.

Moss, Sarah (2012). "Four-Dimensionalist Theories of Persistence." *Australasian Journal of Philosophy* 90(4): 671–686.

Plantinga, Alvin (1976). "Actualism and Possible Worlds." *Theoria* 42: 139–60. Reprinted in Loux 1979: 253–73.

Putnam, Hilary (1978). *Meaning and the Moral Sciences*. Boston: Routledge and Kegan Paul.

— (1980). "Models and Reality." *Journal of Symbolic Logic* 45: 464–82. Reprinted in Putnam 1983: 1–25.

— (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

— (1983). *Realism and Reason: Philosophical Papers, volume 3*. Cambridge: Cambridge University Press.

Quine, W. V. O. (1953). "Reference and Modality." In *From a Logical Point of View*, 139–59. Cambridge, MA: Harvard University Press.

Ramachandran, Murali (1989). "An Alternative Translation Scheme for Counterpart Theory." *Analysis* 49: 131–41.

Rea, Michael (ed.) (1997). *Material Constitution*. Lanham, MD: Rowman & Littlefield.

Rosen, Gideon (1990). "Modal Fictionalism." *Mind* 99: 327–54.

Sider, Theodore (1996*a*). "All the World's a Stage." *Australasian Journal of Philosophy* 74: 433–53. Reprinted in Haslanger and Kurtz 2006: 91–117.

— (1996*b*). "Naturalness and Arbitrariness." *Philosophical Studies* 81: 283–301.

— (2000). "The Stage View and Temporary Intrinsics." *Analysis* 60: 84–88. Reprinted in Haslanger and Kurtz 2006: 443–48.

— (2001*a*). *Four-Dimensionalism*. Oxford: Clarendon Press.

— (2001*b*). "Maximality and Intrinsic Properties." *Philosophy and Phenomenological Research* 63: 357–64.

— (2002). "The Ersatz Pluriverse." *Journal of Philosophy* 99: 279–315.

Sorensen, Roy (1988). *Blindspots*. Oxford: Oxford University Press.

Stalnaker, Robert (1976). "Possible Worlds." *Noûs* 10: 65–75. Reprinted in Loux 1979: 225–34.

— (1981). "A Defense of Conditional Excluded Middle." In William L. Harper, Robert Stalnaker and Glenn Pearce (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, 87–104. Dordrecht: D. Reidel.

Weatherson, Brian (MS). "Growing Individuals and Intrinsic Properties." Available online at `http://brian.weatherson.org/GIs.pdf`.

Williamson, Timothy (1994). *Vagueness*. London: Routledge.