Counterpart theory and the paradox of occasional identity

Wolfgang Schwarz
Draft, 19 June 2011

Counterpart theory is often advertised by its track record at solving metaphysical puzzles. Much of this puzzle-solving power, however, can be mimicked by alternative views. Here I focus on the paradox of occasional identity, also known as the paradox of fission and fusion, or the paradox of contingent identity. To spell out the counterpart-theoretic solution, Lewis's interpretation rules have to be extended beyond the language of quantified modal logic. I present a more comprehensive semantics that takes into account the multiplicity and sortal-dependence of counterpart relations, allows talking about specific times and worlds, and does not require names to denote present (or actual) individuals. I also explain why this semantics does not commit us to the more controversial metaphysical and semantic aspects of counterpart theory.

1 The paradox of occasional identity

The train from Berlin to Düsseldorf and Cologne passes through a place called Hamm, where it gets divided: the front half continues to Düsseldorf, the rear half to Cologne. Before the division, the two halfs compose a single train. (The announcement on the train says that *this train will be divided*, not that *these trains will be separated.*) After the division, two trains seem to be leaving Hamm – one towards Düsseldorf, the other towards Cologne. (Passengers entering between Hamm and Cologne are entering a train to Cologne, and not a train to Düsseldorf.) So there is one train before Hamm, and two trains after Hamm. Yet if you asked whether the train from Berlin ends at Hamm, you would get a negative answer. This is even more obvious if you travel in the opposite direction, say from Cologne to Berlin: the train you enter in Cologne will take you all the way to Berlin, not just to Hamm, where it gets connected with a train from Düsseldorf. So one train enters Hamm, two trains leave Hamm, and the original train doesn't end there. It would seem to follow that the original train is one of the two trains that leave Hamm – the one to Cologne perhaps. But no. The train to Düsseldorf has as much claim to coming from Berlin as the train to Cologne. Again, in the other direction you can enter a train to Berlin not just in Cologne, but also in Düsseldorf.

This is a paradox of occasional identity. The single train entering Hamm seems to be identical to both trains leaving Hamm. The problem, of course, is that identity is euclidean: if $x = y$ and $x = z$, then $y = z$.

A more familiar, albeit more far-fetched instance of the paradox involves fissioning people. Imagine a duplication machine that splits a human body lengthwise, creates perfect copies of both sides, and combines each half of the original body with a copy of the missing other side. All this happens very fast, so that the result are two fully functional, qualitatively identical copies of the original body. One would think that humans can survive half of their body being instantaneously replaced by an exact copy. Hence a person entering the duplication machine should survive the procedure; for why should it kill her if the removed half is being completed to another body? But then the original person would have to be identical to both of the persons leaving the machine.

Another familiar example is Theseus's ship, which at the end of the story seems to be identical both to the repaired ship and to the ship reconstructed from the old planks. There are also spatial and modal cases, where something at one place or possible world seems to be identical to multiple things at another place or world. For a spatial example, consider the river Rhine, which extends upstream both to Dissentis and to Splügen, although the river in Dissentis is not identical to the river in Splügen. For a modal example, let x and y be identical twins and consider a world w where the zygote that actually became x and y did not split; w contains a single individual that may seem to be identical to both of the twins in the actual world. For the sake of concreteness, I will focus on temporal cases in this paper, but most of what I say can also be applied to the other dimensions.[1]

In general, a paradox of occasional identity arises whenever the following statements are all true (where t_1 and t_2 are different times, or places or worlds).

(1) At t_1, there is a single object x of a certain kind.
(2) At t_2, there are exactly two objects, x_1 and x_2, of that kind.
(3) The two objects x_1 and x_2 have equal claim to be x.
(4) At t_2, x still exists (and is still an object of the relevant kind).

If (1)–(4) contradict the logic of identity, there cannot be any paradox of occasional identity. In any alleged example, at least one of (1)–(4) must be false. In the case of personal fission, the most popular response is probably to reject (4). Thus [Parfit 1984] takes the paradox to show that the person entering the duplication machine is killed and replaced by two new persons. Alternatively, one might reject (3) and postulate an asymmetry that privileges one of the successors over the other. [Swinburne 2004] suggests

1 [Gallois 1998: ch.1] gives an overview over a wide range of possible paradoxes of occasional identity. See also [Burgess 2010] for a reminder that our topic has practical, political implications.

that only one of the bodies leaving the duplication machine will carry the original person's immaterial soul. (This solution might look less attractive for trains and rivers.) [Lewis 1976] rejects (1) and claims that there were two people already before the duplication. One might also reject (2) and say that the apparently two objects at t_2 are really one and the same, or that there are actually three objects at t_2: x_1, x_2 and the original object x.

I have no general objection to these proposals. Rejecting one of (1)–(4) may indeed be the best option in this or that concrete example. Consider an instance of asymmetric teleportation, where a human body on Earth is duplicated on a remote planet, without destroying the original body. It is natural to think there is then one person on Earth for the whole time, and a second, new person on the other planet. More controversially, I will argue in section 5 that when a lump of clay constitutes a statue, the two are not identical; the fact they are also not identical at certain other times (or worlds) therefore does not give rise to a paradox of occasional identity. So not everything that from a distance resembles a paradox of occasional identity is a genuine instance. However, in *some* cases, (1)–(4) do all seem true, and rejecting any one of them is a serious cost. The above examples of the train, the ship, the river, the zygote and the fissioning person look like good candidates to me – but feel free to substitute your own examples if you disagree.

Instead of rejecting a particular component of (1)–(4), some authors have appealed to indeterminacy (e.g. [Johnston 1989]). Perhaps, so the idea, our concepts or linguistic conventions do not settle what to say about far-fetched fission scenarios, and hence leave open which of (1)–(4) is false. Again, this may well be right for certain cases, but it is not a satisfactory answer in cases where at least three of (1)–(4) seem determinately true. (I'd also say that not all examples are especially far-fetched or rare. There's a train from Berlin to Düseldorf and Cologne once every hour.)

I will defend a view on which (1)–(4) can all be true, without giving up the euclideanness of identity or relying on controversial metaphysical assumptions. My explanation will be boringly semantic rather than metaphysical. The paradox is resolved by paying close attention to the interaction between singular terms and modal operators.

My solution is thus in line with the widespread conviction – especially among non-metaphysicians – that the puzzle raised by the above examples is in some sense "merely verbal". A test from [Chalmers 2011] may help to bring this out. We could describe what is going on in the train example by talking about the individual carriages, how they are connected to one another at various times, and so on, without ever using the word 'train'. Nothing puzzling or contradictory would be said on this level. But intuitively, we also wouldn't have left anything out. It's not like there are fundamental *train facts* over and above facts about attached carriages etc. that would still have to be settled once all the "lower-level" facts are in place. This suggests that the paradox concerns the application of concepts like 'train' or 'same train' to a situation that, in itself, is not paradoxical at all.

If this is on the right track, it has important consequences for what counts as an adequate solution. Consider the account of personal fission in [Lewis 1976]. Lewis argues that persons and other material objects are four-dimensional aggregates of instantaneous (or very short-lived) *stages* which are the primary bearers of many properties. Just as a wall that is white here and red over there has a white part here and a red part there, a person who sits at t_1 and stands at t_2 has a sitting (temporal) part at t_1 and a standing part at t_2. The stages that compose a person are related to one another by a certain *unity relation* R, which according to Lewis is largely a matter of psychological continuity and connectedness. Persons are identified with maximal aggregates of stages that all stand in the R-relation to one another. A fission case therefore involves two persons: every post-fission stage is R-related to every pre-fission stage, but the post-fission stages aren't all R-related to one another. It follows that each pre-fission stage is part of two different persons – two different, but overlapping, aggregates of R-interrelated stages. This is why Lewis claims that there are two persons already before the fission.

Now this may be a helpful metaphysical description of the scenario, but if our puzzle concerns the application of ordinary terms and concepts, then it is not much of a solution. We would not ordinarily say (or think) that *two* persons entered the duplication machine, that *two* trains left Berlin, or that Theseus owned *two* ships at the beginning of the story. In fact, it would be very inconvenient to speak in the Lewisian manner, for we would never know how many things we face. Standing at a bridge between Berlin and Hamm, we couldn't tell how many trains just passed unless we know what happens further up and down the tracks. (Among other things, we would have to know whether there will be an accident down the track: if the splitting is called off, what we saw is one train, otherwise it's two.)

Lewis could insist that he *is* analysing our ordinary notions of trains and ships and persons, and that we are simply mistaken when we think we know how many trains crossed the bridge. What Lewis actually does is more interesting. He offers a rule for interpreting our ordinary talk in such a way that 'a single train crossed the bridge' comes out true as long as all trains that actually crossed the bridge share a single stage at the relevant time. The rule says that in ordinary language, we count objects not "by identity", but "by stage-sharing" – i.e., we count the cells in the partition effected by the stage-sharing relation.[2]

So there are two parts to Lewis's account. One is the four-dimensional description in terms of aggregates and stages. This is what is really, fundamentally, going on. On this level, (1) is rejected. The second part is an interpretation of our ordinary language in the model provided by the fundamental description. Here (1) comes out true.

Unfortunately, the second part of Lewis's account is at best an incomplete sketch. How,

2 This is Lewis's proposal in [Lewis 1976]; in [Lewis 2002], he instead suggests that we count temporal parts of the relevant objects. The effect is the same.

for example, do names and definite descriptions work in a case of fission? When a person (or train, or ship) is baptised 'Mary' and later fissions, which of the two persons have been called 'Mary'? Who does the name pick out after the fission? What happens when the name is embedded in operators that shift the time of evaluation?[3]

More seriously, Lewis's interpretation rules are quite revisionary. When we think we see a single train on the bridge, or a single amoeba in a dish, we are, according to Lewis, really looking at two trains and possibly thousands of amoebae. Our thoughts (or words) are only true because 'there is a single X' doesn't really mean that there is a single X. But the goal isn't to come up with *some* interpretation that renders our ordinary thought and talk true. We would like an interpretation that we can accept as capturing what we really meant. It would be nice to have an account on which 'there is a single X' is true at the pre-fission time because there really is, strictly and literally and fundamentally, just a single X there.

Before I present my own proposal, let me review another proposal that promises to achieve this goal. It is based on the *stage theory* defended in [Sider 1996] and [Sider 2001], which in turn is based on Lewis's *counterpart theory* ([Lewis 1968], [Lewis 1986]).

2 Stages and counterparts

Sider, in [Sider 1996] and [Sider 2001], agrees with Lewis's four-dimensional picture of reality, but disagrees about the identification of ordinary objects in this picture. According to Sider, ordinary objects are the temporally short-lived stages that Lewis regards as mere temporal parts of ordinary objects. When we say 'this train' at the departure time in Berlin, we refer to a momentary stage that does not exist in the past or the future.

This suggests that Sider rejects condition (4) in our paradox: the object x that exists at t_1 does not exist any more at t_2. The train that leaves Berlin never leaves Hamm; indeed, it disappears immediately after t_1. Since this fits our ordinary thought and talk even less than Lewis's proposal, Sider, like Lewis, offers conciliatory rules for interpreting ordinary language into his stage-theoretic metaphysics.

Consider the train stage x at time t_1. x does not extend temporally beyond t_1. Nevertheless, according to Sider, 'at some time, x arrives at Hamm' is true. This is because 'at some time, x is F' is true iff x has a *temporal counterpart* that is F. The temporal counterpart relation is what we previously met as the unity relation. For trains, it is probably not a matter of psychological continuity and connectedness, and I will not attempt an informative analysis here. (I will assume, however, that unity relations are generally reflexive and symmetric.) It is clear that the train stage x should count as unity-related to a later stage in Hamm. So x has a temporal counterpart at Hamm.

3 [Robinson 1985] looks bit further into questions like these.

Hence we can truly say that the train that leaves Berlin will at some point arrive at Hamm.

This rule for interpreting statements of the form 'at some time, x is F' is modeled on Lewis's counterpart-theoretic interpretation of 'possibly, x is F'. According to Lewis, ordinary objects do not extend across different possible worlds. We have temporal parts at other times, but we do not have "modal parts" at other worlds, nor do we somehow exist wholly and identically at other worlds, perhaps in the way Moskow is wholly and identically both in Russia and in Europe. Still, Lewis doesn't conclude that we have all our actual properties essentially. Instead he suggests that 'possibly, x is F' is true as long as some (modal) counterpart of x at some possible world is F.

On this account, we have to distinguish two senses in which an object can have a property relative to a time (or world, or place). In the fundamental sense, property instantiation is a private affair between a stage and a property. No time or place or world is involved. To say that x is F at t simply means that x is located (wholly and entirely) at t, and that it is F. In the other, derivative sense, one can truly say that x is F at t even if x isn't located at t, as long as x has a counterpart at t which is F (in the fundamental, non-relative sense). In the fundamental sense, the train that leaves Berlin does not exist at t_2. In the derivative sense, it does. The ordinary sense is clearly the derivative sense.

How does the stage-theoretic account compare to Lewis's account in terms of aggregates? That there is only one train crossing the bridge now comes out strictly and literally true, but lots of other things we would normally consider true can be rescued only by forceful re-interpretation: that the train is several years old, or that there has been only one train on this platform during the last 10 minutes. In the end, the stage-theoretic interpretation is at least as revisionary as the Lewisian rules.

Again, it is a bit of an exaggeration to speak of "interpretation rules" here, as nobody has yet attempted a counterpart-theoretic interpretation for a significant fragment of English. [Lewis 1968] gives rules for interpreting the standard language of quantified modal logic into his counterpart-theoretic (modal) metaphysics, but the language of quantified modal logic is a very impoverished language. It does not allow formalising claims like (1)–(4). At a minimum, we here need operators like '*at* t_2' that consider not only whether something is true at *all* times or at *some* times (or worlds), but whether something is true at a particular time t_2. Modal logics that include operators like 'at t_2' are known as *hybrid logics*.

In the remainder of this section, I will now sketch a counterpart-theoretic semantics for (quantified) hybrid logic on which (1)–(4) can all be true, without abandoning the euclideanness of identity. I offer this as a gift to the stage theorist, but the offer isn't entirely altruistic. As will become clear in the next section, essentially the same semantics will figure in my own proposal.

As usual in the semantics for modal languages, sentences are interpreted at a particular point, the "utterance time" t_0. Let's assume for now that all names and variables denote stages located at t_0. Predicates express relations between stages. Operators like 'at some time' or '*at t*' shift the time of evaluation: '$at\ t, Fx$' is true (at the utterance time t_0) iff 'Fx' is true relative to the time denoted by 't'. But when we evaluate 'Fx' relative to t, we do not use the original interpretation on which 'x' denotes a stage located at t_0. Rather, when we shift the time of evaluation from t_0 to another time t, we also shift the reference of all singular terms: if 'x' originally denotes a certain stage x located entirely at t_0, then relative to t, the term denotes a different stage x' located at t – namely x's counterpart at t. Similarly, 'at some time, Fx' is true (at t_0) iff 'Fx' is true relative to some time t, i.e. iff there is a time t such that x's counterpart at t has the property expressed by 'F'.

All this is straightforward as long as x has a unique counterpart at t. What if there are several counterparts, or none? The case of zero counterparts isn't directly relevant to our puzzle, so I will set it aside for the moment. If x has several counterparts at t, then shifting the reference of 'x' from x to its counterparts at t leaves the term ambiguous, in a sense: it now denotes several things at once. To motivate the following decisions, let's have a quick look at ordinary ambiguity.[4] (I will return to this analogy in section 4.)

There are two Londons, one in the UK and one in Canada. So 'London is in Canada' is true on one interpretation of 'London' and false on another. It might be best to leave it at that, but suppose we want to assign truth-values even to ambiguous sentences. How could we do that? Mimicking Russell's theory of descriptions, we could decree a sentence false whenever it contains an ambiguous term. Alternatively, we could say that a sentence is true iff it is true on every disambiguation. Either way, both 'London is in Canada' and 'London is in the UK' would come out untrue. Another option is to say that an ambiguous sentence is true as long as it is true on *some* disambiguation. One can argue that this existential reading, on which both sentences are true, better fits our ordinary usage.

Another question is whether disambiguations can be *mixed*: is there a disambiguation of 'we traveled from London to London' on which the first 'London' refers to London, UK, and the second to London, Ontario? Ordinary usage seems to allow for this. It may not be very *helpful* to say 'we traveled from London to London', but in a suitable context, this may well be considered a somewhat curious truth. (When asked what state New York belongs to, 'New York is in New York' may even be true and helpful.)

The same options arise in our semantics for hybrid logic with multiple counterparts, and I suggest we make the same choices. Thus if 'x' denotes a stage with several counterparts at t, then '$at\ t, Fx$' is true if *at least one* of those counterparts satisfies 'F'. Different

4 See [Lewis 1982], [Priest 1995], [Ripley 2011] for a defense of the following remarks. There are, of course, alternative treatments of ambiguity, see e.g. [Frost-Arnold 2008].

counterparts may be assigned to different occurrences of a term: '$at\ t, (Fx \wedge Gx)$' is true if at least one t-counterpart of x is F and at least one is G.[5]

So if t_0 is the train's departure time at Berlin, then you could truly say 'in five hours, this train will be in Cologne'. The term 'this train' denotes a present train stage x; 'in five hours' shifts the denotation to the relevant counterparts of x, of which there are two. Your utterance is true because one of those counterparts is in Cologne. You could also have truly said, 'in five hours, this train will be in Düsseldorf', and even 'in five hours, this train will be in Cologne and this train will be in Düsseldorf' – although, like 'London is in Canada and London is in the UK', this would rarely be a very helpful statement.

Now we can explain why (1)–(4) do not contradict the logic of identity. Here they are again.

(1) At t_1, there is a single object x of a certain kind.
(2) At t_2, there are exactly two objects, x_1 and x_2, of that kind.
(3) The two objects x_1 and x_2 have equal claim to be x.
(4) At t_2, x still exists (and is still an object of the relevant kind).

The apparent conflict with the logic of identity arises as follows. Focus on t_2. By (2) and (4), we then have two objects x_1 and x_2, and the original object x still exists. Since x_1 and x_2 have equal claim to be x, it can't be that x is identical to x_1 unless x is also identical to x_2, and conversely. So

(5) $$x = x_1 \leftrightarrow x = x_2.$$

By (2), x_1 and x_2 are the only objects of the relevant kind. Together with (4), this means that x is either x_1 or x_2 or both:

(6) $$x = x_1 \vee x = x_2.$$

By elementary propositional logic, (5) and (6) entail that $x = x_1 \wedge x = x_2$. But identity is euclidean, i.e.

(7) $$x = x_1 \wedge x = x_2 \supset x_1 = x_2.$$

So by Modus Ponens, $x_1 = x_2$. And now we have a contradiction with (2), which says that

(8) $$x_1 \neq x_2.$$

5 The translation rules of [Lewis 1968] effectively require truth on all assignments of counterparts for box formulas and truth on some assignments for diamond formulas. The existential and universal readings are often mentioned as possible interpretations for the 'actually' operator, which in the hybrid language might be understood as 'at t_0'; see e.g. [Hazen 1979]. Lewis is not clear on the choice between mixed and uniform assignments; see [Woollaston 1994] for a brief discussion. Mixed assignments are effectively used in [Forbes 1982], [Ramachandran 1989] and follow-up work by these authors, uniform assignments are used in [Sider 2008] and [Russell 2011].

Here I have conveniently dropped the prefix 'at t_2'. But this is important to the counterpart-theoretic solution. If we put it in, we get

(5*) $at\ t_2, (x = x_1 \leftrightarrow x = x_2)$.

(6*) $at\ t_2, (x = x_1 \lor x = x_2)$.

(7*) $at\ t_2, (x = x_1 \land x = x_2 \supset x_1 = x_2)$.

(8*) $at\ t_2, x_1 \neq x_2$.

(5*)–(8*) still look inconsistent: assuming that whatever logically follows from something that is true at a time is itself true at that time, (5*) and (6*) imply that $at\ t_2, (x = x_1 \land x = x_2)$, which together with (7*) means that $at\ t_2, x_1 = x_2$, contradicting (8*).

However, on the semantics given above, (5*)–(8*) are perfectly consistent. They are true whenever the stage x at t_1 has two counterparts, x_1 and x_2, at t_2. In the scope of 'at t_2', each occurrence of 'x' then denotes x_1 on one assignment of counterparts and x_2 on another. Consider, for example, an assignment that maps each occurrence of 'x' to x_1. On this choice of counterparts, '$x = x_1$' is true and '$x = x_2$' false; so '$x = x_1 \lor x = x_2$' and '$x = x_1 \land x = x_2 \supset x_1 = x_2$' are both true. Since truth on *some* assignment of counterparts is sufficient for truth, this verifies (6*) and (7*). More generally, (6*) and (7*) are verified by every uniform assignment of counterparts. (5*), on the other hand, is verified only by mixed assignments. For example, if the first occurrence of 'x' is mapped to x_1 and the second to x_2, then both sides of '$x = x_1 \leftrightarrow x = x_2$' are true. (8*), finally, is true on any assignment, since it doesn't even contain the problematic term 'x'.

Where, then, did the above reductio go wrong? (5*) and (6*) do entail that $at\ t_2, (x = x_1 \land x = x_2)$. (This is made true by the assignment that maps the first occurrence of 'x' to x_1 and the second to x_2.) However, from the fact that $at\ t_2, (x = x_1 \land x = x_2)$ and that $at\ t_2, (x = x_1 \land x = x_2 \supset x_1 = x_2)$, we *cannot* infer that $at\ t_2, x_1 = x_2$. Truth at a time is not closed under modus ponens:

$$at\ t, A;\ at\ t, (A \supset B)\ \not\models\ at\ t, B.$$

In a case of occasional identity, this inference involves something like the fallacy of equivocation. If A is true on one reading and false on another, then we cannot infer from the fact that A is true (on one reading) and that $A \supset B$ is true (on another reading of A) that B is true.

I said that I would like to offer my semantics as a gift to stage theorists like Sider, but I'm afraid he won't accept it. The reason is that my semantics violates certain principles of "classical" hybrid logic, and I take it from [Sider 2008] that Sider would want to preserve those principles. Unfortunately, this is pretty much impossible within a counterpart-theoretic framework. Sider's proposal (like those of [Russell 2011] and [Ramachandran 1989], who share his motivation) manages to stay close to the classical principles only by essentially ruling out the possibility of multiple counterparts. In my view, this robs counterpart theory of its main advantages; it certainly blocks any solution to the paradox of occasional identity. I will address worries about classicality in section 4.

3 Counterparts of persisting objects

Equipped with something like the counterpart-theoretic semantics just outlined, stage theory offers an attractive escape from the paradox of occasional identity. The apparent inconsistency in statements like (1)–(4) is explained as the effect of a harmless semantic phenomenon akin to ambiguity. On the other hand, we have seen that the stage theoretic solution involves a rather revisionary interpretation of ordinary thought and talk. It also presupposes controversial metaphysical doctrines. Many philosophers think that persons and trains persist through time without dividing into temporal parts, and therefore deny the existence of person stages or train stages. Even if you are happy with stages, you may agree that it is a contingent matter that material objects are composed of stages, and that versions of the paradox could still arise in worlds where things are otherwise. The stage theoretic solution then won't be a general solution. Remember also that there are modal versions of the paradox. If possible worlds are maximal propositions and a proposition may directly involve objects like Hubert Humphrey, then you may not want to analyse Humphrey's winning at some world w in terms of world-bound individuals and counterparts.

Can we keep the counterpart-theoretic solution to the paradox without taking on board the whole counterpart-theoretic metaphysics and semantics? We can. What solves the paradox is the idea that singular terms may undergo reference shift in the scope of intensional operators. This does not require that the referents are stages.

So here's my proposal. Terms for trains, ships and persons denote temporally extended objects, which may or may not divide into temporal parts. However, when we evaluate a sentence of the form 'at t, x is F', what matters is not whether the object denoted by 'x' is F relative to t. Rather, what matters is whether the t-counterpart of this object is F at t. Normally this makes no difference, because x will be its own sole counterpart at any time at which it exists, so we can ignore the counterpart-theoretic complication. But in a case of fission, for example, a single person x may have two counterparts at a future time t, and then the complication makes a difference.

To render all this more precise, let's return to the traditional four-dimensional picture, where ordinary objects are taken to be aggregates of stages. We saw that Lewis identified persons with maximal aggregates of stages all of which are unity-related to one another. I prefer an alternative proposal due to [Perry 1972] on which a person is a stage s together with all stages standing in the unity relation to s. Unlike Perry, however, I would like to exclude other stages at the time of s. So for any time t, define a *person at t* to be an aggregate of some person stage s at t together with all stages at non-t times that stand in the relation of personal unity to s. (Similarly for trains, ships, etc.) When x is a person at t and s is its t-stage, we say that s *determines* x.[6]

In easy cases, where no stage is unity-related to multiple stages at a single time t, this account gives the same verdict as Lewis's. But consider a case of fission. Every pre-fission stage s is unity-related to every other pre-fission stage as well as every post-fission stage. So there is only one person at the time t_1 of s. It is the Y-shaped aggregate consisting of s together with all its ancestors and descendants. At t_2, after the fission, there are two persons, corresponding to the two branches (including the stem) of the Y. The whole Y also exists at this point, but it is no longer a person. It has turned into a fusion of two persons.

Now we don't want to say that the person entering the duplication machine will be a fusion of two persons later. You might also want to say that personhood is not the kind of property that things can gain or lose: if x is a person at t_1 and still exists at t_2, then x must be a person at t_2. This is where the counterpart-theoretic semantics comes in. Suppose 'x' denotes the Y-shaped object that is a person at t_1, in the relational sense of 'person' just defined. In the scope of '*at* t_2', reference shifts to the two counterparts of this object, which are the two branches of the Y. These counterparts are persons at t_2. So in the scope of '*at* t_2', the term 'x' is referentially indeterminate between two things both of which count as persons at t_2.

The relevant counterpart relation can be defined in terms of unity: an individual y at t_2 is a counterpart of x at t_1 iff x is determined by a stage s_x at t_1, y is determined by a stage s_y at t_2, and s_y is unity-related to s_x. Note that counterparthood has become a

6 I exclude other stages at the time of s mainly to account for time-travel scenarios – see below. The more substantial disagreement between Perry and me is that he does not combine his metaphysics with a counterpart-theoretic semantics. Perry suggests that as a sentence operator, 'at t' shifts the referent of singular terms 'x' to the unique object determined at t by a stage unity-related to the present stage associated with 'x'. In a statement like 'x is F at t', however, 'at t' is treated as a *predicate modifier*, which is subject to different rules; the statement is true iff *some* t-stage of the object picked out by 'x' satisfies 'F'. [Moyer 2008] defends a simplified version of Perry's account on which 'at t' always works in this latter way. The Perry-Moyer account and mine largely agree on what is true in a case of fission, but they differ in a number of important respects. For one thing, I do not assume that singular terms are associated with stages, and my account is not limited to predicates that express properties of stages. (I dare say there are very few such predicates.) In particular, I do not interpret identity as identity of stages. If 'Goliath' denotes a certain statue and 'Slice' its stage at t (assuming there is such a stage, on which my account is officially neutral), then 'at t, Goliath = Slice' is false on my account, but true on the Perry-Moyer account. Statues, on my view, are not identical to temporally unextended stages. Relatedly, while the Perry-Moyer account entails that Leibniz' Law is generally invalid when temporal predications are involved, my account renders it valid, as long as the relevant terms are associated with the same counterpart relation. Finally, unlike the Perry-Moyer account, my solution to the paradox of occasional identity also carries over to Chisholm's paradox and other related puzzles.

doubly tensed relation involving two individuals and two times. The basic idea is simple: starting with a persisting individual x at t_1, you find the t_2-counterparts by collecting all t_2-stages unity-related to the t_1-stage of x, and expanding those stages by everything that's unity-related to them.

In the fission case, the Y-shaped object that is a person at t_1, before the fission, is its own sole counterpart at any other pre-fission time. Relative to a time t_2 after the fission, however, it is no longer its own counterpart, because it is not determined by any t_2-stage. The t_2-stages unity-related to the t_1-stage instead determine the two branches of the Y, which are therefore the t_2-counterparts of the person at t_1. Ordinary, non-fissioning objects are their own sole counterparts at all times. Here counterparthood reduces to identity.

Having defined a counterpart relation between temporally extended objects, we can simply take over the rest of the stage-theoretic semantics. Formally, the solution to the paradox remains the same. (1)–(4) will be true whenever the individual x at t_1 has two counterparts at t_2.

In contrast to stage theory, the new interpretation is not based on a revisionary identification of ordinary objects in fundamental reality. We no longer have to say that strictly speaking, the Earth only exists at this very moment. Planets, persons and trains are, strictly and literally, persisting objects that exist at various points in time. More importantly perhaps, what we normally think of as properties of persisting objects – being a philosopher, believing in unrestricted composition, being 50 years old – really are properties of persisting objects. We no longer have to explain in what sense a momentary person stage can be a 50 year old philosopher, or can be the subject of beliefs and desires.

The interpretation of modal and temporal operators also becomes more intuitive. If the present Humphrey stage is unity-related to a single stage at some time t at which Humphrey is winning an election, then 'Humphrey' denotes the same person inside and outside the scope of 'at t'; 'Humphrey wins at t' is true because at the relevant time, *this very person here* has the property of winning – just as [Kripke 1980: 45] and many others intuit. It doesn't even matter if Humphrey is subject to fission and fusion at other times. As long as the present Humphrey has only one counterpart at t, that counterpart will be Humphrey himself. The counterpart-theoretic machinery only springs into action when we directly deal with puzzle cases where, for example, an episode of fission has left behind two Humphreys at the relevant time t.

Even in such puzzle cases, the basic ideas behind the proposed interpretation are quite natural. Let's switch back to the more mundane example of the train. If our train x will turn into two trains at t_2, and we try to evaluate the claim that x is in Cologne at t_2, it really does seem like 'x' has acquired something like multiple denotations, or candidate denotations, relative to t_2. This is why it is tempting to "disambiguate" and introduce new terms for the two candidates in the scope of the relevant operator.

Unlike some other attempts to reconcile (1)–(4), notably [Myro 1985] and [Gallois 1998], my proposal does not tamper with identity. '=' denotes the familiar, untensed, two-place relation that figures in '$\sqrt{9} = 3$'. If '*at* $t_1, x = y$', and '*at* $t_2, x \neq y$' are both true, this is not because identity has become a three-place relation linking individuals and a time. It is because the operators '*at* t_1' and '*at* t_2' shift the reference of 'x' and 'y'. Coreference relative to t_1 therefore does not guarantee coreference relative to t_2.

So far, I have assumed a four-dimensionalist picture of stages and aggregates. It should be clear that much of this is dispensable. What matters is only that we have persisting objects that stand in counterpart relations to one another. In an endurantist framework, where ordinary objects do not divide into temporal parts, counterparthood cannot be defined in terms of unity between stages. The unity relation must now be understood as a relation between persisting objects at different times. For example, an endurantist might suggest that x at t_1 *is the same person as* y at t_2 iff there is a suitable chain of psychological and biological continuity leading from x's state at t_1 to y's state at t_2. As before, whether an informative analysis of personal unity is possible is immaterial to the present project. It's enough that we can somehow understand what it means to say that x at t_1 is the same person as y at t_2, without reducing it to a relation between stages. The counterpart relation, in this framework, is then simply the unity relation. Our counterpart-theoretic semantics will say that when an operator shifts the time of evaluation from t_1 to t_2, the reference of singular terms shifts to whatever stands at t_2 in the unity relation to the previous referents at t_1. Endurantists can therefore maintain that the pre-fission person at t_1 is unity-related to both persons at t_2, without clashing with the logic of identity.

To avoid terminological confusion, I use 'unity relation' only for the four-dimensionalist unity relation between stages (except in the preceding paragraph). The endurantist's doubly tensed surrogate for the unity relation is a counterpart relation, not a unity relation. For example, when I said that I assume the unity relation to be symmetrical, this does not apply to the endurantist unity relation, i.e. the counterpart relation.

I have argued that we can have the stage-theoretic solution to the paradox of occasional identity without committing ourselves to stages, nor to the stage-theoretic interpretation of ordinary terms and predicates. My proposal also has some other advantages, of which I want to briefly mention three.

First, in contrast to both stage theory and Lewis's aggregate theory, we immediately get the right answer to many counting questions. *One* person enters the duplication machine at t_1, *two* persons leave it at t_2; *one* person was in the preparation room in the ten minutes leading up to the duplication; *one* amoeba was be in the petri dish for the last hour. Stage theory and Lewis's aggregate theory need elaborate epicycles to deliver these verdicts. On the present account, they automatically come out right.[7]

7 Things get tricky when we try to count across an interval that involves fission or fusion. Lewis [1976:

Compared to stage theory, the present account also has fewer problems with names for non-present objects. Since Albert Einstein, for example, has no present stage, it is hard to see which of the uncountably many Einstein stages is supposed to be the referent of 'Einstein', and why names for present persons must denote present stages, while other names don't. On my account, 'Einstein' can simply denote the whole persisting person that existed from 1879 to 1955. (I will return to this topic in section 6.)

Another advantage is that my solution to the paradox of occasional identity nicely generalises to a number of related puzzles, including the Methusaleh puzzle discussed in [Lewis 1976] and its modal counterpart, "Chisholm's paradox" (from [Chisholm 1967]). It can also handle time travel cases. Suppose at some future time t, a time-traveler x is visiting her younger self. We can then say that in the scope of '*at* t', 'x' multiply refers to both persons determined by the time traveler's two stages at t. Here the difference to Perry comes into play: on Perry's account, there is only a single person at t, comprising both stages. The same is true for Lewis, because the two t-stages are part of one and the same maximal unity-interrelated aggregate. This makes it hard to interpret the kinds of things people say about time-travel cases: that the time-traveler *arrived at* t, *went to her old house*, *opened the door to her bedroom*, etc. On the account I have defended, these are handled much like in a scenario where the hero has fissioned into two people at t, one

73] asks how many persons there are in an interval during which x fissions into x_1 and x_2, and intuits that the answer should be 'two'. One difficulty here is that the Y-shaped object x ceases to be a person at the time of the fission, at which point its two branches x_1 and x_2 begin to be persons. Another complication is that x is composed of x_1 and x_2, and the standards for counting overlapping objects are murky. (If you own a building that is made of two smaller buildings, how many buildings do you own: one? two? three?) *Pace* Lewis, I do not think that there is an obviously correct answer to his question. Still, let's see what our account predicts.

Consider first a slightly simpler case. Suppose x is in room 101 before the fission, and afterwards x_1 is in 101 and x_2 in 102. How many persons are there in room 101 during the whole episode? One might think the present account predicts the answer to be 'two', because $x \neq x_1$. However, one might just as well say that predicted answer is 'one', on the grounds that '$x = x_1$' is true relative to every time in the interval. More precisely, the following statement is true for all t_1 and t_2 in the interval: 'at t_1, there is exactly one person x in room 101 such that at t_2, every person in room 101 is identical to x.' Arguably, this is sufficient for it to be true that there is exactly one person in the room during the interval.

Now suppose x_2 is also in the room after the fission. If certain details of the present account are filled in correctly, it then comes out true that at any pre-fission time, there is a single person in the room that is identical to every person in the room at any other time, while at post-fission times, there are two persons in the room such that any person in the room at other times is identical to one of them. The former would seem to entail that the total number of persons is one; the latter that it is two. Perhaps one should therefore say that the answer to Lewis's question is predicted to be indeterminate between 'one' and 'two'. One might also include 'three' as a legitimate answer by the reasoning that this is the total number over objects (including overlapping objects) that are persons at some time during the interval.

of which comes visiting the other.

So much for advantages. On the flip side, I suspect some will feel uneasy about the revisionary logic generated by my counterpart-theoretic semantics. I will try to defuse this worry in the next section. A more serious problem, in my view, is that the semantics sketched in section 2 is still inadequate in some crucial respects. Officially, it still rules out names for non-present objects. It also does not take multiple counterpart relations into account. I have mentioned in passing that the unity relation for persons may not be the same as that for trains. Different unity relations give rise to different counterpart relations, and such differences play an important role e.g. in Gibbard's [1975] puzzle of Lumpl and Goliath. These issues will be the topic of sections 5 and 6.

4 Occasional identity and classical logic

On the counterpart-theoretic semantics I have outlined, truth at a time is not closed under logical consequence: $at\ t, A$ and $at\ t, (A \supset B)$ do not entail $at\ t, B$. This may come as a surprise. It certainly contradicts standard principles of quantified hybrid logic. In standard hybrid logic, this entailment might be established as follows. Assume $at\ t, A$ and $at\ t, (A \supset B)$. By conjunction introduction, it follows that $at\ t, A \wedge at\ t, (A \supset B)$. Now the distribution principle of hybrid logic states that

$$at\ t, A \wedge at\ t, (A \supset B) \supset at\ t, B.$$

By modus ponens, it follows that $at\ t, B$. The counterpart-theoretic semantics must therefore invalidate either the distribution principle, or conjunction introduction, or modus ponens.

Which of these must go depends on certain details of the semantics that I haven't yet settled. The most important one is this. Suppose A is true at t on some, but not all, assignments of counterparts. I have suggested that we should then regard both $at\ t, A$ and $at\ t, \neg A$ as true. What about $\neg at\ t, A$? One could say that this is the negation of something true and hence false. But one could also classify it as true, on the grounds that there is an assignment of counterparts that makes $at\ t, A$ false and hence $\neg at\ t, A$ true, and truth on some assignment is enough for truth. On the first option, the distribution principle is invalid, as is (for example) the self-duality principle of hybrid logic: $at\ t, A \leftrightarrow \neg at\ t, \neg A$. Conjunction introduction and modus ponens are valid. On the second option, the distribution principle is valid, as is self-duality and conjunction introduction; but modus ponens becomes invalid: if some choice of counterparts verifies A and some choice verifies $\neg A$ and thus $A \supset B$, it may well be that no choice verifies B. This option, on which $A \wedge \neg A$ is satisfiable, determines a (quantified hybrid) version of Priest's "Logic of Paradox" LP (see e.g. [Priest 2006]).

Other principles of hybrid logic are invalid no matter how the details are filled in. The reduction principle '$at\ t_1, at\ t_2, A \leftrightarrow at\ t_2, A$', for example, fails because '$at\ t_1, at\ t_2, x$ is F' only requires that some t_2-counterpart of some t_1-counterpart of x is F, which doesn't entail that any t_1-counterpart of x is F.

In section 2, I mentioned a few alternatives to my proposal. For example, one might require truth on *all* assignments instead of truth on *some* assignments, or one might disallow mixed assignments, requiring that different occurrences of the same term in the scope of $at\ t$ must always be assigned the same object. Most of these alternatives would still allow for a resolution of the paradox. For example, if you dislike mixed assignments and think that truth should require truth on *all* assignments, then (6*)–(8*) still come out true. (5*) is false, but you might argue that it doesn't follow from (3), the claim that x_1 and x_2 have equal claim to be x: what this rules out, you might argue, is merely that $at\ t_2, x = x_1 \wedge x \neq x_2$ or that $at\ t_2, x = x_2 \wedge x \neq x_1$, and these do both come out false. In general, that x_1 and x_2 have equal claim to be x is validated in one way or another by any semantics that treats x_1 and x_2 symmetrically.

Would these alternatives secure a more conservative logic? Not really. For example, if truth of $at\ t, A$ requires truth on every uniform assignment (and $\neg at\ t, A$ is true iff $at\ t, A$ is not true), then self-duality and the reduction principle, among other things, are invalid. Likewise for a Russellian analysis that counts $at\ t, \Phi(x)$ as false whenever x has multiple counterparts at t.

What should we make of this logical deviance? In recent discussions of counterpart theory, it is often considered a fatal flaw. Fara and Williamson [2005], for example, argue that by invalidating principles like self-duality or distribution, counterpart theory allows 'contradictions' to be true, and is therefore unacceptable. As I mentioned above, proponents of counterpart theory such as Ramachandran [2008], Sider [2008] and Russell [2011] have consequently gone to great lengths looking for counterpart-theoretic interpretations that avoid the logical deviance, typically by ruling out the possibility of multiple counterparts.

This reaction strikes me as misguided. First of all, the non-classicality is precisely what solves the paradox of occasional identity. In standard quantified hybrid logic, (5*)–(8*) are inconsistent, as is every other sensible formalisation of (1)–(4). We would be back at the point where we have to reject one of the four assumptions. I think the logical deviance should be considered a *virtue*, as it mirrors what we find in ordinary thought and talk. And remember, our puzzle essentially concerns ordinary thought and talk. I fully agree that we *could* say everything that is worth saying about the puzzle cases in a completely classical language. At it may well be advisable to talk in such a "tidied-up" regimentation of English if we're doing systematic metaphysics. But none of this has any bearing on the fact that in ordinary English, (1)–(4) can all be true.

More importantly, even a staunch proponent of classical logic should be happy to admit that in a language with ambiguous terms, statements of the form $A \wedge \neg A$ may be counted as true – not because the world both is and is not a certain way, but because the first part of the sentence is made true by one reading of A and the second by another. [Lewis 1982], for example, defends LP as a logic of ambiguity, and Lewis is not generally regarded as a card-carrying dialethist.

A case of multiple counterparts is in many ways like a case of ambiguity. If we opt for the LP-style account that rejects modus ponens, we will accept the satisfiability of "blatant contradictions" like $at\ t, Fx \wedge \neg at\ t, Fx$ in situations where x has two counterparts at t. But before we call the logic police, we should check what this sentence is supposed to mean. Does it commit us to the existence of an object x with truly contradictory properties? Not at all. The sentence is true because the term 'x' has multiple denotations in the scope of '$at\ t$', one of which verifies $at\ t, Fx$ and another that verifies $\neg at\ t, Fx$. The contradiction is as harmless and superficial as that between 'London is in Canada' and 'London is not in Canada', which may well both be true.

In the face of ambiguity, classical logic can be restored by disambiguating the relevant terms: London_1 is in Canada, London_2 is in the UK. In practice, disambiguation is often achieved by conversational context and speaker intentions. When someone says 'London is in Canada', it is usually clear which London is meant. But not always. If someone who has never heard of either London watches a news report about London, Ontario, and then reads an article about London, UK, without realising that the two cities are distinct, their use of 'London' may well be ambiguous between the two Londons, and neither context nor speaker intention will resolve the ambiguity. Arguably, 'jade' and 'mass' were ambiguous in everyone's usage until scientific investigation revealed that the supposed referent is actually two, rather different, things (see [Field 1973]). Multiplicity of counterparts is more like those latter examples in that it is rarely resolved by context or speaker intentions. To "disambiguate" a sentence like 'this train will be in Cologne', you would have to replace 'train' by an artificial sortal like 'train-qua-turning-South', associated with a different unity relation.

I'm not saying that the case of multiple counterparthood *is* a case of ambiguity. 'This train' (uttered in Berlin) is not ambiguous between 'this-train-qua-turning-South' and 'this-train-qua-turning-North'. 'This train' unambiguously denotes the Y-shaped object, while the other two terms pick out its branches. It is therefore not out of the question that ambiguity should receive a different treatment than multiple counterparthood. If you prefer, say, a paraconsistent account of ambiguity and a supervaluationist account of multiple counterparts, I haven't said much to dissuade you. What matters for the present point is merely that both phenomena involve some kind of multiple reference and thereby give rise to non-standard logics.[8]

8 Another phenomenon that is in some ways *like* ambiguity is vagueness, where multiplicity of reference shows up as multiple ways of rendering vague expressions precise. Paraconsistent accounts of vagueness

Principles like modus ponens or self-duality are uncontestable only if we ignore cases of multiple reference. When multiply referring terms are involved, we should expect principles of classical logic to break down. This semantic explanation is crucial to my defense of occasional identity. Here it contrasts with the accounts of Gallois [1998] and Myro [1985], who also reject various parts of standard modal logic in their defense of occasional identity, but without offering any defusing semantic explanation. We have to take it as a brute fact that those principles fail. On the present account, you may well say that reality itself is completely classical. Nothing is ever both F and non-F. The non-classicality only arises in certain idioms of ordinary language where terms can take on multiple reference.

5 Statues, lumps, and multiple unity relations

Remember the story of Lumpl and Goliath, in its temporal version. Lumpl, a piece of clay, has been formed into a statue, Goliath. Lumpl and Goliath now occupy the very same space on a shelf. Some philosophers conclude that Lumpl has become identical to Goliath, for different material objects should not be at the same place at the same time. On the other hand, Lumpl is older than Goliath, since it already existed before it was shaped into the statue. And if x is older than y, then it is hard to see how x and y could be identical.

Within a four-dimensional metaphysics, identifying Lumpl and Goliath goes most naturally with stage theory, while aggregate theories tend to treat them as two. On the aggregate view, the present stage of Lumpl is identical to the present stage of Goliath, which explains how the two can be at the same place at the same time. Lumpl and Goliath are not identical, but also not *distinct*, mereologically speaking: Goliath is a temporal part of Lumpl. In stage theory, on the other hand, the identity of present stages suggests that Lumpl and Goliath really are a single object. The fact that Lumpl is older than Goliath then has to be explained by appealing to different counterpart relations associated with the two names. The name 'Lumpl', so the idea, goes with a counterpart relation that emphasises continuity of matter (in some sense), while the counterpart relation associated with 'Goliath' puts more weight on continuity of (say) shape. If we then analyse 'x is older than y' as 'x has earlier temporal counterparts than y', it follows that Lumpl is older than Goliath, despite the fact that they are identical.

My own account sides with the aggregate view. Assuming four-dimensionalism, the present Lumpl-and-Goliath stage is linked to the early stages of Lumpl by the unity relation for pieces of clay, but not by the unity relation for statues. So the *lump* determined

have been defended e.g. in [Hyde 1996] and [Ripley 2011]. It is not a coincidence that the only major account of vagueness that retains classical logic, epistemicism, denies that vagueness involves any sort of multiple reference.

by the present stage is not identical to the *statue* determined by that stage. Since ‘Lumpl’ denotes the lump and ‘Goliath’ the statue, the two are not identical, although they share their present stage.

The case of Lumpl and Goliath is therefore not a case of occasional identity. In a genuine case of occasional identity, one and the same unity relation relates a stage at one time to multiple stages at another time. In the example of the train, the train’s Berlin-stage is train-united both with a later stage in Cologne and with one in Düsseldorf; in the case of personal fission, the person stage entering the duplication machine is person-united with both stages emerging from the machine. There is *one train* and *one person* at the initial time t_1, and several at t_2. Not so in the case of Lumpl and Goliath. To be sure, we have *one statue* and *one piece of clay*. But this doesn’t mean that the statue is identical to the piece of clay.

Admittedly, we might want to say that there is only *one material object* on the shelf. However, this would seem to suggest not only that Goliath is identical to Lumpl, but also that Goliath’s left half is identical to his right half – both of these are material objects, and both are on the shelf. When we say that the number of material objects is one, we are somehow ignoring the two halfs. Perhaps what we count are *maximal (ordinary)* material objects. In this case, Goliath, being a proper part of Lumpl, also doesn’t count. Or perhaps we count aggregates of matter that in some sense “compose a whole” at the relevant time. Relatedly, one might argue that ‘there is one material object’ comes out true for the same reason as ‘there is one piece of clay’ and ‘there is one statue’: whatever unity relation is associated with ‘material object’ (one may suspect this is not a very determinate matter), there is only one object determined by any given stage along this relation. The general point is that counting material objects is a messy business – much more so than counting people or trains. If Lumpl and Goliath are two material objects, then *in some sense*, the number of material objects on the shelf must be greater than one. But there clearly is such a sense, as witnessed by the statue’s many material parts. There is also a sense in which the number of objects on the shelf is one, and I’ve just sketched a few ideas on what that sense might be. The fact that there is only one material object on the shelf therefore doesn’t undermine the non-identity of Lumpl and Goliath.

In recent discussions, Lumpl and Goliath often figure as the standard example of occasional identity. This is unfortunate, not only because the example isn’t one of occasional identity at all. It also makes the puzzle appear much easier than it is. There are plenty of good solutions to *this* puzzle: appealing to multiple counterpart relations, Carnapian intensions (as in [Gibbard 1975]), or rejecting the identity in favour of stage-sharing or “constitution”. None of this looks very promising when it comes to fissioning trains or people or amoebae.

Nevertheless, the case of Lumpl and Goliath is instructive, because it draws attention to the fact that there are different, and sometimes competing, unity relations. These are often associated with specific sortals: 'train', 'statue', 'person'. But there is no good reason to think that the supply of sortal nouns in English exhausts the unity relations. Perhaps when we say that Pierre believes that Hesperus is not Phosphorus, 'Hesperus' and 'Phosphorus' are associated with different modal unity relations. For an object in one of Pierre's doxastically possible worlds to count as Hesperus it must play the Hesperus role; for it to count as Phosphorus it must play the Phosphorus role. These roles do not correspond to different sortals in English. When dealing with cases of occasional identity, it seems that we can also introduce new sortals, speaking of the ship-*qua*-assembly-of-those-planks, or the train-*qua*-turning-South, as I did in the previous section. These pick out specific branches of the Y-shaped object even before the fission.

Since counterparthood can be defined in terms of unity, the multiplicity of unity relations gives rise to a multiplicity of counterpart relations. Often the relevant counterpart relation can be read off from the denoted objects. If something consists of all stages lump-united with a present stage, but not of all stages statue-united with that stage, then it is a lump and not a statue, and so the relevant counterpart relation should be the relation for lumps and not the one for statues. However, this doesn't work if an object's stages happen to be united by several unity relations. (More precisely, if the object is determined by a given stage along several unity relations.) For example, the train that arrives in Cologne at t_2 consists of the very same stages as the train-qua-turning-South that arrives in Cologne; but only the former has the Y-shaped train as counterpart at t_1. Similarly, one might argue that the unity relation for human persons puts greater emphasis on psychological continuity than the relation for human bodies, but that the two only come apart in unusual cases involving things like brain transplantation. The fact that a name picks out an ordinary person then doesn't tell us whether the counterpart relation associated with the name is the one for persons or the one for bodies. This difference can matter e.g. for the evaluation of doubly shifted predications '$at\ t_1, at\ t_2, x$ is F'.[9]

So to evaluate the truth-conditions of sentences involving intensional operators, we need to know which unity or counterpart relation goes with which referring expression. Lewis ([1971], [2003]) and Sider [2001] seem to suggest that is a matter of pragmatics: 'this person' and 'this body' have the very same semantic value, but the choice of words makes a different counterpart relation salient. To interpret 'this person, but not this

9 If my metaphysical proposal is extended to the modal dimension (which might require a higher degree of modal realism than I am comfortable with), the idea that the unity or counterpart relation associated with a term can be determined by its reference becomes more plausible. The reference then includes the relevant object at other possible worlds, which will pull apart 'this person' and 'this body'. It doesn't follow that we couldn't still have co-reference with different unity relations. The relevant unity relations wouldn't have to be co-intensional, for they might come apart for other individuals. Nevertheless, a purely referential account is not out of the question once we include the modal dimension. We could then define counterparthood as a relation between ordinary objects (referents of singular terms), unrelativised by times or worlds or sorts.

body, could survive destruction of its brain' we have to postulate somewhat peculiar mid-sentence switches in utterance context.

A rather more systematic approach, defended by [Geach 1962], [Gibbard 1975] and others, treats the unity relation associated with a term as part of the term's semantic value. The semantic value of a term like 'Lumpl' or 'this person' thus has at least two components: a *referent* and a *sort*. The referent is the object (if any) picked out by the term; the sort tells us whether this object is picked out *qua* statue, *qua* body, *qua* person, etc. It doesn't really matter how the sort achieves this, or what kind of entity it is; the simplest proposal is perhaps to identify sorts with unity relations (or counterpart relations). Like Fregean senses and Carnapian intensions, sorts affect a term's reference at other times and worlds. Unlike senses and intensions, however, sorts fall far short of *determining* reference: merely knowing that a name denotes a person is not enough to figure out its referent, neither here nor at other worlds or times. As Gibbard [1975: 197f.] points out, associating names with sorts as well as referents goes quite naturally with a Kripkean picture of names.[10]

Having associated singular terms with sorts and thereby with counterpart relations, it is straightforward to adjust the counterpart-theoretic semantics. When an intensional operator shifts the evaluation time from t_1 to t_2, the reference of any term in its scope shifts to the counterpart of the previous referent(s), along the counterpart relation associated with the term.

We can also reformulate the semantics in terms of an untensed, two-place counterpart relation. Relative to any given sort, say that y is a *counterpart (simpliciter)* of x iff there are stages s_x, s_y such that s_x determines x, s_y determines y, and s_x stands in the relevant unity relation to s_y. This is just the doubly tensed definition from section 3 with the two times bound by existential quantifiers. (Hence the concept is easily adopted to endurantism. For example, a person y is a counterpart of a person x iff there are times t_1, t_2 such that y at t_2 is the same person as x at t_1.)

In a case of personal fission, the Y-shaped pre-fission person has three counterparts simpliciter: the two branches and itself. To ensure that in the scope of '*at* t_2' only the

10 Gibbard himself defends an "intensional" semantics, in which names are associated with non-trivial functions from worlds or times to referents. This kind of approach is often considered one of the main alternatives to counterpart-theoretic semantics (see e.g. [Fitting 2004], [Kracht and Kutz 2007]). Note that within a four-dimensional setting, my counterpart-theoretic semantics determines *two* intension-like entities. First there is the mapping from times to stages, which is folded into a term's extension. Second, there is the mapping from times to counterparts. This is not part of the term's extension, but determined by the extension and the sort. Neither of these "mappings" are functions from times to individuals: a person can have multiple stages as well as multiple counterparts at t. The intensional account rules this out, which is why it cannot handle genuine paradoxes of occasional identity (although it nicely handles the case of Lumpl and Goliath). Intensional accounts also tend to collapse '*at* t_1, *at* t_2' into '*at* t_2', which blocks the counterpart-theoretic solution to Chisholm's paradox.

two branches count, we stipulate that if a term x is associated with a particular sort, then it always denotes objects of that sort, even in the scope of temporal operators. At t_2, the Y-shaped object is therefore ineligible, because it is no longer a person. That is, it is not determined along the unity relation for persons by any stage at t_2 (because there is no t_2-stage such that the Y-shaped object consists of that stage together with everything unity-related to it).

In general, let's say that an object *belongs to* a sort σ at a time t iff it is determined by a t-stage along the unity relation for σ. Thus relative to any sort σ, y is an eligible counterpart of x at t_2 iff (i) y is a counterpart of x, and (ii) y belongs to σ at t_2. If a temporal operator shifts the time of evaluation from t_1 to t_2, then the reference of singular term shifts from objects that belong to the relevant sort at t_1 to those of their counterparts that belong to that sort at t_2. This is exactly the shift generated by the doubly tensed counterpart relation from section 3. The two formulations are equivalent. The reformulated semantics will, however, make it easier to add names for non-present objects – as we will see next.

6 Things that aren't present

So far I have assumed that names can only denote presently existing objects. This is a common assumption in counterpart-theoretic semantics, as well as modal logic more generally, since names for non-present (or non-actual) objects cause all sorts of trouble.[11] Nevertheless, the assumption severely limits the applicability of the account, as ordinary language is replete with names like 'Albert Einstein' for objects in the past.

As I mentioned in section 3, one notorious problem raised by such *non-local* names in counterpart semantics disappears on my proposal: we are not forced to choose a particular stage as the referent of 'Einstein'. 'Einstein' can refer to a complete persisting individual in the past. But that's not the end of the story. We also want to say, for example, that in 1922, Einstein won a Nobel price. So 'Einstein' should refer to Einstein in the scope of 'in 1922'. On the counterpart-theoretic semantics, 'in 1922' shifts the reference of singular terms to the 1922-counterparts of their initial reference. So if 'Einstein' refers to Einstein outside the scope of any such operator, Einstein must count as a 1922-counterpart of himself today. That is, we have to allow for things to be counterparts at t_2 of themselves at t_1 even if they do not exist at t_1. Perhaps the easiest way to achieve this would be to extend the doubly tensed definition of counterparthood so that if no ordinary, existing individual at a time t_1 has x at t_2 as its counterpart, then x itself counts as having, at t_1,

11 Indeed, names by themselves already cause trouble; this is why [Lewis 1968], like [Kripke 1963] and many others, considers only languages without individual constants. Some problems specific to names for non-present (or non-actual) objects in the context of counterpart-theoretic semantics are discussed in [Ramachandran 1990a], [Ramachandran 1990b] and [Forbes 1990].

x at t_2 as its counterpart. If counterparthood is symmetrical, this amounts to stipulating that if something doesn't have any existing counterparts at some time, then it shall count as its own counterpart at that time. (See [Forbes 1982], [Forbes 1985] for a proposal along these lines.)

What if Einstein had undergone fission or fusion? Suppose Einstein fissioned into two people in 1955, using a duplication machine secretly developed by John von Neumann. There are then three candidates for the reference of 'Einstein': the Y-shaped object that was a person before the fission, and the two branches that were persons after the fission. Which of these does the name pick out? I suppose it would be the Y-shaped object, although we might of course also have a name for one of the branches.

So far, so good. But now suppose the two Einstein successors are still alive. Then there *are* ordinary, existing objects at the present, namely the two branches, which have the original Einstein as their counterpart in 1922. So the above extension doesn't kick in. If 'Einstein' is to denote the Y-shaped person in the scope of 'in 1922', it now has to denote one of its two successors. We seem forced to conclude that there can be names for genuinely past persons only if they do not have present counterparts. If Einstein's successors all happen to have deceased, 'Einstein' can refer to the person in 1922. It cannot refer to such a person if one of the successors is still around. This doesn't sound right.

The problem is the assumption that 'Einstein' must denote something that *presently* represents Einstein via the counterpart relation. Let's drop this assumption, and with it the artificial stipulation that everything is its own counterpart at times when it doesn't exist. Recall the two-place counterpart relation defined at the end of the previous section. On this definition, the Y-shaped Einstein of 1922 has three counterparts in the fission scenario: himself and his two branches. In the scope of '*at t*', reference shifts to those counterparts of the previous reference that belong to the right sort at t. So if 'Einstein' denotes the Y-shaped object, then in the scope of 'in 1922', it denotes those counterparts of the Y-shaped object that are persons in 1922. The only such counterpart is the Y-shaped object itself. So 'Einstein' denotes the 1922 person both inside and outside the scope of 'in 1922' – irrespective of whether he has present counterparts. That's exactly what we wanted.

What's true for names is also true for variables. 'In 1922, there was a person x who entered a duplication machine in 1955 ...' – here the variable x is introduced in the scope of 'in 1922' to stand for a certain person in 1922. The value of x here is Y-shaped person in 1922.

What happens when we predicate something in present tense of a term denoting a past object, as in 'Einstein is dead' or 'Einstein is asleep'? One possibility is to treat the present tense as a temporal operator that shifts reference to present counterpart(s). Present-tense predications would then be interpreted as if prefixed by 'at the present' (or ‘now’). Names for entirely past objects, like ‘Einstein’, would be empty in statements like ‘Einstein is dead’. Worse, suppose in the fission scenario, one of the two Einstein successors comes to live in Cologne and the other in Düsseldorf, and ‘Ed’ is introduced for the successor in Düsseldorf. Then ‘in 1960, Ed lives in Cologne’ would be true when uttered before the fission, because in the present-tense statement, ‘Ed’ would shift to denote the pre-fission person. So I’m inclined to say that reference doesn’t shift for unembedded names. (On the other hand, this means that ‘Einstein lives in Düsseldorf’ wouldn’t be true when uttered in 1960, nor would ‘Einstein is a person’. One might consider allowing both the shifted and unshifted interpretation, leaving the choice to conversational context.)[12]

If you’re into that kind of thing, the appendix gives a rigorous statement of the full semantics, in a form that is neutral on metaphysical issues such as four-dimensionalism. The point of this exercise is mainly to illustrate how the ideas sketched in this paper can be put together into a precise semantics for a rich fragment of ordinary language. Of course, much more would need to be said to defend this particular interpretation, and I have no doubts that it could be improved in various ways. More would also need to be said to apply my proposal directly to English, rather than the language of quantified hybrid logic.[13] But we should not lose sight of the big picture. In some form or other, a counterpart-theoretic semantics along the lines I have defended can provide an attractive solution to the paradox of occasional identity, without committing us to controversial metaphysics or twisting our words.

12 [Priest 1995] gives a semantics for a (non-modal) first-order language with ambiguous terms that closely resembles my own semantics for sentences in the scope of intensional operators. Priest also suggests an application to occasional identity: if ‘x’ denotes an amoeba that fissions into y and z, then after the fission, the name denotes both y and z. Although Priest doesn’t consider the question, it is natural to assume that operators like ‘at t_2’ would likewise render occurrences of ‘x’ in their scope ambiguous, even before the fission. This would lead to something much like the semantics defended here. Priest’s fission puzzle is different from the puzzles I have discussed in that he looks at unembedded sentences involving ‘x’ uttered at times after the original referent has fissioned. Priest assumes that reference here shifts to the post-fission amoebae. The name becomes genuinely ambiguous. On the account I favour, ‘x’ still refers to the Y-shaped object that was an amoeba before the fission. This difference is responsible for the fact that Priest’s account, but not mine, invalidates principles of LP (and classical logic) such as existential generalisation; see [Priest 1995: 365f.].

13 Some philosophers, e.g. [King 2003], seem to suggest that in English, expressions like ‘always’ (and perhaps also ‘at t’) are not modal operators, but quantifiers over times. This might cast doubt on whether my proposal for hybrid logic carries over to English. I do not understand this alleged distinction, insofar as it is supposed to go beyond pure syntax. From a semantic perspective, the operators of temporal logic *are* of course quantifiers over times. Indeed, the standard modal semantics for sentences like ‘$\Box p$’ essentially *is* also an extensional semantics for the corresponding language with sentences ‘$\forall x Px$’ (see [van Benthem 2001]). Likewise for the semantics defended here. The model theory in the appendix applies just as much to languages in which temporal operators are written as first-order quantifiers binding a temporal variable in the complement sentence.

7 Whither counterpart theory?

Counterpart theory (including its temporal incarnation, stage theory) can explain how seemingly paradoxical statements like (1)–(4) may all be true, and thereby solves the puzzle of occasional identity. But the solution comes at a high price. I have argued that we can get away cheaper by hijacking the core idea in the counterpart-theoretic semantics and embedding it into more conservative semantic and metaphysical assumptions.

Now counterpart theory is not advertised merely, or even primarily, by its solution to the paradox of occasional identity. Among other things, it is also said to solve the problem of accidental and temporary intrinsics, explain the indeterminacy and context-dependence of essentialist judgments, and solve paradoxes of coincidence like the case of Lumpl and Goliath ([Lewis 1986: 198–204, 248–259], [Sider 2001: chs.4–5]). I have focused on the paradox of occasional identity because I think it provides the strongest support specifically for counterpart theory. The other advantages can more easily be matched by alternative accounts.

With respect to paradoxes of coincidence, I have argued in section 5 that a view on which ordinary objects are extended and Lumpl is not identical to Goliath is at least as plausible as the counterpart-theoretic alternative. I won't say much about the problem of accidental and temporary intrinsics. If there is a genuine problem here (which is far from obvious), it may indeed require the existence of stages and world-bound individuals to serve as primary bearers of certain intrinsic properties. But this doesn't entail that ordinary trains and people should be identified with stages or world-bound individuals.

The vagueness and context-dependence of essentialist judgments, it seems to me, can just as well be explained by providing a multitude of potential referents, rather than a multitude of potential counterparts. When counterpart theory says that my use of 'this cup' determinately denotes a particular object, but doesn't fully settle whether this or that object at other worlds or times qualifies as its counterpart, we can instead say that the term is indeterminate between various cross-world or cross-time individuals.

This is not to say that other, more metaphysical considerations couldn't swing the balance back in favour of counterpart theory. For example, if you think that material objects are nothing but aggregates of (time-slices of) particles, and that there are no Lewisian possible worlds that could provide particles beyond those provided by actuality, then you may want to identify statues and lumps if they coincide at every time, as there would then seem to be only one aggregate of relevant particle slices. This might support some kind of counterpart theory about modality, perhaps along the lines of [Sider 2002]. I suppose extreme presentists and solipsists could run a similar argument for the temporal and spatial dimensions.

My aim in this paper was not to sell you my metaphysics. I mostly want to sell you my semantics. Even if you're a committed counterpart theorist, you might like to swap

Lewis's original rules for a more comprehensive semantics that can handle talk about specific times and worlds, non-local names, and multiple counterpart relations. The semantics presented in the appendix is compatible with genuine counterpart theory. More importantly, however, I also want to sell you my semantics if you're *not* a counterpart theorist. You can have almost all the advantages at almost none of the costs.

Appendix: Details of the semantics

I will give a precise formulation of the semantics described in the paper. To keep the task manageable, it is not a semantics for English, but for the language $\mathcal{L}$ of quantified hybrid logic, into which the relevant statements of English first have to be regimented.

Sentences of $\mathcal{L}$ are constructed in the usual manner from infinitely many *individual variables and constants* x, y, z, etc., *predicates* $=, F, G, H$, etc. (each associated with an *arity* $\in \mathbb{N}$), *nominals* a, b, c, etc., the boolean connectives $\neg$ and $\supset$, the quantifier $\forall$, the monadic sentence operator $\Box$ and the binary sentence operator *at* that takes a nominal as its first argument and an arbitrary sentence as its second argument.

Expressions of $\mathcal{L}$ are interpreted in *$\mathcal{L}$-structures*, or simply *structures*. A structure has the following components:

1. a non-empty set W (the *points of evaluation*, e.g. worlds or times),
2. a designated member α of W (the actual world, present time, etc.),
3. a binary relation R on W (the *accessibility* relation),
4. a non-empty set S (the *sorts* or *atomic types*),
5. a family D of sets, indexed by $W \times S$ (the world-relative *domains* of each sort); I write D_w for $\cup_{s \in S} D_{w,s}$ (the total domain of w) and D^s for $\cup_{w \in W} D_{w,s}$ (the global domain of sort s),
6. a family C of relations, indexed by S, such that $C_s \subseteq D^s \times D^s$ (the *counterpart relation for sort s*).

An *interpretation* of $\mathcal{L}$ on such a structure is a function V that assigns

1. to each singular term x a pair of a sort and an individual of that sort, i.e. a member of $\{\langle s, d \rangle : s \in S, d \in D^s\}$; I write $[x]^V$ for the individual and $\{x\}^V$ for the sort,
2. to each $w \in W$ and n-ary predicate F, an n-ary relation $V(w, F)$ on D_w, subject to the restriction that $V(w, =) = \{\langle d, d \rangle : d \in D_w\}$,
3. to each nominal a a member of W.[14]

14 Traditionally, nominals are treated as sentences, i.e. zero-place predicates. To this end, the third clause would be replaced by a further restriction on the interpretation of predicates so that for any nominal a, $V(w, a) = 1$ for exactly one world w. Here $1 = \{\emptyset\} = \{\langle\rangle\}$ is the unit set of the empty tuple, which is the only zero-tuple from D_w. (In general, a zero-ary predicate denotes, relative to a world w, either $\emptyset = 0 =$ *False*, or $\{\langle\rangle\} = \{\emptyset\} = 1 =$ *True*, which are the two zero-ary relations on

An ($\mathcal{L}$-)*model* is a structure $\mathcal{S}$ together with an interpretation on $\mathcal{S}$.

Note that sorts are like one-place predicates in that a model associates with each sort s and world w a subset of D_w. You could add that every sort should correspond to a predicate (the corresponding "sortal") with the same range of extensions, but I haven't built that into the semantics.

The division of labour between interpretations, structures, and syntactic properties of the language is a bit arbitrary and not meant to carry any great significance. For example, the assignment of sorts to variables (and therefore the set S) could be moved into the syntax, yielding a more traditional many-sorted language. Since the nature of the sorts is unimportant, we could also identify the sorts with the associated counterpart relations.[15]

Singular terms are non-local ("possibilist"): they may refer to individuals outside the domain of the designated point α. Points of evaluation may have empty domains, but at least one point must harbour at least one individual, to serve as referent of every singular term. Predicates, unlike names, are local in the sense that if something satisfies an atomic predicate at a point w, then it must be in the domain of w. This means that we can't have atomic predicates for properties like 'having existed in 1879', though by adding lambda-abstraction, we could still have a corresponding non-atomic predicate. If you prefer, you may lift the locality restriction on predicates; it doesn't do much work.

To complete the semantics, we need to specify under which conditions a sentence is true in a model. To this end, let's first define a couple of relations between interpretations.

Let V, V' be interpretations on a structure $\mathcal{S} = \langle W, \alpha, R, S, D, C \rangle$. For any variable x and world $w \in W$, V' is an *x-variant of V on w* (for short: $V \overset{x}{\underset{w}{\sim}} V'$) if V' differs from V at most with respect to $[x]$, and $[x]^{V'} \in D_w$. V' is a *w-image of V* (for short: $V \overset{w}{\rightsquigarrow} V'$) if (i) V and V' agree on all predicates, and (ii) for every singular term x with $s = \{x\}^V$, if there is an individual $d \in D_{w,s}$ such that $[x]^V C_s d$, then $V'(x) = \langle s, d \rangle$ for some such d, otherwise $V'(x)$ is undefined. So an x-variant of V on w is like V except that it may assign some other individual (of the same sort) in D_w to x. A w-image of V is like V except that the reference of all singular term is shifted to a counterpart at w (of the right sort) of their reference under V. If there is no such counterpart, the term becomes empty.

If the counterpart relation is *functional* in that nothing has multiple counterparts at

D_w.) For simplicity, I identify $V(a)$ with the unique world at which a is true and ignore occurrences of nominals anywhere else than at the first argument-place of *at* .

15 Entirely skipping the sorts and merely associating terms with a counterpart relation wouldn't suffice, unless we make the counterpart relations tensed. For example, in a fission case, Y has three counterparts Y, L and R. If x initially denotes Y, and we want to evaluate it at a post-fission time, we need to find things at that time which are (i) counterparts of Y, and (ii) of the right sort. Condition (ii) rules out Y, and for this, we need to know which things are of which sort at t, even if there is no predicate associated with the sort. Making counterparthood tensed would help, because we could then say that Y has only L and R as counterparts at t_2, and itself as counterpart at t_1.

any world w, then there is just one w-image of V. We could then say that *at* w, A is true under interpretation V iff A is true under the w-image V' of V.[16] If there are several w-images, we face the choice of whether to require truth on *some* of them or truth on *all* of them, and whether the multiplicity of denotations within modal operators can spread to yield multiplicity of semantic values for the embedding expression. Here I give a semantics that validates an LP-style logic, from which it is easy to generate other variations.

On this account, if there are several w-images of V that assign different referents to x, so that x has multiple semantic values in the scope of *at* w, then *at* w, Fx can also have multiple semantic values: it may be true on some choice of counterparts and false on others. For the definition of truth in a model, we therefore define two relations $\Vdash_1$ and $\Vdash_0$ between a structure $\mathcal{S} = \langle W, \alpha, R, S, D, C \rangle$, a point $w \in W$, a set $\mathbb{V}$ of interpretations on $\mathcal{S}$, and a sentence A. $\Vdash_1$ captures truth on some choice of counterparts, $\Vdash_0$ captures falsehood on some choice of counterparts, and $\mathbb{V}$ may contain various images of an interpretation function. The clauses for atomic sentences reflect the choice to allow for "mixed disambiguations".

$\mathcal{S}, \mathbb{V}, w \Vdash_1 Fx_1 \ldots x_n$	iff there are (not necessarily distinct) $V_1 \ldots V_n \in \mathbb{V}$ such that $\langle [x_1]^{V_1}, \ldots, [x_n]^{V_n} \rangle \in V_1(w, F)$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0 Fx_1 \ldots x_n$	iff there are $V_1 \ldots V_n \in \mathbb{V}$ such that it is not the case that $\langle [x_1]^{V_1}, \ldots, [x_n]^{V_n} \rangle \in V_1(w, F)$.[17]
$\mathcal{S}, \mathbb{V}, w \Vdash_1 \neg A$	iff $\mathcal{S}, \mathbb{V}, w \Vdash_0 A$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0 \neg A$	iff $\mathcal{S}, \mathbb{V}, w \Vdash_1 A$.
$\mathcal{S}, \mathbb{V}, w \Vdash_1 A \wedge B$	iff $\mathcal{S}, \mathbb{V}, w \Vdash_1 A$ and $\mathcal{S}, \mathbb{V}, w \Vdash_1 B$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0 A \wedge B$	iff $\mathcal{S}, \mathbb{V}, w \Vdash_0 A$ or $\mathcal{S}, \mathbb{V}, w \Vdash_0 B$.
$\mathcal{S}, \mathbb{V}, w \Vdash_1 \forall x A$	iff $\mathcal{S}, \mathbb{V}', w \Vdash_1 A$ for $\mathbb{V}' = \{V' : V \underset{w}{\overset{x}{\sim}} V' \text{ for some } V \in \mathbb{V}\}$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0 \forall x A$	iff $\mathcal{S}, \mathbb{V}', w' \Vdash_0 A$ for $\mathbb{V}' = \{V' : V \underset{w}{\overset{x}{\sim}} V' \text{ for some } V \in \mathbb{V}\}$.
$\mathcal{S}, \mathbb{V}, w \Vdash_1$ ***at*** a, A	iff $\mathcal{S}, \mathbb{V}', V(a) \Vdash_1 A$, for V any member of $\mathbb{V}$ and $\mathbb{V}' = \{V' : V \overset{V(a)}{\rightsquigarrow} V' \text{ for some } V \in \mathbb{V}\}$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0$ ***at*** a, A	iff $\mathcal{S}, \mathbb{V}', V(a) \Vdash_0 A$, for V any member of $\mathbb{V}$ and $\mathbb{V}' = \{V' : V \overset{V(a)}{\rightsquigarrow} V' \text{ for some } V \in \mathbb{V}\}$.

16 For ease of discussion, I assume in informal remarks like this that the nominal w in $\mathcal{L}$ picks out the same world as the term w in the meta-language.

17 This covers cases where some of the $[x_i]^{V_i}$ are undefined; hence the awkward formulation with 'it is not the case that'.

$\mathcal{S}, \mathbb{V}, w \Vdash_1 \Box A$	iff $\mathcal{S}, \mathbb{V}', w' \Vdash_1 A$ for all $w', \mathbb{V}'$ where wRw' and $\mathbb{V}' = \{V' : V \overset{V(a)}{\rightsquigarrow} V'$ for some $V \in \mathbb{V}\}$.
$\mathcal{S}, \mathbb{V}, w \Vdash_0 \Box A$	iff $\mathcal{S}, \mathbb{V}', w' \Vdash_0 A$ for some $w', \mathbb{V}'$ where wRw' and $\mathbb{V}' = \{V' : V \overset{V(a)}{\rightsquigarrow} V'$ for some $V \in \mathbb{V}\}$.

Finally, a sentence A is *true* in a model $\langle \mathcal{S}, V \rangle$ iff $\mathcal{S}, \{V\}, \alpha \Vdash_1 A$.

To get a feeling for how this works, let's apply it to the case of the fissioning train. W is the set of times, including the departure time t_1 and some time t_2 when the train has arrived in Cologne and Düsseldorf. The relevant individuals are the Y-shaped object Y that is a train before the division, and its two branches L (towards Cologne) and R (towards Düsseldorf) which are trains after the fission. Assuming that being a train corresponds to sort s, we therefore have $D_{t_1,s} = \{Y\}$ and $D_{t_2,s} = \{L, R\}$ (ignoring other trains). Let α be the departure time t_1. Assume that on the intended interpretation V, 'x' denotes the train at α, i.e. Y, the nominal 't_2' denotes the arrival time t_2, and the predicates 'B', 'C', 'D' apply to things in Berlin, Cologne, and Düsseldorf, respectively. Then 'Bx', for example, is true. That is,

$$\mathcal{S}, \{V\}, \alpha \Vdash_1 Bx.$$

This is because Y is in Berlin at t_1, and so $[x]^V \in V(\alpha, B)$. The sort plays no role here. Sorts are only used for the interpretation of *de re* modality. In fact, as long as we stay away from modal operators, the semantics is completely classical and we could simply identify $\Vdash_0$ with $\nVdash_1$.

More interestingly,

$$\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, Cx.$$
$$\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, Dx.$$
$$\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, \neg Bx.$$

Take the first of these. By the 1-clause for '*at*', $\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, Cx$ requires $\mathcal{S}, \mathbb{V}', t_2 \Vdash_1 Cx$, where $\mathbb{V}'$ is the set of t_2-images of V. There may be many such images, but what matters here is only whether they assign L or R to x, which are the two counterparts of Y of sort s at t_2. So we may assume that $\mathbb{V}' = \{V_L, V_R\}$, where V_R is an arbitrary t_2-image of V with $[x]^{V_L} = L$ and V_R is an arbitrary t_2-image of V with $[x]^{V_R} = R$. By the 1-clause for atomic formulas, $\mathcal{S}, \{V_L, V_R\}, t_2 \Vdash_1 Cx$ iff there is some member of $\{V_L, V_R\}$ that verifies Cx at t_2. And there is: V_L. So $\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, Cx$.

But there is also a member of $\{V_L, V_R\}$ that does not verify Cx. Thus

$$\mathcal{S}, \{V\}, \alpha \Vdash_1 at\ t_2, \neg Cx.$$
$$\mathcal{S}, \{V\}, \alpha \Vdash_1 \neg at\ t_2, Cx.$$

The last example uses the 0-clause for *at*, which renders the logic dialethic.

References

John Burgess [2010]: "Could a Zygote Be a Human Being?" *Bioethics*: 61–70

David J. Chalmers [2011]: "Verbal Disputes and Philosophical Progress". Forthcoming in *The Philosophical Review*

Roderick Chisholm [1967]: "Identity through Possible Worlds: Some Questions". *Noûs*, 1: 1–8

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

Hartry Field [1973]: "Theory Change and the Indeterminacy of Reference". *Journal of Philosophy*, 70: 462–481

Melvin C. Fitting [2004]: "First-order intensional logic". *Annals of Pure and Applied Logic*, 127: 171–193

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

— [1985]: *The Metaphysics of Modality*. Oxford: Oxford University Press

— [1990]: "Counterparts, Logic and Metaphysics: Reply to Ramachandran". *Analysis*, 50: 167–173

Greg Frost-Arnold [2008]: "Too much reference: Semantics for multiply signifying terms". *Journal of Philosophical Logic*, 37: 239–257

André Gallois [1998]: *Occasions of Identity*. Oxford: Clarendon Press

Peter Geach [1962]: *Reference and Generality*. Ithaca: Cornell University Press

Allan Gibbard [1975]: "Contingent Identity". *Journal of Philosophical Logic*, 4: 187–221

Allen P. Hazen [1979]: "Counterpart-Theoretic Semantics for Modal Logic". *Journal of Philosophy*, 76: 319–338

Dominic Hyde [1996]: "From Heaps and Gaps to Heaps of Gluts". *Mind*, 106: 641–666

Mark Johnston [1989]: "Fission and the Facts". *Philosophical Perspectives*, 3: 369–397

Jeffrey C. King [2003]: "Tense, Modality, and Semantic Values". *Philosophical Perspectives*, 17: 195–246

Markus Kracht and Oliver Kutz [2007]: “Logically possible worlds and counterpart semantics for modal logic”. In Dale Jacquette (Ed.) *Philosophy of Logic,* Amsterdam: North Holland, 943–995

Saul Kripke [1963]: “Semantical considerations on modal logics”. *Acta Philosophica Fennica*, 16: 83–94

— [1980]: *Naming and Necessity*. Oxford: Blackwell

David Lewis [1968]: “Counterpart Theory and Quantified Modal Logic”. *Journal of Philosophy*, 65: 113–126

— [1971]: “Counterparts of Persons and Their Bodies”. *Journal of Philosophy*, 68: 203–211

— [1976]: “Survival and Identity”. In Amelie O. Rorty (Hg.), *The Identities of Persons*, University of California Press, 17–40,

— [1982]: “Logic for Equivocators”. *Noûs*, 16: 431–441

— [1986]: *On the Plurality of Worlds*. Malden (Mass.): Blackwell

— [2002]: “Tensing the Copula”. *Mind*, 111: 1–13

— [2003]: “Things qua Truthmakers”. With a postscript by David Lewis und Gideon Rosen. In Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra (Hg.), *Real Metaphysics: Essays in Honour of D.H. Mellor*, London: Routledge, 25–38.

Mark Moyer [2008]: “A survival guide to fission”. *Philosophical Studies*, 141: 299–322

George Myro [1985]: “Identity and Time”. In R. Grandy and R. Warner (Eds.) *Grounds of Rationality: Intentions, Categories and Ends,* Oxford: Clarendon, 383–409

Derek Parfit [1984]: *Reasons and Persons*. Oxford: Clarendon Press

John Perry [1972]: “Can the Self Divide?” *The Journal of Philosophy*, 69: 463–488

Graham Priest [1995]: “Multiple Denotation, Ambiguity, and the Strange Case of the Missing Amoeba”. *Logique et Analyse*, 150–152: 361–373

— [2006]: *In Contradiction: A Study of the Transconsistent*. Oxford: Clarendon Press, 2nd edition

Murali Ramachandran [1989]: “An Alternative Translation Scheme for Counterpart Theory”. *Analysis*, 49: 131–141

— [1990a]: "Contingent Identity in Counterpart Theory". *Analysis*, 50: 163–166

— [1990b]: "Unsuccessful Revisions of CCT". *Analysis*, 50: 173–177

— [2008]: "Kripkean Counterpart Theory". *Polish Journal of Philosophy*, 2: 89–106

David Ripley [2011]: "Sorting out the Sorites". In F. Berto, E. Mares and K. Tanaka (Eds.) *Paraconsistency: Logic and applications,* Springer

Denis Robinson [1985]: "Can Amoebae Divide Without Multiplying?" *Australasian Journal of Philosophy*, 63: 299–319

Jeff Russell [2011]: "Actuality for Counterpart Theorists". Draft

Theodore Sider [1996]: "All the World's a Stage". *Australasian Journal of Philosophy*, 74: 433–453

— [2001]: *Four-Dimensionalism*. Oxford: Clarendon Press

— [2002]: "The Ersatz Pluriverse". *Journal of Philosophy*, 99: 279–315

— [2008]: "Beyond the Humphrey Objection". Manuscript

Richard Swinburne [2004]: "The Possibility of Life after Death". In Peter Cave and Brendan Larvor (Eds.) *Thinking About Death,* Oxford: BHA, 38–42

Johan van Benthem [2001]: "Correspondence Theory". In D. M. Gabbay and F. Guenthner (Eds.) *Handbook of Philosophical Logic (Second Edition), Vol. 3,* Dordrecht: Kluwer, 325–408

Lin Woollaston [1994]: "Counterpart Theory as a Semantics for Modal Logic". *Logique et Analyse*, 147–148: 255–263