

# Actuality for Counterpart Theorists

Jeffrey Sanford Russell

April 25, 2012

## Abstract

The counterpart theorist has a problem: there is no obvious way to understand talk about actuality in terms of counterparts; Fara and Williamson have charged that this obstacle cannot be overcome. Here I defend the counterpart theorist by offering systematic interpretations of a quantified modal language that includes an actuality operator. Centrally, I disentangle the counterpart relation from the related notion of a representation relation. The relation of possible things to the actual things they represent is variable; an adequate account of modal language must keep track of the way it is systematically shifted by modal operators. I apply my account to resolve several puzzles about counterparts and actuality. In technical appendices, I prove some important logical results about this “representational” counterpart system and its relationship to other modal systems.

In this essay I answer what I take to be one of the more serious challenges to a counterpart-theoretic account of possibility: the problems of actuality raised by Allen Hazen (1979) and subsequently pressed by Michael Fara and Timothy Williamson (2005). The essay has two parts. In the main part, I defend my proposal and apply it to some problem cases. The second part consists of technical appendices in which I

---

I am especially grateful to Ted Sider for many helpful discussions and extensive comments on several drafts. I also owe thanks to Andrew Bacon, Dylan Dodd, Cian Dorr, Kit Fine, Jim Pryor, Jennifer Wang, Seth Yalcin, the participants in the 2009 Phlox workshop on modality at Humboldt University, and three extremely helpful anonymous reviewers.

An earlier version of this essay appears as the first chapter of my dissertation, *Possible Worlds and the Objective World*.

characterize the counterpart theorist's logic of actuality and some of its connections to other modal systems.

## 1 Two languages

All of us who speak about what is possible fall (more or less begrudgingly) into using **possibilist** language, speaking of possible worlds inhabited by various possible individuals. We tell various stories about what we really *say* with this language: maybe it's about isolated concrete universes, or abstract states of affairs or propositions or properties, or sets of sentences, or featureless “bare possibilia”; or maybe it isn't really *about* anything, and possibilist language should be translated away not (so to speak) one thing at a time, but rather whole sentences at a time; or maybe it involves some kind of pretense.<sup>1</sup> Whatever story we tell, this way of talking is too useful to give up altogether (c.f. Lewis 1986, ch. 1). But this essay is not about what our possibilist talk really means; it is about counterpart theory, and counterpart theory is not tied to any one of these stories.

Besides the possibilist language—**Possibilese**—we also have another way of talking about what is possible: **Modalese**. In this other language we say what *could have been* or *must be*, the sort of thing we familiarly formalize with boxes and diamonds. When in Possibilese we say, “There is a possible talking donkey”, in Modalese we say, “There could have been a talking donkey”. I have no official view as to whether one locution is more basic than the other; but unofficially I will slip into talking as if the possibilia account for what is possible.

Since both languages concern possibilities, there ought to be some correspondence between the commitments we express in each. The form of this correspondence is the main question of this essay. But much of it is unproblematic. Here's how one version goes—a counterpart theorist's.

For a start, Modalese has a **categorical** fragment that involves no modal operators. This says how things are, remaining silent on what might have been. In Possibilese we speak of a special world—the **actual** world,  $w_{@}$ —which is an image of the categorical truths. If in Modalese we say something is  $F$ , then in Possibilese we say something in the actual world is  $F^*$ , where  $F^*$  is a Possibilese predicate corresponding to  $F$ . In Modalese we say Donnie is a donkey; in Possibilese we say Donnie\*, who is a possible individual in the actual world, is a donkey\*. (Officially, we may wish to distinguish the predicates of the two languages, and similarly names of actual things from corresponding names of the possible things in the actual world. But usually I'll drop the stars to lighten the burden of notation.)

---

<sup>1</sup>Some representatives of these different views: Lewis 1986; Plantinga 1974; Sider 2002; Williamson 1998; Fine 1985; Rosen 1990.

More generally, possible worlds correspond to true possibility claims: at a first pass, what's possible is what's true at some possible world. (I'll ignore accessibility relations, except in the appendices.) There could have been a talking donkey; there is a talking\* donkey\* in some possible world. This will do for *de dicto* possibility—possibility statements without any free variables or other singular terms. But *de re* possibility is a bit trickier. Humphrey could have been president; at least, then, there is a possible world where somebody is president\*. But which possible individuals, by being president\*, secure the possibility of being president for Humphrey? Call these Humphrey's *representatives*. Humphrey could have been president iff he has a representative inhabiting some world who is president\*.

Humphrey has more than one representative. He could have been president, so he has a representative which is president\*. But he also clearly could have been *not* president, and so he has a representative which isn't president\*. These are different. (Being president\* isn't a *relative* matter.) In general, I'm supposing possible individuals are **world-bound**: each possible individual is one specific way, and inhabits exactly one possible world. But even so, Humphrey could have been many different ways, so he has many representatives, in many different worlds.

We don't just attribute *de re* possibilities to actual things; we also sometimes attribute merely possible *de re* possibilities. (If there were a gold mountain, it would essentially have atomic number 79.) So in Possibilese we should speak not just of representatives for actual things, but more generally **counterparts** for each individual in the pluriverse. Humphrey\*'s counterparts present possibilities for Humphrey, *their* counterparts present Humphrey's possible possibilities, and so on.

Some of what I've said so far may sound very controversial. But it isn't really, taken in the right spirit. Suppose Kristin thinks that a single possible object Humphrey\* inhabits many worlds—she believes in “trans-world individuals”. And she thinks that what could be true of Humphrey is just what is true at some world of Humphrey\*. She can still accept everything I've said, by appropriately reinterpreting it. I said that individuals are world-bound. Kristin can accept this, by regarding my “individuals” as her individual-world pairs, and reinterpreting the predicates accordingly. (She might take me to be putting things unnaturally, but naturalness is not at issue.) Likewise, Kristin can accept that possibility claims correspond to claims about counterparts; she just has a special story about what the counterpart relation is: the “individual”  $\langle b, w \rangle$  is a counterpart of  $\langle a, v \rangle$  just in case  $a$  and  $b$  are the very same thing. So Kristin can happily say everything that the counterpart theorist does.

What sets the *serious* counterpart theorist (let's call him Lucius) apart from Kristin is not what he says, but what he *doesn't* say. Kristin has a special account of the counterpart relation, in terms of identity. This has consequences for shape it can take. Kristin's counterpart relation must be an equivalence relation: reflexive, transitive, and symmetric. A single thing can have only one counterpart in any particular world.

Two distinct things in a world must not have the same counterpart. Nothing is a counterpart of anything else in its own world. But Lucius isn't committed to the counterpart relation having these properties. The counterpart relation should probably be reflexive, since, necessarily, things could be the way they are. But these other features are far from obviously desirable constraints, and there are various reasons for rejecting them.<sup>2</sup> Here's one kind of reason: you might want to explain the counterpart relation in terms of *qualitative resemblance* (perhaps in service of some general ambition to reduce *de re* modality to the *de dicto*). If counterparthood is qualitative, then cases of qualitative symmetry will give rise to cases of multiple counterparts.

Whatever his reasons, Lucius has weaker commitments about the structure of counterparthood than Kristin. This reticence is a *ceteris paribus* virtue, since it requires less from a theory of possibilia. But it also raises a worry: the sparser structure of Lucius's counterparts might not stand up to all the pressures we put on a theory of possibilia. In particular, there might turn out to be important modal claims that Kristin can interpret possibilistically but Lucius cannot.

To address this worry, we can start by regimenting the general story about the correspondence between modal talk and counterpart talk, turning it into a system for interpreting a formalized fragment of Modalese in possibilist terms. Doing this will show that this fragment, at least, does not outstrip counterpart theory's resources.

The interpretations go like this. (I'll just deal with the case of formulas of a single variable for now, since dealing with multiple objects brings up a bit of extra complication that isn't immediately important. I discuss the generalization in Section 3.) We'll interpret formulas of quantified modal logic by inductively defining what it takes for a modal formula  $\phi$  to be **true of** a certain possible individual  $a$  **at** an evaluation world  $v$ —abbreviated  $\llbracket \phi \rrbracket^{v,a}$ . (I henceforth drop Possibilese stars.)

$$\begin{aligned} \llbracket Fx \rrbracket^{v,a} &\text{ iff } Fa \\ \llbracket \neg \phi \rrbracket^{v,a} &\text{ iff not } \llbracket \phi \rrbracket^{v,a} \\ \llbracket \phi \wedge \psi \rrbracket^{v,a} &\text{ iff } \llbracket \phi \rrbracket^{v,a} \text{ and } \llbracket \psi \rrbracket^{v,a} \\ \llbracket \exists x \phi \rrbracket^{v,a} &\text{ iff for some } b \text{ in } v, \llbracket \phi \rrbracket^{v,b} \\ \llbracket \diamond \phi \rrbracket^{v,a} &\text{ iff } a \text{ has a counterpart } b \text{ in a world } w \text{ such that } \llbracket \phi \rrbracket^{w,b} \end{aligned}$$

(The other standard connectives can be defined in terms of  $\neg$ ,  $\wedge$ ,  $\exists x$  and  $\diamond$ . In particular,  $\square$  is dual to  $\diamond$ .) A modal sentence  $\phi$  with no free variables is true *simpliciter* iff it is true of an arbitrary actual individual at the actual world. (We can also think of

<sup>2</sup>Lewis discusses some of these reasons (1986, 231–2, 243–6, 258–9). Interesting motivations for multiple counterparts have also arisen in the philosophy of physics (see for instance Butterfield 1989, §6; Maudlin 1990, 550; Belot 1995; Pooley n.d., §2.3). Accessibility relations can help the trans-world identity theorist relax some of these constraints, but not all of them.

free  $x$  as a name; if  $x$  refers to an actual individual  $a$ , then  $\phi$  is true iff  $\phi$  is true of  $a$  at the actual world.)<sup>3</sup>

David Lewis (1968) provided essentially this account. I should point out one difference in my presentation: Lewis gave a *translation manual*, a way of systematically transforming modal formulas into formulas of a regimented counterpart-theoretic language. I think that this extra level of semantic ascent makes the main discussion harder to follow, so in my interpretations I'll continue to *use* informal Possibilese sentences, rather than *mentioning* their formalizations. But once we've carefully worked out the informal interpretations, it isn't difficult to turn them into formal translations, and I show explicitly how to do this in [Appendix D](#). (I also do a third thing, in [Appendix A](#), which is to give *model-theoretic* interpretations for meta-logical purposes.)

So far, so good: the bit of Modalese we just interpreted, at any rate, can be systematically interpreted in terms of counterparts.

## 2 The problems of actuality

But our modal commitments go beyond what can be expressed in this formal fragment: quantified modal logic is expressively weak. For instance, it cannot express modal claims like this one (Hazen 1976, 34; Hodes 1984c, 23):

There could have been things that don't actually exist. (1)

To capture this meaning, we need to add to our modal repertoire, alongside “possibly” and “necessarily”, a new sentence operator  $@\phi$  with the sense “It is actually the case that  $\phi$ ”. With it, we can render (1) as

$$\diamond \exists x \neg @ Ex$$

( $Ex$  abbreviates  $\exists y y = x$ .)

But there is no obvious way to extend Lewis's interpretations to this new operator—and so, in general, there is no obvious way for the counterpart theorist to systematically interpret the full range of Modalese. Allen Hazen (1979) pointed out this gap, and more recently Michael Fara and Timothy Williamson (2005) have argued that it cannot be filled.

To interpret an actuality clause, it isn't enough simply to reset the evaluation world to  $w_@$ ; we also need to choose actual individuals to stand in for the objects designated

---

<sup>3</sup>Since I said that the trans-world identity theorist Kristin could accept the counterpart theorist's claims, I ought to mention one problem she might have with this  $\diamond$ -clause: It requires that we restrict our attention to individuals that exist at the evaluation world, whereas Kristin may wish to make claims about non-existents. I return to this issue in [Section 4](#).

by free variables. On the model of the  $\diamond$ -clause, we might try

$$\llbracket @ \phi \rrbracket^{v, a} \text{ iff } a \text{ has a counterpart } b \text{ in } w_{@} \text{ such that } \llbracket \phi \rrbracket^{w_{@}, b}$$

But this has the wrong consequences. For a possible object may very well have *two* counterparts in the actual world, one happy and one unhappy; in that case the suggested interpretation would render this formula true (Hazen 1979, 330):

$$\diamond \exists x (@ Ex \wedge @ Fx \wedge @ \neg Fx) \quad (2)$$

There could have been something which actually exists and is actually happy and is actually not happy.

This sounds contradictory. On the other hand, we might try to model “actually” on “necessarily”, and instead require that *every* counterpart of  $a$  in  $w_{@}$  satisfy  $\phi$ . This is no better, since then the following is satisfied:

$$\diamond \exists x (@ Ex \wedge \neg @ Fx \wedge \neg @ \neg Fx) \quad (3)$$

There could have been something which actually exists and is not actually happy and not actually not happy.

Again (ignoring the vagueness of “happy”) this is an intuitive contradiction. This is **the problem of multiple counterparts**.

Since the counterpart relation need not be symmetric, we may also look to **the wrong counterpart**: the existential and universal approaches both let us satisfy this unacceptable sentence (Sider n.d., 24):

$$\exists x (Fx \wedge \diamond @ \neg Fx) \quad (4)$$

There is something happy which could have been such that actually it is not happy.

Moreover, a possible object might have **no counterparts** in the actual world. In that case the existentially and universally quantified approaches both satisfy this (Fara and Williamson 2005, 9):

$$\diamond \exists x (@ Fx \leftrightarrow @ \neg Fx) \quad (5)$$

There could have been something which is actually happy just in case it is actually not happy.

These simple interpretations yield an unacceptable logic of actuality. Other candidates have been offered, but thus far each has foundered (Forbes 1982, 1990; Ramachandran 1989, 1990; Sider n.d.). Moreover, Fara and Williamson argue that the challenge cannot be met: in particular, that there is no coherent way of systematically

extending Lewis’s interpretations of quantified modal logic to include actuality, and that “the prospects for *any* coherent translation scheme are dim” (2005, 23).

One might wonder how serious a problem this is.<sup>4</sup> Though we want to eventually provide systematic semantics for *natural* language, that is by no means the same as the formal modal language under discussion. The rigid operator @ is no synonym for the English word “actually” (c.f. Lenk 1998, 158ff). So, first, why should the counterpart theorist have any interest in interpreting this operator? And second, why be moved by some philosophers’ judgments that certain sentences involving it are inconsistent? Since “actually”, as we use it here, isn’t quite English, it isn’t in our capacity as English-speakers that we make judgments on the problem sentences. To what other authority can we appeal?<sup>5</sup>

We care about @ not because it is *English*, but because it is *useful*, by playing an important role in a certain regimentation of possibility talk. This regimented language is philosophically useful, for framing theories and discerning logical consequences, and it is semantically useful, providing a ladder rung on the way to natural language. Our judgments on the problem sentences, then, are best thought of not as linguistic data, but rather as test conditions we (in our philosophical sophistication) take as important qualifications for an operator playing a certain theoretical role—“the @-role”.

The @-role’s importance comes in part from the fact that it is a special case of a more general apparatus. In natural and philosophical speech alike, we freely shift our attention between many different salient possibilities, and compare how things are in several of them in a single clause. For example:

There could have been something that doesn’t actually exist, and in fact *necessarily*, whatever existed, there could have been something that in that case didn’t exist. (6)

How can we express this claim formally? The actuality operator is not enough here, but we can capture (6) with a more general “modal anaphor” operator, glossed “in that case” (Hodes 1984b, 426). It isn’t difficult for a trans-world identity theorist (like Kristin) to give a possibilist interpretation for such an operator; but the counterpart theorist runs against all the same problems as with actuality. Our concern with identifying individuals in different possibilities goes well beyond the simple case that Lewis (systematically) accounts for.

In the face of failure to give a counterpart-theoretic account of the @-role, it isn’t out of the question for the counterpart theorist to give it up as a lost cause—or at

<sup>4</sup>Thanks to Jim Pryor for discussion on this point.

<sup>5</sup>“What is the correct counterpart-theoretic interpretation of the modal formulas of the standard language of quantified modal logic? — Who cares? We can make them mean whatever we like. We are their master. We needn’t be faithful to the meanings we learned at mother’s knee—because we didn’t” (Lewis 1986, 12).

least reject some of the constraints on that role imposed by the problem sentences. He can resign himself to finding other resources to do its philosophical work, and he can hope that there is another route to natural language that doesn't go through this regimentation. While the "high-road" response may be painful, it isn't out of the question. But my "low-road" response is better: by providing a counterpart-theoretic account of an actuality operator, I spare the philosopher and the semanticist the loss of a useful theoretical tool.

One might still wonder why we should care about *systematic interpretations* of "actually". After all, it looks like each of the problem sentences considered so far can be given some kind of ad hoc counterpart-theoretic gloss. For instance, (4) looks to be paraphrasable thus:

For some  $a$ ,  $a$  is happy and  $a$  has a counterpart  $b$  in a world  $w$  such  
that  $a$  is not happy.

This paraphrase is contradictory, as it should be. If off-the-cuff interpretations are always available, why fuss over systematizing them? I see four main reasons.

First, systematic interpretations allay any worry that our ad hoc creativity will eventually run out, putting to rest the charge of expressive weakness.

Second, *ad hoc* paraphrases will be of no help in the project of giving systematic semantics for natural language.

Third, systematic interpretations help decide logical questions. One of the chief benefits of possible worlds semantics is that it transforms questions about modal logic into easier questions about the extensional logic of Possiblese. Without a systematic account of "actually", the counterpart theorist cannot take advantage of this strategy, and so he is in a weak position to settle questions about the logic of actuality. Indeed, puzzles like the one raised by Delia Graff Fara (Section 5) show that reasoning about actuality without a systematic theory of its behavior quickly leads into brambles. Maybe the brambles come with the terrain: perhaps "actually" cannot be systematized and has no simple logic. But if this defeatist presumption turns out to be wrong, we stand to gain.

Finally, systematizing actuality yields philosophical insights: it illuminates what we say with possibilist language, and how best to use it as we navigate our theory of possibilities. The evidence for this is in the account itself.

### 3 Representing actuality

Humphrey's counterparts represent him in other worlds. Humphrey could have been president; some possible president represents Humphrey. When we ask what is *actually* the case concerning that president, we are interested in what is true of *Humphrey*: for



instance, that he lost. In general, what is actually true of a possible individual is what is true of the actual individual that it represents. In order to evaluate a formula governed by an actuality operator, then, we need to decide what actual thing (if any) each of its terms represents. We are not looking for a *new* representative for Humphrey—as we would do by searching the actual world for counterparts of his counterparts. Rather, we want the individual we were already discussing, Humphrey himself. “Actually” effectively undoes representation *de re*, sending representatives back to the actual individuals for which they interceded.

Note that Kristin (who talks like a counterpart theorist but believes in trans-world individuals) has a ready way to do this: she can define a function  $A$  that maps each ordered pair  $\langle a, w \rangle$  to the corresponding pair  $\langle a, w_{@} \rangle$ . On her understanding of representation *de re*, this is the only actual “individual” that  $\langle a, w \rangle$  could represent. So her interpretation looks like this:

$$\llbracket @ \phi \rrbracket^{v, a} \text{ iff } \llbracket \phi \rrbracket^{w_{@}, A(a)}$$

Seeing this, the counterpart theorist might also try to define an “actuality function” in terms he accepts. This idea immediately runs into trouble. Romulus and Remus are identical twins. Romulus founded Rome (after killing his twin). Their common zygote might not have split: there is a possible world in which only one boy is born—call him Primo, and let “Primo-ish” be a predicate that applies to Primo uniquely among all possible individuals. (Perhaps this is a complete centered qualitative description of Primo’s world.) Romulus could have been Primo-ish, and Remus could have been Primo-ish. So the following holds:

$$\diamond \exists x (Px \wedge @ Fx) \wedge \diamond \exists x (Px \wedge @ \neg Fx) \quad (7)$$

There could have been someone Primo-ish who actually founded Rome, and there could have been someone Primo-ish who actually did not found Rome.

But suppose there were an “actuality function”  $A$  from possibilia to the actual-world individuals they represent. For (7)’s first conjunct to be true,  $A(\text{Primo})$  would have to be Romulus. (No one else founded Rome.) But then the second conjunct would be false. How then can (7) be true? Clearly the answer is that Primo does not once and for all represent any *one* actual thing. To make (7)’s first conjunct true, Primo represents Romulus, and for the second, the very same Primo represents Remus.

Here’s another example to illustrate the same point. Lewis allows that someone can have a counterpart in her own world; this lets him say things like “I could have been Fred (and everything else have been just the same)”, without postulating a duplicate world for their role reversal to take place in.<sup>6</sup> Say poor, dull Fred is David

<sup>6</sup>Lewis rules this out in his original 1968 treatment, but he allows it in Lewis 1986 (230ff). Section 5 takes this up further.

Lewis’s counterpart in the actual world. Thus  $\diamond(\text{DL is dull})$  is true, and in making it true, Fred represents DL. But in making true  $\diamond(\text{Fred is dull})$ , Fred represents *himself*. In general, the counterpart theorist must recognize that *representation is variable*: there is no single actual thing that is represented by a given possible individual (even fixing a particular counterpart relation). Possible things do represent actual things, but which things they represent is not fixed once and for all.

A counterpart-theoretic treatment of “actually” requires an account of how representation relations vary. The first task is to rewrite Lewis’s interpretation of the modal language without “actually” in a way that brings the representation relation to the surface. Once this is done, extending the interpretation to actuality is straightforward.

Here’s the basic idea: in the interpretations I already presented, we keep track of an evaluation world, which we shift whenever we interpret a  $\diamond$ -clause. In the new system, we also keep track of a shiftable **representation relation**, which pairs representatives in the evaluation world with what (if anything) they represent in the actual world. Whenever we shift the evaluation world and choose new counterparts at the new world, we also shift the representation relation accordingly. Here’s a picture: we start by tying strings to the individuals of the actual world; as we evaluate possibilities, we carry the ends of those strings from world to world, attaching them to the counterparts we find at each stage. When we then wonder what is *actually* true of certain things, we trace back along those strings to their anchors in the actual world.

Before I can make this precise, I need to address a point about counterparts that I put off from [Section 1](#). So far I’ve only said how to interpret modal formulas of a single variable, and it isn’t obvious how to generalize this. In Lewis’s original 1968 treatment, he told us to find a counterpart for each thing individually:  $\diamond Rxy$  is true of  $a_1$  and  $a_2$  iff there is a possible world  $w$  in which  $a_1$  has a counterpart  $b_1$  and  $a_2$  has a counterpart  $b_2$  and  $Rxy$  is true of  $b_1$  and  $b_2$  at  $w$ . That is:

$$\begin{aligned} \llbracket \diamond \phi \rrbracket^{v, a_1, \dots, a_n} \text{ iff there are } b_1, \dots, b_n \text{ in some world } w \text{ such that} & \quad (\text{Old}) \\ b_1 \text{ is a counterpart of } a_1 \text{ and } \dots & \\ \text{and } b_n \text{ is a counterpart of } a_n \text{ and } \llbracket \phi \rrbracket^{w, b_1, \dots, b_n} & \end{aligned}$$

But this proposal overlooks the fact that sometimes the fates of different individuals are linked.<sup>7</sup> For example, suppose a certain essentialist claim is correct: Chelsea couldn’t have had any mother but Hillary. But suppose also there is a symmetric possible world which contains two duplicate mother-daughter pairs, each very like Hillary and Chelsea, and so each of the daughters,  $\text{daughter}_1$  and  $\text{daughter}_2$ , is a counterpart of Chelsea, and each of the mothers,  $\text{mother}_1$  and  $\text{mother}_2$ , is a counterpart of Hillary. Then since in particular  $\text{daughter}_1$  is Chelsea’s counterpart and

<sup>7</sup>Hazen (1979) made this point (328–9), and Lewis (1986) accepted it with Hazen’s suggested solution (233–4). See also Dorr n.d.

her worldmate  $\text{mother}_2$  is Hillary's counterpart, according to Lewis's original instructions (Old),  $\diamond(\text{Chelsea's mother is not Hillary})$  comes out true, which contradicts the essentialist claim.

The solution: in cases of joint possibility we also have *joint counterparts*. Don't look for a counterpart of Hillary and a counterpart of Chelsea *independently*: look for a *pair* of things which are collectively counterparts of the pair of Hillary and Chelsea. In our example,  $\langle \text{mother}_1, \text{daughter}_1 \rangle$  and  $\langle \text{mother}_2, \text{daughter}_2 \rangle$  are two such counterpart pairs, but the crossed pair  $\langle \text{mother}_2, \text{daughter}_1 \rangle$  is not. So we can respect the essentialist claim.

In general, instead of a two-place counterpart relation, we'll have a kind of counterpart relation which tells us which things in a world  $w$  *collectively* present a possibility for the things in a world  $v$ . There is more than one way to implement this idea precisely; here is one. Our basic notion is of a **counterpart-link** between the inhabitants of a world  $w$  and those of another world  $v$ . Each  $w$ - $v$  link is a two-place relation  $S$  such that if  $bSa$  then  $a$  inhabits  $v$  and  $b$  inhabits  $w$ . We say that  $b_1, \dots, b_n$  are **jointly counterparts** of  $a_1, \dots, a_n$  iff there is a link  $S$  such that  $b_1Sa_1$  and ... and  $b_nSa_n$ . The old two-place counterpart relation is a special case:  $b$  is a counterpart of  $a$  iff  $b$  bears some link to  $a$ . And we say:

$$\begin{aligned} \llbracket \diamond \phi \rrbracket^{v, a_1, \dots, a_n} \text{ iff there are } b_1, \dots, b_n \text{ in some world } w \text{ such that} & \quad \text{(Joint)} \\ b_1, \dots, b_n \text{ are jointly counterparts of } a_1, \dots, a_n, & \\ \text{and } \llbracket \phi \rrbracket^{w, b_1, \dots, b_n} & \end{aligned}$$

In this version of counterpart theory, what does the job of the counterpart relation is a family of relations which tell us which things collectively present possibilities for other things.<sup>8</sup>

<sup>8</sup>Why not take the joint counterpart relation as basic, instead of these links? This is possible, but there are some complications. One thought is that the basic machinery should be a family of  $2n$ -place counterpart relations:  $b_1, \dots, b_n$  are jointly counterparts of  $a_1, \dots, a_n$ . To get sensible results we'll need to put some conditions on permutations and contractions of the sequences of arguments, which adds some complication. But the deeper problem is that—at least once we consider actuality—I don't think this will easily handle possibilities for infinitely many individuals. For example, supposing there are infinitely many actual things, we may want to say:

Possibly: only actual things exist, everything has a cause, nothing is its own cause,  
and causing is transitive.

(This is easy to regiment in the language of quantified modal logic with actuality.) But any merely finite choice of counterparts in such a world will leave out some things, not linking them up to any actual things. (You might try to deal with this by adjusting the interpretations of the quantifiers, so each quantifier tries to *extend* the operative link. I suspect that something along these lines may work, but I haven't worked out the details.)

On the other hand, we could use an infinitary relation:  $a_1, a_2, \dots$  are jointly counterparts of  $b_1, b_2, \dots$ . This should work, but it isn't totally straightforward. There are some technical worries

I should pause to say something about the heavy machinery: this way of interpreting possibility claims goes beyond the resources of the simple first-order possibilist theory which Lewis uses in his original 1968 treatment, in that it quantifies over *relations*. This could be understood in terms of primitive higher-order quantification, or plural quantification (over pairs), or (perhaps setting aside potential size issues) first-order quantification over something set-like. For what it's worth, *Lewis* would happily take on board some member of this family. In his 1986 presentation he freely appeals to classes, sequences, etc. But other counterpart theorists might balk. Maybe the relation-quantifiers can be dispensed with, by replacing them with finite sequences of individual quantifiers.<sup>9</sup> I suspect that this is possible, but there are some tricky details. So the counterpart theorist might really be stuck with relation quantifiers in the possibilist theory used to interpret possibility claims. Buyer beware.

Facts about modal logic put constraints on what links are like. For instance, since (necessarily) things could be the way they are, links should obey a corresponding condition: for each world  $w$ , there is a  $w$ - $w$  link which is the identity relation on  $w$ 's inhabitants.

Another logical issue about joint possibility is the contingency of identity. Is distinctness possible for an identity pair  $\langle a, a \rangle$ ? Is being identical possible for a pair of distinct things  $a_1$  and  $a_2$  (see Lewis 1986, 259)? Using (Joint) these questions become: if  $b_1$  and  $b_2$  are jointly counterparts of  $a_1$  and  $a_2$ , is it ever the case that  $a_1 = a_2$  but  $b_1 \neq b_2$ , or conversely that  $a_1 \neq a_2$  but  $b_1 = b_2$ ? Or in other words, are there any links which are not one-to-one relations?<sup>10</sup> In short, the **necessity of identity and distinctness**,

$$\Box \forall x \forall y (\Diamond x = y \rightarrow \Box x = y) \quad (\text{NID})$$

is (using our modal interpretation) equivalent to the claim that every link is one-to-one.

Note that this claim about joint counterparts is compatible with one thing having *multiple counterparts* in the same world (as opposed to Lewis 1968, 124). For instance, in the case of Chelsea and Hillary there is more than one link between the symmetric

---

about the cardinality of this relation's sequences of arguments, and of course again we would have to impose conditions on permutations and contractions of the sequence of counterparts to get sensible results. (Furthermore, if you didn't like the link version because it quantifies over relations, then this version which involves quantification over infinite sequences probably isn't much help.)

I think the "link" formulation is the most straightforward way to go. It has the extra benefit of fitting neatly into an account of actuality. Note also that if we took joint counterparts as basic, we could still define the links. (If  $b_1, b_2, \dots$  are jointly counterparts of  $a_1, a_2, \dots$ , then the relation which holds just between each  $b_i$  and  $a_i$  is a link, and these are all of them.) So my account could still proceed just as it does.

<sup>9</sup>Kit Fine suggested this to me. See the preceding footnote.

<sup>10</sup>Just to be clear, a relation  $R$  on  $X \times Y$  is **one-to-one** iff: whenever  $bRa$  and  $b'Ra'$ ,  $a = a'$  iff  $b = b'$ . On my use, it does *not* imply that that every member of  $X$  bears  $R$  to something, nor likewise for  $Y$ .

world with two mother-daughter pairs and the actual world—more than one way of choosing things that present a possibility for Chelsea and Hillary. In that case, Chelsea has multiple counterparts in one world. But even though  $daughter_1$  and  $daughter_2$  are *each* counterparts of Chelsea, they need not be *jointly* counterparts of the identity pair  $\langle Chelsea, Chelsea \rangle$ . Each link between the symmetric world and ours may well be one-to-one, never simultaneously choosing different counterparts for the same thing, or the same counterpart for different things. So the doctrine of (NID) is compatible with cases of multiple counterparts. It also allows cases like Remus and Romulus, or like David Lewis and poor Fred. In particular, it is perfectly consistent with the idea that counterparthood is a qualitative matter, and that there are qualitatively symmetric worlds. It is only under Lewis’s old treatment, which reduces joint possibilities to single-thing possibilities, that the distinct questions of multiple counterparts and contingent identity collapse together.

It’s also worth noting that affirming (NID) does not close off the counterpart theorist’s solution to the problems of contingent coincidence (See Lewis 1971; Gibbard 1975; Lewis 1986, 256; Fara n.d.; Sider n.d., 27; Dorr n.d.). For example: a certain statue is identical to the lump of clay it is made of; but if the statue had been headless, both the statue and the lump would still have existed, and yet they would not in that case have been identical. (The statue would have been a different, smaller lump.) It might seem that (NID) forbids this. I don’t think it does, though, if properly understood. In particular, we should affirm (NID) only if there is no equivocation on the sense of possibility involved. (The most banal logical truths will turn out false if we let you change the sense of your words mid-sentence.) But a plausible thing to say about statue-lump cases is that there are two different modalities involved: possibility-for-statues and possibility-for-lumps. We explain these in terms of different counterpart relations, (that is) different kinds of counterpart-links. What we want to say is that a single object could have been statuewise a different object than it would in that case have been lumpwise. But this claim is *compatible* with (NID) on any unequivocal interpretation. What (NID) rules out is that anything could have been distinct from what it then would have been, where “could” and “would” are used *in exactly the same sense*.

In what follows, I’m going to take a stand and assume that (NID) is true, and thus that links are one-to-one. This is for two reasons. First, I suspect the doctrine is correct, for Kripke’s reason: since  $a$  is necessarily  $a$ , if  $b$  is  $a$  then by Leibniz’s Law  $b$  is necessarily  $a$  as well. Second, the assumption makes what follows technically much simpler, both when it comes to stating the interpretation of actuality claims, and when it comes to proving logical facts about them. This assumption is, of course, controversial. I am confident that my account of actuality can be extended to apply even without this assumption, but for now you just get a promissory note: future work. If you believe in contingent identity, you can take what follows as a provisional theory, relying on an unrealistic but useful simplifying assumption.

Now that I've addressed these preliminaries about joint possibilities, let's return to the main issue. We want to rewrite the interpretation of possibility clauses so that we explicitly keep track of which actual things are represented by the possible individuals under consideration. This just takes a little bit of massaging. We do it by inductively defining when a modal formula  $\phi$  is **true of** individuals  $a_1, \dots, a_n$  **at** a world  $v$  **with** a representation relation  $R$ . For short:  $\llbracket \phi \rrbracket^{v, R, a_1, \dots, a_n}$ . A representation relation is any two-place, one-to-one relation between things in a world  $v$  and things in the actual world  $w_{@}$ . It gives us strings reaching back to things in the actual world, which we can use to find our way home to them when we need to.

The non-modal clauses are the same as before, with the new arguments tacked on. For example:

$$\llbracket \neg \phi \rrbracket^{v, R, a_1, \dots, a_n} \text{ iff not } \llbracket \phi \rrbracket^{v, R, a_1, \dots, a_n}$$

The possibility clause needs to be slightly rewritten. A point of notation:  $S \circ R$  denotes the composition of  $S$  and  $R$ :  $c(S \circ R)a$  iff for some  $b$ ,  $cSb$  and  $bRa$ .<sup>11</sup> Then:

$$\begin{aligned} \llbracket \diamond \phi \rrbracket^{v, R, a_1, \dots, a_n} \text{ iff there are } b_1, \dots, b_n \text{ in some world } w \\ \text{and there is some } w\text{-}v \text{ link } S \text{ such that} \\ b_1Sa_1 \text{ and } \dots \text{ and } b_nSa_n, \text{ and } \llbracket \phi \rrbracket^{v, S \circ R, b_1, \dots, b_n} \end{aligned}$$

This clause says the same thing as (**Joint**): it yields exactly the same modal truths. The only difference is that we explicitly keep track of *which* counterpart-link we used to get from the  $v$ -things to the  $w$ -things. This makes it possible to retrace our steps. We compose the link with the old representation relation because the counterpart we choose for  $a$  represents the same actual thing (if any) that  $a$  does.

As in [Section 1](#), we can use this technical notion of truth of ... at ... with ... to define truth simpliciter: a modal sentence  $\phi$  is true iff it is true of the empty sequence of things, at the actual world, with the relation  $\text{Id}_{@}$ , which is the relation that holds between each actual thing and itself. (We can also treat free variables as names as before.)

And at last, the interpretation of actuality clauses:

$$\llbracket @ \phi \rrbracket^{v, R, a_1, \dots, a_n} \text{ iff there are individuals } b_1, \dots, b_n \text{ in } w_{@} \text{ such that} \\ a_1Rb_1 \text{ and } \dots \text{ and } a_nRb_n, \text{ and } \llbracket \phi \rrbracket^{w_{@}, \text{Id}_{@}, b_1, \dots, b_n}$$

What is actually true of some things is what is true of the actual things they represent. The relation  $R$  tethers actual things to their possible representatives. An actuality operator follows back along  $R$  from each possible object to what it represents, the object at the beginning of its string. Note that here we write  $a_iRb_i$ , whereas in the  $\diamond$ -clause we wrote  $b_iSa_i$ : the order of the arguments is reversed, and thus the representing

<sup>11</sup>Note that some common conventions put  $R$  and  $S$  in the opposite order from my notation.

done by  $\diamond$ -clauses is undone by the  $@$ -clause. We also reset the evaluation world to  $w_{@}$  and tie off new strings, so each actual object now represents itself.<sup>12</sup>

## 4 The problems solved

Let's check how this solves the problems of actuality. Consider first Hazen's sentence:

$$\diamond \exists x (@ Ex \wedge \neg @ Fx \wedge \neg @ \neg Fx) \quad (2)$$

The diamond posits a world  $w$  and a link  $S$ . (We don't have to worry about new individuals, since we are considering whether a closed formula is true of zero individuals.) The representation relation is updated to  $S \circ \text{Id}_{@} = S$ . The quantifier introduces an individual  $a$  in  $w$ , so the interpretation so far is:

For some world  $w$  and some  $w$ - $w_{@}$  link  $S$ , there is some  $a$  in  $w$  such  
that  $\llbracket @ Ex \wedge \neg @ Fx \wedge \neg @ \neg Fx \rrbracket^{w, S, a}$

Take the three conjuncts one at a time. The first conjunct says that there is an individual  $b_0$  in the actual world such that  $aSb_0$ . The second conjunct says that there is no individual  $b$  such that  $aSb$  and  $Fb$ . So it implies in particular that  $b_0$  is not  $F$ . The final conjunct says that there is no  $b$  such that  $aSb$  and not- $Fb$ . So it implies that  $b_0$  is  $F$ . But this means (2) is contradictory—as it should be.

Of course, the naïve translation that treats “actually” as an existential quantifier over counterparts also gets (2) right. The real test is whether the translation also renders its partner sentence contradictory:

$$\diamond \exists x (@ Ex \wedge @ Fx \wedge @ \neg Fx) \quad (3)$$

In Possibile, (3) says that there is a world  $w$ , a  $w$ - $w_{@}$  link  $S$ , and a  $w$ -inhabitant  $a$ , such that (i) there is some  $b_1$  such that  $aSb_1$ ; (ii) there is some  $b_2$  such that  $aSb_2$  and  $Fb_2$ ; and (iii) there is some  $b_3$  such that  $aSb_3$  and not- $Fb_3$ . But  $S$  is one-to-one, so  $b_1 = b_2 = b_3$ , and this individual must be both  $F$  and not- $F$ —a contradiction again. The problem of multiple counterparts is solved.

<sup>12</sup>Hazen originally proposed something very similar in spirit (1979, 333ff): for simple cases, he uses a set of “representative functions”, which are one-to-one partial functions from the actual world's domain to the domains of other worlds—the same kind of thing as my representation relations. The main difference between his version and mine is that he takes these functions as basic machinery *replacing* counterparts, whereas in my version representation relations are built up “on the fly” out of (joint) counterparts. Furthermore, when he generalizes his version to handle iterated modalities, the representative functions he takes as basic turn out to be complicated objects that link up the objects from the domains of arbitrary finite sets of worlds.

We also solve the problem of the wrong counterpart. Recall:

$$\exists x(Fx \wedge \diamond @ \neg Fx) \quad (4)$$

The first step of the interpretation says: there is some actual  $a$  which is  $F$ , and  $\diamond @ \neg Fx$  is true of  $a$ . The second clause is interpreted to say: there is a world  $w$ , a  $w$ - $w_@$  link  $S$ , and a  $w$ -inhabitant  $b$  such that  $bSa$ , and  $\llbracket @ \neg Fx \rrbracket^{w, S, b}$ . And this last bit says: there is an individual  $c$  such that  $bSc$  and not- $Fc$ . Again, since  $S$  is one-to-one,  $a = c$ , and again one individual is both  $F$  and not- $F$ . So (4) too is contradictory.

The problem of *no* counterpart isn't quite so nice. I'll lead up to it with a general observation. The “actually” clause existentially quantifies over actual things. If something mentioned in the scope of an actuality operator does not represent *any* actual thing—its string dangles loose—then the actuality claim comes out false. This isn't clearly wrong: should we positively say such-and-such is actually the case concerning a non-actual thing? But this position has some awkward consequences.

For instance,

$$\Box \forall x(\neg @ \phi(x) \leftrightarrow @ \neg \phi(x))$$

is not a logical truth:  $@$  does not commute with negation.  $@ \phi(x)$  is false of any individual that does not represent anything actual, and so  $\neg @ \phi(x)$  is true of such an individual. But  $@ \neg \phi(x)$  is also false of such a thing, and so the biconditional fails. On the other hand, we do get a logical truth by adding an existence condition:

$$\Box \forall x(@ Ex \rightarrow (\neg @ \phi(x) \leftrightarrow @ \neg \phi(x)))$$

As a special case, we must be careful when expressing actual non-existence, since the clause  $@ \neg Ex$  is guaranteed to be false of any possible thing: instead we should say  $\neg @ Ex$ .

Though it is awkward, this take on non-existents fits with the general Lewisian approach.<sup>13</sup> The Lewisian interpretation of the modal operators incorporates a kind of “very serious actualism”. The **serious actualist** holds that an object can have no *properties* in a world where it does not exist—but it can still satisfy “conditions” such

<sup>13</sup> That is, the approach of 1968, particularly p. 119. Lewis's endorsement of this way of doing things in 1986 is mixed (see p. 10) but he doesn't offer a precise alternative. At one point he proposes using “gappy sequences” to represent the possibility of non-existence (p. 233), and I think something along those lines is the counterpart theorist's best bet for dealing with contingent existence; but there are tricky details about iterated possibilities which are unaddressed. (For example: it could have been that I didn't exist, and yet in that case I still *could have* existed, though I couldn't have been, say, a fish. To get this to work out, it looks like some things but not others count as counterparts of the “gap” which represents me in that possibility.) So while it's a promising thought, and while the old approach leaves much to be desired, the details await future work.



as non-existence (Plantinga 1983, 4ff.). The **very serious actualist** goes further: *nothing whatsoever* is true of an object in worlds where it doesn't exist:

$$\phi(x) \wedge \neg Ex$$

is false of every possible thing, for any modal formula  $\phi$  whatsoever. So it is not possibly true of anything:  $\Diamond(\phi(x) \wedge \neg Ex)$  is never satisfied; and so the very serious actualist affirms

$$\Box \forall x \neg \Diamond(\phi(x) \wedge \neg Ex)$$

It is impossible that something might  $\phi$  without existing.

As an immediate consequence,  $\Box \forall x \Box Ex$  is a logical truth. Everything exists necessarily, according to the Lewisian understanding of “necessarily”. Effectively, the Lewisian modal box really means “essentially”—that is, necessity given the existence of each thing in question. Understanding it this way commits the Lewisian to a zero-tolerance stance toward non-entities under modal operators.

If we take this stance on possibility, even less should we tolerate non-existents under the actuality operator. The very serious actualist should hold that

$$@(\phi(x) \wedge \neg Ex)$$

is false of every possible thing, and so should affirm

$$\Box \forall x \neg @(\phi(x) \wedge \neg Ex)$$

A consequence of this principle (with the principle that contradictions aren't actually true and substitution of logical equivalents) is this:

$$\Box \forall x \neg @ \neg Ex$$

Necessarily, nothing actually fails to exist.

For (just as in the previous case), if something is not actual, then nothing is actually true of it—not even non-existence.

Here's another consequence. Since in general the very serious actualist takes  $@\phi(x)$  to be equivalent to  $@(\phi(x) \wedge Ex)$ , in particular

$$\neg @Fx \text{ is equivalent to } \neg @(Fx \wedge Ex)$$

$$@\neg Fx \text{ is equivalent to } @(\neg Fx \wedge Ex)$$

But these are not equivalent to each other. (Note the scope of the negation.) In fact, they come apart whenever  $@Ex$  fails. So this is true of each possible thing:

$$\neg @Ex \rightarrow (@Fx \leftrightarrow @\neg Fx)$$

So this is true:

$$\diamond \exists x \neg @ Ex \rightarrow \diamond \exists x (@ Fx \leftrightarrow @ \neg Fx)$$

Now recall the problem of no counterparts: Fara and Williamson claim that

$$\diamond \exists x (@ Fx \leftrightarrow @ \neg Fx) \tag{5}$$

is contradictory. But the very serious actualist will affirm (5), as long as there could be something that doesn't actually exist—for in that case,  $@ Fx$  and  $@ \neg Fx$  will both be false! So (5) is not *so* clearly bad: it follows naturally from a commitment to very serious actualism, together with the possibility of non-actuals. Fara and Williamson defend their rejection of (5) by appealing to “the fact that [ $@$ ] commutes with every truth-functional connective” (Fara and Williamson 2005, 17); but this kind of very serious actualist should deny this “fact” in general. There might have been a gold mountain, though that mountain does not actually exist, whether as an otherwise-constituted mountain or anything else. Is it actually in Texas? Certainly not. Then is it actually *not* in Texas? Again, no, for it is neither elsewhere nor non-spatial. Of course, it is not the case that it is actually in Texas, but that is another thing altogether. Thus the possible gold mountain is actually in Texas if and only if it is actually not in Texas—for it is neither. The very serious actualist should conclude that Fara and Williamson's sentence is not contradictory after all.

I don't seriously mean to defend very serious actualism; indeed, I think it is a commitment that counterpart theory should be freed of. My point is that it is not a commitment peculiar to an account of *actuality*, but rather a thoroughgoing doctrine of the Lewisian approach to contingent existence. Contingent existence is a problem for the counterpart theorist; it was a problem before the actuality problems were raised, and it remains a problem after the actuality problems are dispelled. But it is a separate problem from the problem of actuality, and thus a project for a different essay. (But see note 13.)

This sort of very serious actualism does not entail the claim in the *possibilist* language that everything is actual—i.e., that every possible individual inhabits  $w_@$ , or that every possible individual represents something in  $w_@$ . The Lewisian will generally accept neither of these claims. This simultaneous acceptance of the *modal* doctrine of very serious actualism and rejection of the *possibilist* doctrine that everything represents something actual is what gives rise to such logical oddities as the discrepancy between  $\neg @ \phi$  and  $@ \neg \phi$ . Similarly, when speaking modally the Lewisian says that everything exists necessarily, but when speaking possibilistically he says there are worlds in which some actual things have no counterparts: loosely, some things exist contingently. There is an intuitive mismatch between the modal and possibilist commitments; this mismatch dooms Lewisian modal logic to some complications.

To escape this—at least as it applies to actuality—you might try alternative interpretations that replace the @-clause’s existential quantification with something else. *Universal* quantification is no improvement. That amounts to making any actuality claim about a non-entity *true*—so something that doesn’t actually exist is both actually happy *and* actually not-happy. In particular, Fara and Williamson’s sentence (5) is still satisfied. A better option is to use definite descriptions. The variant clause is this:

$$\llbracket @\phi \rrbracket^{v, R, a_1, \dots, a_n} \text{ iff for the individuals } b_1, \dots, b_n \text{ in } w_{@} \text{ such that} \\ a_1 R b_1 \text{ and } \dots \text{ and } a_n R b_n, \llbracket \phi \rrbracket^{w_{@}, \text{Id}_{@}, b_1, \dots, b_n}$$

The consequences of this approach hinge on how we unpack the definite description. If we treat them as Russellian “incomplete symbols” taking scope *in situ*, then the new clause is equivalent to the old. If the descriptions are not Russellian, though, we have to do something about non-denoting terms. Maybe Possibilese is governed by some species of free logic; or maybe the definite description is analyzed some other way (maybe as quantification over “gappy sequences” as in note 13). The status of Fara and Williamson’s (5) will vary with the details. For instance, if we say every *atomic* formula that includes a non-denoting term is false, (5) indeed comes out contradictory. But I won’t build this idea into my official system; I’ll wait for a more uniform solution to the counterpart theorist’s problems of contingent existence.

I’ve given an account of the actuality operator in terms of counterparts. It does pretty well, but it’s limited in a few important ways. First, relation quantifiers: I rely on some heavier possibilist machinery than Lewis’s original system. Second, contingent identity: I’ve ruled it out, which some will find disappointing. (But, let me say again, this is not to rule out multiplicity of counterparts.) Third, contingent existence: I didn’t exactly rule this out, but we get a deviant logic for such cases. Each of these calls for further technical improvements to the system here, and in each case I think there are some promising lines to pursue. But despite these limitations, at this point we have something good enough to be useful. So next I’ll put it to work on some puzzle cases.

## 5 Some related problems and their solutions

Our new understanding of actuality helps us reply to a general challenge that Fara and Williamson raise against the counterpart theorist (23). Consider a possible object Penny such that two actual things, one Heads and the other Tails, are equally good qualitative matches for Penny. Fara and Williamson claim that in this case the counterpart theorist must affirm

$$@H(\text{Penny}) \leftrightarrow @T(\text{Penny}) \tag{8}$$

For as far as counterparts go, being Heads and being Tails are completely on a par for Penny. We can also suppose that being Heads and being Tails are mutually exclusive and exhaust Penny’s options:

$$\neg(@H(\text{Penny}) \wedge @T(\text{Penny})) \quad (9)$$

$$@H(\text{Penny}) \vee @T(\text{Penny}) \quad (10)$$

All this ought to be compatible with Penny actually existing, so we affirm  $@E(\text{Penny})$  as well, to avoid very serious actualist complications. But in this case,  $@$  commutes with the connectives, and so (8), (9), and (10) are jointly inconsistent. So it seems the counterpart theorist has a quite general problem: “therefore, we have in  $[(8)\wedge(9)\wedge(10)]$  a *de re* modal sentence that should not be understood in terms of counterparts” (Fara and Williamson 2005, 28).

The misstep is at the beginning: the right counterpart-theoretic account of actuality renders (8) false. Fara and Williamson reason by symmetry that if two worldmates are equally good qualitative matches for Penny, then she must represent both of them, if either. This is a plausible thought, given that the counterpart relation fixes what actual thing Penny represents—which *prima facie* seems like what the counterpart theorist must say. But as we have seen, what Penny represents is *not* fixed by the counterpart relation; it is *variable* (though its range of variation is determined by the counterpart relation). So Penny can represent Heads, and also represent Tails, without thereby being coerced into representing both of them at once.

There are two ways to make this response precise, corresponding to two ways of disentangling the possibilist name “Penny” from the modal actuality operator in Fara and Williamson’s argument. The first way is semantic: speaking Possibilese, we might take  $\phi(\text{Penny})$  to mean that the modal formula  $\phi(x)$  is true of Penny (at Penny’s world—but since Penny is world-bound we can elide this). But there isn’t enough information here to settle the question of whether  $@Hx$  or  $@Tx$  is true of Penny. It depends on *which actual thing Penny represents*, and this is not fixed once and for all. Penny can represent Heads, and thus satisfy  $@Hx$ , and Penny can represent Tails, and thus satisfy  $@Tx$ —but there is *no* representation relation with which Penny represents *both* Heads *and* Tails, and thus no representation makes the biconditional (8) true of Penny.

The second way is to introduce a predicate “Pennyish” that picks out Penny uniquely among all possible objects. Then we can rewrite (8) (and the rest of the argument) in Modalese:

$$\diamond \exists x (Px \wedge (@Hx \leftrightarrow @Tx)) \quad (8')$$

But once again, symmetry considerations do not compel the counterpart theorist to accept this sentence. Rather, those considerations recommend this:

$$\diamond \exists x (Px \wedge @Hx) \leftrightarrow \diamond \exists x (Px \wedge @Tx) \quad (11)$$

For indeed, if there is a possibility wherein a Penny (i.e., something Pennyish) represents Heads, then by symmetry there should just as well be a possibility wherein a Penny represents Tails. The critical point is that (8') does not follow from (11). Take the positive half of the biconditional (11):<sup>14</sup>

$$\diamond \exists x(Px \wedge @ Hx) \wedge \diamond \exists x(Px \wedge @ Tx) \quad (12)$$

A Penny could be actually Heads, and a Penny could be actually Tails.

Even though we know that there is just one possible Penny, we must not conclude from (12)

$$\diamond \exists x(Px \wedge @ Hx \wedge @ Tx)$$

A Penny could be actually Heads *and* actually Tails.

For Penny, a single possible individual, nonetheless presents two possibilities: one of being actually Heads, another of being actually Tails. These are two *different* possibilities, not a single possibility in which she is *both*.

This story is in the same spirit as Lewis's account of haecceitistic possibilities:

Here am I, there goes poor Fred; there but for the grace of God go I; how lucky I am to be me, not him. Where there is luck there must be contingency. I am contemplating the possibility of being poor Fred in a world just like this one. ... I suggest that the possibility I have in mind is not a world that is like ours qualitatively but differs from ours haecceitistically. Instead it is a possible individual, in fact an actual individual, namely poor Fred himself (Lewis 1986, 231).

A single way for a world to be may capture two ways for David Lewis to be: he could be the way he actually is, or the way poor Fred is. Say "It's actualish" is a sentence which describes the actual world and no other—this could be a complete qualitative description of the way things actually are. And say "Freddish" is a complete qualitative description of Fred. We can affirm the haecceitistic possibility like this:

$$\diamond(\text{It's actualish} \wedge \text{DL is Freddish}) \wedge \diamond(\text{It's actualish} \wedge \text{DL is not Freddish}) \quad (13)$$

(This is analogous to (12).) And yet we cannot conclude

$$\diamond(\text{It's actualish} \wedge \text{DL is Freddish} \wedge \text{DL is not Freddish}) \quad (14)$$

Explaining this, Lewis says: "*Possibilities are not always possible worlds*. ... I say that *any* possible individual is a possibility" (230). The diamond may range twice over the same

<sup>14</sup>That is, the first disjunct of  $(p \wedge q) \vee (\neg p \wedge \neg q)$ .

world, so to speak, picking out different counterparts for DL each time, thereby making (13) true but not (14).

In Penny’s case we have on our hands a single possible individual which nonetheless captures more than one possibility. Here is the possibility of a Penny which is actually Heads; here is the possibility of a Penny which is actually Tails; and here is lone Penny, carrying both possibilities on her shoulders. What we must say is that Lewis did not go far enough: possibilities are not always possible worlds, indeed, and neither are they possible individuals. Rather, a possibility includes a representation relation, tying possible individuals to whatever actual individuals they are possibilities *for*. Thus the diamond can range twice over the same individual, so to speak, picking it out now as a possibility for Heads, then as a possibility for Tails.

Delia Graff Fara (2009) raises another actuality problem, which arises when an individual has counterparts in its own world as in the case of David Lewis and poor Fred. (Fara uses a different example.) She reasons thus: Fred is a counterpart of DL “in the sense that’s relevant for determining what’s *possible* for [DL]” (8); so since Fred is actually quite dull, actually being dull is a possibility for DL: DL could have been actually dull. But what is possibly actual is actual: so we can conclude that DL is actually dull—which is clearly false. What went wrong?

The key premise of Fara’s argument is this

If a counterpart of DL is actually dull, then it is possible for DL to be  
actually dull. (15)

This certainly looks plausible: as Fara points out, DL’s counterparts are supposed to *determine what’s possible* for DL. And usually this kind of principle works fine: Fred is dull, and Fred is a counterpart of DL, so it is possible for DL to be dull. What goes wrong in the case of actuality?

Principle (15) is delicate: it combines possibilist talk about counterparts with modal statements using the actuality operator. Let’s disentangle them; again, one way to do this is semantic. We can unpack (15) into these three principles:

If Fred is actually dull, then  $@(x \text{ is dull})$  is true of Fred. (16a)

If  $@(x \text{ is dull})$  is true of a counterpart of DL, namely Fred, then  
 $\diamond @(x \text{ is dull})$  is true of DL. (16b)

If  $\diamond @(x \text{ is dull})$  is true of DL, then it is possible for DL to be  
actually dull. (16c)

But (again) something is missing from each of these claims. Whether Fred is actually a certain way (whether an actuality claim is true of Fred) is a matter of *which actual thing Fred represents*. This is not settled just by which things are counterparts of which. Remember what we observed in Section 3: Fred accounts for the possibility of DL

being dull—in so doing he represents DL. And Fred also accounts for the possibility of *Fred* being dull—then Fred represents *himself*. To settle the actuality claims we need to say which thing Fred represents. The claims (16a–c) are indeterminate, because of a missing parameter: the representation relation. Each claim is correct for *some* way of filling this parameter in; but the correct ways to fill them in are different from one claim to the next.

Start with (16a). Fred is actually dull: this is a statement about how things are—that is, how things *actually* are—that is, how things are if everything is just as it actually is. So it is true iff @(*x* is dull) is true of Fred *with every actual thing representing itself*. This is the correct way of filling in (16a):

If Fred is actually dull, then @(*x* is dull) is true of Fred *with Fred* (16a\*)  
*representing Fred*.

For the same reason, this is the correct way of filling in (16c):

If  $\diamond$ @(*x* is dull) is true of DL *with DL representing DL* then it is (16c\*)  
possible for DL to be actually dull.

But what about the step between them? The correct form is a special case of our clause for interpreting possibility statements:

$\diamond\phi(x)$  is true of *a*, with *a* representing *c*, iff  $\phi(x)$  is true of some  
counterpart of *a*, namely *b*, with *b* representing *c*.

When a counterpart presents a possibility for *a*, it represents whatever actual object *a* does (if any). So in particular, this is the correct way of filling in (16b):

If @(*x* is dull) is true of a counterpart of DL, namely Fred, *with Fred* (16b\*)  
*representing DL*, then  $\diamond$ @(*x* is dull) is true of DL *with DL representing*  
*DL*.

But what the argument requires to support Fara’s premise (15) is this:

If @(*x* is dull) is true of a counterpart of DL, namely Fred, *with Fred*  
*representing Fred*, then  $\diamond$ @(*x* is dull) is true of DL with DL  
representing DL.

And this conditional just isn’t true. Even though Fred is a counterpart of DL, when actuality is involved *there is more to being a possibility than just being a counterpart*. The counterpart also has to represent the right actual thing. Since Fred is a counterpart of DL, it follows that Fred *representing DL* is a possibility for DL (representing himself). But Fred *representing Fred* is *not a possibility for DL* (representing himself). DL’s counterparts do determine what’s possible for DL; but this doesn’t mean that whatever is true of his

counterparts is possible for him. What is possible for DL is what is true of his counterparts when they are doing the job of representing what *he* represents—in this case, DL himself.

We can semantically descend to put the point another way: unlike the case of “dull” (or even “possibly dull”) there is just no unequivocal fact of the matter whether a counterpart is *actually dull*.<sup>15</sup> This follows from the central observation of this essay: a single possible object can represent many different actual objects, some dull and others not. Whether it is actually dull depends on which one it represents. This goes for actual counterparts just as much as any others.

We can also try a different way of disentangling counterpart-talk from modal-talk, getting a different way of unpacking Fara’s argument. If Fred is the unique Freddish thing among all possible objects, the following principle is true:

If something Freddish is dull, then necessarily everything Freddish is  
dull. (17)

There is just one way to be Freddish, and it is a dull way. Furthermore, since DL has a Freddish counterpart, DL could have been Freddish. So, since Fred is Freddish and dull, DL could have been dull. This argument is sound. But the principle analogous to (17) breaks down when it comes to “actually dull”:

If something Freddish is actually dull, then necessarily everything  
Freddish is actually dull.

If this were correct, then (since DL could have been Freddish and Fred is Freddish and actually dull) the same reasoning would imply that DL could have been actually dull. As in Fara’s argument, what could have been actual is actual, so this would imply that DL is actually dull, which is wrong.

If anything were qualitatively just as Fred is, it would have to be dull; there is no other way for a Freddish thing to be. But there is more than one way a Freddish thing could *actually* be. Although the Fred we know is actually dull, there could have been someone Freddish who is actually DL, and not dull at all. A possibility’s qualitative character does not fix how it represents actuality.

This comes down to the same point as before: even if qualitative features suffice to pick out a single *possible individual*, they do not suffice to pick out a single *possibility*. Even though Fred is the only Freddish possible individual, Fred shoulders more than one possibility. The possibilities are qualitatively alike, but they differ in how they represent actuality: in one, the Freddish person is actually Fred; in another, the Freddish person

<sup>15</sup>What about the fact of being actual (i.e. inhabiting  $w_{@}$ ) and also dull? This is determinate enough, but it doesn’t fit our Modalese use of “actually dull”. For instance, we can’t conclude from “*a* could have been clever but actually dull” that *a* has a counterpart which is clever, actual, and also not dull.



is actually DL. Only one of these possibilities is *possible for DL*: namely, the possibility in which the Freddish person is actually DL.

Possibilities *de re* are relational: they are possibilities *for* this or that—possibilities that represent particular things as being a certain way. Since, for the counterpart theorist, possible worlds and possible individuals do not by themselves settle what they represent, they are unsuitable for playing the relational role of a possibility. Relations themselves must help fill that role.

## A Model theory

In the following appendices I prove several important technical results about my “representational” system of interpretations **R**. The headline results:

- In [Appendix B](#), I show that these counterpart interpretations are equivalent to a variant of the standard Kripkean modal semantics, one that builds in certain assumptions about existence—the consequences of “very serious actualism”.
- In [Appendix C](#), I present a sound and complete axiomatic characterization of Kripkean actuality systems and, with the result of [Appendix B](#), use this to characterize **R**. I go on to discuss some of its important logical features. In particular, it turns out that the logic of **R** is similar in some important ways to that of constant-domain models.
- Finally, I show how to turn the informal interpretations of the main text into a manual for syntactic translations from a quantified modal language into an extensional possibilist language, along the same lines as Lewis’s original 1968 treatment.

Before I take on these tasks, though, in this appendix I give a model-theoretic version of both Kripke and counterpart interpretations of modal logic with actuality.

First, let’s define the quantified modal language with actuality. Suppose we have a set  $\text{Var}$  of variables and a family of sets  $\text{Pred}^n$  of  $n$ -place predicate symbols. (I assume that each of these is countably infinite.) Let  $\mathcal{L}$  be the smallest set of formulas such that

- For  $x_1, \dots, x_n \in \text{Var}$  and  $F \in \text{Pred}^n$ , the atomic formulas  $Fx_1 \dots x_n$  and  $x_1 = x_2$  are in  $\mathcal{L}$ .
- For  $\phi, \psi \in \mathcal{L}$  and  $x \in \text{Var}$ ,  $\neg\phi$ ,  $(\phi \wedge \psi)$ ,  $\exists x \phi$ ,  $\diamond \phi$ , and  $@ \phi$  are in  $\mathcal{L}$ .

The additional connectives  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$ ,  $\perp$ ,  $\top$ ,  $\forall x$ , and  $\square$  are defined in the usual ways. Parentheses are dropped as usual.

Next I’ll give two different model theories for this language. First, the Kripke interpretation. A **K-model** has the following five components:

- A non-empty set  $W$ : the **worlds**.
- For each  $w \in W$ , a set  $Dw$ :  $w$ ’s **domain of individuals**. Let  $D = \bigcup_{w \in W} Dw$ .
- For each  $n$ -place predicate  $F$  and world  $w$ , a subset of  $D^n$ ,  $[F]_w$ :  $F$ ’s **extension** at  $w$ .

- An element of  $W$ ,  $w_{@}$ : the **actual world**.
- A reflexive relation on  $W \times W$ ,  $\overset{\diamond}{\rightarrow}$ : the **accessibility** relation for possibility.

A **K-point** in a given model is a sequence of a world and  $n$  individuals. We can recursively define truth at a **K-point**  $\mathbf{a} = \langle v, a_1, \dots, a_n \rangle$  as follows. (The model is implicit, as is an ordering of variables, to keep things a bit cleaner.)

$$\begin{array}{ll}
\mathbf{a} \models_{\mathbf{K}} Fx_{i_1} \dots x_{i_n} & \text{iff } \langle a_{i_1}, \dots, a_{i_n} \rangle \in [F]_v \\
\mathbf{a} \models_{\mathbf{K}} x_{i_1} = x_{i_n} & \text{iff } a_{i_1} = a_{i_n} \\
\mathbf{a} \models_{\mathbf{K}} \phi \wedge \psi & \text{iff } \mathbf{a} \models_{\mathbf{K}} \phi \text{ and } \mathbf{a} \models_{\mathbf{K}} \psi \\
\mathbf{a} \models_{\mathbf{K}} \exists x_{n+1} \phi & \text{iff for some } b \in Dv, \langle v, a_1, \dots, a_n, b \rangle \models_{\mathbf{K}} \phi \\
\mathbf{a} \models_{\mathbf{K}} \diamond \phi & \text{iff for some } w \text{ such that } v \overset{\diamond}{\rightarrow} w, \langle w, a_1, \dots, a_n \rangle \models_{\mathbf{K}} \phi \\
\mathbf{a} \models_{\mathbf{K}} @ \phi & \text{iff } \langle w_{@}, a_1, \dots, a_n \rangle \models_{\mathbf{K}} \phi
\end{array}$$

A formula  $\phi$  is (strongly) **K-valid** iff it is true at every **K-point** (which is **long enough**, in the sense that it has enough objects for all of  $\phi$ 's free variables, with respect to the implicitly given ordering) in every **K-model**. Let **K** denote the set of **K-valid** formulas.

With an extra bit of notation we can write some of our clauses in a more uniform way. Say  $v \overset{@}{\rightarrow} w$  iff  $w = w_{@}$ . We can also define accessibility relations not just for worlds, but also for evaluation points:

$$\begin{array}{l}
\langle v, a_1, \dots, a_n \rangle \overset{\diamond}{\rightarrow} \langle w, b_1, \dots, b_n \rangle \text{ iff } v \overset{\diamond}{\rightarrow} w \text{ and } a_i = b_i \text{ for each } i \\
\langle v, a_1, \dots, a_n \rangle \overset{@}{\rightarrow} \langle w, b_1, \dots, b_n \rangle \text{ iff } v \overset{@}{\rightarrow} w \text{ and } a_i = b_i \text{ for each } i
\end{array}$$

Then the following clauses are equivalent to what I wrote above:

$$\begin{array}{l}
\mathbf{a} \models_{\mathbf{K}} \diamond \phi \text{ iff there is some } \mathbf{K}\text{-point } \mathbf{b} \text{ such that } \mathbf{a} \overset{\diamond}{\rightarrow} \mathbf{b} \text{ and } \mathbf{b} \models_{\mathbf{K}} \phi \\
\mathbf{a} \models_{\mathbf{K}} @ \phi \text{ iff there is some } \mathbf{K}\text{-point } \mathbf{b} \text{ such that } \mathbf{a} \overset{@}{\rightarrow} \mathbf{b} \text{ and } \mathbf{b} \models_{\mathbf{K}} \phi
\end{array}$$

A point  $\mathbf{a}$  is **diagonal** iff  $\mathbf{a} \overset{@}{\rightarrow} \mathbf{a}$ —its world is the actual world. We say a formula is **diagonally K-valid** iff it is true at every (long enough) diagonal **K-point**. Let  $\Delta \mathbf{K}$  be the set of diagonally **K-valid** formulas. Intuitively, a strongly valid inference is acceptable even in hypothetical reasoning about alternative possibilities, whereas a diagonally valid inference is only acceptable in categorical reasoning about the way things are. So, for instance, necessitation preserves strong validity, but not diagonal validity:  $\phi \rightarrow @ \phi$  is diagonally valid, but  $\Box(\phi \rightarrow @ \phi)$  is not valid in any sense.

Next, the counterpart system. A **counterpart model** is like a **K-model**, with a few changes.

- Individuals are world-bound: if  $Dv \cap Dw$  is non-empty, then  $v = w$ .
- Predicates have extensions which are not world-relative:  $[F]$  is a subset of  $D^n$ . We can let  $[F]_w = [F] \cap Dw$  to keep notation uniform.
- Instead of an accessibility relation, there are **counterpart-links**: for each pair of worlds  $w$  and  $v$ ,  $C(w, v)$  is a set of relations on  $Dw \times Dv$ , the  $w$ - $v$  links. I assume each link is one-to-one (as discussed in Section 3). Furthermore, for each world  $v$ ,  $C(v, v)$  includes the identity relation on  $Dv$ .

An **R-point** (in a given counterpart model) is a sequence  $\langle v, R, a_1, \dots, a_n \rangle$ , where  $v$  is a world,  $a_1, \dots, a_n$  are elements of  $Dv$  (note the “very serious actualist” condition), and  $R$  is any one-to-one relation on  $Dv \times Dw_{\textcircled{a}}$ . We can define truth at an **R-point** similarly to truth at a **K-point**. Very similarly indeed, if we define relations  $\xrightarrow{\diamond}$  and  $\xrightarrow{\textcircled{a}}$  on **R-points**:

$$\begin{aligned} \langle v, R, a_1, \dots, a_n \rangle &\xrightarrow{\diamond} \langle w, R', b_1, \dots, b_n \rangle \\ &\text{iff for some link } S \in C(w, v), R' = S \circ R \text{ and } b_i S a_i \text{ for each } i \\ \langle v, R, a_1, \dots, a_n \rangle &\xrightarrow{\textcircled{a}} \langle w, R', b_1, \dots, b_n \rangle \\ &\text{iff } w = w_{\textcircled{a}}, R' = \text{Id}_{\textcircled{a}}, \text{ and } a_i R b_i \text{ for each } i \end{aligned}$$

( $\text{Id}_{\textcircled{a}}$  is the relation that holds between  $a$  and  $b$  iff  $a = b \in Dw_{\textcircled{a}}$ .) Using these definitions, the definition of **R-truth** takes exactly the same form as the definition of **K-truth** (in its second formulation), just replacing each **K** with an **R**. This matches the interpretation I gave in the main text.

A formula is (strongly) **R-valid** iff it is true at every (long enough) **R-point** in every counterpart model. We can also define diagonal **R-points** the same way as for **K**:  $\mathfrak{a}$  is diagonal iff  $\mathfrak{a} \xrightarrow{\textcircled{a}} \mathfrak{a}$ , and a diagonally **R-valid** formula is one that is true at every diagonal **R-point**. Note that diagonality now requires not just being at the actual world, but also that actual objects represent themselves. **R** denotes the set of **R** valid formulas and  $\Delta \mathbf{R}$  the set of diagonally **R-valid** formulas.

## B Representations and Kripke

In this section I sketch an equivalence between the representational semantics **R** and a version of the Kripkean semantics which builds in certain “very serious actualist” provisos about existence. The upshot is that any *logical* objection to the system I present must come down to an objection either to the Kripkean logic of actuality, or else to very

serious actualism. This equivalence has the extra benefit of allowing the application of standard completeness results (Appendix C).

The first step is to set out the actualist Kripkean semantics, which I'll call **KE**. This system uses the same models as **K**, but like the counterpart system adds an extra condition on its evaluation points: a **KE**-point is a **K**-point  $\langle v, a_1, \dots, a_n \rangle$  such that  $a_1, \dots, a_n$  are all in  $Dv$ . The truth conditions are the same as **K**, replacing each **K** with **KE** (in particular, in the  $\diamond$  and  $@$  clauses). Strong and diagonal validity are defined in the obvious way.

There is clearly a close relationship between **K** and **KE**. In fact, **KE**-logic is just what you get by thinking of modal operators as carrying implicit existence conditions, and interpreting the fully explicit formulas the ordinary Kripkean way. To spell this out precisely, the first step is to say how to translate formulas to make the implicit existence requirements explicit. For any finite set of variables  $X = \{x_1, \dots, x_n\}$ , let  $EX$  abbreviate the formula  $Ex_1 \wedge \dots \wedge Ex_n$ , and let  $\varepsilon^X$  be a translation function from  $\mathcal{L}$  to  $\mathcal{L}$  such that, for each  $\phi \in \mathcal{L}$ ,

$$\begin{aligned} \varepsilon^X(\diamond \phi) &\text{ is } \diamond(EX \wedge \varepsilon^X \phi); \\ \varepsilon^X(@ \phi) &\text{ is } @(EX \wedge \varepsilon^X \phi). \\ \varepsilon^X(\exists x \phi) &\text{ is } \exists x(\varepsilon^{X \cup \{x\}} \phi). \end{aligned}$$

The translation makes no changes to the other clauses.<sup>16</sup> If you take any validity of **KE** and translate it by adding explicit existence conditions, you get a **K**-validity. Conversely, any formula that has a **K**-valid translation is **KE**-valid.

**Theorem 1.** Let

$$E^-(\Gamma) = \{\phi \in \mathcal{L} : EX \rightarrow \varepsilon^X \phi \in \Gamma \text{ for } X \text{ including } \phi\text{'s free variables}\}$$

Then **KE** =  $E^-(\mathbf{K})$  and  $\Delta \mathbf{KE} = E^-(\Delta \mathbf{K})$ .

*Proof.* We can prove by a straightforward induction argument on the complexity of  $\phi$  that for any **KE**-point  $\mathbf{a} = \langle v, a_1, \dots, a_n \rangle$  and  $X = \{x_1, \dots, x_n\}$  (which includes all of  $\phi$ 's free variables)

$$\mathbf{a} \Vdash_{\mathbf{KE}} \phi \text{ iff } \mathbf{a} \Vdash_{\mathbf{K}} \varepsilon^X \phi.$$

Furthermore,  $\psi$  is true at every (length  $n$ ) **KE**-point iff  $EX \rightarrow \psi$  is true at every (length  $n$ ) **K**-point—and the same is true for diagonal points. The first result immediately follows.  $\square$

<sup>16</sup>That is:

$$\varepsilon^X Fx_1 \dots x_n \text{ is } Fx_1 \dots x_n \qquad \varepsilon^X(\phi \wedge \psi) \text{ is } \varepsilon^X \phi \wedge \varepsilon^X \psi \qquad \varepsilon^X(\neg \phi) \text{ is } \neg \varepsilon^X \phi$$

Next I'll spell out the relationship between **KE** and **R**, the representational counterpart system. In fact, **R** is nearly equivalent to **KE**, except that it makes one extra demand. In **R**, representation relations are built up step by step as possibilities are iterated. Actual objects don't latch onto their representatives directly; instead, a chain of intervening objects reaches through the various possible worlds that were visited along the way. This means in particular that once an actual object is lost, because it isn't attached to any representative, it is gone forever. If nothing represents actual Alice at some possibility  $\mathfrak{a}$ —that is, if Alice does not exist, according to  $\mathfrak{a}$ —then  $\mathfrak{a}$  cannot see any further possibilities for Alice. This is another side effect of very serious actualism: otherwise, we could allow Alice to be *represented* by a possibility even if not by anything that *exists* in that possibility.

In the Kripkean framework, where representation is identity, this amounts to requiring the **Middle Property**:

$$\text{For } v, w \in W, \text{ if } v \xrightarrow{\diamond} w, \text{ then } Dv \supseteq Dw \cap Dw_{\mathfrak{a}}$$

What might exist and actually exists, exists. A **KMid-model** is a **K**-model with the Middle Property. The **KEMid**-validities are the formulas which are true at each (long enough) **KE**-point in each **KMid** model. I will now prove in several steps that **R** is equivalent to **KEMid**.

**Lemma 1.**  $\mathbf{R} \subseteq \mathbf{KEMid}$  and  $\Delta \mathbf{R} \subseteq \Delta \mathbf{KEMid}$ .

*Proof.* To prove that **KEMid** is as strong as **R**, it suffices to show that for any (long enough) **KEMid**-point  $\mathfrak{a}$  there is a corresponding **R**-point  $\mathfrak{a}^*$  which is **equivalent** to  $\mathfrak{a}$  in the sense that, for each formula  $\phi \in \mathcal{L}$ ,

$$\mathfrak{a} \left| \frac{}{\mathbf{KEMid}} \phi \text{ iff } \mathfrak{a}^* \left| \frac{}{\mathbf{R}} \phi \right.$$

Starting with a **KMid**-model  $W, D$ , etc., we'll construct a counterpart model  $W^*, D^*$ , etc. The basic idea of the construction came up in [Section 1](#): we get world-bound individuals using the pair construction,

$$\begin{aligned} \langle a, w \rangle \in D^*w &\text{ iff } a \in Dw. \\ \langle \langle a_1, w \rangle, \dots, \langle a_n, w \rangle \rangle \in [F]^* &\text{ iff } \langle a_1, \dots, a_n \rangle \in [F]_w \end{aligned}$$

The worlds are unchanged:  $W^* = W$  and  $w_{\mathfrak{a}}^* = w_{\mathfrak{a}}$ . The counterpart-links are determined by identities: if  $v \xrightarrow{\diamond} w$  then there is one link in  $C(w, v)$ , the relation  $S$  such that

$$\langle b, w \rangle S \langle a, v \rangle \text{ iff } a = b \in Dv \cap Dw$$

Otherwise  $C(w, v)$  is empty. Since  $\overset{\diamond}{\rightarrow}$  is reflexive,  $C(v, v)$  includes the identity relation on  $Dv$ . So this is our counterpart model.

Next, given a Kripkean point  $\mathbf{a} = \langle v, a_1, \dots, a_n \rangle$  we construct the **R**-point

$$\mathbf{a}^* = \langle v, R, \langle a_1, v \rangle, \dots, \langle a_n, v \rangle \rangle$$

where the relation  $R$  is again determined by identities:

$$\langle a, v \rangle R \langle b, w_{\textcircled{a}} \rangle \text{ iff } a = b \in Dv \cap Dw$$

Finally, we need to check by induction on the complexity of  $\phi \in \mathcal{L}$  that

$$\mathbf{a} \overset{\text{KE}}{\models} \phi \text{ iff } \mathbf{a}^* \overset{\text{R}}{\models} \phi$$

The only non-trivial steps concern modal clauses; the  $\diamond$ -clause, in particular, requires the Middle Property to ensure that the representation relations “line up”. I won’t go into the details. For the diagonal validities, we just need to check that if  $\mathbf{a}$  is diagonal then so is  $\mathbf{a}^*$ , which is clear.  $\square$

**Lemma 2.**  $\text{KEMid} \subseteq \mathbf{R}$  and  $\Delta \text{KEMid} \subseteq \Delta \mathbf{R}$ .

*Proof.* This time our goal is to show that, given a counterpart model and an **R**-point  $\mathbf{a}$ , there is an equivalent point  $\mathbf{a}^*$  in a Kripke model. The central idea here<sup>17</sup> is that we can regard the counterpart theorist’s individuals as *roles* played by a fixed set of Kripkean individuals; these Kripkean individuals keep their identities from world to world, but in different worlds they may play different counterpart roles. The implementation of this idea requires some care.

A few points of notation. The cardinality of the set  $X$  is written  $|X|$ . The range of the function  $f$  is  $\text{ran } f$ . I freely interchange functions and one-to-one relations, by identifying the function  $f$  with the relation that holds between  $b$  and  $a$  iff  $b = fa$ . (So the relation is identified with a function from its *right* argument to its *left*. This is the opposite of the convention sometimes used.)

Here is how to make a Kripke model from a counterpart model. First, choose some set  $D^*$  of infinite cardinality  $\kappa \geq |D|$ . These are the Kripkean individuals. Next, pick a one-to-one function  $f_{\textcircled{a}}$  from  $Dw_{\textcircled{a}}$  into  $D^*$  such that

$$|D^* \setminus \text{ran } f_{\textcircled{a}}| = \kappa$$

This is the actual world’s cast list, which tells each Kripkean individual which role it plays in  $w_{\textcircled{a}}$ . The cardinality condition will ensure that we never run out of individuals in our construction.

<sup>17</sup>Thanks to Kit Fine for pointing it out to me.

A  **$v$ -admissible function** is a one-to-one function  $f$  from  $Dv$  into  $D^*$  that obeys this cardinality condition:

$$|D^* \setminus (\text{ran } f \cup \text{ran } f_{\text{@}})| = \kappa$$

The Kripkean worlds  $W^*$  are the pairs  $\langle v, f \rangle$  such that  $v \in W$  and  $f$  is a  $v$ -admissible function. Clearly  $f_{\text{@}}$  is  $w_{\text{@}}$ -admissible, so  $\langle w_{\text{@}}, f_{\text{@}} \rangle$  is in  $W^*$ . This is the Kripkean actual world. The domain of  $\langle v, f \rangle$  is  $\text{ran } f$ , and the extension of  $F$  at  $\langle v, f \rangle$  is  $\{fa : a \in [F]_v\}$ . Finally, we need an accessibility relation:

$$\langle v, f \rangle \xrightarrow{\diamond} \langle w, g \rangle \text{ iff } g^{-1} \circ f \in C(w, v) \text{ and } \text{ran } f \supseteq \text{ran } g \cap \text{ran } f_{\text{@}}$$

Since  $f^{-1} \circ f$  is the identity relation on  $Dv$ , which is in  $C(v, v)$ ,  $\xrightarrow{\diamond}$  is reflexive. This makes a **K**-model, which is easily checked to have the Middle Property.

Say  $f$  **fits** a given **R**-point  $\mathbf{a} = \langle v, R, a_1, \dots, a_n \rangle$  iff  $f$  is a  $v$ -admissible function such that

$$f^{-1} \circ f_{\text{@}} = R$$

This says that considering  $f$  as a “cast list” for  $v$  lines it up correctly with the actual world’s assignment of roles. If  $f$  fits  $\mathbf{a}$ , let

$$\mathbf{a}_f = \langle \langle v, f \rangle, fa_1, \dots, fa_n \rangle$$

which is a **KE**-point. The next step is to inductively prove that

$$\mathbf{a} \Vdash_{\mathbf{R}} \phi \text{ iff for each } f \text{ that fits } \mathbf{a}, \mathbf{a}_f \Vdash_{\mathbf{KE}} \phi$$

I’ll just sketch the possibility step. For this, it’s enough to show:

$$\text{If } f \text{ fits } \mathbf{a}, \text{ then } \mathbf{a} \xrightarrow{\diamond} \mathbf{b} \text{ iff } \mathbf{a}_f \xrightarrow{\diamond} \mathbf{b}_g \text{ for some } g \text{ which fits } \mathbf{b} \quad (*)$$

(That this suffices turns on the fact that every **K**-point in the constructed model is equal to  $\mathbf{b}_g$  for some **R**-point  $\mathbf{b}$  and  $g$  which fits  $\mathbf{b}$ . This can be checked from the definitions.) The proof of  $(*)$  uses these two facts:

**Fact 1.** If  $\text{ran } f \supseteq \text{ran } g \cap \text{ran } h$ , then  $h^{-1} \circ g = h^{-1} \circ f \circ f^{-1} \circ g$ .

**Fact 2.** If  $f$  is  $v$ -admissible and  $R \subseteq Dw \times Dv$  is a one-to-one relation, then there is a  $w$ -admissible function  $g$  such that  $g^{-1} \circ f = R$  and  $\text{ran } f \supseteq \text{ran } g \cap \text{ran } f_{\text{@}}$ .

To prove **Fact 2**, first pick some one-to-one function  $h$  from  $Dw \setminus \text{ran } R$  into  $D^* \setminus (\text{ran } f \cup \text{ran } f_{\text{@}})$ . (The cardinality condition guarantees that there is one.) Then let

$$gb = \begin{cases} f(R^{-1}b) & \text{for } b \in \text{ran } R \\ hb & \text{for } b \in Dw \setminus \text{ran } R \end{cases}$$



Let  $\mathbf{a} = \langle v, R, a_1, \dots, a_n \rangle$  and  $\mathbf{b} = \langle w, R', b_1, \dots, b_n \rangle$ , and suppose  $f$  fits  $\mathbf{a}$ . For the first direction of (\*), suppose  $\mathbf{a} \xrightarrow{\diamond} \mathbf{b}$ ; that is,  $R' = S \circ R$  where  $S$  is some member of  $C(w, v)$ , and  $b_i S a_i$  for each  $i$ . By Fact 2 there is a  $w$ -admissible function  $g$  such that  $g^{-1} \circ f = S$  and  $\text{ran } f \supseteq \text{ran } g \cap \text{ran } f_{\textcircled{a}}$ . Then Fact 1 implies that

$$g^{-1} \circ f_{\textcircled{a}} = g^{-1} \circ f \circ f^{-1} \circ f_{\textcircled{a}} = S \circ R = R'$$

This means that  $g$  fits  $\mathbf{b}$ . Furthermore, since  $S = g^{-1} \circ f$  and  $b_i S a_i$ ,  $g b_i = f a_i$ . So  $\mathbf{a}_f \xrightarrow{\diamond} \mathbf{b}_g$ .

Conversely, suppose  $g$  fits  $\mathbf{b}$  and  $\mathbf{a}_f \xrightarrow{\diamond} \mathbf{b}_g$ . This means that  $g^{-1} \circ f \in C(w, v)$ ,  $\text{ran } f \supseteq \text{ran } g \cap \text{ran } f_{\textcircled{a}}$ , and  $g b_i = f a_i$  for each  $i$ . So, if we let  $S = g^{-1} \circ f$ , then  $b_i S a_i$ , and furthermore (by the fact that  $g$  fits  $\mathbf{b}$  and Fact 1)

$$R' = g^{-1} \circ f_{\textcircled{a}} = g^{-1} \circ f \circ f^{-1} \circ f_{\textcircled{a}} = S \circ R$$

This tells us that  $\mathbf{a} \xrightarrow{\diamond} \mathbf{b}$ .

With the other steps filled in, this proves that if  $f$  fits  $\mathbf{a}$  then  $\mathbf{a}_f$  is equivalent to  $\mathbf{a}$ . There are two last things to check. First, each  $\mathbf{R}$ -point has some function that fits it; this follows from Fact 2. Second, if  $f$  fits  $\mathbf{a}$  and  $\mathbf{a}$  is diagonal then so is  $\mathbf{a}_f$ ; this is because if  $f^{-1} \circ f_{\textcircled{a}} = \text{Id}_{\textcircled{a}}$  then  $f = f_{\textcircled{a}}$ .  $\square$

**Corollary 1.**  $\mathbf{R} = \mathbf{KEMid}$  and  $\Delta \mathbf{R} = \Delta \mathbf{KEMid}$ .

**Theorem 2.** If  $E^-$  is defined as in Theorem 1,

$$\begin{aligned} \mathbf{R} &= E^-(\mathbf{KMid}) \\ \Delta \mathbf{R} &= E^-(\Delta \mathbf{KMid}) \end{aligned}$$

*Proof.* By Corollary 1 and Theorem 1. (The proof of Theorem 1 trivially extends to the Middle Property.)  $\square$

To sum up: the logic of  $\mathbf{R}$  is precisely the same as the standard Kripkean actuality system, once we add in all the very serious actualist's existence conditions.

## C The logic of actuality

Next I present a sound and complete axiomatic characterization of the Kripkean logic of actuality, and use this to characterize the logic of the counterpart system  $\mathbf{R}$ .<sup>18</sup> I'll

---

<sup>18</sup> This generalizes a result of Hodes (1984a). Hodes does not consider modal logics weaker than  $\mathbf{S5}$ , which my application requires. If the logic I present for  $\mathbf{K}$  is strengthened with the addition of a pair of

begin by characterizing a “minimal” quantified modal logic which says nothing distinctive about actuality.

A **theory** is a set of formulas closed under the consequence relation of classical propositional logic. For sets of formulas  $\Gamma$  and  $\Delta$ , let  $\Gamma + \Delta$  denote the smallest theory containing  $\Gamma \cup \Delta$ . We’ll start with the following basic axioms:

$$\begin{aligned} \Box(\phi \rightarrow \psi) &\rightarrow \Diamond \phi \rightarrow \Diamond \psi && \text{(K}\Diamond\text{)} \\ \neg @ \neg(\phi \rightarrow \psi) &\rightarrow @ \phi \rightarrow @ \psi && \text{(K}@) \\ \forall x(\phi \rightarrow \psi) &\rightarrow \exists x \phi \rightarrow \exists x \psi && \text{(K}\exists\text{)} \end{aligned}$$

(The ugly  $\neg @ \neg$  in (K@) can be replaced by just @ in our final system, since @ is self-dual under the **K** interpretation. But I use this form for now since some intermediate results don’t assume self-duality.) Let the label for an axiom schema denote the set of its instances: for example,

$$\text{(K}\Diamond\text{)} = \{\Box(\phi \rightarrow \psi) \rightarrow \Diamond \phi \rightarrow \Diamond \psi : \phi, \psi \in \mathcal{L}\}$$

Let  $\Box \Gamma = \{\Box \gamma : \gamma \in \Gamma\}$ , and similarly for other operators. Then we say a theory  $T$  is **normal** iff

$$T \supseteq \Box T + \neg @ \neg T + \forall x T + \text{(K}\Diamond\text{)} + \text{(K}@) + \text{(K}\exists\text{)}$$

In other words, a normal theory includes all of the **K** axioms and is closed under necessitation, “actualization” (again using the dual form for the time being), and universal generalization. Let  $\text{Normal}(\Gamma)$  denote the smallest normal theory containing  $\Gamma$ .

Next, consider the following principles of free logic with identity.

$$\begin{aligned} \forall x \phi \rightarrow Ey \rightarrow \phi(y/x) &&& \text{(UI}_E\text{)} \\ \phi \rightarrow \forall x \phi \quad \text{if } x \text{ is not free in } \phi &&& \text{(VQ)} \\ \forall x Ex &&& \text{(UE)} \\ x = x &&& (=1) \\ x = y \rightarrow \phi(x/z) \rightarrow \phi(y/z) &&& (=2) \end{aligned}$$

---

**S5**-style principles,

$$\begin{aligned} \Diamond \phi &\rightarrow \Box \Diamond \phi \\ \Diamond \phi &\rightarrow @ \Diamond \phi \end{aligned}$$

then the resulting logic is equivalent to Hodes’.

The axiomatization here is generally based on Cresswell and Hughes 1996, ch. 16–17, but see note 20.

(Here  $\phi(y/x)$  is the result of replacing each instance of  $x$  in  $\phi$  with  $y$ .) Also, two principles governing the interaction of identity with the modal operators:<sup>19</sup>

$$\begin{aligned} \diamond x = y &\rightarrow x = y && (\diamond=) \\ @ x = y &\rightarrow x = y && (@=) \end{aligned}$$

These axioms generate our *minimal* theory of quantified modal logic with actuality,

$$\text{QML}_0 = \text{Normal}((\mathbf{UI}_E) + (\mathbf{VQ}) + (\mathbf{UE}) + (=1) + (=2) + (\diamond=) + (@=))$$

To go with this minimal theory, we need a minimal kind of model, which places no special conditions on the accessibility relations for possibility or actuality. Let a **K<sub>0</sub>-model** be just like a **K** model, except the relation  $\overset{\diamond}{\rightarrow}$  isn't required to be reflexive, and instead of an actual world we just have a primitive accessibility relation  $\overset{@}{\rightarrow}$  on  $W \times W$  with no constraints. The truth conditions are just the same as those I gave for **K**-models before (in the general form). In other words, according to the **K<sub>0</sub>** interpretation @ and  $\diamond$  are each plain old normal modal operators, nothing special. **K<sub>0</sub>**-validity is defined in the obvious way. Our first completeness result applies to this minimal logic.

**Lemma 3** (Henkin's Lemma). If  $T$  is a normal theory which includes  $\text{QML}_0$ , then there is a **canonical  $T$ -model**  $\mathfrak{M}$  such that for  $\phi \in \mathcal{L}$ ,

$$\phi \in T \text{ iff } \mathfrak{a} \overset{\vDash}{\underset{\mathbf{K}_0}{\vDash}} \phi \text{ for every } \mathbf{K}\text{-point } \mathfrak{a} \text{ in } \mathfrak{M}$$

For each world  $v$  in  $\mathfrak{M}$ , let  $|v|$  be the set of formulas  $\phi$  such that  $\mathfrak{a} \overset{\vDash}{\underset{\mathbf{K}_0}{\vDash}} \phi$  for some point  $\mathfrak{a} = \langle v, a_1, \dots, a_n \rangle$  (with respect to a fixed variable order). Then furthermore,

$$\begin{aligned} v = w &\text{ iff } |v| = |w| \\ v \overset{\diamond}{\rightarrow} w &\text{ iff } \diamond|w| \subseteq |v| \\ v \overset{@}{\rightarrow} w &\text{ iff } @|w| \subseteq |v| \end{aligned}$$

Since we are dealing with a normal modal logic (with two modal operators) this can be proved by standard completeness methods.<sup>20</sup> The basic idea is to identify

<sup>19</sup> $(\diamond=)$  is only required for logics that don't validate the **B**-schema  $\phi \rightarrow \Box \diamond \phi$ . Similarly  $(@=)$  is only required in the absence of the principle  $\phi \rightarrow @ \diamond \phi$  (assuming @ is self-dual).

<sup>20</sup>See for instance, Cresswell and Hughes 1996, Theorems 16.1 and 16.2, pp. 296–302. The proof given there relies on a complicated extra proof rule (UGLV<sup>m</sup>). This can be dispensed with if instead we use a proof method like that in Fine 1978, §3. (Fine does not carry out the proof for systems weaker than **S5**, though he asserts that it can be done (p. 11). The main things that need to be adjusted are Fine's definitions of "consistent diagram" and " $\diamond$ -complete" to take accessibility relations into account, and correspondingly the  $\diamond$ -step in the construction.)

worlds with certain maximal  $T$ -consistent sets (such that every  $T$ -consistent formula appears somewhere) and individuals with sets of variables, ensuring along the way that every existential and modal claim has a witness.

**Corollary 2.**  $\mathbf{K}_0 = \text{QML}_0$ .

*Proof.* To show that  $\text{QML}_0 \subseteq \mathbf{K}_0$  (**soundness**) it is simple to check that each of the axioms is  $\mathbf{K}_0$ -valid and the normality conditions are validity-preserving. The converse (**completeness**) follows from Henkin’s Lemma: if  $\phi$  is  $\mathbf{K}_0$ -valid, then in particular  $\phi$  is true at every point in the canonical  $\text{QML}_0$ -model, so  $\phi \in \text{QML}_0$ .  $\square$

We can extend this result to stronger systems by giving extra axioms to characterize the extra model constraints, and then showing that if a theory includes the axioms then its canonical model obeys the constraints. For example, the condition that  $\xrightarrow{\diamond}$  is reflexive is characterized by the axiom schema

$$\phi \rightarrow \diamond \phi \tag{T}$$

If a theory  $T$  includes  $\text{QML}_0 + (T)$ , then (T) is true at each point in the canonical  $T$ -model, so for each world  $v$ , if  $\phi \in |v|$  then  $\diamond \phi \in |v|$ ; so  $v \xrightarrow{\diamond} v$ .

We want to do something similar to characterize the standard interpretation of actuality. A nice way to do this is by way of an intermediate kind of model. The “global” condition that there is just one actual world—that is, exactly one world which is @-accessible from every world—turns out to be equivalent to some simple “local” conditions on accessibility relations. This is convenient, because local rules are easier to enforce using modal axioms. (The situation is analogous to **S5**: the global condition that accessibility is a *total* relation can be reduced to the local condition that it is an *equivalence*.) Let a **weak actuality model** be a  $\mathbf{K}_0$ -model with  $\xrightarrow{\diamond}$  reflexive and which obeys the following additional conditions:

$$\text{For each } v \in W, \text{ there is exactly one } w \text{ such that } v \xrightarrow{@} w \tag{A1}$$

$$\text{If } u \xrightarrow{\diamond} v \text{ and } u \xrightarrow{@} w, \text{ then } v \xrightarrow{@} w \tag{A2}$$

$$\text{If } u \xrightarrow{@} v \text{ and } u \xrightarrow{@} w, \text{ then } v \xrightarrow{@} w \tag{A3}$$

(Note the further analogy with **S5**: (A2) and (A3) have the same form as the Euclidean condition that characterizes **S5**.) Apply the usual definitions of validity, etc.

**Lemma 4.** For  $\phi \in \mathcal{L}$ ,  $\phi$  is (diagonally)  $\mathbf{K}$ -valid iff  $\phi$  is true at each (diagonal) point in each weak actuality model.

*Proof.* A model is **connected** iff no proper subset of its worlds is closed under the two accessibility relations. A formula is (diagonally) valid iff it is true at each (diagonal) point in each connected model. (If  $\phi$  is true at  $\mathfrak{a}$  in a model  $\mathfrak{M}$ , then it is also true at the model obtained from the smallest connected set of worlds which includes  $\mathfrak{a}$ 's world.) Furthermore, a connected  $\mathbf{K}_0$  model contains exactly one world which is  $@$ -accessible from any other. (By induction on the number of worlds. It suffices to show that if  $v \xrightarrow{@} w$  and  $v' \xrightarrow{@} w'$ , and either  $v \xrightarrow{\diamond} v'$  or  $v \xrightarrow{@} v'$ , then  $w = w'$ . The first case follows from (A1) and (A2), and the second case follows in the same way from (A1) and (A3).) So a connected weak actuality model becomes a standard  $\mathbf{K}$ -model just by letting  $w_@$  be the unique  $@$ -accessible world.  $\square$

The ‘‘local’’ actuality conditions are easy to characterize using modal axioms, and by Lemma 4 this suffices to characterize the standard logic of actuality as well. These are our actuality axioms:

$$\begin{aligned} \neg @ \phi &\leftrightarrow @ \neg \phi && (@1) \\ \diamond @ \phi &\rightarrow @ \phi && (\diamond @) \\ @ @ \phi &\rightarrow @ \phi && (@@) \end{aligned}$$

Let  $\text{QML}_@ = \text{Normal}(\text{QML}_0 + (\mathbf{T}) + (@1) + (\diamond @) + (@@))$ .

**Theorem 3.**  $\mathbf{K} = \text{QML}_@$ .

*Proof.* Using arguments in the same pattern as the case of  $(\mathbf{T})$ , it is straightforward to show that the canonical  $\text{QML}_@$ -model is a weak actuality model.  $\square$

Note also that the diagonal points of the canonical model are precisely those at which each instance of the following schema is true:

$$\phi \rightarrow @ \phi \tag{\Delta}$$

This gives us a characterization of diagonal validity.

**Theorem 4.**  $\Delta \mathbf{K} = (\Delta) + \text{QML}_@$ .

Note that  $\Delta \mathbf{K}$  is not a normal theory. The axiom  $(\Delta)$  is a ‘‘weak axiom’’: even though it is a kind of logical truth, it is not a *necessary* truth, and accordingly the diagonal validities aren’t closed under necessitation.

Recall that the version of the Kripke semantics that is equivalent to the representational system has one more constraint, the Middle Property. It can be checked that if a theory includes this further axiom

$$\diamond Ex \rightarrow @ Ex \rightarrow Ex \tag{\text{Mid}}$$

then its canonical model has the Middle Property. So

$$\mathbf{KMid} = \text{Normal}(\text{QML}_{@} + (\mathbf{Mid}))$$

Putting this together with the results of [Appendix B](#) characterizes the logic of our counterpart system **R**: simply make the existence conditions fully explicit, and then reason by Kripkean logic.

**Theorem 5.** Let  $E^-$  be as in [Theorem 1](#).

$$\begin{aligned} \mathbf{R} &= E^-(\text{Normal}(\text{QML}_{@} + (\mathbf{Mid}))) \\ \Delta \mathbf{R} &= E^-(\Delta + \text{Normal}(\text{QML}_{@} + (\mathbf{Mid}))) \end{aligned}$$

I'll now summarize a few important logical facts about **R**.

It has been observed that Lewis's interpretation of necessity admits exceptions to the K-schema (Cresswell and Hughes 1996, 356; c.f. Hazen 1979, 326ff.):

$$\Box(\phi \rightarrow \psi) \rightarrow \Diamond \phi \rightarrow \Diamond \psi \quad (\mathbf{K}\Diamond)$$

Suppose all of  $a$ 's counterparts are beautiful. Whenever  $a$  exists, not everything is ugly, so this is true of  $a$  (reading  $F$  as "ugly"):

$$\Box(\forall y Fy \rightarrow \neg Ex)$$

Even so, there are dismal worlds where  $a$  has no counterparts, so this is true:

$$\Diamond \forall y Fy \quad (\mathbf{Ugly})$$

But since  $\Box Ex$  is a consequence of very serious actualism it looks like an instance of  $(\mathbf{K}\Diamond)$  fails.

My interpretations are subtly different from Lewis's. Lewis's original translation manual is sensitive to *which variables occur free* in the formula being translated. All and only the things which are explicitly mentioned in a formula get assigned counterparts in the modal translations. This is essential to the failure of  $(\mathbf{K}\Diamond)$ : intuitively, the necessity of a conditional depends only on certain worlds, those including counterparts for the referents of all its terms; but the antecedent might be possible because of some other world, if it has fewer free variables. (Indeed,  $(\mathbf{K}\Diamond)$  holds in Lewis's system as long as the antecedent  $\phi$  has all of the free variables that the consequent  $\psi$  does.)

My version of counterpart theory isn't sensitive to the syntactic accident of free variables. For  $\Diamond \phi$  to be true of some things, *all* of those things must get counterparts, whether or not  $\phi$  explicitly mentions them.<sup>21</sup> This is logically nicer in some ways:

---

<sup>21</sup>This seems to be the way Lewis does things in (1986, 10–11), though he doesn't note the difference with his earlier version.

for instance,  $(\mathbf{K}\Diamond)$  is *valid* in the sense that the conclusion is true of anything all the premises are true of. When we add explicit existence conditions to  $(\mathbf{K}\Diamond)$ , we add them uniformly:

$$\Box(EX \rightarrow \phi \rightarrow \psi) \rightarrow \Diamond(EX \wedge \phi) \rightarrow \Diamond(EX \wedge \psi)$$

This is  $\mathbf{K}$ -valid, so  $(\mathbf{K}\Diamond)$  is  $\mathbf{R}$ -valid. (The same goes for  $(\mathbf{K}@)$ .)

But we don't escape logical oddities scot-free. In the counterexample,  $(\mathbf{Ugly})$  is *true*, even though, considered as a trivial one-place predicate,  $(\mathbf{Ugly})$  isn't *true of a*. For  $(\mathbf{Ugly})$  to be true of  $a$ , there would have to be an ugly world *where a has a counterpart*—and of course there is no such thing. This is a strange result: we are used to thinking that if a predicate is **syntactically trivial** in the sense that it has no free variables, it is also **semantically trivial** in the sense that it is true of anything if it is true at all. In our counterpart system, this is not so. It isn't just *terms* that semantically “depend on” particular things. Modal operators do too—since they require counterparts (and having a counterpart at a world is not trivial).

I think the best general way to handle these complications is to introduce some explicit “substitution” operators for shifting variable dependences in the modal language—for instance, relating  $Fx$  considered as a formula of just  $x$  to  $\iota_y Fx$ , a formula of  $x$  and  $y$ . Using these we can effectively bring syntactic and semantic dependence back together, without giving up principles like  $(\mathbf{K}\Diamond)$ . (This approach is also helpful for characterizing the logic of contingent identity, which I explore in other unpublished work.) But this would require significant changes; the way I've done things here, we should just make sure to distinguish the variables which are *free* in  $\phi$  from those whose values *make a difference* to  $\phi$ . We can inductively define  $\text{Var } \phi$ , the **non-trivial variables in**  $\phi$ :

$$\begin{aligned} \text{Var } Fx_1 \dots x_n &= \{x_1, \dots, x_n\} \\ \text{Var } \neg\phi &= \text{Var } \phi \\ \text{Var } \phi \wedge \psi &= \text{Var } \phi \cup \text{Var } \psi \\ \text{Var } \exists x \phi &= \text{Var } \phi \setminus \{x\} \\ \text{Var } \Diamond \phi &= \text{Var } @ \phi = \text{the set of all variables in } \mathcal{L} \end{aligned}$$

Here is a useful equivalent formulation:

$$x \notin \text{Var } \phi \text{ iff } \varepsilon^{Y \cup \{x\}} \phi \text{ is the same as } \varepsilon^Y \phi \quad (\dagger)$$

This distinction makes a difference to the quantificational logic of  $\mathbf{R}$ . Because of its very serious actualism,  $\mathbf{R}$  very nearly satisfies standard quantificational logic, as opposed to free logic—even though counterpart models have variable domains,

and even though the corresponding Kripke system uses free logic. For a start, it is closed under universal generalization, and in fact it validates a general version of this principle:

$$\text{If } \phi \rightarrow \psi \text{ is valid and } x \notin \text{Var } \phi, \text{ then } \phi \rightarrow \forall x \psi \text{ is valid.} \quad (\text{UG})$$

The difference from standard systems is just that  $\text{Var } \phi$  must be understood in the sense I just gave:  $x$  must be trivial in  $\phi$ , which is a stronger condition than simply not occurring free in  $\phi$ . Suppose  $x \notin \text{Var } \phi$ ,  $X = Y \cup \{x\}$  includes all of  $\phi$  and  $\psi$ 's free variables, and  $\phi \rightarrow \psi$  is in  $\mathbf{R}$ . Then these are in  $\text{QML}_@$ :

$$\begin{array}{ll} EX \rightarrow \varepsilon^X(\phi \rightarrow \psi) & \text{by Theorem 5} \\ EY \rightarrow \forall x \varepsilon^X \phi \rightarrow \forall x \varepsilon^X \psi & \text{by free logic} \\ EY \rightarrow \forall x \varepsilon^Y \phi \rightarrow \forall x \varepsilon^X \psi & \text{by } (\dagger) \\ EY \rightarrow \varepsilon^Y \phi \rightarrow \forall x \varepsilon^X \psi & \text{by (VQ)} \\ EY \rightarrow \varepsilon^Y(\phi \rightarrow \forall x \psi) & \text{definition of } \varepsilon^Y \end{array}$$

So by Theorem 5,  $\phi \rightarrow \forall x \psi$  is in  $\mathbf{R}$ .

Furthermore, the standard principle of universal instantiation is  $\mathbf{R}$ -valid—no existence condition is required as in  $(\text{UI}_E)$ :

$$\forall x \phi \rightarrow \phi(y/x) \quad (\text{UI})$$

Proof: let  $Y$  contain  $\phi$ 's free variables and also  $y$ , and let  $X = Y \cup \{x\}$ . Then these are in  $\text{QML}_@$ :

$$\begin{array}{ll} Ey \rightarrow \forall x \varepsilon^X \phi \rightarrow (\varepsilon^X \phi)(y/x) & \text{by } (\text{UI}_E) \\ Ey \rightarrow \forall x \varepsilon^X \phi \rightarrow \varepsilon^Y \phi(y/x) & \text{since } (\varepsilon^X \phi)(y/x) \text{ is } \varepsilon^Y \phi(y/x) \end{array}$$

So by the definition of  $\varepsilon^Y$  and Theorem 5,  $(\text{UI})$  is in  $\mathbf{R}$ .

The converse Barcan formula is also  $\mathbf{R}$ -valid:

$$\Box \forall x \phi \rightarrow \forall x \Box \phi \quad (\text{CBF})$$

For in fact

$$\Box \forall x \psi \rightarrow \forall x \Box (Ex \rightarrow \psi)$$



is **K**-valid—in particular in the case where  $\psi$  is  $EY \rightarrow \varepsilon^{Y \cup \{x\}}\phi$ . The  $\varepsilon^Y$ -translation of (CBF) follows.<sup>22</sup> (The analogous fact about the dual of @ holds as well.)

Furthermore, the Middle Property guarantees that a weakened form of the Barcan formula itself is **R**-valid, namely:

$$\forall x \Box \phi \rightarrow \Box \forall x (@ Ex \rightarrow \phi) \quad (\mathbf{BF}^@)$$

This can be confirmed by a straightforward model-theoretic argument. To put it another way, the Barcan formula holds in **R** for any formula that entails actual existence. (This version of the Barcan formula is also *diagonally* valid in **K**.)

On each of these points, the counterpart theorist’s *disjoint*-domain semantics turn out to yield a very similar logic to that of *constant*-domain semantics. The reason is that each in its own way is committed to very serious actualism.

## D Translations

So far, I’ve informally explained how to interpret Modalese in terms of counterparts, and I’ve also given a model-theoretic version of the interpretations. In Lewis’s original discussion he does something different from either of these, giving a manual for transforming modal formulas into possibilist formulas. So in this appendix I’ll similarly formalize my interpretations as a syntactic translation manual. I’ll also show that the translations in this section agree with the model theory of [Appendix A](#) in a natural sense.

The possibilist language  $\mathcal{L}^P$  is a sorted first-order language with identity. (We could instead use second-order logic for our relation quantifiers, but nothing really turns on it—we don’t need very *many* relations.) There are three sorts of terms: one sort for possible individuals ( $a, b, c$ ), one for worlds ( $v, w$ ), and one for binary relations ( $R, S$ ). The language includes a world name  $w_@$ , a two-place predicate  $aIw$  (“ $a$  is in  $w$ ”), and a three-place predicate  $CwvS$  (“ $S$  is a  $w$ - $v$  counterpart-link”)—as well as all of the predicates in the modal language  $\mathcal{L}$ . There is also a three place instantiation predicate  $\epsilon$ , where  $abc\epsilon S$  intuitively means that  $a$  and  $b$  are related by  $S$ ; I’ll notationally cheat a little and just write this as  $aSb$ . It’s also convenient to include a two-place

<sup>22</sup>I should note that despite **R**’s nearly standard quantificational and modal logic, the usual argument for (CBF) doesn’t go through in **R**:

$$\begin{array}{ll} \forall x \phi \rightarrow \phi & (\mathbf{UI}) \\ \Box \forall x \phi \rightarrow \Box \phi & \text{necessitation and K} \\ \Box \forall x \phi \rightarrow \forall x \Box \phi & (\mathbf{UG}) \end{array}$$

The reason is that  $x$  isn’t *trivial* in  $\Box \forall x \phi$ , because of the bare modal operator.

function symbol  $S \circ R$  on relations and a relation name  $\text{Id}_@$ , though these could be dispensed with.

Now I'll define a translation function from  $\mathcal{L}$  into  $\mathcal{L}^P$ . Lewis's original translation manual was given by way of an inductively defined function  $[\cdot]$  that takes a modal formula  $\phi$  together with a world term  $v$  to a first-order formula, the translation  $[\phi]^v$ . The final translation of  $\phi$  is  $[\phi]^{w@}$ , which we denote  $[\phi]$ .

Compared to Lewis's, my translation function has sprouted some extra arguments. First, whereas Lewis relies on the variables that occur free in  $\phi$  to decide how to translate it (in particular, to decide how many counterparts we'll need) and does variable substitutions along the way, I prefer to keep track of the relevant terms explicitly. (See the related discussion in [Appendix C](#).) So the translation function takes as arguments, in addition to the  $\mathcal{L}$ -formula  $\phi$  and an  $\mathcal{L}^P$ -term  $v$  for the world, also a sequence  $a_1, \dots, a_n$  of  $n$  individual terms from  $\mathcal{L}^P$ . And naturally we'll also need an argument  $R$  to represent the representation relation. This will generally be a complex relation term of the form  $S_1 \circ \dots \circ S_n \circ \text{Id}_@$ . So, for an  $\mathcal{L}$ -formula  $\phi$  and some  $\mathcal{L}^P$  terms  $v, R$ , and  $a_1, \dots, a_n$ , we inductively define the translation  $[\phi]^{v, R, a_1, \dots, a_n}$ . Writing down the definition is essentially a matter of regimenting the interpretations I gave informally in the main text, or for that matter the model-theoretic version in [Appendix A](#). To keep the clauses more legible, let  $\mathbf{a}$  abbreviate the sequence of terms  $v, R, a_1, \dots, a_n$ . (I'm also assuming an implicit variable ordering like before.)

$$\begin{aligned}
[Fx_{i_1} \dots x_{i_n}]^{\mathbf{a}} &\text{ is } Fa_{i_1} \dots a_{i_n} \\
[\neg\phi]^{\mathbf{a}} &\text{ is } \neg[\phi]^{\mathbf{a}} \\
[\phi \wedge \psi]^{\mathbf{a}} &\text{ is } [\phi]^{\mathbf{a}} \wedge [\psi]^{\mathbf{a}} \\
[\exists x_{n+1} \phi]^{\mathbf{a}} &\text{ is } \exists b (bIv \wedge [\phi]^{v, R, a_1, \dots, a_n, b}) \\
[\diamond\phi]^{\mathbf{a}} &\text{ is } \exists w \exists S \exists b_1 \dots \exists b_n (CwvS \wedge \bigwedge b_i S a_i \wedge [\phi]^{w, S \circ R, b_1, \dots, b_n}) \\
[@\phi]^{\mathbf{a}} &\text{ is } \exists b_1 \dots \exists b_n (\bigwedge a_i R b_i \wedge [\phi]^{w@, \text{Id}_@, b_1, \dots, b_n})
\end{aligned}$$

(Here  $\bigwedge \psi_i$  abbreviates the obvious finite conjunction.) The final translation  $[\phi]$  of a closed formula  $\phi$  is  $[\phi]^{w@, \text{Id}_@}$ . (We can also treat free variables as names for actuals if we associate each one with an individual term in  $\mathcal{L}^P$ ; if  $x_i$  is associated with  $a_i$  for each  $i$ , then  $[\phi]$  is  $[\phi]^{w@, \text{Id}_@, a_1, \dots, a_n}$ .)

These translations agree with the model-theory **R** of [Appendix A](#), in a sense I will now make precise. Consider these postulates for counterpart theory (which cor-

respond to the conditions on counterpart models) (c.f. Lewis 1968, 114):

$\forall v \forall w \forall a (aIv \rightarrow aIw \rightarrow v = w)$	Individuals are world-bound.
$\forall v \forall w \forall S (CvwS \rightarrow \text{In}_{w,v} S)$	$w$ - $v$ links relate $w$ -things to $v$ -things.
$\forall v \forall w \forall S (CvwS \rightarrow 1-1 S)$	Links are one-to-one.
$\forall v \exists S (CvvS \wedge \forall a (aIv \rightarrow aSa))$	Each world has a reflexive link.
$\forall a \forall b (a \text{Id}_{\textcircled{a}} b \leftrightarrow (a = b \wedge aIw_{\textcircled{a}}))$	Definition of $\text{Id}_{\textcircled{a}}$ .
$\forall a \forall c \forall R \forall S (c(S \circ R)a \leftrightarrow \exists b (cSb \wedge bRa))$	Definition of composition.

These use the abbreviations

$$\begin{aligned} \text{In}_{w,v} S \text{ is } & \forall a \forall b (bSa \rightarrow (aIv \wedge bIw)) \\ 1-1 S \text{ is } & \forall a \forall b \forall a' \forall b' (bSa \rightarrow b'Sa' \rightarrow (a = a' \leftrightarrow b = b')) \end{aligned}$$

Call the set of all these postulates  $\mathcal{C}$ .

There's a pretty obvious way in which (sorted first-order) models of  $\mathcal{C}$  correspond to the counterpart models described in Appendix A (c.f. Fine 1978, 130). If we have a counterpart model  $\langle W, w_{\textcircled{a}}, D, [\cdot], C \rangle$ , then we can construct a corresponding model of  $\mathcal{C}$ . The domain of worlds is  $W$ ; the domain of individuals is  $D$ ; the domain of relations is the set of one-to-one relations on  $W \times W$ . The pair  $\langle a, w \rangle$  is in the extension of  $I$  iff  $a \in Dw$ ; the extension of  $C$  is the set of triples  $\langle w, v, S \rangle$  such that  $S \in C(w, v)$ ; the extension of the instantiation predicate  $\epsilon$  is the set of triples  $\langle b, a, S \rangle$  such that  $bSa$ . And we can also go the opposite direction: a model of  $\mathcal{C}$  corresponds to a counterpart model by a similar sort of construction.

Furthermore, each  $\mathbf{R}$ -point  $\langle v, R, a_1, \dots, a_n \rangle$  corresponds to a variable assignment in the corresponding  $\mathcal{L}^P$  model: one which takes the world variable  $v$  to the point's world, the relation variable  $R$  to its relation, and the individual variables  $a_1, \dots, a_n$  to its individuals. This assignment is guaranteed to satisfy the following formulas (with free variables  $v, R, a_1, \dots, a_n$ ):

$$\text{In}_{v, w_{\textcircled{a}}} R \quad 1-1 R \quad a_1 Iv \quad \dots \quad a_n Iv$$

Conversely, any assignment function that satisfies these formulas corresponds to an  $\mathbf{R}$ -point. It's straightforward to check by induction that if the  $\mathbf{R}$ -point  $\mathbf{a}$  corresponds this way to the assignment  $g$  in the sorted first-order model  $\mathfrak{M}$ , then for each formula  $\phi \in \mathcal{L}$ ,

$$\mathbf{a} \models_{\mathbf{R}} \phi \text{ iff } \mathfrak{M}, g \models [\phi]^{v, R, a_1, \dots, a_n}$$

What's more,  $\mathbf{a}$  is a diagonal point iff  $g$  takes  $v$  to the actual world and  $R$  to the extension of  $\text{Id}_{\textcircled{a}}$ .

This establishes a precise sense in which the translations and model theory agree:

**Theorem 6.** For  $\phi \in \mathcal{L}$ ,  $\phi$  is **R**-valid iff

$$\mathcal{C}, \text{In}_{v, w@} R, 1-1 R, a_1 I v, \dots, a_n I v \models [\phi]^{v, R, a_1, \dots, a_n}$$

Furthermore,  $\phi$  is diagonally **R**-valid iff  $\mathcal{C} \models [\phi]$ .

Consequently, the logical results in [Appendix C](#) can also be applied to the translations.

## References

- Belot, Gordon. 1995. New work for counterpart theorists: Determinism. *British Journal for the Philosophy of Science* 46, no. 2 (June): 185–195.
- Butterfield, Jeremy. 1989. The Hole Truth. *British Journal for the Philosophy of Science* 40 (1): 1–28.
- Cresswell, M.J., and G.E. Hughes. 1996. *A New Introduction to Modal Logic*. 1st ed. Routledge.
- Dorr, Cian. n.d. Counterparts. Manuscript.
- Fara, Delia Graff. n.d. Possibility relative to a sortal. Manuscript.
- . 2009. Dear Haecceitism. *Erkenntnis* 70:285–297.
- Fara, Michael, and Timothy Williamson. 2005. Counterparts and Actuality. *Mind* 114, no. 453 (Jan.): 1–30.
- Fine, Kit. 1978. Model Theory for Modal Logic Part I—The De Re/De Dicto Distinction. *Journal of Philosophical Logic* 7, no. 1 (Jan.): 125–156.
- . 1985. Plantinga on the Reduction of Possibilist Discourse. In *Alvin plantinga*, ed. James Tomberlin and Peter Van Inwagen, 145–186. Profiles. Dordrecht: Reidel.
- Forbes, Graeme. 1982. Canonical Counterpart Theory. *Analysis* 42, no. 1 (Jan.): 33–37.
- . 1990. Counterparts, Logic and Metaphysics: Reply to Ramachandran. *Analysis* 50, no. 3 (June): 167–173.
- Gibbard, Allan. 1975. Contingent Identity. *Journal of Philosophical Logic* 4, no. 2 (May): 187–221.

- Hazen, Allen. 1976. Expressive Completeness in Modal Language. *Journal of Philosophical Logic* 5, no. 1 (Feb.): 25–46.
- . 1979. Counterpart-Theoretic Semantics for Modal Logic. *The Journal of Philosophy* 76, no. 6 (June): 319–338.
- Hodes, Harold. 1984a. Axioms for Actuality. *Journal of Philosophical Logic* 13 (Feb.): 27–34.
- . 1984b. On Modal Logics Which Enrich First-Order S5. *Journal of Philosophical Logic* 13, no. 4 (Nov.): 423–454.
- . 1984c. Some Theorems on the Expressive Limitations of Modal Languages. *Journal of Philosophical Logic* 13 (Feb.): 13–26.
- Lenk, Uta. 1998. *Marking Discourse Coherence: Functions of Discourse Markers in Spoken English*. Language in Performance. Tübingen: Gunter Narr Verlag.
- Lewis, David K. 1968. Counterpart Theory and Quantified Modal Logic. *The Journal of Philosophy* 65, no. 5 (Mar.): 113–126.
- . 1971. Counterparts of Persons and Their Bodies. *Journal of Philosophy* 68 (7): 203–211.
- . 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- Maudlin, Tim. 1990. Substances and Space-Time: What Aristotle Would Have Said to Einstein. *Studies in History and Philosophy of Science* 21, no. 4 (Dec.): 531–561.
- Plantinga, Alvin. 1974. *The Nature of Necessity*. Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.
- . 1983. On Existentialism. *Philosophical Studies* 44, no. 1 (July): 1–20.
- Pooley, Oliver. n.d. Substantivalism and Haecceitism. Manuscript chapter.
- Ramachandran, Murali. 1989. An Alternative Translation Scheme for Counterpart Theory. *Analysis* 49, no. 3 (June): 131–141.
- . 1990. Unsuccessful Revisions of CCT. *Analysis* 50, no. 3 (June): 173–177.
- Rosen, Gideon. 1990. Modal Fictionalism. *Mind* 99, no. 395 (July): 327–354.
- Sider, Theodore. n.d. Beyond the Humphrey Objection. Manuscript.
- . 2002. The Ersatz Pluriverse. *The Journal of Philosophy* 99, no. 6 (June): 279–315.
- Williamson, Timothy. 1998. Bare Possibilia. *Erkenntnis* 48 (2 & 3): 257–273.