Counterparts and Qualities

Manfred Kupffer

February 18, 2010

Dissertation
zur Erlangung des akademischen Grades
des Doktors der Philosophie (Dr. phil.) an der
Universität Konstanz,
Geisteswissenschaftliche Sektion,
Fachbereich Philosophie

vorgelegt von Manfred Kupffer

Tag der mündlichen Prüfung: 18. März 2000
Referent: Professor Wolfgang Spohn
Referent: Professor Arnim von Stechow

Preface

This book presents, in draft form, a revised and streamlined form of my thesis. There has been a long history of successive versions, and I still feel that it will take some time till the ideas expressed here have reached a form that deserves to appear in print.

David Lewis proposed to deal with the semantics of sentences that state what is possible for an individual in terms of possible individuals that are in ways the first individual might have been, so called *counterparts* of the individual. In this book, I defend counterpart semantics as an approach to the semantics of modality and natural language semantics in particular.

Counterpart semantics has a rival, the standard Kripkean semantics that deals with the same sentences in terms of an accessibility relation between possible worlds. While counterpart semantics appears to be more flexible and carries less metaphysical presuppositions (it does not presuppose that I myself exist in various possible worlds, e.g.), Kripkean semantics has some advantages of its own. Most importantly, it incorporates the core assumptions of what may be called standard possible-worlds semantics; i.e. the core assumptions of the most successful framework within linguistic semantics. Counterpart semantics, on the other hand, does not conform to these assumptions, or so I will argue.

In this book I opt for a synthesis of these two approaches that allows to combine their advantages. In Chap.1 I introduce the basic idea of counterpart semantics. I defend counterparts against the charge, made by Kripke, that counterparts are irrelevant for the analysis of modality. In Chap. 2 I present the classical version of counterpart semantics, i.e. David Lewis's translation of the language of quantified modal logic into counterpart theory. I discuss some well-known problems and opt for a revision of that translation that is able to solve the problems. Lewis's translation indirectly defines semantical values for expressions of quantified modal logic. In Chap.3 I evaluate counterpart semantics with respect to basic features of possible-worlds semantics. I conclude that the semantical values indirectly assigned to sentences by Lewis's translation are not sets of possible worlds, and that they do not meet certain intuitive demands on notions of meaning and content. (The revisions proposed in Chap.2 do not change this.) In Chap.4 I discuss various versions of a combination of counterpart semantics with Kripkean semantics found in the literature. But I conclude that here, the benefits of the latter are achieved by forfeiting the advantages of the former. Chap.5 finally contains my own proposal how to combine the two approaches in a more satisfactory way. The core semantical notion is *truth according to a maximal complete representation*, where representations now play the role of worlds.

It is an important feature of the counterpart relation that it is *qualitative*, i.e. a relation in which things stand in virtue how they are. In Chap.1 I point out why it is qualitative, and why it ought to be. In Chap.6 I show how to make the notion of a qualitative property or relation precise. Finally, in Chap.7 I show how to extend this and related notions to predicates and propositions. For counterpart semantics to work at all we have to assume that everything that can be said at all can in principle be expressed in a language with qualitative predicates only (and perhaps, additionally, names, treated directly referentially). In Chap.7 I show how this assumption allows to maintain that the representations of our semantics deserve their name.

I started working on a thesis in 1993 when I joined the Graduiertenkolleg Integriertes Linguistikstudium at the University of T”ubingen with a grant by the "Deutsche Forschungsgemeinschaft". In these years I had the opportunity to work under the supervision of Arnim von Stechow. I thank him for his constant encouragement and his willingness to listen to my all too varied, and premature, ideas (although he may have expressed more urgently that I should concentrate on some definite thing). Arnim has influenced me deeply in his admiration of the philosophy of David Lewis.

In 1997, when we moved to Konstanz, Wolfgang Spohn took over as supervisor of the thesis. I want to thank him for his help and encouragement, and for his somewhat premature praise. I am also indebted to him for his patience. This version has benefited a lot by suggestions and criticisms in Arnim's and Wolfgang's reviews of my thesis. Max Cresswell, André Fuhrmann, Ivan Kasa, Luc Schneider, Magda Schwager and Ede Zimmermann read and commented on material that went into the book. I am indebted to Ede and to

André, for all too many things.

Earlier versions of the book profitted from help by Eric Dalton (language) and Radu Dudau (TeX). I had the ability to present material from this book at many occasions, the one I enjoyed most was at Phileas, the student association for philosophy in Geneva in 2007.

Contents

Chapter 1

Introduction

1.1 Possibility *De Re*

We often express possibilities or necessities for particular individuals. E.g. we say things like

(1) Possibly, Humphrey wins

(2) Humphrey is necessarily human

to express that it is possible *for Humphrey* to be winning or that it is necessary for him to be human.

There are two different ways to define truth-conditions for such *de re* statements in terms of possible worlds.

(3) a. (1) is true iff "Humphrey wins" is true at some world accessible from ours.

b. (2) is true iff "Humphrey is human" at every world accessible from ours.

(4) a. (1) is true iff some counterpart of Humphrey wins.

b. (2) is true iff every counterpart of Humphrey is human.

The clauses in (3) are characteristic of the classical, Kripkean semantics for such sentences. Kripke used accessibility relations in his mathematical study of modal logic, Kripke (1963). In this book, "Kripkean semantics" refers to the use of accessibility relations in natural language semantics, as e.g. in Hintikka (1969).[1]

The clauses in (4) are typical specimen of counterpart semantics, see Lewis (1968). Counterpart semantics takes care of the idea that (1) and (2) express possibility and necessity *for* Humphrey. Both parties agree that some cases of modality are correctly analysed in terms of quantification over (accessible) possible worlds. While Kripke-semantics simply extends this treatment to the *de re*, counterpart semantics generalises the idea of an accessibility relation between worlds in a different way: counterpart relations are relations of accessibility between possible individuals. In the case of the *de re*, counterpart semantics takes "possibly" and "necessarily" to quantify over counterparts, accessible possible individuals.

This book tries to defend counterpart semantics. It will do so by trying to solve a couple of problems, old and new, that counterpart semantics faces.

1.2 The Underlying Metaphysics

Both kinds of semantic approach use the notion of possible world. And both have to assume that there are possible worlds galore. Neither Kripkean semantics nor counterpart semantics could be correct, e.g., if there were only one possible world. In addition, the counterpart semantics of (1) and (2) involves possible individuals, individuals in other worlds. Again, in order for the above truth-conditions to be correct, counterpart semantics has to assume that there are a lot of individuals.

This is expressed by the following postulate.

> absolutely every way that a world could possibly be is a way that some world is, and
>
> ... absolutely every way that a part of a world could possibly be is a way that some part of some world is.

(Lewis (1986a), p.86)

In the following I will refer to this postulate as the principle of plenitude. The second conjunct talks about possible individuals addressed as "parts of possible worlds".

[1] These two enterprises are different, but obviously related. Logicians define a space of models. They are not committed to any claim that the ingredients of their models refer to an outside reality, whereas the claim that (A) or (B) express the truth-conditions of (1) and (2) obviously involves such a commitment. On the other hand, semanticists could be portrayed as investigating the one true model.

The principle bears its name because, given our intuitions about what is possible, it may be used to derive the consequence that there are very many possible worlds and possible individuals. E.g. since being a world in which there are talking donkeys is a way the world might be, the principle allows to infer that there is a world in which there are talking donkeys (clearly, this is not our world). It is important that the existence of such a world does not follow from the principle of plenitude alone, but from the principle together with assumptions about what is possible and what is actually the case.[2]

I will not try to defend possible worlds and possible individuals, or only indirectly. Possibilia have to be justified in terms of their usefulness, and of course their success in the semantics of modality (i.e. the semantics of notions akin to possibility and necessity) is part of a proof of their usefulness. In general, possible worlds semantics is able to deliver what is, up to day, still the most fruitful elucidation of the notion of content. We will be concerned with details of the abovementioned applications of possibilia, here. There are many other applications (indeed possibilia provide a "paradise for the philosopher" as David Lewis puts it), most of them we will not touch upon here, of course. For an extensive discussion, see Lewis (1986a).

I will make some additional assumptions about the underlying metaphysics. I will assume that possible individuals have the same kind of properties that actual individuals might have. E.g. possible individuals win, are human, and so on. That means, e.g., that Humphrey's counterparts are as concrete as he is. Furthermore I will assume that every individual is in exactly one world, such that worlds do not overlap.

These assumptions come in handy, otherwise we could not state the counterpart semantics of (1) and (2) in the same simple way. E.g. since abstract objects do not win or loose, an abstract counterpart of Humphrey could not be said to win, although perhaps we might say it represents Humphrey as winning. And if Humphrey's counterpart were in many different worlds, it would not suffice merely to say that he wins, we would also have to say where.

I will not be concerned with a further defense of the above metaphysical assumptions either. For a comparison of different versions of modal meta-

[2]For Lewis, the principle is a trivial consequence of his identification of ways worlds could be with non-empty sets of worlds (see Lewis (1986a)). As a postulate of his metaphysics Lewis tries to replace it with a principle of recombination, see Lewis (1986a), pp.87–88, a principle to roughly the same effect as the combination of plenitude with the intuitions about the vast realm of possibilities I have alluded to above. Divers and Melia (2002) argue that this is only partly successful.

physics see Lewis (1986a); Divers (2002); Melia (2003). It has to be noted, though, that the above assumptions are not shared by all possible worlds frameworks. They are derived from one such framework, David Lewis's.

According to Lewis, possible worlds are precisely like the actual world. To understand the answer, consider the actual world. True, it may be seen as "everything that is the case", a collection of facts, c.f. Wittgenstein (1922). But we might also consider it as a big *thing* instead. Wherever you go in space, every individual you meet is part of the same world. Past and future individuals, likewise, are only in some temporal distance from present individuals. But however great the temporal distance, a thing in my past or future is still part of the same world. This is why there is a sense of world in which the world is just what comprises all individuals that are spatio-temporally related to me at all. So the sum of all these individuals, a certain big individual may justly be called "the world".

Within this picture of things, other possible individuals are just like individuals in this world. They only happen to be spatio-temporally disconnected to us. Likewise, other possible worlds are just like this world. Worlds are maximal sums of spatio-temporally connected individuals.

It is this idea which motivates our above assumptions. Because possible worlds are maximally connected, no thing is in two different worlds. And because possible worlds are sums of individuals, they and their parts are not abstract. Just like parts of our world, parts of other worlds win and are human.

I said that, according to Lewis metaphyiscs, there are a lot of spatio-temporally maximal sums of individuals, hence there are a lot of concrete individuals that are spatio-temporally disconnected from us. This consequence is breathtaking, especially given Lewis' assumption that possible worlds are things exactly like our world, only that all of them except the actual world are causally isolated from us. These features of his metaphysics have caused what Lewis calls the "incredulous stare". Fortunately, counterpart semantics could be adapted to different versions of the metaphysics of modality, and so could most of my solutions to the problems of counterpart semantics.—I hope that this remark is able to sustain your attention even if you do not share Lewis' metaphysical views.

1.3 The Role of the Counterpart-Relation

1.3.1 The Humphrey Objection

There is a famous objection by Kripke against counterpart semantics:

> [I]f we say "Humphrey might have won the election (if only he had done such-and-such)." we are not talking about something that might have happened to Humphrey but to someone else, a 'counterpart'. Probably, however, Humphrey could not care less whether someone else, no matter how much resembling him, would have been victorious in another possible world. (Kripke (1972), p.344)

When Kripke wrote these words they were quite influential. Nowadays, the objection seems to be less than compelling.[3] E.g. Kripke claims that according to counterpart semantics "Humphrey might have won" is not about what might have happened to Humphrey. But contrary to what Kripke says, the sentence *is* about what might have happened to Humphrey, according to counterpart semantics. It is about the existence of a victorious counterpart of Humphrey, and, according to the counterpart-theoretic analysis the existence of such a counterpart makes it true that Humphrey might have won (see Lewis (1986a), p.195–196).

Kripke also complains that the sentence is about what might have happened to someone else, and here he has a point: according to Lewis, Humphrey's victorious counterpart is a person different from Humphrey.—But even this point still does not amount to a compelling objection. It is a feature of every interesting analysis that it possesses features that go beyond our untutored intuitions (see Sider (2008)).

While it might not be terribly damaging, the Humphrey objection raises an important issue. What is the connection between what is possible for Humphrey and what holds of his counterpart? Why an analysis of modality *de Humphrey* in terms of counterparts of Humphrey rather than, say, in terms of cane-toads? Maybe it is this question of the relevancy of counterparts which motivates Kripke's dictum that Humphrey couldn't care less about what happens to a counterpart.

[3] c.f. Sider (2008), The Humphrey objection has often been discussed, beginning with Hazen (1979).

But the question of the relevancy of counterparts still does not provide any reason in favour of a Kripkean analysis. First note that the mere existence of the question is nothing which is special to counterpart semantics. Every possible-worlds analysis of modality should be able to address questions like these. E.g. Kripke should be able to explain what is the connection between what is possible for Humphrey *here* and what happens to Humphrey *at another (*accessible*) possible world*.

In the remainder of this section I will first consider how a vindication of accessibility could look like. I will then show that analogous considerations may serve to vindicate the counterpart relation.

1.3.2 Accessibility

It is an idea that goes back at least to Leibniz that modality is quantification over possible worlds. Possible worlds analyses of modality implement (variants of) this idea. In Lewis (1986a), p.5, Lewis starts the section about modality with an attempt to justify the quantificational analysis. Lewis thinks that the relation between *worlds* and *ways* could be helpful for giving such a justification of the Leibnizian semantics:

> If there are many worlds, and every way that a world could possibly be is a way that some world is, then whenever such-and-such might be the case, there is some world where such-and-such is the case. Conversely, since it is safe to say that no world is in any way that a world could not possibly be, whenever there is some world at which such-and-such is the case, then it might be that such-and-such is the case. So modality turns into quantification: possibly there are blue swans iff, for some world W, at W there are blue swans. (Lewis 86, p.5)

The premises of his argument are the principle of plenitude "every way that a world could possibly be is a way that some world is" and its converse "no world is in any way that a world could not possibly be". Together these principles form a biconditional that connects modality and quantification, namely that (for every way W and every world w) W is a way w might be iff there is a possible world that is in way W. Since, possibly, there are blue swans iff to be a world where are blue swans is a way the world might be, it follows that "possibly there are blue swans iff, for some world [w], at [w] there are blue swans."

So what Lewis shows is that, given assumptions taken from his metaphysics, the quantificational analysis is correct. I take the correctness of an analysis to be a strong vindication of the relevancy of its ingredients, at least for those ingredients that are not redundant.

Kripke-semantics generalises the quantificational analysis by adding the notion of an accessibility relation (see Kripke (1963)). This allows to apply the possible worlds analysis to other kinds of modality than just metaphysical modality, see below Sec.1.4.1.

While we want to replace the Kripkean account of the *de re* by counterpart semantics, counterpart semantics is not in conflict with a semantics in terms of accessibility for what is called modality *de dicto*, i.e. modal sentences that are not *de re*. So it might at least be instructive to see how a Lewis-style defense of the Kripkean semantics for such cases would look like. We can show as much: if an accessibility relation plays a certain specifiable role, then we can give a semantics in terms of accessibility that is correct. We need the following premises

(5) For every way W that a world w could be in there is a world v, such that (i) every way v is in is a way w could be and (ii) v is in W.

(6) v is accessible from w iff every way v is in is a way w could be. (In other words: a world accessible from w is in no way w could not be)

(5) is reminiscent of Lewis's principle of plenitude for possible worlds.

> absolutely every way that a world could possibly be is a way that some world is

(Lewis (1986a), p.86)

To plenitude it adds condition (i) above. But given plenitude, (5) is very plausible.

Let's say a way w might be is fully specifiable iff it belongs to a set of ways S such that (i) there is a world s such that S comprises all and only the ways s is in, (ii) all ways in S are ways w might be. It seems as if ways the world could be are always in principle fully specifiable. E.g. if being a P world or a Q world is a way for the world to be, then either being a P world is a way the world could be or being a Q world is a way the world could be. Imagine (5) fails, but plenitude holds.

Then there would be a way some world might be which is not fully specifiable. But this is implausible. It is to say that something is possible but we cannot specify (even in principle, even in an infinite language) the conditions under which it is possible.

(6) specifies the role of accessibility in the present context: accessiblity coincides with a certain relation of the ways worlds are with ways other worlds might be.

Together (5) and (6) imply

(7) for every way a world could be there is an accessible world which is that way.

(6) alone implies

(8) if there is a world accessible from w that is in way W, then W is a way that w could be in

(7) and (8), taken together, finally yield

(9) W is a way that world w could be iff there is a world accessible from w that is in way W

Again, we assume that being a world in which there are blue swans is a way our world might be iff it is possible that there are blue swans. Then we could finally derive

(10) it is possible that there are blue swans exactly if there is an *accessible* world at which there are blue swans.

Summing up, we have derived (10) from (5) and (6) above. I.e. if the accessibility relation used in a semantical analysis of sentences like "it is possible that there are blue swans" is of the kind specified by (6), then a semantics of modality *de dicto* in terms of accessibility is correct. And this is why such a relation would be relevant to modality.

1.3.3 Counterparts

Since worlds are a kind of individuals, the counterpart relation, which is very similar to accessibility, could in fact be seen as a generalisation of the accessibility. Hence, we can now give a parallel vindication of the counterpart relation. We need the following premises:

(11) for every way W that an individual A could be in there is an individual B, such that (i) every way B is in is a way A could be in and (ii) B is in W

(12) B is a counterpart of A iff every way B is in is a way A could be

(12) determines the role of counterparthood in the present context: the counterpart-relation coincides with a certain relation of ways a counterpart of A is with ways A might be.

Again, (11) is reminiscent of the principle of plenitude for individuals.

> (2) absolutely every way that a part of a world could possibly be is a way that some part of some world is.

(11) could be defended in a way completely analogous to the defense of (5) above.

Together (11) and (12) imply

(13) for every way an individual could be there is a counterpart of that individual which is that way.

(12) alone implies

(14) (for every A and every W) if there is a counterpart of A that is in way W, then W is a way that A could be in

(13) and (14), taken together, finally yield

(15) (for every A and every W) W is a way individual A could be iff there is a counterpart of A that is in way W

It is obvious that it is possible for Humphrey to win exactly if being a winner is a way Humphrey could be. Hence, we can finally derive the following truth-conditions for sentence (1) above.

(4a) (1) is true iff some counterpart of Humphrey wins

Again, we have derived (4a) from (11) and (12) above. I.e. if the counterpart relation used in a semantical analysis of sentences like "Possibly, Humphrey wins" is of the kind specified by (12), then a semantics of modality in terms of counterparts is adequate. And this is why such a relation would be relevant to modality.

1.3.4 Qualitative Properties

Successful as our vindication of Kripke-semantics and the parallel justification of counterpart semantics in terms of ways might look, there is a certain qualification to make. We have to be careful what we mean by "way". Ways simply seem to be properties. But our principle

(11) for every way W that an individual A could be in there is an individual B, such that (i) every way B is in is a way A could be in and (ii) B is in W

is false for certain properties!

- The property of being numerically different from A is a property every thing numerically different from A has, but it is a property A could not have. So here is a way every thing different from A is in, which is not a way A might be in. With (11) we conclude that the only ways A might be in are ways A is in. But this conclusion is blatantly false!

- To be bald-haired is a way Ede is not, but might be in. So is the property of being bald-haired and numerically identical to Ede. But, simply because Ede is confined to the actual world, there is no individual B, such that (i) every way B is in is a way Ede could be in and (ii) B is bald-haired and numerically identical to Ede, in contradiction to (11).

- Ede is Tom's father. To be in a world with talking donkeys is a way Ede is not, but might be in. So is the property of being Tom's father in a world with talking donkeys. But, simply because Tom is confined to the actual world, there is no individual B, such that (i) every way B is in is a way A could be in and (ii) B is Tom's father and in a world with talking donkeys, in contradiction to (11).

The last two examples are counterexamples against plenitude, too. And since, according to our current metaphysics, worlds are a kind of individuals this argument directly implies that (5) is incorrect, too.

These counterexamples involve identity-properties and other impure, non-qualitative properties. Fortunately, we can save (11) and (5) from such counterexamples. We can do so by ruling out non-qualitative properties from being ways. Let, in the following, "way" always be understood as meaning "qualitative way".—In Chap.6 we will try to make it more precise what is meant by "qualitative".

This restriction of way-talk has a direct impact on (12), our characterisation of the role of the counterpart relation. It now only says that a counterpart of A is any individual which is in no *qualitative* way A could not be. I.e. being a counterpart of A turns out to be a qualitative property.[4]

Lewis calls this idea, the idea that "representation *de re* is determined by qualitative character" *Anti-Haecceitism.* His reasons for Anti-Haecceitism are different from the above. Lewis asks "what the non-qualitative determinants of representation *de re* are, and how they do their work" (Lewis (1986a), p.229).

Another thing we should say is that the modal properties of A are determined by the way A is. Whether something is possible in our world is determined by the way our world is, and whether something is possible for Humphrey seems to be determined by the way Humphrey is. Representation *de re* "supervenes on qualitative character" and so does modality in general. But if the modal properties of A are determined by the way A is, then being something that has B as its counterpart turns out to be a qualitative property, too. I will call this the *Supervenience of Essence.*

If we combine Anti-Haecceitism and the Supervenience of Essence, it follows that the counterpart-relation is a qualitative relation. Lewis, likewise, characterises the counterpart relation as a qualitative relation, see Lewis (1986a), p.229.

But aren't there cases of modality where non-qualitative ways appear to be involved? What about the following:

(1) Possibly, Humphrey wins

(16) Ede might be bald-haired and identical to Ede

(17) Ede might be bald-haired and father of Tom

(1) seems to state that there is an accessible world which is in an impure, non-qualitative way, *viz.* being a world in which Humphrey wins; (16) seems to state that some counterpart of Ede has an impure, non-qualitative property, *viz.* the property of being bald and identical to Ede. And (17) seems to

[4]Similar things hold for (6), our characterisation of the role of the accessibility-relation. This is the reason why our justification of the use of accessibility in a semantics of modality above only covers the case of qualitative ways of worlds, and cannot be used to justify, on the background of our metaphysics, a Kripkean semantics for the whole of modality.

state that some counterpart of Ede has the impure, non-qualitative property of being bald-haired and father of Tom.

If we were compelled to analyse the above examples as above, then we would be in trouble: there is no accessible world in which Humphrey wins because Humphrey is only in a world where he loses; there is no counterpart of Ede who is bald and identical to Ede, because the only counterpart of Ede who is identical to Ede is Ede himself, and he has still some hair; there is no counterpart of Ede who is bald and father of Tom, because the only counterpart of Ede who is father of Tom is Ede himself, and, as I said, Ede has still some hair.

Does this mean that the semantics of modality is not able to deal with these examples? Fortunately not: we can treat such cases in terms of qualitative ways and counterparts, instead. E.g. instead of saying (1) is true iff Humphrey wins in some accessible world, we say that it is true iff some counterpart of Humphrey wins. I.e. instead of analysing the example in terms of a non-qualitative property of worlds, we analyse it in terms of a qualitative properties of individuals. And instead of saying that (16) is true iff some counterpart of Ede has the property of being bald and identical to Ede, we might say that (16) is true iff some counterpart of Ede has the property of being bald and standing in the relation of identity to a counterpart of Ede[5] ; here, again, instead of analysing it in terms of a non-qualitative property of individuals, we analyse it in terms of a qualitative relation, likewise for the last example.

Of course we cannot simply rest our case on some examples. Above we have learned the limits of how counterparts may represent. In order to be able to maintain that all cases of modality can be treated by counterpart semantics, we now have to assume that every proposition expressible in English is, at least in principle, expressible by a sentence that does not use impure, non-qualitative predicates. So, every sentence is translatable in principle into one that uses only qualitative predicates and, perhaps, names.

1.3.5 The Problem of Multiple Counterparts

Because counterparthood is qualitative, there might be multiple counterparts of one and the same individual in a given world.

The paradigm case for a provider of multiple counterparts is a world of

[5]Certain inessential details of this analysis may have to be revised, see Chap.2.

what Lewis calls "two-way eternal recurrence". A world of eternal recurrence is composed of an infinity of epochs which are identical except for the numerical identity of the individuals involved, and nothing besides these epochs. E.g. in a future-directed *one-way* world of eternal recurrence, after the end of every epoch, exactly the same things will begin to happen again. In addition, if right before the beginning of every epoch, just the same things have happened, we speak of *two-way* recurrence. But then, two such epochs are totally identical except for the numerical difference of the individuals involved.[6] And, because counterpartness is a qualitative relation, a thing which has a counterpart in one epoch of such a world will have counterparts in every other epoch. Think also of worlds which consist of two or more totally point-symmetrical spatial parts (parts which are symmetrical through all of time). A thing which has a counterpart in one of these parts will have counterparts in all the other parts as well.

All of the above cases are cases where different counterparts are qualitatively identical or *indiscernible*. This might suggest that multiple counterparts only occur if the individuals in question are indiscernible. But this does not seem to be true.

If the counterpart relation is as in (12) above, then ordinary individuals will sometimes have many distinguishable counterparts in one and the same world, because sometimes different maximal specific ways a given individual might be in could be realised within one and the same world.

The following argument for the existence of distinguishable maximal counterparts is based on an example from Feldman (1971): Suppose I am the prince and you are the pauper. Now it seems to be completely accidental that I am in my place and not in yours: I am "there but for the grace of god". It could have been otherwise. I might have been the pauper and you the prince in a world w which is qualitatively exactly like ours. Hence, the pauper of w is my counterpart. Now, supposing the counterpart-relation is qualitative, since the prince in our world is my counterpart, he is my counterpart in w as well. Since the pauper is my counterpart in w, too, I have two distinguishable counterparts in w.[7]

---

[6] In a world of one-way recurrence this does not hold, because e.g. the purely qualitative property of living in an epoch which is n epochs distant from the endpoint of time will distinguish an individual of the n-th epoch from all its predecessor and successors.

[7] Lewis discusses a version of this example where one of the possibilities in question is actual, but this feature seems to be inessential to me. Feldman (1971) and Hazen (1979) think that Lewis has a problem with that special case, but see the postscript to Lewis

1.4 Specific Relations

1.4.1 . . . of Accessibility

Modal locutions are highly ambiguous. "Possibly" and "necessarily", or "can" and "must" may mean a lot of different things in different contexts. E.g. if we say "Pirmin must do his homework at noon" we may either express that Pirmin is obliged to do his homework at noon or that, as far as we know, he does it at noon.

The first is what we call a deontic necessity, the second an epistemic necessity (for us). There are many different kinds of modalities, e.g. metaphysical, nomological, deontic, bouletic, epistemic, and doxastic possibilities and necessities (and so on). At least some of these kinds of modalities have many different instances, e.g. something may be deontically possible in view of the by-laws of one city, but ruled out by bylaws of Chicago (which forbid, e.g., to carry hatpins, or to bring french poodles to the opera), see Montanarelli and O'Gorman (2005).

Such flavours all seem to be cases of *restricted modality* where the realm of possible worlds taken into account is narrower than in the case of metaphysical modality. They are also cases of *relative* modality, where what is possible or necessary depends on what is the case. E.g. in the case of nomological necessity, whether a statement is nomologically necessary depends on what the laws of nature are. If these laws were different, then we would also have different nomological necessities.

If we regard all the different modal flavours, mere quantification over worlds appears to be not flexible enough. It seems to apply to what is called "metaphysical" or "logical" possibility only, but not to other kinds.

Fortunately, the advent of Kripke-semantics for modal logic provided a way to generalise the simple semantics in order to deal with the other kinds of modality, too. Different modal flavours now correspond to different notions of accessibility, i.e. there are relations of metaphysical, deontic, epistemic, doxastic, nomological accessibility, and so on. E.g. in our world, it is an epistemic possibility for me that Pirmin does his homework at noon iff he does his homework at noon in some world accessible from our world, where the accessibility relation is epistemic accessibility for me. Table 1.1 contains some further samples of such an analysis.

(1983d).

v is ... accessible from w	every way v is in is a way w might be, ...
metaphysically	
nomologically	in view of the laws of w
deontically	in view of the X-rules in w
doxastically	in view of the beliefs of the agent in w
epistemically	in view of what the agent knows in w
bouletically	in view of the desires of the agent in w

Table 1.1:

The above analysis of belief and knowledge derives from Hinttikka, for an overview and a refined theory of the modalities above, see Kratzer (1981, 1991).

Details of these examples of different accessibility-relations are not important in the following, but it should be mentioned that there are also certain difficulties for some of the above analyses. E.g. the above implies that metaphysical possibility is the broadest kind of possibility, such that if something is possible in any sense, it is also metaphysically possible. This view was challenged in the mid-seventies, when Kripke drew our attention to cases of seemingly *a posteriori* necessities, cases where not p is epistemically possible but metaphysically impossible. E.g. to take an example of Putnam (1975), while water is necessarily H_2O, some persons do not know that. For such persons it is epistemically possible that water is not H_2O. The usual reaction to such cases is not to deny that metaphysical accessibility is unrestricted, it is rather to present different analyses for the kind of knowledge or belief involved in such examples (see e.g. the two-dimensionalist school, e.g. Stalnaker (1978); Davies and Humberstone (1980); Haas-Spohn (1995); Chalmers (2006)).

1.4.2 ... of Counterparthood

There are also various kinds of possibilities for particular persons, corresponding to different *counterpart relations*. In most of the above cases it is even more natural to assume that the possibilities in question are possibilities for persons, rather than possibilities for worlds. E.g. a bouletic necessity for *me* is a way *I* want to be. The same thing applies to deontic, epistemic, doxastic necessities and possibilities.

Now counterpart semantics is able to treat these different possibilities in terms of different counterpart relations.[8] E.g. I want to be rich iff every bouletic counterpart of me is rich, where a bouletic counterpart of me is a person that is in no way I don't want to be. And it is doxastically necessary for Kaplan in w to have pants that are on fire (in other words Kaplan believes that his pants are on fire) iff every doxastic counterpart of Kaplan has burning pants. Here a doxastic counterpart of Kaplan is a person, such that every way that person is is a way Kaplan could be, according to his beliefs.

A metaphysical counterpart of me is someone who shares my essential properties, i.e. someone who is not in any way I could not be in view of my essence. Table 1.2 contains the above and some further characterisations of specific accessibility relations.

B is a . . . counterpart of A	every way B is in is a way A might be, . . .
metaphysical	in view of the essence of A
deontic	in view of what the X-rules say for A
doxastic	in view of the beliefs of A
epistemic	in view of what A knows
bouletic	in view of the desires of A

Table 1.2:

Again an analysis of the relevant phenomena in terms of these counterpart relations need not be the final word about the modalities in question. Both Kratzer's refinements of the original analysis as well as the above considerations about *a posteriori necessities* are relevant here.

Again, the categories above only correspond to kinds of accessibility relations, not to specific accessibility relations. This time, there also seem to be many different metaphysical counterpart relations. E.g. is it essential to Oxford that it is north of London? That seems to depend on what we want to keep constant about Oxford when talking about various possibilities. Essences and, therefore, metaphysical counterpart-relations seem to vary with different contexts.

> [A] counterpart of Oxford is similar to Oxford in its origins, or in its location vis-â-vis (counterparts of) other places, or in the arrangement and nature of its parts, or in the role it plays in the life of a nation or a discipline. Thus Oxford might be noted more for the manufacture of

[8] C.f. Lewis (1986a), p.28–36.

> locomotives than of motor cars, or might have been a famous centre for the study of paraconsistent hermeneutics, iff some other-worldly counterpart of our Oxford, under some suitable counterpart relation, enjoys these distinctions. (Lewis (1986a), p. 8)

1.4.3 Similarity

Lewis sometimes says the counterpart relation is a relation of similarity, e.g. in Lewis (1968), p.115.

Similarity is usually similarity in a certain respect, and the notion of a respect does not seem to be constrained at all. Therefore, there are very many relations of similarity. Hence, to say a relation is one of those does not say very much. But it says something about the formal properties of the relation. A relation of similarity is at least reflexive (everything is similar to itself) and symmetric (if A is similar to B, then B is similar to A). A relation of similiarity does not have to be transitive: even if A is similar to B and B is similar to C, then A and C may not be similar to each other.

At other places, Lewis characterises counterparthood as a relation of *maximal comparative similarity*:

"Your counterparts resemble you closely in content and context in important respects. They resemble you more closely than do the other things in their worlds." (Lewis (1968) p. 114)

B is maximally similar to A iff B is similar to A and there is no individual C in B's world, such that C is more similar to A than B is. Just like similarity, maximal similarity is reflexive, and does not have to be transitive. Maximal similarity is not necessarily symmetrical: if B is maximal similar to A, there can still be a C in A's world such that C is more similar to B than A is.

The use of maximal comparative similarity rather than simple similarity is perhaps inspired by the use of comparative similarity in the semantics of counterfactuals (see Stalnaker (1981); Lewis (1973)). A motive for the choice of maximal similarity might be that, with regard to this-worldly counterparts, Lewis wants to cut down on the number of counterparts. He wants to say that the counterpart of Humphrey in our world is just Humphrey himself, and no one else. He wants the counterpart relation to have the following property:

(B) A is the only counterpart of A in A's world. [9]

[9] Together reflexivity and (B) form the Axiom of Centering, familiar from the semantics

Indeed it may seem that if we are looking for the best double of Humphrey, and he enters the contest himself, he is bound to win. This is a reason for thinking that

(B') only A is maximally similar to A in A's world.

holds. Since the analogue of (B') does not hold for ordinary relations of similarity, at this point it seems as if maximal similarity was preferable.

But the alleged advantage is spurious. First (B') is not necessarily true. E.g. Humphrey himself and a truly perfect duplicate of Humphrey should be tied first in the similarity contest, contrary to (B'). Moreover, we should not demand that (B) holds either, see the above case of the prince and the pauper. Lewis contends: "In 'Counterpart Theory and Quantified Modal Logic' I took it as axiomatic that nothing can have any counterpart besides itself in its own world. I would now consider that requirement appropriate under some but not all resolutions of the vagueness of the counterpart relation" (Lewis (1986a), p. 232, footnote 22).

If (B') is no longer assumed, then the only remaining formal difference between maximal comparative similarity and an ordinary relation of similarity seems to be that the latter is symmetrical. It turns out the counterpart relation is symmetrical exactly if the following implication is valid: if something is in a certain way, then it is necessarily possibly in this way, i.e. in terms of modal logic, $Px \rightarrow \Box\Diamond Px$ is logically valid. It is difficult to understand nested modalities out of context, but I would say the implication *is* valid. Hence, to say that the counterpart relation is simply a relation of similarity seems to be preferable.

At other places Lewis is still more cautious wrt. to similarity. In Lewis (1986a), Lewis merely observes that often, counterpart relations involve *similarity* relations.

> As quantification over possible worlds is commonly restricted by accessibility relations, so quantification over possible individuals is commonly restricted by counterpart relations. In both cases, the restrictive relations usually involve similarity. (Lewis (1986a), p.8)

I guess the reason for this restraint is that some cases of counterpart relations are no relations of similarity at all, because they are neither reflexive nor symmetrical. Of course these cases are no relations of maximal comparative

---

of counterfactuals, see e.g.Lewis (1973).

similarity, either. E.g. a doxastic counterparthood relation is not reflexive, because what holds in every world that is compatible with what I believe may nevertheless not be true and a nomological accessibility relation may not be symmetrical because if all the laws of our world hold in another world, the other world could still have additional laws that do not hold in our world.

Summing up, it can at best be maintained as a hypothesis about *metaphysical* counterpart relations, that counterparthood is a relation of similarity.

1.5 The Virtues of Counterpart Semantics

There are a many problems counterpart semantics is able to deal with succesfully. These belong to broadly three classes.

First there are some natural counterpart semantic accounts of particular cases a Kripkean could not have in the same simple way. There are arguably also a few cases a Kripkean could not account for at all. Second, the Kripkean semantics has, from the standpoint of a formal logician, some undesirable technical features. And third, the Kripkean analysis seems to possess metaphysical presuppostions that are at least debatable. So, while there does not seem to be a knockdown argument in favour of counterpart semantics, there are some reasons to prefer it over its Kripkean rival. Let me now present these applications of counterpart semantics in some more detail.

1.5.1 Reductive Analysis of Modality

It is often claimed that the main advantage of counterpart semantics in combination with its Lewisian background metaphysics is a reductive explanation of modality.[10]

A reductive explanation of modality would provide a translation of every modal statement into a statement that is not modal (such that the theory that characterises the primitive notions involved in a non modal way). Here, a modal statement is one that involves modal expressions like “possibly”, “necessarily”, and related ones like “essence”.

The question arises whether the counterpart relation is a modal one. But this depends on the particular use of a particular counterpart-relation.

[10]For a critical examination of whether the background metaphysics fulfills this duty, see Divers and Melia (2002).

It should be clear from Table 1.2 on p.16 that, for many flavours of modality, we are not in the possession of a non-modal characterisation of the counterpart-relations involved. Sometimes we can only characterise the relevant counterpart relations by using the very modal expressions we try to analyse; we characterise the notion of a doxastic counterpart in terms of the notion of belief, sometimes we seem to exchange just one modal notion for another; e.g. when we define the nomological counterparts used in the semantics of nomological necessity in terms of the notion of a law. The same verdict would be due if we defined metaphysical necessity in terms of ways the world might be, or in terms of essence.

Only if we define metaphysical modality in terms of relations of similarity no clearly modal notions are used on the right hand side of semantical definitions. But this still does not mean that metaphysical counterparthood is clearly non-modal.

Suppose you are presented with the following.

(18) "Possibly, Humphrey wins" is true iff a possible individual that is similar to Humphrey wins.

(18) is true if context provides an appropriate resolution of the vagueness of "similar". Say it has to provide a respect of similarity.

The following questions arise: could this be termed an analysis at all? It is incredibly vague! And if the relevant aspect of similarity is provided by context, rather than by some explicit expression; does this already imply that it contributes in a non-modal way, simply because it is not expressed in modal terms?—Since I am far from able to answer such questions, in this book I will not take sides wrt. the question whether metaphysical counterpart-relations provide a non-modal account of modality.

1.5.2 Shifty Intuitions, Indeterminate Truth-Values

Counterpart semanticists usually think that counterpart relations are normally provided by context and that different contexts provide different counterpart relations. This would account for the fact that judgements about possibility *de re* or essential properties often seem to contradict each other, and that often people are easily persuaded to reverse their judgements. Different judgements may simply be due to different contexts which provide different counterpart relations.

That contexts may pick from several possible counterpart-relations could also be used to explain why sometimes there does not seem to be a determinate answer to questions of possibility *de re* at all. E.g. it seems to be clear that it is possible for Theseus' ship to exist with one "new" plank in place of one of its actual planks. At least many people are convinced that it is not possible for Theseus ship to be made out of totally new planks. Now ask the same people whether it possible for Theseus' ship to exist with having "new" planks in place of 23 of its actual planks? — The counterpart theorists can explain why sometimes there does not seem to be a clear answer to such questions by saying that in such a case context failed to determine the counterpart relation completely such that there are still different contextually admissible counterpart relations, which differ about the maximal number of planks that may be exchanged without destroying the identity of the ship.

1.5.3 Qua-Talk

A good reason to back the claim that there is context dependence in a certain locution is when we can show that the contribution of the context may also be made explicit. E.g.

(19) It is raining

is context-dependent, where context adds time and place. The proposition the sentence expresses depends on the time and the place we are talking about. This claim can be made plausible by showing that we may use the following sentence to say the same thing.

(20) It is raining in most of Frankfurt on November 19, 2009 during the interval from 8:30-8:31am

What goes on here seems to be that what context has to supply to the interpretation of (19) is made explicit in (20).

Can we find the same kind of indirect evidence for the existence of counterpart-relations? In Kratzer (1977, 1978, 1981, 1991), Angelika Kratzer points out that we are usually able to make informations about the relevant flavour of modality explicit. In a sentence like

(21) In view of what is known, the ancestors of the Maoris must have arrived from Tahiti. (Kratzer (1977))

the locution "in view of" serves to disambiguate "must". "In view of what is known" makes it clear that we are dealing with an epistemic "must". In terms of accessibility, "in view of what is known" determines an epistemic *conversational background*, which in turn could be shown to determine an epistemic accessibility relation.

If epistemic possibility had better be analysed in terms of epistemic counterpart relations, as it is argued in Lewis (1983a), then these observation at least sometimes pertain to the counterpart relation; indeed at least sometimes informations about the latter relation may be made explicit by the "in view of"-locution.

Sometimes such paraphrases are less natural, e.g. in the case of metaphysical counterpart-relations. I don't think

(22) In view of being made of wood, the table could not be made of ice.

is a very natural thing to say.

In Lewis (2003), Lewis points to the "qua" locution which seems to be more natural, here. We may say things like '

(23) Qua being made of wood, the table could not be made of ice.

And it may be used as an argument for the existence of different metaphysical counterpart relations, that we may indeed use different qua-locutions to claim different incompatible sets of modal properties for the same thing "qua being a red rose, this is necessarily red" "qua being a rose, this could have been white".

1.5.4 Representation *De Re* not Transitive?

Finally the counterpart-relation need not have all formal properties of identity. E.g. it has been claimed that possibility *de re* is not transitive, i.e. if something is possibly possible for an individual we cannot infer that it is possible for that individual. E.g. it is possible for Theseus' ship to be made out of a heap of planks that differs from the actual heap of planks wrt. to the identity of one single plank. But similar things hold for that ship Theseus' ship might be and the heap of planks that ship is composed of, so it is possibly possible for Theseus' ship to be made out of a heap of planks that differs from the actual heap of planks wrt. to the identity of two planks. By a repeated application of that manoeuvre we end up with predicting that it is possibly possibly . . . possible for Theseus' ship to be made out of a totally different heap of planks. If the counterpart relation is transitive, then we can infer that it is simply possible for Theseus' ship to be made out of a totally different heap of planks. But the latter is something many people would want to deny, hence the scenario is called a *paradox*.[11]— The counterpart semanticist can simply deny that the counterpart relation is transitive, and thus block the inference to the paradoxical conclusion.

1.5.5 Contingent Identity and Non-Identity

So far I have only claimed that the Kripkean cannot have the same effects the same way. But perhaps she can have these effects her own way, in terms of the relation of accessibility, without having to add any counterpart-theoretic machinery. Clearly, the Kripkean can adapt the idea that accessibility is context-dependent, shifty and sometimes indeterminate. Of course, she could have non-transitive relations of accessibility as well. (There is a certain price to pay here, because people tend to think that metaphysical accessibility is that relation that holds between any two arbitrary worlds, and one cannot have a non-transitive relation that is total at the same time.) The following cases are more difficult to deal with for the Kripkean.

If there are different counterpart-relations it is at least possible that different counterpart-relations could co-occur within one and the same sentence. And indeed, it seems to be natural to apply this idea to cases like

(24) I dreamt that I was Brigitte Bardot and that I kissed me.[12]

(25) I and my body might not have been identical today.[13]

(24) apparently has a reading where, according to a dream I am Brigitte

[11]See Chisholm (1973) and Chandler (1975).The paradox consist of the intuition that while nothing could consist of entirely different matter, it could consist of slightly different matter, together with the observation that, as Quine put it "... you can change anything to anything by easy stages through some connecting series of possible worlds" Quine (1976), p.861. For further discussion of counterpart-theoretic solutions of the paradox, see Forbes (1984).

[12]The example, as well as the analysis in terms of two counterpart-relations is due to G. Lakoff (1970) "Linguisics and Natural Logic." Synthese p.245, for a discussion of related examples see Heim, Irene: Puzzling reflexive pronouns in de se reports, Handout 1994.

[13]For Lewis, as an identity theorists, I and my body are identical. But (4) seems to be true as well. Hence, in Lewis (1983c), Lewis suggests that the sentence should be analysed in terms of a bodily counterpart and a person-counterpart of mine.

Bardot, and kiss a different person, which is, nevertheless, referred to by a first-person pronoun, similarly for (25). A counterpart-theoretic analysis is able to describe this in terms of different counterpart-relations that are employed for "I" and "me" ("my body").

These mixed cases present a difficulty for the Kripkean, witness the following two sentences.

(26) If I met me at a party, I would have a wonderful conversation.

(27) If I met me at a party, I would have a dull evening.[14]

If said by Paula, (26) seems to be true, and (27) false. In order for a counterfactual to be false, the antecedent proposition has to be possible at least. Now there are no possible worlds in which somebody is (so to speak) two different persons who meet (in the literal sense that seems at stake here). So, in order for the above two counterfactuals to be non-trivially true on a Kripkean analysis, there has to be a world in which Paula is two different persons who meet. But there is no such world. The best the Kripkean can do with cases like that seems to be to deny that taken literally, they present real possibilities.[15]

1.5.6 Modal Logic

The use of counterpart-relations also presents considerable technical advantages. It has been noted that in quantified modal logic based on the classical Kripkean semantics, no general completeness proofs are available, often completeness proofs are not known, and sometimes adding the axioms of quantification theory turns a set of axioms for propositional modal logic which is complete wrt. some class of frames into a set of axioms which is incomplete, i.e. not complete wrt. to any class of frames. In fact, it turns out that incompleteness is widespread and examples of incomplete logics comprise some very natural and simple examples of quantified modal logics. (This distinguishes quantified modal logic from propositional modal logic, where there is also incompleteness, but the examples all seem to be somewhat artificial.)

Now, the use of counterpart-theoretic methods promises to provide general, all-purpose completeness proofs, and indeed the problem has been solved

[14] The first example is due to Paula Menendez-Benito.

[15] Cases like (24), on the other hand, could perhaps also be dealt with in terms different relations of acquaintance in the sense of the traditional *de re* analysis.

in Skvortsov and Shehtman (1993), a paper that employs what appear to be counterpart relations between (abstract) n-tuples.

It is of course open to discussion whether completeness is always desirable. If you are presented with an arbitrary calculus, there is no guarantee that the set of theorems generated by that calculus really constitutes a logic in any natural sense. If it does not, then no harm done if there is no notion of logical validity that coincides with theoremhood in the calculus. So a defense of the kind of completeness the counterpart-theorists can offer presupposes an answer to the question what consititutes a logic.—We will not attempt to enter in this debate here.

1.5.7 Haecceitist Differences

Lewis claims that counterpart semantics is able to account for examples of seemingly *haecceitist differences*. Lewis defines the notion of a haecceitist difference in Lewis (1986b).

> If two worlds differ in what they represent *de re* concerning some individual, but do not differ qualitatively in any way, I shall call that a haecceitistic difference. p.221

(This is called an haecceitist difference because it seems to rest on mysterious *haecceities* (thisnesses), non-qualitative features of worlds.)

Here are some scenarios that are *prima facie* examples of haecceitist differences.

We might live in the 17th epoch of a world of one-way eternal recurrence with a first epoch. We might as well live in the 137th rather than the seventeenth.

I might have been either one of a pair of twins. I.e. I might have been the first-born one and I might have been the second-born one.

Take the first example. Intuitively, the following two sentences should both come out true.

(28) Possibly, I live in the 17th epoch (and not in the 137th) of a world of one-way eternal recurrence

(29) Possibly, I live in the 137th epoch (and not in the 17th) of a world that is qualitatively exactly like the first

This is an example of a prima facie haecceitist differences, because here, there seems to be a difference in representation between two possible worlds without a corresponding difference in qualitative character.—But note that in counterpart semantics, the unit of representation *de re* is not the world, but the possible individual.

According to counterpart semantics

(11) is true iff there is a world of one-way eternal recurrence w and a counterpart A of me in w such that

1. A lives in the 17th epoch and

2. A does not live in the 137th epoch.

and (11) is true iff there is a world v qualitatively identical to the first and a counterpart B of me in v such that

1. B lives in the 137th epoch and

2. B does not live in the 17th epoch.

A suitable counterpart relation is able to deliver counterparts A and B as above. It will do so without presupposing a difference in representation without a difference in qualitative character: while A and B are qualitative duplicates of me, both are representatives (counterparts of me). Still counterpart semantics is able to account for (11) and (12), simply because it only considers one representative at a time.

Because w and v above are treated exactly alike by the counterpart relation, the analysis does not presuppose that w and v are different. We need not assume that for any world there is more than one qualitative duplicate. On the other hand we also do not have to assume that there is only one duplicate of every world. A Kripkean analysis of these examples, on the other hand, has to posit two qualitatively identcal, but numerically different accessible worlds w and v, such that (11) is true in w and (12) is true in v. I.e. Kripkean semantics has to presuppose haecceitist differences, here. This brings me to the topic of metaphysical costs.

1.5.8 Metaphysical Costs

Jim wonders whether he is, after all, human. Instead of saying, e.g. that it is doxastically possible for Jim to be an android robot iff there is a doxastic counterpart of Jim who is, the Kripkean could put forward the following analysis:

"it is doxastically possible for Jim to be an android iff there is a doxastically accessible world, at which Jim is an android."

On the face of it, this alternative analysis is as good as the first one. But notice that it makes stronger metaphysical presuppositions. *Viz.* it presupposes that there is a world at which Jim is an android. Needless to say that some people, Kripke included, believe that, provided Jim really is human, there is no world in which Jim himself is an android.

Worse still, the content of belief or of dreams seems to be virtually unconstrained as to the essences of the subjects in these dreams, the alternative analysis really requires a metaphysics of possible individuals that allows unrestricted recombination of individuals and properties. Such a metaphysics is discussed under the label of "extreme haecceitism" in Lewis (1986a). According to it, e.g. Bertrand Russell might have been a boiled egg.—At least for some people, extreme haecceitism is very hard to swallow.

Counterpart semantics does not presuppose extreme haecceitism, neither does the counterpart semantic analysis of our example presuppose that at some world compatible with Jim's beliefs, he is an android. That analysis only presupposes that some individual at some world is an android.

It is part of the abovementioned metaphysical costs of the Kripkean analysis that it requires individuals to inhabit different worlds. This might be worrisome if you adhere to the metaphysics of possible worlds proposed by David Lewis according to which individuals inhabit one world only. So Lewisians are *forced* to adopt counterpart semantics. (The alternatives would be to let reference be a relation not to possible individuals, but to individual concepts, or sums or sets of possible individuals from different worlds. Now the question arises what it is that glues together such individual concepts, or sums or sets of possible individuals. If it is the counterpart-relation, the alternative does not seem to differ substantially from counterpart semantics. But what else could be taken to unify the referent of a name? (see Lewis (1986a), p.216)

1.6 Problems

1.6.1 Conceptual Problems

Above I have pointed out that counterpart semantics has some features which may be deemed advantages over the Kripkean semantics. Nevertheless, there are also a couple of problems that seem to threaten the whole enterprise. For certain kinds of properties counterpart semantics invariably seems to deliver wrong results. In Chap.2 I will address, and try to solve, the two most important problems.

Perhaps the most important problem is how to deal with the property of non-existence. How could A's possible non-existence ever be represented by a counterpart of A *in some other possible world*?

The second important range of problems concerns putative examples of essential relations like identity, non-identity, or fatherhood. As has been observed early on by Allen Hazen, given the semantics in Lewis seminal paper all the above relations come out as not being essential. This is odd, because in these cases there are strong essentialist intuitions, and counterpart semantics was made to be able to accommodate essentialist intuitions.

1.6.2 Semantical Problems

A further problem of counterpart semantics appears to be that *it does not go together well with the standard version of possible-worlds semantics.* By "standard possible-worlds semantics" I mean a set of assumptions shared by most practitioners of possible-worlds-semantics. These assumptions derive from the possible-worlds account of the content of a sentence in terms of *circumstances* under which the sentence is true. Standard possible worlds semantics is both a theory about the nature of content as well as up to day the most successful framework for doing linguistic semantics. It would be a grave problem if counterpart semantics required a change in the basic assumptions, worse still if counterpart semantics had nothing to replace them.

An instance of these difficulties is the question how counterpart semantics could be combined with a treatment of the locution "actually" as in "it is possible that someone who is actually rich was poor". While there are straightforward truthconditions for such a kind of example within the Kripkean semantics, it is not clear how to come up with an extension of counterpart semantics. Recently, Fara and Williamson (2005) have even claimed that this is impossible and that counterpart semantics should be rejected for that very reason.

Another instance pertains to impure, non-qualitative predicates like "being identical to Ede" or "being the son of Ede". I argued above that such predicates are always analysable in terms of predications of pure, qualitative ones. But such predicates sometimes are syntactic constituents and should then be given meanings of their own.

Finally, the question arises how counterpart semantics can be squared with the possible-worlds account of the content of a sentence as a set of circumstances.

1.7 What I will not Discuss

There are a couple of interesting problems I will not address here.

- I will not address the question how to analyse *de re* occurrences of names or variables in attitude ascriptions.
- I will only briefly touch upon the problem how to combine different metaphysical counterpart relations in one sentence, and I will not discuss how to combine metaphysical and other modalities.
- In fact, I will not address the semantics of modalities other than metaphysical modality and counterpart relations other than metaphysical ones.
- I will limit myself to quantification over individuals as opposed to e.g. sets.
- I will not discuss so-called temporal counterpart relations at all.

Chapter 2

Lewis' Counterpart Semantics

In this chapter I will discuss Lewis' counterpart-theoretical translation scheme (here called "T") from the language $\mathcal{L}_{QML}$ of quantified modal logic into the language $\mathcal{L}_{PL}$ of counterpart theory (CT). The scheme appears in his seminal paper Lewis (1968). It is Lewis's only exposition of counterpart semantics that goes into technical detail.

Some features of T are fundamental to counterpart semantics, others do not follow from the basic ideas themselves. I will point out these additional assumptions and discuss their impact. In the end I will argue for a revision of counterpart semantics, which, incidentally, is able to address some of the most pressing objections.

These problems include the question whether counterpart semantics is able to make sense of its own possibilist metalanguage; whether it can do for essential relations what it does for essential properties; the very non-standard logic of de re modal formulas presented by Lewis' scheme; and its inability to provide a sense in which you and I may be said to exist contingently.

2.1 Translation into Counterpart Theory

2.1.1 CT

Counterpart theory (CT) is a system of postulates in the language $\mathcal{L}_{PL}$ of first-order predicate-logic that constrain the interpretation of the predicates W (for "is a world"), I (Ixy stands for "x is in (world) y") and C (Cxy stands for "x is a Counterpart of y").

Counterpart theory consists of the following postulates P1–P6, see Lewis (1983b). Lewis has two additional postulates governing the predicate A (for "actual").

P1 $\forall x \forall y (Ixy \rightarrow Wy)$
(Nothing is in anything except a world)

P2 $\forall x \forall y \forall z (Ixy \ \wedge Ixz \rightarrow y = z)$
(Nothing is in two worlds)

P3 $\forall x \forall y (Cxy \rightarrow \exists z Ixz)$
(Whatever is a counterpart is in a world)

P4 $\forall x \forall y (Cxy \rightarrow \exists z Iyz)$
(Whatever has a counterpart is in a world)

P5 $\forall x \forall y \forall z (Ixz \ \& Iyz \ \& Cxy \rightarrow x = y)$
(Nothing is a counterpart of anything else in its world)

P6 $\forall x \forall y (Ixy \rightarrow Cxx)$
(Anything in a world is a counterpart of itself)

Since the name "counterpart theory" is often used to refer to the semantical use of counterparts altogether, one might think that it is essential to counterpart semantics that its core assumptions form a first order theory. But this is not so. The postulates only express those respects of counterpart relation and background metaphysics which are relevant for counterpart semantics. Anyway, a book about counterparts wouldn't be complete without introducing these core postulates.

P1 seems to be a stipulation that governs the use of I in the present context: let it only be applied to pairs of individuals and worlds. P2 states that worlds do not overlap. Given the intended interpretation of I and W, P3 and P4 restrict counterpart relations to hold between parts of worlds. Since, according to Lewis, every arbitrary sum of possible individuals is an individual, there are lots of individuals which neither are nor have counterparts.

In Lewis (1968), Lewis tries to derive P5 and P6 from the hypothesis that counterpart theory is a relation of maximal comparative similarity. Above in Chap.1 I discussed this hypothesis. I pointed out that P6 would already follow from the simpler assumption that the counterpart relation is a relation of similarity. P5 on the other hand, though suggestive, does neither follow from Lewis's hypothesis nor from the suggested replacement.

I pointed out that P6 does not hold for all flavours of modality. It holds for metaphysical modality, though. Since, in the following I will only be concerned with metaphysical modality, we can accept P6, here. I also said Lewis discarded P5 "for some applications", anyway. Since these applications include cases of metaphysical modality, this means we cannot take it to be axiomatic for metaphysical modality any longer, and we should drop it.

2.1.2 T

Counterpart-theory is applied in the following translation T of the language $\mathcal{L}_{QML}$ into the language of counterpart theory (i.e. the language $\mathcal{L}_{PL}$ with the counterpart-theoretical vocabulary).

Let $w = x_0$ a variable of $\mathcal{L}_{PL}$; suppose it does not occur in $\mathcal{L}_{QML}$. For formulae φ of $\mathcal{L}_{QML}$ (the language of modal predicate logic), φ^w [φ holds in w] is defined recursively as follows. Let $\varphi[u_1/v_1 \ldots u_n/v_n]$ be the result of the substitution of every free occurrence of $v_1 \ldots v_n$ in φ by $u_1 \ldots u_n$, resp.

1. $\varphi^w = \varphi$, if φ is atomic
2. $(\neg\varphi)^w = \neg\varphi^w$
3. $(\varphi \wedge \psi)^w = (\varphi^w \wedge \psi^w)$
4. $(\exists x\varphi)^w = \exists x(Ixw \wedge \varphi^w)$
5. if φ contains n free individual variables $v_1 \ldots v_n$ (listed in their order of appearance), then

$$(\Diamond\varphi)^w = \exists w \exists x_{i+1} \ldots \exists x_{i+n} (Ww \wedge Cx_{i+1}v_1 \wedge Ix_{i+1}w \wedge \ldots \wedge Cx_{i+n}v_n \wedge Ix_{i+n}w \wedge \varphi[x_{i+1}/v_1 \ldots x_{i+n}/v_n]^w),$$

where x_i is the greatest variable that occurs in φ.[1]

[1] Lewis simply says "variables introduced in translation are to be chosen in some systematic way that prevents confusion of bound variables" (p. 30), whereas I have tried to implement some such way.

Since $\Box\varphi$ is defined as $\neg\Diamond\neg\varphi$,

$$(\Box\varphi)^w = \forall w \forall x_{i+1} \ldots \forall x_{i+n} \\ ((Ww \wedge Cx_{i+1}v_1 \wedge Ix_{i+1}w \wedge \ldots \wedge Cx_{i+n}v_n \wedge Ix_{i+n}w) \rightarrow \\ \varphi^w[x_{i+1}/v_1]\ldots[x_{i+n}/v_n]).$$

E.g.

$$Px^w$$

$$(\Diamond Px)^w$$

$$(\Box Px)^w$$

are translated into

$$Px$$

$$\exists w \exists y((Cyx \ \wedge Iyw) \ \wedge Py)$$

and

$$\forall w \forall y((Cyx \ \wedge Iyw) \rightarrow Py)$$

respectively.[2] If φ is closed,

$$(\Diamond\varphi)^w = \exists w(Ww \wedge \ \varphi^x)$$

and

$$(\Box\varphi)^w = \forall w(Ww \rightarrow \ \varphi^x)$$

Hence, possibility *de dicto* is just truth in some possible world, and necessity *de dicto* is just truth in all possible worlds, as it should be.

Let's suppose @ is a constant of $\mathcal{L}_{PL}$ that refers to the actual world.[3] We say that φ is true in a model of counterpart-theory wrt. to an assignment iff $\varphi^w \wedge (w{=}@)$ is. E.g.

$$(\Diamond Rxy)^w = \exists w \exists x_3 \exists x_4(Cx_3x \ \wedge Cx_4y \ \wedge Ix_3w \ \wedge Ix_4w \ \wedge Rx_3x_4)$$

so $\Diamond Rxy$ is predicted to be true iff there are a world and counterparts of the values assigned to x and y, resp. which are R-related.

The main idea of counterpart semantics is, of course, satisfaction by a counterpart. There are three features that T adds to the main idea. First

[2] I simplify, where possible. E.g. I omit Ww, where it is redundant.

[3] Lewis works with a *description* of the actual world, instead.

its treatment of quantifiers: quantifiers range over individuals in the relevant world. Additionally, there are two features pertaining to the semantics of sentences that are multiply *de re*: second, the multiple *de re* is interpreted in terms of a *sequences of counterparts*, third, sequences of individuals taken from one and the same world. All three features will be subject to a critical examination below.

2.1.3 S

T could be used in two different but related ways. First one could regard it as an indirect definition of validity (i.e. logical truth) for the language of QML. A formula φ of QML is valid, then, exactly if its T-translate is true in every model of counterpart theory.

Secondly, given some fixed model of counterpart-theory, the translation scheme provides an *indirect interpretation* of QML. We have seen how T defines a truthvalue for every formula of QML. One could also define other semantical values, e.g. one might define the *proposition* expressed by φ as the set of worlds that satisfy φ^w.

If QML is taken as a formal language, then there is no sense in which either the derived definition of validity or the derived semantics for QML could be wrong: we can take formal languages to mean whatever we want. But we could understand QML expressions to stand for certain expressions of English. Indeed most discussions presuppose some tacitly understood translation S from English$^-$, a part of English, into QML. The net result (ST)

$$\text{English}^- \overset{S}{\rightsquigarrow} \text{QML} \overset{T}{\rightsquigarrow} \text{CT}$$

provides an indirect indirect interpretation of English$^-$ and hence, a derived definition of logical validity and a derived semantics for English$^-$. It is in this sense that Lewis's translation scheme can be evaluated. Does ST provide a correct definition of logical validity? Does it assign correct truth-values in the intended model of counterpart theory? And do the semantical values assigned in the intended model satisfy our demands on such values? We will also ask whether ST could be extended to to the analysis of expressions of English that did at least not receive a separate meaning by the translation so far.

In this chapter I will be concerned both with logical validity and with ST's predictions about truth-values. The next chapter will be concerned with an inspection of the compositional semantics of ST.

2.1.4 Characteristics of ST

ST incorporates the basic form of counterpart semantics presented in Chap.1. E.g.

$$\text{"possibly, } x \text{ wins"} \overset{S}{\rightsquigarrow} \Diamond Px^w \overset{T}{\rightsquigarrow} \exists w \exists y((Cyx \ \wedge Iyw) \ \wedge Py)$$

So possibly, Humphrey satisfies "possibly, x wins" iff there is a winning counterpart of Humphrey in some world.

Other ingredients of the semantics follow naturally. Because $\wedge$ translates "and", there can be no question that the T-translation of $\wedge$ is correct, the same holds for the other connectives.

Some features of ST are of a different kind. They are neither forced upon us by the general idea behind counterpart semantics nor by the intended interpretation. One of these features is

(A) Quantifiers are actualist.

Possibilist quantifiers range over all possible individuals. *Actualist* quantifiers only range over individuals in any given evaluation-world. The name "actualist" is probably a misnomer, because, if the evaluation world is not the actual one, the quantifier ranges over non-actual individuals. Lewis prefers to speak about *unrestricted*quantification and quantification restricted to a world.

Anyhow, clearly, T interprets the quantifiers of QML as being restricted to a world or actualist. This feature of T is independent from the basic motivation of counterpart semantics: a semantics with possibilist quantifiers would also be viable.

Two other such independent features of Lewis' implementation of the basic idea of counterpart semantics pertain to the multiple *de re*, i.e. possibilities for more than one individual. We might e.g. say that

(1) Peter could have married Paul

to express that it is true of Peter and Paul that they could have married. ST treats this in the following way:

$$\text{“}x \text{ could have married } y\text{”}$$
$$\overset{S}{\rightsquigarrow} \Diamond Rxy^w \overset{T}{\rightsquigarrow} \exists w \exists x_3 \exists x_4 (Cx_3x \ \wedge Cx_4y \ \wedge Ix_3w \ \wedge Ix_4w \ \wedge Rx_3x_4)$$

So, Peter could have married Paul iff there is a world with two counterparts of Peter, and Paul, resp. such that the counterpart of Peter marries the counterpart of Paul.

According to counterpart semantics to it is possible for a pair (A, B) to stand in relation R if there is another pair (C, D) defined in terms of the counterpart relation that stands in that relation itself. E.g. in the case of ST, to have married is possible for (Peter, Paul) iff some pair of a counterpart of Peter and a counterpart of Paul is a married couple. We could call such a pair (C, D) a *counterpart of the pair* (A, B).

Now put in terms of counterparts of pairs the analysis of the multiple *de re* incorporated in ST makes the following requirements on the notion of a counterpart of a pair.

(B) Every pair of a counterpart of A and a counterpart of B in the same world is a counterpart of the pair (A, B).

(C) If (C, D) is a counterpart of (A, B), then C and D are in the same world.

Again, (B) and (C) are not forced upon us by the basic conception of counterpart semantics. One could as well have a counterpart relation between pairs that is not defined in terms of the counterpart relation between individuals. Such a relation need neither satisfy (B) nor (C).[4]

In the following I will scrutinize (A) - (C). I will argue that they should be given up: the translation should employ possibilist quantifiers and it should not require (B) and (C) to hold.

---

[4](B) and (C) are also part of the so-called functional semantics for quantified modal logic.

> As to the necessity operator, we should say that an n-tuple of individuals necessarily enjoys the property A iff all the n-tuples of their respective counterparts enjoy it. (Corsi and Ghilardi (1992), p. 187)

.

2.2 Possibilist Language

2.2.1 Possibilism and Semantical Reflection

You may give a semantics for the English language that is totally fictitous, e.g. you may define the term "red" to mean blue. The project of a semantical analysis of English is instead to define meanings for expressions that are faithful to the meanings these expressions actually possess. This enterprise has an important consequence for the relation between object- and metalanguage: it should conform to the following principle.

The principle of semantical reflection When the object-language is part of the meta-language, the meanings defined should agree with the meanings the same terms have in the meta-language. Furthermore, the semantics should be extendable to other (non-semantical) vocabulary of the object-language.

Counterpart semantics is a case in point. In statements like "some counterpart of Humphrey wins" or metaphysical principles like plenitude, quantifiers like "some" can clearly be seen to range over arbitrary possible individuals. If counterpart semantics treats such statements at all, it should interpret these quantifiers to range over arbitrary possible individuals. On the other hand, it should be able to treat such statements, because it should be extendable to other non-semantical vocabulary of the object-language.

Furthermore, clearly, in our meta-language, some relations are trans-world relations. It is therefore that sentences like

(2) There are pairs of individuals which are in different worlds

(3) Some counterpart of Humphrey wins

(11) for every way W that an individual A could be in there is an individual B, such that (i) every way B is in is a way A could be in and (ii) B is in W

are true in the meta-language (and in fact necessarily so). Because of semantical reflection, when treating such sentences as part of the object-language they should receive an interpretation that allows them to be true as well.

2.2.2 Unrestricted Quantification

In Lewis (1986a), Lewis contends that normally, natural language quantifiers range only over actual individuals. But when the modal realist says, e.g. that there are things in different worlds, then his quantifiers range over arbitrary possible individuals, including worlds. The quantifiers in CT are possibilist, likewise. Lewis calls this kind of quantification "unrestricted".[5]

Ordinary uses of quantification, then, are *restricted* uses: the domain of quantification only contains a subset of all possibilia. Restricted quantification is a frequent phenomenon; quantifiers may be restricted in many different ways. If, during a lecture, the professor notices "everyone is asleep" he is most probably restricting quantification to all persons in the audience; if someone in a restaurant pronounces that noone smokes anymore he may well refer only to people in restaurants in Italy. Sometimes quantifiers range over all the individuals in the actual world. E.g. "there are no honest politicians" is understood as "there are no honest politicians *in the world*". If you do not believe in possibilia, this may be unrestricted quantification to you. In the present possibilist framework it is interpreted as a case of restricted quantification.

The interpretation to QML, provided by T is restricted in the same way: quantifiers are interpreted to range over individuals in a world.

This means that according to T, the quantifiers in QML are not fit to represent unrestricted quantification; if S translates unrestricted "every" into $\forall$ in QML, then the net result (the ST-translation of "every") is incorrect. Better assume that S only translates implicitly restricted uses of "every" by $\forall$. Then, unrestricted quantification is not dealt with at all.

So depending on how S treats unrestricted "every", the semantics of quantification provided by ST is either not correct or at least not complete, at least as far as David Lewis's own views about quantification are concerned, views that are presupposed by CT itself.

If we want to give a complete and correct semantics of English, we should give up (A) above and alter T. We should first translate "every" into unrestricted quantification over all possibilia, likewise for other quantifiers. Then we should express restricted quantification by restricting unrestricted quantifiers. Since the language of CT does not contain *implicit* restrictions, whenever a quantifier is implicitly restricted, our S-translation has to make the

[5]Unrestricted quantification in this sense is not *absolutely* unrestricted, e.g. the domain of quantification does not contain sets.

restriction *explicit*. If "actualist" quantifiers are implicitly restricted, they have to recieve a translation where the restriction is made explicit, too.

Consequently, we should translate restricted existential claims into QML as follows:

(4) There are P $\overset{S}{\rightsquigarrow} \exists x(Ix@ \wedge Px)$

2.2.3 Expressive Completeness?

One may classify versions of QML wrt. to the following characteristics:

Weak necessity : $\Box\varphi$ is true according to w wrt. to an assignment iff φ true according to every world in which all the individuals assigned to the open variables in φ exist.

Strong necessity : $\Box\varphi$ is true according to w wrt. to an assignment iff φ true according to every world.

Actualist Quantification : quantifiers range over the individuals in the evaluation world.

Possibilist Quantification : quantifiers range over all possible individuals.

Actually : The language contains an operator A, such that $A\varphi$ is true according to w iff $A\varphi$ is true (according to the actual world).

In Hazen (1976), Allen Hazen presents a hierarchy of kinds of languages, defined in terms of the above characteristics, with respect to their expressive power. E.g. languages with weak necessity, actualist quantification, and without A cannot express certain things languages with strong necessity, possibilist quantification and "actually" can. Finally there are certain things that can only be said with explicit quantification over possible worlds.

Here I will focus on the following kind of expressive weakness: the language of QML, under the interpretation induced by T, cannot properly express trans-world relations. Suppose a closed formula of QML contains some relational predicate R which we take to express some trans-world relation, like "is in a different worlds than" or "counterpart". The problem is that, provably, nothing we can say with the help of R transcends what we can say with the help of certain inner-world relations, here the empty relation, and the relation that holds between things which are similar *and worldmates*, resp.

(For details, see Appendix B.) Now, the use of such terms by possibilists is clearly intended to express things that are not expressible with inner-world relations. Sentences in which replacing a trans-world relation by an inner-world one is supposed to make a difference, are not correctly translatable by S. They are either not translated at all or translated incorrectly.

Closer analysis reveals that the culprits are features (A) and (C) of our translation T. They conspire to ensure that relational predicates in closed formulas are always satisfied by pairs of worldmates.

2.3 Essential Relations

2.3.1 Essential Relatives

For the problem of essential relations, see Hazen (1977) and Lewis (1983b). I consider Lewis's example here. Dee and Dum are a pair of twins. Essentialists want to say that they are essentially twins. For the sake of simplicity, we will assume that both in $\mathcal{L}_{QML}$ and $\mathcal{L}_{PL}$ A and b are names of Dee and Dum, respectively.[6]

(5) Essentially, Dee and Dum are twins

$\Box Rab$

$\forall w \forall x \forall y((Ixw \ \wedge Cxa \ \wedge Iyw \ \wedge Cyb) \rightarrow Rxy)$

is false if there are worlds in which there are two counterparts Dee', and Dee" of Dee and two counterparts Dum' and Dum" of Dum such that Dee' and Dum' are twins and Dee" and Dum" are twins. In that case, Dee' and Dum" are not twins, of course. If w refers to such a world,

$$\forall x \forall y((Ixw \ \wedge Cxa \ \wedge Iyw \ \wedge Cyb) \rightarrow Rxy)$$

is false, and because of that (5) comes out false, too. But given the metaphysical framework of this book we have to admit that there are such worlds. This is so because it is possible that there are symmetrical worlds, which contain indiscernible duplicates of individuals and even duplicate pairs of a pair of twins.

[6]CT does not contain individual constants. But we could add names. The use of names in the present chapter serves only for purposes of exposition and could be entirely avoided.

The precise nature of the example is not important. It is only meant to illustrate a general feature of ST, *viz.* that relational essentialist theses such as the above invariably fail because of the existence of multiple counterparts. Now it was one of the main selling-points of counterpart semantics that it is able to accommodate essentialist intuitions. It would be strange if this advantage were limited to essential properties. But, so far, the same cannot be achieved in the case of essential relations. So, even if you do not agree with the above essentialist thesis yourself, that T is unable to accommodate the intuition should be judged as a fault.

But we can achieve for essentialism about relations precisely the same that counterpart theory does for essentialism about properties. Simply replace the counterpart-relation by a counterpart-relation that relates n-tuples of the same length; do not require (B) above to hold for this relation. Then, the semantics of $\Diamond$ and $\Box$ should be altered, and put in terms of this generalised counterpart relation, e.g. as

(6) if φ contains n free individual variables $v_1 \ldots v_n$ (listed in their order of appearance), then

$$(\Diamond\varphi)^w = \exists w \exists x_{i+1} \ldots \exists x_{i+n}(Ww \wedge C(x_{i+1}, \ldots, x_{i+n})(v_1, \ldots, v_n) \wedge Ix_{i+1}w \ldots \wedge Ix_{i+n}w) \wedge \varphi^w[x_{i+1}/v_1] \ldots [x_{i+n}/v_n]),$$

where i is the greatest natural number such that x_i occurs in φ.

The resulting translation of (5) is

(7) $\forall w \forall x \forall y((Ixw \wedge Iyw \wedge C(x,y)(a,b)) \rightarrow Rxy)$

In a postscript to Lewis (1968) in Lewis (1983e), Lewis reaction is slightly different. He accepts that we need counterparts of pairs, but thinks we can combine them with something very much like the original semantics of $\Diamond$. His solution is to say that (5) QML is not the correct way to formalize (5). Instead, we should say that "twin" is a predicate of a pair of individuals and render (5) as

(8) $\Box Rc$

where c refers to the pair of Dee and Dum. (If there are nonsymmetrical essential relations, such pairs have to be ordered.) And only the pair of Dee' and Dum', or the pair of Dee" and Dum", are eligible candidates for being counterparts of Dee and Dum, but not the pair consisting of Dee' and Dum" because this pair is less similar to Dee and Dum than the first two pairs.

Perhaps (8) *is* a correct translation of (5), because of the plural "are" in (5). But other cases are different. E.g. if we say that Ede is Tom's father, S should translate "father" into a relation between two individuals rather than into a property of the pair; the same holds if we say that, necessarily Ede is Tom's father. To translate this sentence into something analogous to (8) would be *ad hoc*.

Likewise, to say that "possibly, Dee is not related to Dum", should be translated as $\Diamond\neg Rc$ would be an *ad hoc* modification of S. If we do not make this modification, on the other hand, Lewis' semantics still predicts that "possibly, Dee is not related to Dum" is true. Now this is precisely what essentialists about twinhood want to deny, so Lewis cannot claim to accommodate the essentialist intuitions other than by ad hoc modifications of the way we render English in QML.

2.3.2 Identity

The following examples of essentialist theses are even more entrenched. They concern the essentiality of identity and non-identity. Under CT neither (NI) nor (NNI) below receive valid translations.

(NI) $x = y \rightarrow \Box x = y$

(NNI) $x \neq y \rightarrow \Box x \neq y$

By CT these formulae are translated into

$$x = y \rightarrow \forall w \forall x_2 \forall x_3 ((Cx_2x \wedge Cx_3y \wedge Iwx_2 \wedge Iwx_3) \rightarrow x_2 = x_3)$$

and

$$x \neq y \rightarrow \forall w \forall x_2 \forall x_3 ((Cx_2x \wedge Cx_3y \wedge Iwx_2 \wedge Iwx_3) \rightarrow x_2 \neq x_3),$$

respectively. All of these formulae may come out false because the single individual assigned to x and y may nevertheless have different counterparts in another world.

Again, we should adapt the above semantics in terms of multiple counterparts. Simply require that a counterpart of a pair is an identity pair iff the original pair is. This is, essentially, the solution in Hazen (1977).

These remarks suggest that the necessity of identity and the necessity of non-identity is something we might have or might not have, depending on whether we are essentialists about (non-)identity. Instead I will opt for a different solution: I propose that the requirement should always be in force. Seeming examples of contingent identities and contingent non-identities always involve a mixture of different counterpart relations, as discussed in Sec.1.5.5; therefore they do not require a counterpart-relation which enables contingent identities and non-identities.

2.4 The Logic of T

2.4.1 Principle T

There is a feeling that modality *de re* is not very important and that wide parts of its logic are negotiable, a case of "spoils to the victor". E.g. in Divers (2002), pp.142-3, Divers claims there is no problem with logic. According to him, Lewis' analysis only fails to validate those theorems of classical modal logic which are negotiable, anyway. He goes on:

> It might be argued that certain modal principles are no more negotiable than those of non-modal predicate logic. If we seek an applied semantics for alethic modal logic, two prominent candidates are:
>
> **(viii)** $\Box A \rightarrow A$
>
> **(ix)** $A \rightarrow \Diamond A$
>
> But the validity of these central modal principles is not at risk in CT semantics.

I agree with Divers that defenders of translation T have to care for core principles of modal logic at most. These include the following:

(T) $\Box\varphi \rightarrow \varphi$

(K) $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$

(LL) $x = y \rightarrow (\Diamond\varphi \leftrightarrow \Diamond\varphi[x])$
(where $\varphi[x]$ is the result of replacing some free occurrences of y in φ by x)

But I think it is also true that if you want to defend translation T, then you have to care for these examples. Why, these axiom schemes are intuitively valid for metaphysical modality; this applies to open instances of them as well. E.g. if A_1 and A_2 are necessarily R-related, then *of course* they are themselves R-related.

Unfortunately, contrary to what Divers says, the validity of principle (T) and the equivalent principle of possibility introduction $\varphi \rightarrow \Diamond\varphi$ *are* at risk under CT semantics. E.g. these things come out invalid under translation scheme T. Take e.g. possibility introduction. Let R be the relation that holds between any x and y exactly if they are in different worlds. Suppose A_1 and A_2 are R-related. (8.b) predicts a_1, a_2 are possibly R- related iff there is a world w, and counterparts B_1, B_2 in w of A_1, A_2 resp. such that Rb_1, b_2, i.e. b_1, b_2 are in different worlds. Since every individual is only in one world, this condition is impossible to satisfy, so A_1 and A_2 are not possibly R-related. Hence, they do not satisfy "if x_1, x_2 are R-related, then x_1, x_2 are also possibly R-related", either; i.e. possibility introduction is invalid according to Lewis' counterpart theoretic translation. What is worse: since the intended model of counterpart theory is one where there are different worlds, and hence, pairs of things that are in different worlds, possibility introduction is predicted to be false in the intended model.[7]

The culprit is requirement (C) above on the notion of a counterpart of a pair. If (C, D) is a counterpart of (A, B), then (C, D) should not be required to be worldmates, things taken from the same world. (C) is built into Lewis semantics of $\Diamond$, but it is also built into our revised semantical rule for the diamond, (6) above, so we should drop the requirement from that rule, too.

2.4.2 More on Quantified Modal Logic

I will now briefly show that the other core principles of modal logic mentioned above also fail to be valid.[8]

[7]It has to be mentioned that the rule of uniform (second order) substitution, which allows, in a formula φ, to substitute uniformily any formula ψ for every occurrence of some atomic sub-formula χ, for details see Bauer and Wansing (2002), fails as well: while $\Box Px \rightarrow Px$ is valid, $\Box Rxy \rightarrow Rxy$ is not. Bauer and Wansing (2002) argue that this rule is essential to the notion of a logic because it incorporates the idea that logic is structural, put semantically it incorporates the idea that logical validity is truth in virtue of the meanings of the logical constants.

[8]For an overview of the theorems and non-theorems of Lewis's scheme, as well as of alternative versions of counterpart theoretic translation, see Ramachandran (1989, 1998).

Below you find a consequence of (K) and its T-translation.

(9) $\Box(Px \wedge Py) \rightarrow \Box Px$

$$\forall w \forall x_2 \forall x_3 (Cx_2 x \wedge Cx_3 y \wedge Iwx_2 \wedge Iwx_3 \rightarrow (Px_2 \wedge Px_3) \rightarrow \forall w \forall x_2 (Cx_2 x \wedge Iwx_2 \wedge Px_2)$$

The translation is not valid in CT, indeed it is easy to imagine counterexamples. E.g. let there be just two worlds w, v and two individuals A, B. A is in w and only in w, whereas B is only in v. The counterpart relation is the relation $\{(A, A), (B, B)\}$. Neither individual is P. So, A, B satisfy $\forall w \forall x_2 \forall x_3 (Cx_2 x \wedge Cx_3 y \wedge Iwx_2 \wedge Iwx_3 \rightarrow (Px_2 \wedge Px_3)$ but A does not satisfy $\forall w \forall x_2 (Cx_2 x \wedge Iwx_2 \wedge Px_2)$, hence A, B fail to satisfy (9). By the same token, (K) is predicted to be invalid, (see Hughes and Cresswell, 1996, p.356), so the logic of counterpart theory is not even a classical normal modal logic.

And here is an instance of (LL) which can easily be seen to fail.

(10) $x = y \rightarrow (\Diamond(Px \wedge Qy) \rightarrow \Diamond(Px \wedge Qx))$

$$x = y \rightarrow \exists w \exists x_2 \exists x_3 (Cx_2 x \wedge Cx_3 y \wedge Iwx_2 \wedge Iwx_3 \wedge Px_2 \wedge Qx_3) \rightarrow \exists w \exists x_2 (Cx_2 x \wedge Iwx_2 \wedge Px_2 \wedge Qx2)$$

Suppose w is a world with two counterparts of A in it, one bald and the other one not. Let P be the property of being bald and Q the property of being not bald. Then A satisfies $\exists w \exists x_2 \exists x_3 (Cx_2 x \wedge Cx_3 y \wedge Iwx_2 \wedge Iwx_3 \wedge Px_2 \wedge Qx_3)$ without satisfying $\exists w \exists x_2 (Cx_2 x \wedge Iwx_2 \wedge Px_2 \wedge Qx2)$ at the same time, hence A fails to satisfy the translation of (10) which is thereby predicted to be not valid.[9]

(LLI) could be easily restored by the combination of two things I have proposed above, *viz.* a semantics based on counterpart of pairs together with the requirement that a counterpart pair is a pair of identicals exactly if the original pair is.

[9] In Lewis (1986b) to Lewis (1968)) Lewis claims that nothing is amiss. He claims that he can stick to Leibniz's Law while at the same time admitting the invalidity of (10). How is this possible? He claims that (10) is not an instance of Leibniz's Law at all. Why? Simply look at its translation. This formula is not an instance of Leibniz's Law. Here, the outcome of the translation is allowed to overrule the preexisting intuitions about what are instances of (LL).

None of the formulas discussed in this section presents a problem for the revised counterpart semantics I have argued for. This is a semantics based on counterpart of pairs. In Appendix A I present the details of such a semantics.

2.5 Contingent Existence

2.5.1 The Problem

According to translation T, every possible individual satisfies

(11) $\Box\exists y(y = x)$

because this is translated as

(12) $\forall w\forall y((Cyx \wedge Iwy) \rightarrow \exists z(Iwz \wedge y = z))$,

which is trivially true.

Likewise, according to T, no individual satisfies

(13) $\Diamond\neg\exists y(y = x)$

is , because this is translated as

(14) $\exists w\exists y((Cyx \wedge Iwy \wedge \neg\exists z(Iwz \wedge y = z))$,

which is trivially false.

But surely, there is a sense in which I, e.g., could fail to exist. So, if (13) translates

(15) possibly, x does not exist

in that sense, then T's semantics is inadequate.

Now above I have argued that quantifiers are not treated satisfactorily by ST and that T should take the quantifiers of QML to express unrestricted natural language quantification. Now, if we alter S and T as proposed there, (13) is correctly predicted to be unsatisfiable, because it now says that it is possible that x is not a possible individual, and of course no possible individual is such an x.

But even if I am right this does not rehabilitate Lewis' semantics. For consider the following:

Any sentence of the form of (16) below should be rendered in QML in the form of (17) below. But if we stick to the counterpart translation (v) of $\Diamond$, then that QML-formula will finally be translated in the form of (18).

(16) Possibly, $\ldots x \ldots$

(17) $\Diamond \ldots x \ldots$

(18) $\exists w \exists y \ldots (Cyx \ \wedge Iwy \wedge \ldots y \ldots)$

Applied to our example this means that the CT-translation of (15) will express that x has a counterpart in w which is *in* w and there satisfies a formula that expresses non-existence. But by being in w it seems, such a counterpart would represent x to be existing in the world after all. Summing up, however we express non-existence, T is not able to translate the relevant sense of (15) correctly.

2.5.2 . . . as the Possibility of Having no Counterpart

This problem has motivated a number of people to devise alternatives to T (e.g. Forbes (1982); Ramachandran (1989)). The main feature of these alternatives is

(19) A satsifies (15) iff there is a world w which contains *no counterparts* of A.

In Lewis (1986a), Lewis himself accepts (19). (19) is plausible enough, e.g. what could it mean for Humphrey to fail to exist according to a world than that the world contains no counterparts of Humphrey?

But note that this reasoning for (19) presents a departure from what I have called the main motivation of counterpart semantics: that possibility *de re* is analysed in terms of the properties of a counterpart of Humphrey; in Lewis' terms: although (15) is certainly *de re*, it is not analysed in terms of *vicarious satisfaction*, i.e. satisfaction by a counterpart of Humphrey.

This observation is not in itself a refutation of (19). Indeed, the plausibility of (19) could be taken to indicate that what I have called the main motivation of counterpart semantics has to be given up. It could also be taken to indicate that possibility de re has to be analysed differently in different cases: sometimes Humphrey satsifies "possibly φ" by having a counterpart that satisfies φ, and sometimes he satisfies "possibly φ" because of the existence of a world with no counterpart of Humphrey in it.

Graeme Forbes proposes the following solution. He wants to stick to the truth-conditions for sentences like (15) proposed by Lewis but tries to derive them compositionally.

First let us define that y is a *Forbesian representative* of x at w, if either

1. there is *no* counterpart of x in w and $y = x$, or
2. y is a counterpart of x

Now replace "counterpart" by "Forbesian representative" in Lewisian translations. This has the following effect:

- "Possibly, Humphrey doesn't exist" is true iff there is a world w, and a Forbesian representative of Humphrey at w which does not exist in w

  (i.e. iff Humphrey has no counterparts at some world)

- "Possibly, Humphrey is not human" is true iff there is a world w, and a Forbesian representative of Humphrey at w which is not human in w

  (i.e. iff somewhere, Humphrey has a counterpart which is not human)

Forbes' proposal predicts exactly the truth-conditions Lewis accepts but is unable to derive compositionally. But it does so at a price. The price is a *disjunctive account of representation.* A counterpart represents Humphrey *in virtue of their similarity*[10]; sometimes Humphrey represents himself, but then he does *in virtue of there being no counterpart of Humphrey.*

In Lewis (1986a), fn.6, p.10–11, Lewis claims Forbes has a disjunctive semantics of satisfaction *in absentia* in disguise. In Forbes (1987), the author protests that this allegation is not fair. However, either we take Forbes' non-disjunctive semantics of satisfaction at face value. Then it is his account of representation which is disjunctive. Or we take this treatment of representation merely as definitional trick that aims to hide a disjunctive semantics of satisfaction. Either way, something is treated in a disjunctive way in Forbes' theoretical package.

So if you want to combine (19) with a semantics that analyses other cases of possibility de re in terms of satisfaction by a counterpart, at least one ingredient of your semantics will end up being disjunctive; and since

[10]or in virtue of whatever it is counterparts represent in virtue of

(19) is an exception made for a single case, the modification of counterpart semantics is furthermore *ad hoc*.

But is (19) correct in the first place? There is room for doubt.

Suppose Humphrey has no essential properties worth speaking of, i.e. the only properties he has essentially are trivial properties, so he might be a mountain on the moon. Every possible individual is in a way, Humphrey could be; hence every such individual is a counterpart of Humphrey. According to (19), there is still one thing which is impossible for Humphrey to accomplish: he cannot fail to exist. Because he has counterparts in every world, there is no world in which he has no counterparts. And, according to (19) this means, he exists according to every world.

There may indeed be some sense of "exists" in which Humphrey exists necessarily — the sense in which to exist means being a possible individual. But here, we are dealing with a sense of "exists" in which existence is a contingent matter, the sense where, roughly, to exist means being in the world. Metaphysicians sometimes ponder the idea of a necessary being in the second sense, I admit. But, as far as I know, nobody ever proposed simply a being with a weak essence as a specimen of a necessary being.

There is another problem with (19). Since the counterpart relation is qualitative, (19) says that Humphrey's existence or non-existence is decided by the way the world is. This excludes that both of the following are true: (i) Humphrey could exist under certain circumstances and (ii) Humphrey could fail to exist under the same circumstances. In Chap.1 we have seen that it is an advantage of counterpart-theory that it is able to account for seemingly haecceitist differences. Now here we have an exception: with (19) in place, we are not able to account for seemingly haecceitist differences wrt. to existence.

I conclude that there is at least some reason to doubt that (19) is correct. Furthermore, it does not agree with the rest of counterpart semantics. Therefore, we should try to find a solution to the problem of providing truth-conditions for contingent existence which is not based on (19).

2.5.3 Lewis' Reaction

Lewis already discusses this problem in the original paper. He mentions a proposal by David Kaplan to translate "necessarily" by using the existential instead of the universal quantifier, and a proposal to mix the diamond of his proposal with Kaplan's box. All of these proposals are rejected.

Is something wrong with counterpart semantics? And if, what is? Lewis doesn't put the blame on T. In Lewis (1986a), he apparently puts it on the syntax of $\mathcal{L}_{QML}$.

> Shall we dump the method of counterparts? – That wouldn't help, because we can recreate the problem in a far more neutral framework. Let us suppose only this much. (1) We want to treat the modal operators simply as quantifiers over worlds. (2) We want to grant that Humphrey somehow satisfies various formulas at various other worlds, never mind how he does it. (3) We want it to come out that he satisfies the modal formula "necessarily x is human", since that seems to be the way to say something true, namely that he is essentially human.(4) We want it to come out that he satisfies the modal formula "possibly x does not exist", since that seems to be the way to say something else true, namely that he might not have existed. (5) We want it to come out that he does not satisfy the modal formula "possibly x is human and x does not exist" since that seems to be the way to say something false, namely that he might have been human without even existing. *So he satisfies "x is human" at all worlds* and "x does not exist" at some worlds; so he satisfies both of them at some worlds; yet though he satisfies both conjuncts he doesn't satisfy their conjunction! How can that be? (p. 11, emphasis added by me)

Finally, after having rejected possible repairs, Lewis suggests:

> If this language of boxes and diamonds proves to be a clumsy instrument of talking about matters of essence and potentiality, let it go hang. Use the resources of modal realism *directly* to say what it would mean for Humphrey to be essentially human, or to exist contingently. (p. 13)

I have to admit I find this passage deeply puzzling. The problem he raises does not seem to be a problem about QML but rather about the semantics of "possibly" and "necessarily". So, if successful, rather than establishing that QML is defective, his argument would show that no possible worlds semantics can be given for these core modal notions. If this were true, then this would be devastating for his own philosophical enterprise. What good are possible worlds and individuals, if they are of no use for the semantics of modality?

Lewis's argument, however, does not show what it purports to show; namely, that if we take the constraints on the interpretation of modal talk (or is it the interpretation of QML?) (1)–(5), we have to face the paradoxical conclusion that a conjunction maybe false in a case where both conjuncts are true.

The reason is that there is an additional, but tacit, assumption. That Humphrey satisfies "x is human" at all worlds, and hence at some worlds where he also satisfies "x does not exist", is only true if "necessarily φ" implies that φ is *true at all worlds*. The tacit assumption which Lewis seems to make is that we are dealing with such a semantics. This is somewhat inconsequential, for it does not hold for Lewis' own CT. It is an assumption which anybody who wants to combine (1)-(5) should try to avoid. And it is possible to combine (1)-(5) with a semantics which does not treat necessity as truth at all possible worlds. E.g. one could treat modal operators as quantifiers that range over all the worlds *that meet the presuppositions of the embedded sentences*. Then necessity would be treated as truth at all possible worlds where the presuppositions of the embedded sentence are met and possibility truth at some such world. Now it is plausible that "Humphrey is human" presupposes that Humphrey exists, while "Humphrey exists" and "Humphrey does not exist" do not presuppose this. This would explain why "Necessarily, Humphrey is human" can be true while at the same time there is a world where "Humphrey does not exist" is true. One could implement this proposal by combining any formal treatment of presuppositions with QML, e.g. in a three-valued semantics.

Finally, I want to raise some doubts about Lewis's condition (5). Is it really so odd to say that there is a possible world in which Humphrey is human, and yet does not exist? I don't think so. According to modal realism, there are lots of possible individuals which are human, but do not exist in the sense that they are not part of our world (or in the sense that they have no counterparts in our world). If this is actually true, it is possibly true.

Summing up, Lewis' argument does not show what it purports to show. As far as that argument is concerned, QML still seems to be an adequate medium for an enterprise that aims at a counterpart semantics for natural language. Of course, QML is a rather simplistic language. Because of its simple syntax it could not be expected to be an all-purpose intermediary language for the purpose of an indirect interpretation of natural language. Here, type-logical languages seem to be much more useful. But still, there does not seem to be any problem with QML's treatment of necessity and possibility, which, incidentally, is part of any standard treatment in intensional type logic, see e.g. Montague (1974).

A fortiori, the argument does not show that no possible-worlds semantics can be given for necessity and possibility.

2.5.4 A Proposal

We want to translate

(15) possibly, x does not exist

where existence is meant in the restricted sense of an actualist restricted quantifier.

To that end I propose simply to apply the various changes to ST we have argued for in this chapter.

First, we have proposed to translate quantifiers into possibilist ones and to represent actualist quantifiers by restricting the possibilist quantifiers, in the following way.

(4) There are P $\overset{S}{\rightsquigarrow} \exists x(Ix@ \wedge Px)$

If we apply that to our case, we get that

(20) x exists $\overset{S}{\rightsquigarrow} \exists y(Iy@ \wedge x = y)$

(21) x does not exist $\overset{S}{\rightsquigarrow} \neg\exists y(Iy@ \wedge x = y)$

(22) possibly, x does not exist $\overset{S}{\rightsquigarrow} \Diamond\neg\exists y(Iy@ \wedge x = y)$

Second, we should apply the changes to T we argued for above. So, relative to any assignment of values to variables, $\Diamond\neg\exists y(Iy@ \wedge x = y)$ should be taken to express that there is a counterpart (w, B) of the pair $(@, A)$ of the actual world @ and the value assigned to x, A, such that B is not in w. Now we have argued above that a counterpart of a pair should not be required to consist of a pair of worldmates. And this is why the revisions proposed to T and S above allow

(15) possibly, x does not exist

to have true instances; in accordance with the intuition that Humphrey could have failed to exist.

The solution also conforms to those intuitions that posed problems for the treatment of contingent existence as having no counterparts in some world in

Sec.2.5.2. I argued that a person with no essential properties worth speaking of should be able to fail to exist; and that there are seemingly haecceitist differences wrt. to non-existence. Now, according to our proposal, it turns out that a person with no essential properties worth speaking of could easily fail to exist. Suppose, e.g. that every pair (w, B) is a counterpart of $(@, A)$, provided w is a counterpart of @. Then A satisfies (15) already if there is more than one world.

Furthermore, the proposal allows to account for seemingly haecceitist differences wrt. non-existence. A might exist under circumstances W and fail to exist under the same circumstances (for some complete qualitative way the world might be W), if the following holds: for some W-world w, $(@, A)$ has a counterpart (w, B) with B being part of w and a counterpart (w, C) with C being not in w.

Chapter 3

Possible-Worlds Semantics

Indirectly, S and T provide a counterpart semantics for a part of English. They do so in a possible-worlds framework. Nevertheless, as I will show, this kind of counterpart semantics is not consistent with basic assumptions of possible-worlds-semantics. Even the revised counterpart semantics of the last chapter does not improve the situation.

3.1 Basic Possible-Worlds Semantics

The basic assumptions of possible worlds semantics include:

The possible-worlds account of meaning and content

What we express when we say some declarative sentence is called a proposition. The *content* of a sentence (in a context of use) is the proposition it expresses. The possible worlds-account of the proposition expressed by a sentence identifies propositions with truth-conditions, truth-conditions in turn are defined as sets of possible circumstances, in the following way.

Suppose utterances of two declarative sentences express two propositions p and q. If there is a possible circumstance in which p is true, while q is false, then these propositions cannot be the same. So, the above reasoning shows that propositions are different if they have different truth-values in some possible circumstance. The possible-worlds account of propositions also affirms the converse: two propositions are identical if they are true in precisely the same circumstances. Therefore,

a proposition is identified with the set of circumstances in which *it is true.*

Meaning is then defined in terms of content. For many purposes it suffices to consider the meaning of a declarative sentence as the proposition expressed by arbitrary utterances of that sentence. Sometimes the meaning of a sentence is considered to be a certain function from contexts of utterances of the sentence into the proposition expressed in those contexts. This is the notion of meaning employed in Kaplan's account of the meaning of indexicals and demonstratives. For other purposes we might still need more complicated accounts of meaning, e.g. the meaning of sentences might be understood in terms of their contribution to larger parts of texts, and the meaning of the latter could be understood in terms of propositions, see e.g. Heim (1983). Meanings of other types of sentences could also be defined in terms of propositions; e.g. the meanings of questions are sometimes defined to be sets of propositions.

Compositionality
The content of a complex expression is a function of the contents of its parts (and their mode of combination). What complex expressions and their parts are is to be defined by linguistic theory. E.g. if linguistic theory says that "Ede" and "is identical to Ede" are constituents of the sentence "Ede is identical to Ede", then the content of that sentence is a function of the contents of the two constituents (and their mode of combination). This implies that "is identical to Ede" has to be give a semantical value of its own by every semantical analysis of English.

The standard explication of relations in sense
Let p and q be two propositions. p follows from q iff p is true under every circumstance under which q is; equivalently, iff q is a subset of p. p and q are compatible iff p and q may be true at the same time; equivalently, iff p and q have a non-empty intersection (p and q are incompatible iff they have an empty intersection).

The following semantical clauses are at least natural given the possible-worlds account.

Semantical essentials
For every circumstance w:

'not s' is true in w iff s is not;
's and t' is true in w iff both s and t are;
'possibly s' is true in w iff there is an (accessible) possible circumstance w', such that s is true in w';
'necessarily s' is true in w iff s is true under all (accessible) circumstances.
'actually s' is true in w iff s is true under the actual circumstance.

The following feature is independent of the possible-worlds account by itself, but it is accepted by most people that use that framework. Indeed most take Kripke in Kripke (1972) to have decisively refuted description theories of the meaning of names.

Direct referentiality for names
Names refer – full stop. They are not descriptions in disguise. More technically: the meaning of a name (within a compositional semantic theory) is simply its referent.[1]

3.1.1 . . . and ST

ST provides only an indirect interpretation of English, so the question how it compares to classical possible-worlds-semantics is not easy to answer. It will turn out however that, on closer inspection ST is not a classical possible-worlds semantics.

- ST does not treat names as directly referential,
- meanings of sentences cannot be reconstrued as being sets of possible worlds,
- nor as being sets of possible circumstances in any natural sense of the word.
- ST is not able to deliver separate meanings for constituents like "is identical to Ede" (or for expressions that express singular properties in general).

Furthermore, it is at least not clear how to amend T with a satisfactory treatment of "actually".

[1]For arguments in favour of this doctrine, see Kripke (1972).

In the following we will inspect these problems in some detail.

3.1.2 T+

T targets a variant of QML without individual constants. This may seem to make it unfit to represent natural language names. Lewis proposes to treat names as descriptions, but this departs from standard possible worlds semantics. We will, without further argument, assume that names refer directly, and are not descriptions in disguise.

Therefore let us consider a variant of QML that contains, additionally, individual constants. Now we may redefine clause (5.) of the definition of T, in the following way.

Let a *free term* be either a free variable or an individual constant.

5.' if ϕ contains n free terms $t_1 \ldots t_n$ (listed in their order of appearance), then

$$(\Diamond\phi)^w = \exists w \exists x_{i+1} \ldots \exists x_{i+n} \\ (Ww \wedge Cx_{i+1}t_1 \wedge Ix_{i+1}w \wedge \ldots \wedge Cx_{i+n}t_n \wedge Ix_{i+n}w \wedge \\ \phi[x_{i+1}/v_1 \ldots x_{i+n}/v_n]^w),$$

where either $i = 0$ and no variable occurs in in ϕ, or x_i is the greatest variabe that occurs in ϕ.

I will refer to the resulting system as T+. In the following I will mainly be concerned with T+. Its problems are problems for T, too; at least in the sense that they are problems that T is bound to have if its most obviously fault, the inability to treat names as directly referential, is removed.

3.2 The Semantic Values of Sentences

3.2.1 Sets of worlds?

Is T+ a version of possible worlds semantics? That depends on whether it is based on a notion of proposition that is in accordance with the the possible-worlds account of content, where propositions are sets of circumstances. Now it is not easy to tell whether this is the case, because the semantics is given in form of an indirect interpretation. And even if the semantical values defined were no sets of circumstances, the semantics could still be equivalent to a semantics where they are.

So maybe, the technical complexity of the translation mechanism hides a simple truth: the semantical value of a sentence is simply the sets of circumstances under which the sentence is true. Often, circumstances are equated with possible worlds. Therefore, in this subsection I will ask whether the semantical values of sentences provided by T could be construed as being sets of possible worlds after all. $| \phi |^{g}$ is intended to denote the set of worlds according to which ϕ is true wrt. g, i.e. the set of worlds which may be assigned to w such that ϕ^w is true. In the following I will omit g, if it is redundant.

It will turn out that neither conjunction nor negation are operations on sets of worlds. This will be shown in detail below. In the next section, we will finally ask what the meanings induced by T+ really are.

$\wedge$ and $\neg$ do not receive their classical meanings

There are predicates P and Q, names a and suitable counterpart relations, such that:

(i) $|\neg Pa|$ is not the complement of $|Pa|$.

Firstly, there are worlds where there is no counterpart of a, and hence neither Pa nor $\neg Pa$ are true according to them. Secondly, there are worlds in which there is a counterpart of a which is P, and another one which is not P. I argued in Sec.1.3.5 on p.13 that there are cases like this in logical space. Both Pa and $\neg Pa$ are true according to such worlds.

(ii) $|Pa \wedge Qa|$ is not the intersection of $|Pa|$ and $|Qa|$.

Taje a world w in which there are just two counterparts of a: one which is P and not Q, and another which is Q and not P. According to w, both Pa and Qa are true, but not $Pa \wedge Qa$, since there is no counterpart of a in w which is both P and Q. In order to show (ii), I need only show that there are indeed worlds like w. Again reference to the example in Sec.1.3.5 already suffices, here.

The meaning of $\wedge$ is no operation on sets of worlds at all

That the meanings of $\wedge$ is no operation on sets of worlds may be shown by providing instances of (S) below.

(S) $|\phi| = |\phi'|$ and $|\psi| = |\psi'|$, but $|\phi \wedge \psi| \neq |\phi' \wedge \psi'|$.

Here are two such instances.

- Suppose that a and b refer to the same possible individual A. Hence, $|Pa| = |Pb|$ (provided the counterpart-relation stays the same in both contexts). Let us assume that there is a world w, containing two counterparts of A, say B and C, such that B has the property expressed by P, but lacks the property expressed by Q, whereas the converse is true for C. Since $|Pa| = |Pb|$, any two-place function on sets of worlds will deliver the same value when applied to $|Pa|$ and $|Qa|$, or to $|Pb|$ and $|Qa|$. But there is a world in $|Pb \wedge Qa|$ which is not in $|Pa \wedge Qa|$, namely w. Hence, $|Pa \wedge Qa| \neq |Pb \wedge Qa|$.

- Let P and Q be incompatible properties, such that there are counterparts B and C of A, such that B has P and C has Q. Say, P is the property of being bald and Q the property of having a natural curl, so P and Q are expressible by predicates of English. Then define P' to be

  > the property of being P and inhabiting a world, such that there are counterparts of A in that world which are Q.

  And let Q' be

  > the property of being Q and inhabiting a world, such that there are counterparts of A in that world which are P.

  This proves that English has the means to express P' and Q'. Let P' and Q' be predicates which express P' and Q', resp. Now $|P'a| = |Q'a|$, but $|P'a \wedge Q'a| \neq |P'a \wedge P'a|$. Hence, we have another instance of S.

Hence, there can be no operation on sets of worlds which corresponds to the meaning of $\wedge$. (This does not change, if we revise counterpart semantics as suggested in the last chapter.)

This shows that the semantical values of the connectives provided by T+ (and T) are not operations on sets of worlds. Now, if the semantics of T+ (T) is compositional, it follows that the semantical values of sentences are not sets of worlds in that sense, either.

3.2.2 Induced Interpretations

So what are the semantical values provided by T? T is a mere translation of $\mathcal{L}_{QML}$ into another language. But, given any model of counterpart theory, T indirectly provides an interpretation of the language of quantified modal logic $\mathcal{L}_{QML}$, the *induced interpretation* (see Montague (1974)). In the following I shall examine the induced interpretation.

Let us first discuss T. Above, we have seen what the semantics induced by T isn't. It remains to say what it really is. You can achieve the answer rather mechanically. Simply pair every construction of $\mathcal{L}_{QML}$ with the truth-conditions of its T-translation. Part of the semantics for an arbitrary expression ϕ will be a set of assignments $[\phi]$, such that g in $[\phi]$ iff ϕ is true wrt. g. Second, in the definition of $(\Diamond\phi)^w$, T makes reference to the free variables of and to the variables used in ϕ. But then, instead of a single semantic value, we need three that are defined in parallel. We will later show that this complexity is not essential.

We will first consider the version of $\mathcal{L}_{QML}$ *without* individual constants. For the following, we will consider an arbitrary model of counterpart theory (D, V) (i.e. D is a set of individuals and V an interpretation function, such that the postulates P1–P6 hold). For the sake of simplicity, we will assume that the values of the variable w are invariably worlds ($g(w) \in V(W)$ for every g). Let G be the set of all those assignments.

First of all, we will present the induced interpretation, since you can compute it mechanically from T. Then, I will present the above-mentioned simplification.

3.2.3 The Semantics of T

So here are the definitions of $[\phi]^{<D,V>}$ (the superscript $^{<D,V>}$ will not be required for the moment), $Fr(\phi)$ (the set of free variables of ϕ), and $\mathrm{Var}(\phi)$ (the set of variables used in ϕ). Since all three ingredients are used as input for the semantics of $\Diamond\phi$, they all have to be regarded as to be parts of the semantic value of ϕ:

(i) $g \in [Rv_1 \ldots v_n]$ iff $Vg(v_1), \ldots, Vg(v_n) \in V(R)$ and w is a world

$\mathrm{Var}(Rv_1 \ldots v_n) = \{v_1, \ldots, v_n\}$
$\mathrm{Fr}(Rv_1 \ldots v_n) = \{v_1, \ldots, v_n\}$

(ii) $[\neg\phi] = D - [\phi]$

$\mathrm{Var}(\neg\phi) = \mathrm{Var}(\phi)$
$\mathrm{Fr}(\neg\phi) = \mathrm{Fr}(\phi)$

(iii) $[\phi \wedge \psi] = [\phi] \cup [\psi]$

$\mathrm{Var}(\phi \wedge \psi) = \mathrm{Var}(\phi) \cup \mathrm{Var}(\psi)$
$\mathrm{Fr}(\phi \wedge \psi) = \mathrm{Fr}(\phi) \cup \mathrm{Fr}(\psi)$

(iv) $g \in [\exists v\phi]$ iff there is an a and a g', such that

$(a, w) \in V(I)$ ("a is in w"),
$g' = g[a/x]$ and
$g' \in [\phi]$.

$\mathrm{Var}(\exists v\phi) = \mathrm{Var}(\phi) \cup \{v\}$
$\mathrm{Fr}(\exists v\phi) = \mathrm{Fr}(\phi)$ - $\{v\}$

(v) Suppose ϕ contains n free variables $v_1 \ldots v_n$ (listed in their order of occurrence). If $n \neq 0$, let i be the greatest number such that $x_i \in \mathrm{Var}(\phi)$.

$g \in [\Diamond\phi]$ iff there is a g', such that

(a) for every j $[1 \leq j \leq n]$,
$g'(x_{i+j})$ is a counterpart of $g(v_j)$ and $g'(x_{i+j})$ is in $g'(w)$,

(b) $g' \in [\phi[x_{i+1}/v_1] \ldots [x_{i+n}/v_n]]$.

$\mathrm{Var}(\Diamond\phi) = \mathrm{Var}(\phi) \cup \{x_i, \ldots, x_{i+n-1}\}$
$\mathrm{Fr}(\Diamond\phi) = \mathrm{Fr}(\phi)$

Note that, although $[\Diamond\phi]$ is defined with reference to $[\phi[x_{i+1}/v_1]\dots[x_{i+n}/v_n]]$ and not to $[\phi]$, still the meaning of $\Diamond\phi$ is a function of the meaning of ϕ. This is so because the syntactical substitution of variables in ϕ corresponds to a purely semantical operation on $[\phi]$. Therefore, the first part of (v) is equivalent to the following considerable simplified version:

$g \in [\Diamond\phi]$ iff there is a g', such that

(a) for every $v \neq w$:
$g'(v)$ is a counterpart of $g(v)$ and $g'(v)$ is in $g'(w)$,

(b) $g' \in [\phi]$.

Observe that (v) was the only semantical rule in which Fr(ϕ) and Var(ϕ) have figured at all, so we can get rid of them. Now here is the simplified semantics:[2]

(i') $g \in [Rv_1 \dots v_n]$ iff $g(v_1), \dots, g(v_n) \in V(R)$

(ii') $[\neg\phi] = G^* - [\phi]$

(iii') $[\phi \wedge \psi] = [\phi] \cap [\psi]$

(iv') $g \in [\exists v\phi]$ iff there is an a and a g', such that

$(a, w) \in V(I)$ ("a is in w"), $g' = g[a/x]$ and $g' \in [\phi]$

(v') $g \in [\Diamond\phi]$ iff there are w' and g', such that

(a) for every $v \neq w$:
$g'(v)$ is a counterpart of $g(v)$ and $g'(v)$ is in $g'(w)$,

(b) $g \in [\phi]$.

With these simplifications the semantics does not look very much different from the usual semantics of predicate logic. There the semantical values also turn out to be sets of assignments on closer inspection.

[2] A similar semantics may be found in Bauer and Wansing (2002), the authors refer to Ghilardi (1991) in turn.

3.2.4 The Semantics of T+

What happens if we add the directly referential treatment of names? Now, not only the interpretation of variables gets shifted by $\Diamond$, but also the reference of individual constants. Accordingly, the interpretation of individual constants will be part of the resulting semantical values. Instead of evaluating at assignments, we now have to evaluate at pairs (g, i), where i is an interpretation of the individual constants. Now G^* will be the set of such pairs. ig will denote the union of i and g. Here is the resulting semantics:

(i') $(g, i) \in [Rt_1 \ldots t_n]$ iff $ig(t_1), \ldots, ig(t_n) \in \mathrm{V}(R)$ and w is a world

(ii') $\neg\phi] = \mathrm{D}$ - $[\phi]$

(iii') $[\phi \wedge \psi] = [\phi] \cap [\psi]$

(iv') $(g, i) \in [\exists\phi v]$ iff there is an a and a g', such that

$(a, w) \in \mathrm{V}(I)$,
$g' = g[a/x]$ and
$(g', i) \in [\phi]$

(v') $(g, i) \in [\Diamond\phi]$ iff there are w' and g' and i', such that

(a).1 for every variable $v \neq w$:
$g'(v)$ is a counterpart of $g(v)$ and $g'(v)$ is in $g'(w)$,

(a).2 i' is an interpretation which is exactly like i, only that for every constant c:
$i'(c)$ is a counterpart of $i(c)$and $i'(c)$ is in $g'(w)$,

(b) $(g', i') \in [\phi]$.

3.2.5 The Semantics of Counterpart Frames

What happens if we apply the revsion of T I argued for in the last chapter? For that revision of T , see p.42; a formal semantics based on the same idea may be found in Appendix A. This semantics is again based on sets of assignments. If we add names, again we end up with semantical values (g, i) as for T+. Closer inspection reveals that, within this semantics, we also have to keep track of the sequence of free variables in a formula. So, a sequence Frϕ now turns out to be a non-redundant ingredient of the semantics of ϕ. This presence of further syntactical details in the semantical values only adds to the difficulties that pertain to the counterpart semantic account of meaning, we will discuss in the following subsection.

3.2.6 Critique

In the following, I will compare this semantics with standard possible worlds semantics. In T+, the meaning of a sentence is a set of pairs (g, i).

The semantics for

(1) Fritz is rich

e.g., may now be put as follows:

(2) (1) is true in (g, i) iff i(Fritz) is rich.

I will argue that such pairs are bad candidates for being circumstances in the sense of possible-worlds semantics.

There are two immediate worries this formulation might raise. Doesn't (2) say that (1) is true iff the guy called "Fritz" is rich? This is odd, because then the name would be synonymous with a certain meta-linguistic description. But it has to be kept in mind that "the referent of the name "Fritz"" is a technical term of the metalanguage. E.g., the referent of the name "Fritz" in (g, i) need not be called "Fritz". So "Fritz" does not mean "the guy called "Fritz"", and the semantics is not metalinguistic in the ordinary sense.

You might also object that this semantics of names is no longer rigid designation because the name "Fritz" does not denote the same individual in every circumstance of evaluation. But this objection either, on one hand, applies to every version of a semantics based on disjoint domains, and just says that our semantics of names is wrong because it is based on a metaphysics you don't like. Then it is pointless to raise that objection here. Or you allow for a sensible redefinition of rigid designation within the present metaphysics. Then the semantics is likely to fall under that definition.[3] Simply define a

[3] see Lewis[86a], p. 256

term t to be rigidly designating iff for any (g, i) the denotation of t in (g, i) coincides with $ig(t)$, the referent of t in (g, i).

The next two objections I will present seem much more troublesome than the objections raised so far. They concern the adequacy of the resulting notion of proposition. The problem is not that we *change* the reference of a term, but that meanings are defined in terms of the reference of *terms*. It is therefore, that meanings become much more dependent on how we express things than is healthy.

We can't substitute co-referential terms salva interpretatione

E.g. $[Pa]$ and $[Pb]$ are different, even if $V(a) = V(b)$ for the interpretation V of the model. This is so because Pa is true wrt. triples (g, i), such that $i(a)$ has the property expressed by P, and these need not be such that $i(b)$ has the same property.

This is odd for two reasons: First, it suggests that, contrary to our characterization of direct referentialism, there is *more* to the meaning of a name than its referent. Second, while substitution failures of co-referential terms abound in some contexts (e.g. in propositional attitude ascriptions), they are out of the question in others. Take, for instance, the context "... is true under the same circumstances as ...".

(3) "Cicero sleeps" is true under precisely the same circumstances as "Tully sleeps"

(3) is intuitively true; given T+'s notion of circumstances as pairs of the form (g, i) it comes out false. This suggests that T+'s notion of circumstance is not a natural one.

Sameness of semantical values and intertranslatability

We will now ask, whether the semantical values provided by T+ are good approximations of the notion of meaning (or content). We will assume that sentences are intertranslatable if, and only if they have the same meaning. Hence, if semantical values are good approximations of meaning, the following criterion should hold.

(4) Sentences are intertranslatable if, and only if they have the same semantical values

Unfortunately, (4) can be shown to fail.

Sameness of semantical value is not sufficient

Let $\mathcal{L}$ and $\mathcal{L}'$ be two different interpreted languages of quantified modal logic which use the same vocabulary. It holds that, if $V(P) = V'(P)$, then $[Pa] = [Pa]'$, even if $V(a) \neq V'(a)$ because, as we have characterized circumstances of evaluation (g, i) so far, $i(a)$ is not constrained in any way, particularly it is not tied to the interpretation of a provided by V. Suppose $V(a) \neq V'(a)$. Then, "Pa" in $\mathcal{L}$ is not a correct translation of "Pa" $\mathcal{L}'$, although $[Pa] = [Pa]'$.

We could try to do something about this result by showing that the relevant feature of the above semantics of T+ is inessential. Let's try to limit the class of circumstances of evaluation to those which are related to some actual circumstance (g, V) via some chain of counterpart-related circumstances. Under not too unfavourable circumstances, it will then turn out that if $V(a) \neq V'(a)$, then $[Pa] \neq [Pa]'$. But, in order to show that sameness of meaning is not sufficient, it suffices to show that the combination $V(P) = V'(P)$, $V(a) \neq V'(a)$ and $[Pa] = [Pa]'$ may happen under certain special circumstances. And this is indeed the case. E.g. suppose $V(a)$ and $V'(a)$ are indiscernible individuals of the same world. Because the counterpart relation is qualitative, $V(a)$ and $V'(a)$ are counterparts of each other. (Although this is prohibited by postulate P5 of counterpart theory, above I have argued that P5 should be withdrawn; and Lewis seems to admit this in later writings.) Because of the same reason, the two sets of circumstances related to (g, V) and (g, V') via some chain of accessible circumstances will also be identical. But Pa in $\mathcal{L}'$ is not a correct translation of Pa in $\mathcal{L}$, because, in the two languages a refers to different although indiscernible individuals.

Sameness of semantical value is not necessary

Let $\mathcal{L}'$ be just like $\mathcal{L}$ in syntax and interpretation, only that one single name c has been added. Then $[Pa] \neq [Pa]'$. It even holds that $[Pa] \cap [Pa]' = \emptyset$. No (g, i) in $[Pa]$ can be also in $[Pa]'$! This is so because c is not in the domain of i. For the same reason, *no* expression of $\mathcal{L}$ can have the same semantical value as any expression of $\mathcal{L}'$. But, according to intuition, in such a scenario Pa would still be the perfect translation of itself, indeed every expression of $\mathcal{L}$ can be translated identically.

In this section, I have investigated the interpretation for $\mathcal{L}_{QML}$ induced by T+. I have shown that the interpretation of $\Diamond\phi$, its *semantical value*, makes essential reference to syntactical notions about ϕ and about the language in general. But this makes these semantical values unfit for approximating meanings. It also makes their elements unfit for being circumstances of evaluation. Such a circumstance should be independent of the particular expression we want to evaluate in it. It should be independent on language in general.

3.3 Impure Predicates

What are the denotations of predicates in our counterpart semantics? Sets of possible individuals. Given the background metaphysics of disjoint universes, this works fine for pure, qualitative predicates. But if the predicates are impure, like "being Ede", or "being father of Tom", things are different. Let P denote the property of being R-related to B, the referent of b.

Is there, given T+, a set of possibilia that could be taken to be the denotation of P? Or is there one, given the revised version of counterpart semantics argued for in Chap.2?

A proposed denotation is correct, only if it is able to yield the following equivalences

(5) $\Diamond Pa \Leftrightarrow \Diamond Rab$

(6) $\Box Pa \Leftrightarrow \Box Rab$

In this section I will show there is no set of possibilia with these properties.

3.3.1 . . . and T+

According to T+, $\Diamond Rab$ is true iff a counterpart of A, the referent of a, is R-related to a counterpart of B, the referent of b, in the same world. Therefore, Pa should be satisfied by those individuals which are R-related to some worldmate which is a counterpart of B. But this means that $\Box Pa$ is already true if every counterpart of A is R-related to *some* counterpart of b in the same world, whereas (6) demands that $\Box Pa$ is true exactly if *every* counterpart of A is R-related to *every* counterpart of B in the same world.

If there are multiple counterparts of one and the same individual in the same world, these two conditions cannot be squared. Now we have already seen examples, like the case of the pair of twins which have two pairs of counterparts. It is consistent with the existence of such a world to suppose that every counterpart of Dee is a twin of some worldmate which is a counterpart of Dee, still in the symmetrical world there is a counterpart of Dee which is not the twin of every counterpart of Dum, so it is not the case every counterpart of Dee is a twin of *every* counterpart of Dum in the same world.

So, given T+, there do not seem to be denotations for Pa and impure predicates in general that deliver (5) and (6).

3.3.2 ... and Counterparts of Pairs

One might argue that the semantics of $\Box Rab$ is wrong anyway, precisely because it cannot accommodate essentialism, e.g. essentialism about twinhood. Above we have argued for a revision of T that is able to accommodate this kind of essentialism. Does it also fare better wrt. to the problem of providing a denotation for P? It turns out that (i) if (5) comes out correctly, then the underlying counterpart relation is not essentialist; and (ii) if (5) comes out correctly, then (6) doesn't .

Let R be identity, and P the property of being identical to B, the referent of b.

Since $\Diamond b = b$ is true, there is a counterpart C of B in the denotation of P. But then, $\Diamond Pa$ is true for every a which refers to an individual A, such that C is a counterpart of A. Now there are such individuals, only suppose that a refers a qualitative duplicate of B. In such a case $\Diamond Pa$ gets true, although $b \neq a$ is true, so if we want to save (5) we have to sacrifice the necessity of non-identity. In general, nontrivial relational essentialist claims have to go, contrary to the very motivation of the revision of T. So, (i) if (5) comes out correctly, then the underlying counterpart relation is not essentialist.

Since, according to the revised semantics proposed in Chap.2, $\Diamond Rab$ is true iff a counterpart of (A, B) is R-interrelated, Pa should be satisfied by exactly those individuals C which are R-related to to some D such that (C, D) is a counterpart of (A, B). But this means that $\Box Pa$ is predicted to be true exactly if for every counterpart of A *there is some* D such that (C, D) is R-interrelated and a counterpart of (A, B), whereas (6) implies that $\Box Pa$ is true exactly if every counterpart of (A, B) is R-interrelated.

Returning to our first example, the first clause says that Dee has the property of being necessarily the twin of Dum iff for every counterpart Dee' of Dee there is an individual Dum' such that the pair (Dee', Dum') is a pair of twins which is a counterpart of the pair (Dee, Dum); whereas the second say that Dee has that property iff every counterpart pair of the pair (Dee, Dum) is a pair of twins. The first, but not the second condition rules out that Dee might be an only child, so again, the two conditions do not agree.

Hence, (ii) if (5) comes out correctly, then (6) doesn't.

This means that counterpart semantics, in its various forms discussed here, is not able to provide semantical values for impure predicates like "is identical to Ede" or "is the father of Tom". But these expressions occur as constituents of sentences of the English language. Therefore they should receive semantical values by any compositional semantics.

3.4 "Actually" a Problem

It is possible that some people, who are actually rich, are poor. We will represent "actually" by some operator **A** in QML. Hazen (1977) asks how to represent **A**, and thereby "actually" in T. He ponders some proposals and concludes that there is no way to achieve a correct semantics. Recently Fara and Williamson have tried to strengthen his argument. They show that there is no way to interpret **A** as a (sequence of) counterpart quantifiers(s). Before we can address their objections, let me first address a possible misunderstanding.

In many cases we can understand "actually" as expressing something like "in the actual world". E.g. when someone say "there are actually honest politicians" he or she probably wants to say that there are honest politicians in the actual world. Still we should not conflate "actually" and "in the actual world". At least within our present modal realist metaphysics, according to which there are individuals in other worlds, the two locutions might depart. The following statement is true for arbitrary sentences ϕ, if **A** is taken to translate "actually", but not when it is taken to translate "in the actual world".

(7) $\mathbf{A}\phi \leftrightarrow \phi$

If something is true, it is also actually true and vice versa. At least this is taken as a basic truth about "actually" in most accounts, see e.g. Hodes (1984). "In the actual world", on the other hand, works as a restrictive quantificational modifier just like "in this room"[4] and if we presuppose a metaphysics of many worlds, this restriction is very considerable. E.g. according to the modal realist metaphyiscs, there are talking donkeys because

[4] see Lewis (1986a), p.5

there are talking donkeys in other worlds. And yet, there are no talking donkeys in the actual world. And this is why "actually" and "in the actual world" are not always interchangeable.

Here are some of the examples that have been taken to show that "actually" cannot be translated into counterpart theory. In the following, **E** stands for "exists".

(8) $Fx \wedge (\mathbf{A}\neg Fx \vee \neg\mathbf{A}Fx)$,

(9) $\Diamond\exists x((\mathbf{A}x = y \equiv \mathbf{A}x = z) \wedge \mathbf{A}(x = y \vee x = z) \wedge \neg\mathbf{A}(x = y \wedge x = z)$,

(10) $\Diamond\exists x(\mathbf{AE}x \wedge \neg\mathbf{A}Px \wedge \mathbf{A}Px))$,

(11) $\Diamond\exists x(\mathbf{AE}x \wedge \mathbf{A}\neg Px \wedge \mathbf{A}Px)$.

Fara and Williamson build their critique of counterpart-theoretic semantics on the fact that the first two are bound to turn out satisfiable once we enhance such a semantics with an actuality-operator. Hazen notes that the most straightforward extensions of T with such an operator make either the third or the last example true. All these examples are intuitively false, however.

Here, I discuss the first example. Fara and Williamson suppose that since "actually" is very much similar to "possibly" or "necessarily", just like these it will be translated into a sequence of counterpart quantifiers, such that e.g. $\mathbf{A}Fx$ is of the form $[Qy : Iy@ \wedge Cyx]Fy$ with Q being a (logical) determiner.[5] Logical determiners are relations between two sets. All natural language determiners are logical. If Q is a logical determiner and G is empty, then $[Qx : Gx]Fx$ is either false, or $[Qx : Gx]\neg Fx$ is true, regardless of F.

If **A** is treated like this, (8) is translated into

(12) $Fx \wedge ([Qy : Iy@ \wedge Cyx]\neg Fy \vee \neg[Qy : Iy@ \wedge Cyx]Fy)$.

Now suppose a is an F which has no counterparts in the actual world. Because it is an F it satisfies Fx. Because it has no counterparts in the actual world, $Iy@ \wedge Cyx$ is empty and a satisfies either

$$[Qy : Iy@ \wedge Cyx]\neg Fy$$

or

$$\neg[Qy : Iy@ \wedge Cyx]Fy,$$

[5]For the notion of a logical determiner, see Barwise and Cooper (1981).

so a satisfies

$$[Qy : Iy@ \wedge Cyx]\neg Fy \vee \neg[Qy : Iy@ \wedge Cyx]Fy.$$

Hence, according to our translation a satisfies (8) for any logical quantifier Q. But (8) is contradictory! Because of

(7) $\mathbf{A}\phi \leftrightarrow \phi$

is true for any ϕ, a satisfies $Fx \wedge (\mathbf{A}\neg Fx \vee \neg\mathbf{A}Fx)$ iff a satisfies $Fx \wedge \neg Fx$. But the latter formula *is* contradictory.

We have derived a contradiction from the two assumptions that **A** is a counterpart quantifier and that there is an F which has no counterparts in the actual world. It would be strange to prohibit the latter, so we should drop the first assumption.

Here is another argument to the conclusion that the analysis of **A** as a counterpart quantifier can't be correct. We will use the fact that logical determiners Q are *conservative*; i.e. for arbitrary variables v:

$$[Qv : \phi[v]]\psi[v] \Leftrightarrow [Qv : \phi[v]](\phi[v] \wedge \psi[v])$$

(Here $\phi[v]$ and $\psi[v]$ are arbitrary formulae which have v as a free variable.) Assume that **A** is a counterpart quantifier. Again we will presuppose (7). Let $\phi[x]$ be any formula that contains x as its only free variable and does not contain any names. We may now reason as follows.

	$\phi[x]$	
$\Leftrightarrow$	$\mathbf{A}\phi[x]$	because of (7)
$\Leftrightarrow$	$[Qy : Iy@ \wedge Cyx]\ \phi[y]$	because **A** is a counterpart quantifier
$\Leftrightarrow$	$[Qy : Iy@ \wedge Cyx]\ (Iy@ \wedge Cyx \wedge \phi[y])$	because Q is conservative
$\Leftrightarrow$	$[Qy : Iy@ \wedge Cyx]\ (Iy@ \wedge \phi[y])$	because Q is conservative
$\Leftrightarrow$	$\mathbf{A}(\phi[x] \wedge Ix@)$	because **A** is a counterpart quantifier
$\Leftrightarrow$	$\phi[x] \wedge Ix@$	because of (7)

Therefore, an individual satisfies a formula $\phi[x]$ exactly if it satisfies $\phi[x] \wedge Ix@$. Because every individual satisfies some formula, this implies that every individual is in the actual world. Given our metaphysics of non-overlapping worlds this implies that there is only one world! Hence, the analysis of **A** as a counterpart quantifier is correct only if there is only one world.

The moral from such arguments is clear: we should not treat **A** as a counterpart quantifier. Does that mean that no counterpart semantics for "actually" is feasible? This does not follow. We do not have to treat "actually" as a quantifier. In propositional modal logic, e.g. "actually" is not treated as a quantifier, likewise. There, it is used to refer to the actual circumstances. In Chap.5 we will propose an analysis that may be interpreted to do exactly the same for quantified modal logic.

Chapter 4

Counterpart Kripkeanism

Even if counterpart semantics is not able to provide a standard possible worlds semantics; perhaps the basic ideas of counterpart semantic could be combined with standard Kripkean semantics. In this chapter, we study the idea to define the notion of truth according to a world in terms of couterpart-semantics and to use the result within a standard Kripkean semantics.

4.1 The Basic Idea

Recall that, in contradistinction to counterpart semantics, see (4) below, Kripkean semantics, see (3) below, analyses modality not in terms of quantification over counterparts but in terms of quantification over worlds. I repeat for convenience the examples from Chap.1.

(1) Possibly, Humphrey wins

(2) Humphrey is necessarily human

(3) a. (1) is true iff "Humphrey wins" is true at some world accessible from ours.
    b. (2) is true iff "Humphrey is human" at every world accessible from ours.

(4) a. (1) is true iff some counterpart of Humphrey wins.
    b. (2) is true iff every counterpart of Humphrey is human.

These truth-conditions are not necessarily in conflict. E.g. if "Humphrey wins" is defined to be true at world w exactly if there is a counterpart of

Humphrey in w which wins (and accessibility is universal), then (**??**) and (**??**) are equivalent.

Now some people employ counterparts in a way that is best characterised as a specimen of Kripkean semantics. First they define, in terms of counterparts, the notion of truth according to a world, and then they define modality as in (3). These people will be called Counterpart Kripkeans.[1]

This shows there are two different ways how counterparts may enter recursive semantics. According to both ways, counterparts are introduced by a counterpart quantifier; for the purpose of illustration we will at first assume by an existential quantifier. (We will shortly see other possible choices, too.) But still, counterparts could be introduced *early* or *late*. On the early option, counterparts are introduced by the semantics of atomic predications; such an account might say, e.g., that Humphrey satisfies "x wins" at exactly those worlds where some counterpart of Humphrey wins. Then, possibility is analysed as truth at some world (or at some accessible possible world). Hence, Humphrey is predicted to satisfy "possibly, x wins" precisely if there is a possible world and a winning counterpart of Humphrey in that world, i.e. if, and only if some counterpart of Humphrey wins.

On the late option, counterparts are introduced by modal operators. Atomic predications do not mention them, e.g. Humphrey satisfies "x wins" at any world exactly if Humphrey wins. Departing from what I have called "standard possible worlds semantics", modal operators are no longer simply quantifiers over worlds. They are interpreted as doubly quantifying over worlds *and counterparts*. E.g. suppose that x is the only free variable in $\phi[x]$ (and there are no names in $\phi[x]$). A late introduction account might say, e.g. that A satisfies "possibly $\phi[x]$" exactly if there is a world w and a counterpart of A in w such that the counterpart satisfies $\phi[x]$ in w. If we combine these two features it turns out that Humphrey is predicted to satisfy "possibly, x wins" precisely if there is a possible world and a counterpart of Humphrey in that world which satisfies "x wins" i.e. exactly if some counterpart of Humphrey wins.

The first approach is the approach of Counterpart Kripkeans; the second is counterpart semantics. I noted above that the two options arrive at precisely the same truth-conditions for our sample sentence. Are they equivalent throughout? In Lewis (1986a), Lewis sounds as if he thinks they are:

[1] The arch Counterpart Kripkean is Ramachandran (1989). Kratzer (1989) contains a Counterpart Kripkean treatment of situation semantics.

"When they attach to sentences we can take the diamond and the box as quantifiers ... How to take them when they attach to open formulas ... is more questionable. A simple account would be that in that case also they are quantifiers over worlds." (p.9) The simple account then turns out to be what I have called early introduction, here, whereas the "more complicated account" turns out to be late introduction. Lewis then says: "The simple and complex accounts are not in competition" (p.10).

But the two accounts *are* in competition, as you might see if you turn to other sentences. Take e.g. two atomic predications under one and the same modal operator, as in "possibly, Px and Qx". Again we assume the semantics of such sentences is to be analysed in terms of existential quantification over counterparts. According to early introduction the open sentence is satisfied by A iff there is a world w, such that there is a counterpart of A in w that has the property P and a counterpart of A in w that has the property Q. According to late introduction the open sentence is satisfied by A iff there is a world and some counterpart of A in that world has *both* the property P and the property Q. Differences between these two conditions could arise because, in some world, there might be multiple counterparts of a that differ in some of their properties.—Another case, where the two approaches differ is wrt. to negation. According to early introduction A satisfies "possibly, x is not P" iff there is a world in which there is *no counterpart* of A with the property P, according to late introduction it satisfies the sentence iff there is a world with *some counterpart* of A in it that fails to have P.

So which option does Lewis take? The somewhat disconcerting answer is: sometimes one, sometimes the other. Although T above is an example of counterpart semantics, there are places in Lewis work, where he is a Counterpart Kripkean, most clearly when he advertises the use of counterparts for natural language semantics in Lewis (1986a) p. 44f. Here, Lewis interprets "possibly" as a quantifier over possible worlds alone, and defines the semantic values of atomic sentences in a way that may be illustrated by the following example.

> 'Am' is a basic sentence; its semantic value for any speaker is the function that assigns *truth* to all and only those worlds that contain counterparts of that speaker. Lewis (1986a) p. 45

What Lewis calls "semantic value" here also depends on the speaker of "am". But still it is obvious that the sentence gets typical early introduction truthconditions; together with the appropriate semantics for “not”, “possibly, not am” might come out true, something that couldn’t happen under a late introduction semantics.

Which way to use counterparts is the right one? Is it counterpart semantics or the semantics of the Counterpart Kripkeans? Of course this may depend on details of these accounts. To say that atomic predications are interpreted in terms of a counterpart quantifier does not say much; the question remains which quantifier it is. In the next section we will inspect a couple of choices.

4.2 Predication Rules

Counterpart Kripkean semantics is based on a definition of truth according to a world in terms of a counterpart quantifier. We will consider various candidate counterpart quantifiers. We will conclude that all of them are objectionable. One criterion of adequacy will be whether the resulting semantics is a possible-worlds semantics in the sense of the last chapter. For this aim, it does not suffice that the semantics is in terms of possible worlds; the notion “truth according to a world” defined should also play the role of “truth in a circumstance”.

We can state the truth-condition for atomic sentences of the form Pa, just considered, as

> (T1) Pa is true according to w iff a counterpart of the referent of a in w has the property expressed by P.

(T1) appears to be the truth-condition employed in the above quote from Lewis. Unfortunately, there is something deeply wrong with (T1). No one can be rich and poor at the same time, at least not according to the same standards. But, according to (T1) “Fritz is poor” and “Fritz is rich”, may be true in the same circumstances. This is so because there are worlds in which both sentences are true, due to two different counterparts of Fritz in those worlds, one of which is rich, the other poor. We presuppose here that worlds play the role of circumstances.

I take the outcome to be intuitively unacceptable and a refutation of (T1). “Fritz is poor” and “Fritz is rich” could not be true under the same circumstances. The propositions expressed by the two sentences should be incompatible; but (T1) predicts that they're compatible (this presupposes the possible-worlds explanation of compatibility, see Sec.3.1.).

Furthermore, to say that both "Fritz is poor" and "Fritz is rich" are true according to some possible world means that

(5) Possibly, Fritz is poor and Fritz is rich

is true. Put more generally: according to (T1), $\Diamond(Pa\& \ Qa)$ has true instances, even if the properties expressed by P and Q are incompatible.

So something is wrong with (T1). We should look for an alternative. Take first the universally quantified (T2):

> (T2) Pa is true according to w iff *every counterpart* of A in w has the property expressed by P

Unfortunately, (T2) is not really better than (T1). There are worlds in which Fritz does not exist, and according to (T2), both "Fritz is poor" and "Fritz is rich" are true in these worlds. Hence, there are circumstances under which the two sentences are (simultaneously) true; the two propositions expressed are therefore predicted to be compatible; and (5) is predicted to be true again. These consequences are clearly not acceptable.

The following condition (T3) avoids the problem of vacuous universal quantification:

> (T3) Pa is true according to w iff there is a counterpart of A in w, and every counterpart of A in w has the property expressed by P.

In Ramachandran (1989, 1990), Murali Ramachandran proposes a revison of Lewis's scheme in mainly two respects. First Lewis's late introduction semantics is replaced by an early introduction one and second, the quantifier which introduces counterparts is exchanged. Precisely, he uses (T3) above, i.e. the existential quantifier is replaced by plural "the" under a Russelian analysis. E.g. I might be a prince under circumstances W exactly if there is a W-world where *the counterparts* of me are princes.

But (T3) as well does not seem right to me. For any two predicates, English can form a disjunction. Let $P \vee Q$ denote the property of having the property denoted by P or the property denoted by Q. That is, $P \vee Q$ denotes the union of the sets denoted by P and Q. Now, according to (T3), there may be circumstances in which $P \vee Qa$ is true, although neither Pa nor Qa are; hence

(6) $\Diamond(P{\vee}Qa \wedge \neg Pa \wedge \neg Qa)$

may be true. This looks very odd. How could Hans have the property of being nice or lucky without Hans being nice or Hans being lucky?

An attempt at a justification might appeal to vagueness, or indeterminacy: Hans is medium-sized or tall, because, on every resolution of the vagueness of these predicates, he is either medium sized or tall. But is he tall? Well, on many resolutions of the vagueness of "tall", he isn't, so we'd better not say he is. Is he medium-sized? Again, on many resolutions of the vagueness of "medium-sized", he isn't, so we'd better not say he is. So, vagueness really seems to provide cases where we attribute a disjunctive property without being willing to attribute one of the disjuncts.

But note that this defense only applies to the example I gave, it does not serve to defend (6), because (6) would still have true instances if P and Q were totally precise predicates. Similar things apply to the strategy to excuse (6) by pointing to cases of vague reference. Likewise it would not help to point to alleged cases of vagueness in the objects themselves, where the boundaries of an object are not fixed precisely, for (6) would still have true instances if the boundaries of the referent of a where as clear-cut as can be.

But still there is a kind of vagueness or indeterminacy that may be taken to account for (6). We are asking: who represents Fritz in world w? In a world in which there are two counterparts of Fritz, we have a choice. Now nothing seems to say which counterpart of Fritz to choose; and this could with some right be called indeterminacy of representation. Reconstructed this way, the semantics in (T3) looks very much similar to the supervaluationist approach to truth-values of sentences containing vague expressions: while "Fritz is tall" is said to be (super-) true iff it is true according to every resolution of the vagueness of "tall", we now say that w represents Fritz as being tall iff, according to every choice of a representant of Fritz in w, the representant of Fritz in w is tall.

I conclude that the choice of the counterpart quantifier used in (T3) possesses an internally coherent motivation. Where does this leave our debate? Let's suppose we found an internal justification for the existence of true instances of (6). We cannot persuade the friend of (T3) by appeal to any data to abandon his proposal, for he may now take his justification to reject the data: according to his justification Hans could indeed have the property of being nice or lucky without Hans being nice or Hans being lucky. But we need not accept his justification either, because it presupposes the controversial semantics. The showdown has to be postponed.

Let's turn to negation. Given (T3), Pa is not true according to w iff either there are no representatives of the referent of a in w, or the representatives of the referent of a in w do not agree as to whether a is P. So if the representatives of the referent of a in w do not agree as to whether a is P, then Pa is not true according to w. This is consistent with the above motivation for (T3): in a case where the representatives do not agree, w does not represent a to be P univocally. Compare w to a picture with two representations of the same person in it. Say one representative wears a blue shirt, the other a red one. It makes sense to say that it is not the case, that according to the picture (in its entirety), the person is wearing a blue shirt; why, by another part of the picture, the person is represented to wear a red shirt, too.

The aim of inspecting various atomic predication rules was to find an early introduction semantics in terms of counterparts that is a classical possible-worlds semantics. Therefore, in our discussion of (6) we have assumed that the semantics of negation is classical; that "it is not the case that Pa" is true according to w iff Pa is not true according to w. But should we really say that, in the case above, *according to w it is not the case that Pa*? Rather than representing that it is not the case that Pa, w does not seem to deliver a determinate answer to that question at all. We wouldn't to say that according to our picture, it is not the case that the person is wearing a blue shirt.

So, something seems to be wrong with the combination of (T3) with the classical semantics of negation. This suggests that while the motivation of the particular counterpart quantifier used in (T3) is internally coherent, this counterpart quantifier should rather be employed in a theory of late introduction if we want to retain that motivation.

For this reason, I think that (T3) is not acceptable as an atomic predication rule either.

I now come to the final proposal for such a predication rule:

> (T4) Pa is true according to w iff there is a unique counterpart B of the referent of A in w and B has the property denoted by P.

(T4) does not fall prey to any of the above counterexamples. (T4) seems to be well motivated: if you don't have *the* referent at your disposal, take *the* representative.

Unfortunately, even (T4) makes the wrong predictions, see Hazen (1977), p. 114:

> One could give some definition of a "protocounterpart" relation, and then say that x is a counterpart of y if and only if x is a protocounterpart of y and nothing else in x's world is a protocounterpart of y. Since there are senses of possibility for which "I might have had an identical twin" is not obviously false, however, such a definition could not capture the logic of our ordinary modal concepts.

This proposal is just (T4), put in a slightly different way. Although I think the argument is basically correct, I want to elaborate on it a bit. First let's modify the example slightly:

(7) Fritz might have had an identical twin

Hazen obviously thinks that the plausibility of (7) provides a reason to reject the uniqueness-requirement for counterparts because in a world in which Fritz has an identical twin, Fritz has at least two counterparts. Since, according to (T4) (7) is automatically false in a world where Fritz has more than one counterpart, (7) can never be true.

The argument presupposes that "identical" in (7) means "identical in every qualitative respect". Only then it is guaranteed that the twin identical to the counterpart of Fritz is himself a counterpart. Now, it is implausible that we ever intended to say such a thing by the use of (7). "Identical" in (7) means less than "identical in every qualitative respect". In fact, we often use "identical" in a weaker sense.

We can easily improve on (7). Let a doppelgaenger of A be defined to be an individual indiscernible from A, living in the same world. Now we could use

(8) Fritz might have had a doppelgaenger

instead of (7) for the purpose of our argument.

Similarly, consider

(9) Fritz is smart.

In every world w in which (9) is true, there is a unique counterpart of Fritz. Therefore, w can't be a world of eternal recurrence, because in such a world Fritz has multiple counterparts.[2] Therefore, (9) implies

(10) There is no eternal recurrence.

I can hardly think of a more far-fetched implication.

4.3 Haecceitist Differences Again

Above we have inspected atomic predication rules based on various counterpart-quantificatiers one by one. We have concluded that all of them fail. But you might point out that there are other possible quantifiers we have not inspected. Now I am going to develop a general argument to the extent that *all* early introduction atomic predication rules fail.

The argument is simple. It is that, supposing early introduction is correct, there are examples which demand the existence of worlds that differ wrt. to representation *de re* but not qualitatively, i.e. emphhaecceitist differences. But there can't be haecceitist differences. Therefore early introduction cannot be correct.

E.g. according to (T3) I might be a prince under circumstances W exactly if there is a W-world where *the counterparts* of me are princes and I might be a pauper under circumstances W if and only if there is a W-world where the counterparts of me are poor. In fact I might be a pauper under precisely the same circumstances under which I might be a prince. So, W should be a complete qualitative description of the world, such that I have simply changed places between these two possibilities. So, either the example comes out true under our semantics. This requires there to be two different worlds w and v that are qualitatively perfectly alike, and a counterpart relation that makes all my counterparts be princes in w and be paupers in v. But such a counterpart relation would be utterly mysterious. Or the example is mistakenly predicted to be false.[3]

---

[2]It could be claimed that there is an exception, here, namely the actual world. According to P5, Fritz has a unique counterpart in the actual world, even if it is a world of eternal recurrence. Perhaps I should have chosen an example that is actually false.

[3]The point has been made already in Hazen (1977) against what is in effect a semantics with singular "the".

Given early introduction, "possibly, p" is true iff there is a possible world according to which p is true. Different theories of early introduction only differ wrt. to the precise definition of truth according to a world. We know that such theories can only be correct, if worlds represent *coherently*, i.e. if there is no world w such that, according to w, p is true and according to w, p is false. Otherwise, "possibly, p and not p" would predicted to be true, which is inacceptable. (This last condition already excludes conditions like (T1) and (T2).)

No matter how worlds represent me coherently, the cases at hand they turn out to be cases of haecceitist differences. Take the first example. Intuitively, the following two sentences should both come out true.

(11) Possibly, I live in the 17th epoch (and I do not live in the 137th) of a world of one-way eternal recurrence

(12) Possibly, I live in the 137th epoch (and I do not live in the 17th) of a world that is qualitatively exactly like the first

On theories of early introduction (11) is true iff there is a world of one-way eternal recurrence such that "I live in the 17th epoch" is true according to w and "I do not live in the 137th epoch" is true according to w, equivalently: "I live in the 137th epoch" is not true according to w.

(12) is true iff there is a world v qualitatively identical to the first such that "I live in the 137th epoch" is true according to v and "I do not live in the 17th" epoch is true according to v, equivalently: "I live in the 17th epoch" is not true according to v.

So, if (11) and (12) are true there are two qualitatively identical worlds w and v which have different representational properties (different *de re* propositions are true according to them).[4]

> So it seems that there is a possible world that is qualitatively just like ours - the same infinite sequence of epochs, all exactly alike, and exactly like the epochs of our world - but that represents de re, concerning us, that we live in the 137th epoch rather than the seventeenth. Then the difference between this world and ours is a haecceitist difference. Lewis (1986a), p.227

But there can't be such differences. Counterpart Kripkeans define representation *de re* (truth according to) in terms of counterparthood, and counterparthood is a *qualitative* relation.

[4]In Lewis version of the example we actually live in the 17th epoch.

If I have a counterpart in the 17th epoch of w, I have so in every world that is qualitatively identical to w, and likewise if I have a counterpart in the 137th epoch of v, I have so in every world that is qualitatively identical to v. But this means that the distribution of my counterparts and their properties are the same in every pair of qualitatively identical worlds. If the question is: how do two such worlds represent me, and the answer is defined in terms of quantifying over counterparts, then two such worlds can never represent me differently, regardless of the details of the definition of how worlds may happen to represent me (equivalently, how things about me can be true according to different worlds). Therefore there can't be a world that represents me as living in the 17th epoch (and not in the 137th) and another qualitatively identical world, that represents me as living in the 137th (and not in the 17th) epoch.

Chapter 5

A Semantics for Possibilism

We appear to have reached a dilemma. On the one hand, we want to have a decent notion of content, self-standing semantical values for impure predicates, and a decent semantics of "actually"; we want to have a classical possible-worlds semantics, for short. Therefore we should base our semantics on a notion of "truth according to a circumstance".

On the other hand, we want to be able to have the advantages of counterpart semantics, too. E.g. we want to be able to combine different counterpart relations to be able to account for exceptional cases where something like contingent identity seems to be what is going on and we want to be able to account for seemingly haecceitist differences without having to assume haecceitism. Finally we do not want to be forced to make those metaphysical assumptions one has to make, if a standard Kripkean semantics is assumed. In particular, we do not want to be forced to assume that individuals like me and you are part of more possible worlds other than our own. On the contrary, throughout this book I have assumed that no possible individual is (wholly) part of more than one world. Therefore we should define our semantics in terms of counterparts.

But in the last chapter we saw that any definition of "truth according to a world" in terms of counterparts is bound to disappoint. Does this mean that we have to give up? Are counterparts unable to function properly after all?

This impression is misleading. There could be other, more fine-grained notions of circumstance of evaluation. Indeed there are; and in this chapter I will provide a semantics of modality in terms of the notion of "truth according to a representation". It turns out that this approach combines classical possible worlds semantics with an account of modality *de re* in terms of counterparts. Therefore, it is able to have the benefits of both counterpart semantics and Kripkean semantics.

My solution is inspired by Allen Hazen's notion of a representational stipulation, see Hazen (1977, 1979), Graeme Forbes' idea to deal with non-existence in terms of representation by absent individuals, see Forbes (1982), Lewis's own notion of a joint individual possibility in Lewis (1986a), and, more indirectly by Kit Fine's characterisation of the *de dicto* in terms of automorphisms Fine (1978). The system is then amended with an actually-operator, and a treatment of restricted quantification.

5.1 Representations

In a representation certain entities represent others as having certain properties. (In a representation *by representative*, that is. When I speak of representations in the following, it will always be with this implicit qualification. Generally a person may be represented even without having a representative. E.g. an emperor may be represented as being mourned over by way of an allegoric figure weeping at the emperor's tombstone.) To introduce some bit of terminology: a *representative* represents a *representee* as being a certain *way*. E.g. a part of a certain painting[1] represents Queen Victoria as reading on horseback. Here the Queen is the representee, the representative is the part of the painting, and part of the way is the property of reading on horseback. Of course representatives need not be parts of paintings. Anything may be taken to represent. And of course representees need not be persons. Sometimes they are as large as the entire world. E.g. a suitable representative may represent the world as being a world in which there are talking donkeys. So let us only say that representatives and representees are possible individuals, among them possible worlds.

The above is a specimen of a single representation. A *joint* representation consists in a simultaneous representation of several individuals. E.g. in our painting one part of the picture represents Mr. Brown, one Queen Victoria, and another her horse. A joint representation represents a number of things to be in a certain joint way, to display a certain pattern of properties and relations. Joint ways may be understood as certain complicated relations: e.g. if representation r represents $A_1, \ldots, A_n$ to be in way W, then W is an

[1] Sir Edward Landseer: *Queen Victoria at Osborne*, Royal Collection

n-place relation. E.g., in our painting Mr. Brown, the horse and the Queen are represented to be such that Mr. Brown holds the reins of the horse, with the queen sitting on the horse.[2]

I will call a representation *direct* if and only if representees are represented to be exactly how their representatives themselves are. So, in the case of a direct representation the way is the entire pattern of qualitative properties and relations among representatives. This pattern may be conceived of as a maximally specific complicated qualitative relation.

In a direct representation, representees are represented to play the *roles* of the representatives, but not to be identical to them. The identity of the representatives never enters the content of a representation.

Representations in this sense are not constrained to be representations by entities of the same world. As far as the notion of a joint direct representation goes, an individual could be represented by any such pair.

E.g. an otherworldly counterpart of George represents George as being red-haired exactly if the counterpart himself is red-haired. So, the counterpart represents George directly.

For every representation, we can talk of *truth according to that representation*. Say A represents B as being P. Then we can say that, according to that representation, B is P and, equivalently that, according to that representation, it is *true* that B is P. E.g. according to our painting, it is true that the Queen is reading on horseback. And, according to the abovementioned representation of the world, there are talking donkeys in the world, equivalently, according to that representation it is true that there are talking donkeys in the world.

Generally, according to representation r, $A, \ldots, O$ stand in relation R if, and only if, $A', \ldots, O'$ jointly represent $A, \ldots, O$ as standing in relation R; hence

(1) It is true according to r $A, \ldots, O$ stand in relation R iff according to representation r, $A, \ldots, O$ stand in relation R.

E.g. it is true according to a certain representation that A and B inhabit different worlds exactly if A and B are represented as being in different worlds by that representation.

[2]The idea to work with joint representations in the semantics of modality is due to Hazen Hazen (1977).

Truth according to a representation will indeed be the core semantical notion in what follows. We will define, recursively, what it is for a proposition to be true according to a representation. I.e. (1) will be taken to define the semantic value of atomic sentences, and truth according to a representation of a complex sentence will be defined in terms of the semantic values of its parts. E.g. whether $\exists x Px \wedge Qa$ is true according to r will depend on whether $\exists x Px$ and Qa are, and whether $\exists x Px$ is true according to a representation will depend on whether Px is (for some assignemnt of an object to x). But there are some obstacles to be overcome.

5.2 The Semantics of Representations

5.2.1 Connectives and Quantifiers

First we want the following to come out true:

(2) According to any representation r, *it's not the case* that A is P exactly if it is not true according to r that A is P, generally,

According to any representation r, *it's not the case* that $A_1, \ldots, A_n$ are R-related exactly if it is not true according to r that $A_1, \ldots, A_n$ are R-related.

But this does not seem to be correct: a representation may fail to represent a thing A. In such a case it does not seem to be correct to say that it is true according to the representation that it is not the case that A is P. It is rather, that such a representation does not say anything about A. Therefore, the reverse direction does not seem to hold for arbitrary representations. It only holds for *maximal* representations, representations that do not fail to represent any possible individual.

A joint direct representation is not necessarily finite. It may be a simultaneous representation of any number of things. Indeed, the limiting case of a joint representation is a simultaneous representation of all possible individuals. How can this be? What is left to represent in such a case? The set of possible indviduals can play the role of the set of representatives simultaneously; every possible individual is represented by some other possible individual, then. We do not require the set of representees and the set of representatives to be different. This, at least should be familiar from more mundane sorts of representations, in a role-play participants may simply trade roles, e.g.

Such representations are far beyond our powers to stipulate representational relations. How could they exist, then? Taken as mathematical objects representations are simply functions from some set of possible individuals into another such set. So a representation of all possible individuals is of precisely the same kind as lesser representations. It is another question in virtue of what such functions are taken as representation relations. They could be taken as representation relations by being part of fitting some simple description of a class of functions that we take to be representation relations. Summing up, I see no principled reason to dismiss representations of all possible individuals from being genuine representations.

We can now also say that

(3) according to r, it is true that p and q, iff, according to r, it is true that p and according to r it is true that q,

(4) according to r, it is true that p or q, iff, according to r, it is true that p or according to r, it is true q

Quantification poses a greater problem. Because the semantics of sentences and open formulas like Px is defined in terms of representations, quantifiers like $\forall x$ have to range over representees rather than representatives, e.g.

(5) $\forall x Px$ is true according to r iff for every assignment of a possible object to x, Px is true according to r, i.e. if every possible object is represented by r to be P.

(6) $\exists x Px$ is true according to r iff for some assignment of a possible object to x, Px is true according to r, i.e. if some possible object is represented by r to be P.

But a representation may fail to imply that the individuals represented are really all individuals. In such a case it will sometimes be intuitively wrong to say that a universal statement $\forall x Px$ is true according to a representation iff every individual represented is P. E.g. suppose that our picture, which contains several dogs and a a horse, does not do represent any particular horses or dogs. It only *represents* the Queen and Mr.Brown and represents them, among other things, as being human. Still it would be very wrong to say that according to the picture, everything is human. Therefore (5), while being a straightforward consequence of our treatment of atomic sentences does not seem to be correct.

Still, (5) and (6) might be correct for representations of a very special kind. They are correct once we restrict ourselves to direct representations where every possible individual is a representative. I will call a representation where no possible individual fails to be a representative *complete*. Again, complete representations exist because there are functions which are complete in the above sense and we can take a suitably defined class of such functions to be a class of representations.

In the following we will confine ourselves to representations which are permutations of the set of possible individuals.[3] Such representations are maximal and complete.

5.2.2 Modality

Our approach to modality is counterpart theoretic, i.e. we presuppose a metaphysical counterpart relation with the following property mentioned in the introduction

(12) B is a counterpart of A iff every way B is in is a way A could be

As we have argued in Chap.2, this relation should also connect pairs, quite generally sequences of arbitrary finite length (with the counterpart relation between individuals as the simplest case). So (12) should be generalised to

(7) $(B_1, \ldots, B_n)$ is a counterpart of $(A_1, \ldots, A_n)$ iff every relational way $B_1, \ldots, B_n$ are in (in that order) is is a relational way $A_1, \ldots, A_n$ could be.

Modality *de re* is then be dealt with in terms of this generalised counterpart-relation. E.g. Dee and Dum are necessarily twins if every counterpart of the pair (Dee, Dum) is a pair of twins.

But now, we have to put this analysis of modality *de re* in terms of permutations of the set of possibilia. In Appendix C I will argue, that as far as metaphysical possibility is concerned, this is is indeed possible.

I will argue that there is a set C of permutations of the set of possible individuals such that

[3] A permutation of a set is a 1-1 function from the set onto itself.

(8) If $(B_1, \ldots, B_n)$ is a counterpart of $(A_1, \ldots, A_n)$, then there is a $\pi \in C$, such that $(\pi A_1, \ldots, \pi A_n) = (B_1, \ldots, B_n)$ and

(9) if there is a $\pi \in C$, such that $(\pi A_1, \ldots, \pi A_n) = (B_1, \ldots, B_n)$, then $(B_1, \ldots, B_n)$ is a counterpart of $(A_1, \ldots, A_n)$

But this will suffice to enable me to put the semantics of modality in terms of permutations. Instead of saying that it is possible for A to be P iff there is a counterpart B of A which is P, we can now say that it is possible for A to be P iff there is a $\pi \in C$, such that πA is P.

And instead of saying that it is possible for $(A_1, \ldots A_n)$ to be R-related iff there is a counterpart $(B_1, \ldots B_n)$ of $(A_1, \ldots A_n)$ which is R-related, we can now say that it is possible for $A_1, \ldots A_n$ to be R-related iff there is a $\pi \in C$, such that $\pi A_1, \ldots \pi A_n$ are R-related.

E.g. it is possible that there are talking donkeys iff there is, in C, a representation according to which there are talking donkeys in the world, i.e. a representation $\pi \in C$ such that $\pi @$ is a world with talking donkeys; it is possible that A is a talking donkey, iff there is, in C, a representation according to which A is talking donkey, i.e. a $\pi \in C$ such that πA is a talking donkey; and it is possible that A and B are in different worlds iff there is a $\pi \in C$ such that πA and πB are in different worlds.

5.3 Permutation Semantics

5.3.1 Models

Now let us present our semantics based on permutations in slightly more formal detail. Id_D is the identical permutation of D; if π is any permutation, then π^{-1} is the inverse permutation $(\pi^{-1}x = y \Leftrightarrow \pi y = x)$; and if ρ and π are permutations of D, then $\rho \circ \pi$ is that permutation of D, such that $\rho \circ \pi x = y \Leftrightarrow$ there is a z, such that $\pi x = z$ and $\rho z = y$.

A *permutation-model* for quantified modal logic is a quadruple $(Posse, C, V)$, such that $Posse \neq \emptyset$, C is a set of permutations of $Posse$ (the set of *admissible representations*), and V is a function which meets the following conditions:

(a) $V(c) \in Posse$ for every constant c,

(b) For every permutation π of $Posse$: V_π is a function, such that for every n-place predicate P: $V_\pi(P) \subseteq Posse^n$,

$V_\pi(=) := \{(a,a)|a \in Posse\}$.

Consider some arbitrary but fixed permutation-model $(Posse, C, V)$. Assignments are defined as usual. Likewise, $g[a/v]$ is that assignment that is exactly like g, only that $g(v) = a$. If g is an assignment, Vg is defined to be a function such that $Vg(t) = V(t)$ exactly if t is a constant and $= g(t)$ exactly if t is a variable.

Let us now define $\pi \models \varphi[g]$ (formula φ is true wrt. an assignment g and at a permutation π).[4]

1. $\pi \models Pt_1, \ldots, t_n[g]$ exactly if $(Vg(t_1), \ldots, Vg(t_n)) \in V_\pi(P)$
2. $\pi \models \neg\varphi[g]$ exactly if $\pi \not\models \varphi[g]$
3. $\pi \models (\varphi \wedge \psi)[g]$ exactly if $\pi \models \varphi[g]$ and $\pi \models \psi[g]$
4. $\pi \models \exists x\varphi[g]$ exactly if there is an $a \in Posse$, such that $\pi \models \varphi[g[a/x]]$
5. $\pi \models \Diamond\varphi[g]$ exactly if there is a $\rho \in C$, such that $\rho \circ \pi \models \varphi[g]$

As usual, $\Box\varphi$ is defined to be $\neg\Diamond\neg\varphi$. Finally we can define the notions of truth and validity.

φ is *true* wrt. g, exactly if it is *true* at Id_{Posse} wrt. g.
φ is *valid* exactly if φ is true wrt. all models and assignments.

All of the following formulae are valid in this kind of semantics: $(x = y \rightarrow \Box x = y)$, $(x \neq y \rightarrow \Box x \neq y)$, $(x = y \rightarrow (\Diamond(Px \wedge Qy) \rightarrow \Diamond(Px \wedge Qx)))$.

5.3.2 Comparison to Constant Domain QML

You may have noticed that we have ended up with a completely classical semantics for quantified modal logic. It is a semantics for classical constant domain semantics, with permutations (representational stipulations) playing the role of "possible worlds". The actual "world" is the identical permutation, the representation according to which every possible individual represents itself as having those properties it actually has.

[4] "at" is short for "according to"

A local model of of S5 constant domain quantified modal logic is a quintuple $(W, R, D, V, @)$, such that W and D are non-empty sets, $R \subseteq D \times D$ is an equivalence relation (i.e. it is reflexive, symmetrical, and transitive), $@ \in W$, and V a function which meets the following conditions:

(a) $V(c) \in D$ for every constant c,

(b) For every $v \in W$: V_v is a function, such that

for every n-place predicate P: $V_v(P) \subseteq D^n$,

$V_v(=) := \{(a, a) | a \in D\}$.

Let us now define $w \models \varphi[g]$, formula φ is true wrt. an assignment g and at a world w.[5]

1. $w \models Pt_1, \ldots, t_n[g]$ exactly if $(Vg(t_1), \ldots, Vg(t_n)) \in V_w(P)$

2. $w \models \neg\varphi[g]$ exactly if $w \not\models \varphi[g]$

3. $w \models \varphi \wedge \psi[g]$ exactly if $w \models \varphi[g]$ and $w \models \psi[g]$

4. $w \models \exists x\varphi[g]$ exactly if there is an $a \in Posse$, such that $w \models \varphi[g[a/x]]$

5. $w \models \Diamond\varphi[g]$ exactly if there is a $v \in W$, such that wRv and $v \models \varphi[g]$

Finally we can define the notions of truth and validity.

φ is *true* in a model $(W, R, D, V, @)$ wrt. an assignment g, exactly if it is *true* at @ wrt. g.

φ is *valid* exactly if φ is true wrt. all models and assignments.

Let us impose some further constraints on C.

Refl $Id_{Posse} \in C$,

Symm $\pi \in C$, then $\pi^{-1} \in C$,

Trans $\pi \in C$ and $\rho \in C$, then $\rho \circ \pi \in C$;

[5] "at" is short for "according to"

Now we can show that for every permutation model that obeys these constraints there is a (local) model of S5 constant domain quantified modal logic that makes precisely the same formulae true. Hence, all of the logical truths of what has been termed the simplest quantified modal logic, see Linsky and Zalta (1994) stay true on our semantics.

Let $(Posse, C, V)$ be a permutation model, and Π be the set of all permutations of $Posse$. $(Posse, C, V)*$ is defined to be $(\Pi, R, Posse, V, Id_{Posse})$, where $R \subseteq \Pi \times \Pi$ is that relation such that for all $\pi, \rho \in \Pi : \pi R \rho$ iff there is a $\tau \in \Pi$ such that $\rho = \tau \circ \pi$.

R is reflexive (because of Refl), symmetrical (because of Symm) and transitive (because of Trans); furthermore φ is *true* in $(Posse, C, V)*$ wrt. an assignment g, exactly if it is *true* in $(Posse, C, V)$ wrt. g.

On the other hand, the requirement that worlds be permutations of the domain makes the resulting semantics stronger. E.g. the following formula is logically valid on our permutation semantics, but not in S5 constant domain quantified modal logic.

$(\Diamond \exists x Px \& \Diamond \exists x \neg Px) \rightarrow \exists x \exists y (x \neq y)$.

In the following two sections I will discuss extensions and possible modifications of the basic semantics.

5.4 Applications

5.4.1 Lexical Semantics and Impure Properties

So far, we have dealt with arbitrary models. Let us now talk about the *intended interpretation*. The intended interpretation is one, where the set $Posse$ really *is* the set of possible individuals, and where V gives our predicates their intended meanings. In the specification of the intended representation we will not only refer to by whom a given individual is represented, but also to the properties of the representatives.

Take a look at some examples of the intended interpretation. Who is in the extension of "rich" at π? Those represented as being rich by π. What pairs satisfy "in" at π? Pairs (A, B) such that A is represented to be in B by π. Who satisfies "father of Tom" at π? Those individuals that are represented to be the father of Tom at π.

$A \in V_\pi(R)$ exactly if πA is rich.

$(A, B) \in V_\pi(\leq)$ exactly if πA is in πB.

$(A, B) \in V_\pi(F)$ exactly if πA is the father of πB.

$A \in V_\pi(F_t)$ exactly if πA is the father of πTom.

This suffices to show that our representation relations (permutations) actually play a part in the semantic game. But they play a different part than the one the counterpart relation had to play. *We have moved representation* de re *from semantic composition into lexical semantics.* And this is the reason why we are able to deal with impure properties like "being father of Tom" above. It is easily seen that, relative to any representation π, the pair (Ede, Tom) satisfies "x is father of y" at π exactly if Ede satisfies "father of Tom" at π, as it ought to be.

5.4.2 Restricted Quantification

Normally we do not quantify over arbitrary possible individuals. E.g. if someone were to claim that there are honest politicians, the person would probably want to express more than that some possible individuals are so. What the person wants to say is that there are honest politicians in the actual world. To accomodate this use, and also in order to provide a sense of existence in which individuals may be said to exist contingently, let us assume actualist quantifiers are expressed as proposed in Chap.2. Let @ be a name for the actual world (not to be mixed up with Id, the designated representation relation that plays the role of the actual "world").We introduce the following abbreviations:

$$\bigvee x\varphi := \exists x(Ix@ \wedge \varphi)$$
$$\mathbf{E}x := \bigvee y(x = y).$$

One can easily see that $\Diamond\neg\mathbf{E}x$ is satisfiable and $\Box\mathbf{E}x$ is not valid. $\bigwedge$ is then defined to be the dual of $\bigvee$.

5.4.3 Actually

Fara and Williamson (2005) claim it is impossible to equip counterpart semantics with an "actually"-operator. It is easy to do so in the case of our semantics; just add the following clause.

6. $\pi \models \mathbf{A}\varphi[g]$ exactly if $Id \models \varphi[g]$

According to Fara and Williamson, any semantics of the actually-operator should imply that $\mathbf{A}\varphi$ is true iff φ is true at the actual world; now the above clause does exactly that if "world" is understood in the logical sense, not in the metaphysical sense. (The second reading simply makes no sense, here, because the worlds of the metaphyiscian no longer serve as the things propositions are true at.)

Hazen's critique of counterpart-theoretic semantics, as well as Fara and Williamson's, centers around the fact that an actually-operator, construed as quantifying over counterparts in the actual world, is bound to get certain things intuitively wrong, either because sometimes there are multiple counterparts of a single thing in the actual world, or because sometimes there is no counterpart of such a thing in the actual world. Now, these problems do not concern us any longer, because (i) we do not quantify over representants and (ii) in any world and for any given individual, there is always a unique representant.

What distinguishes the above clause from more familiar ones based on a different metaphysico-semantical outlook is that now $\mathbf{A}Px$ may be true of merely possible individuals. Suppose Fred is a talking donkey from a world different from ours. We predict that one may say that Fred actually talks and that it is possible that Fred actually talks. I think these are totally natural things to say, given that it is *true* that Fred talks. One may feel reluctant to say it, because Fred doesn't talk in the actual world, with "actual world" being understood in the metaphysical sense. But, as I just said, this is not the sense of "world" that seems to be appropriate now.

How does our operator deal with run-of-the-mill examples of the use of "actually" like

(10) It might have been that everyone actually rich was not rich?

We now face an additional difficulty, due to the fact that our quantifiers may either be understood restrictedly or unrestrictedly. Fortunately, understood either way (1) comes out true as it ought to.

(11) $\Diamond\forall x(\mathbf{A}Rx \rightarrow \neg Rx)$

(12) $\Diamond\bigwedge x(\mathbf{A}Rx \rightarrow \neg Rx)$

On the restricted reading (11), (10) could already come true because of the existence of representations that represent the world as containing no actual individual. (Read R instead of "rich" as "entity" and you get a real difference in truthvalue between the two readings.) But this is so, as it ought to be, for (12) might be rephrased as "it might have been that every individual in the world that is actually rich might have been not rich", and now this could easily get true because it might have been that no individual in the world did actually exist.

The second example is even more ambiguous if quantifiers have a restricted and an unrestricted reading.

(13) There might have been individuals which do not actually exist.

(14) $\Diamond\exists x\neg\mathbf{A}\exists y(y{=}x)$
(It is possible that there are possible individuals for which it is not actually true that there are possible individuals identical to them.)

(15) $\Diamond\bigvee x\neg\mathbf{A}\exists y(y{=}x)$
(It is possible that there are individuals in the world for which it is not actually true that there are possible individuals identical to them.)

(16) $\Diamond\exists x\neg\mathbf{AE}x$
(It is possible that there are possible individuals for which it is not actually true that they exist in the world.)

(17) $\Diamond\bigvee x\neg\mathbf{AE}x$
(It is possible that there are individuals in the world for which it is not actually true that they exist in the world.)

The differences between these disambiguations are vast. E.g. if the second quantifier in (13) is read unrestrictedly, as in (14) and (15), (13) is contradictory; (16) and (17) on the other hand, will come out true because there is more than one world. Of all the different disambiguations (17) is perhaps the only one people normally aim at if they say (13). But at least the above paraphrases seem to be exactly what is required to disambiguate (13) if our semantic assumptions about (13) are correct.

5.5 The Virtues of Counterpart-Theory

In Lewis (1986a), Lewis claims that there are applications of counterpart theory that rival approaches to representation *de re* couldn't accommodate. In this section we will show that our semantics could.

5.5.1 Intransitivity

Above we have seen that Chisholms paradox consists of an argument that, wrt. to modality de re, (4) $\Box\varphi \rightarrow \Box\Box\varphi$ sometimes fails. Lewis points out that it comes out invalid in counterpart theory, because of the intransitivity of the counterpart relation. Since representations are supposed simply to encode the counterpart relation, they will inherit this intransitivity.

Transitivity corresponds to the above condition **Trans**, repeated below:

Trans $\pi \in C$ and $\rho \in C$, then $\rho \circ \pi \in C$;

Suppose we want to know whether

(18) Possibly, possibly A is P

is true.

If we represent A directly as B, we represent A to possess the properties B has. Now these properties include B's *modal* properties, properties like being possibly P. On this view (18) is true iff there is an admissible representation that represents A as someone B such that there is an admissible representation of B by some C that represents B as being P. Now admissible representability need not be transitive: i.e. there need not be any admissible representation of A by C. This is why (18) does not imply

(19) Possibly, A is P

is true.

5.5.2 Contingent Identities

Gibbard presents a scenario, where a statue and the lump of clay it is made of exist at precisely the same places in space-time. This suggests that they are identical. But if they are, they are only contingently so, because e.g. the lump could have been formed long before the statue. In Lewis (1983c), Lewis argues that cases of contingent identites should be dealt with by a machinery of multiple counterpart-relations, see p.23 above.

Now, to have multiple representation-relations is open to us as well. At least the question whether we should admit multiple representatives for the same representee, only distinguished by different roles, does not depend on whether one focusses on counterparts, or complete maximal direct representations. In the present framework, this might be achieved by a multiplicity of permutations, e.g. one for every sortal property P. According to Lewis, the choice of the right representation relation is made by context. Of course we could say the same wrt. to such a variant of our theory.

Another instance of the denial of coincident entities is the identification of a world with the sum of its parts. The world and the sum of individuals in it might come apart. E.g., while the world is necessarily a world, the sum of its parts does not necessarily consititute a world; it could also constitute a proper part of a world. Therefore, if you want to combine our semantical apparatus with an identity theory about worlds and certain sums, already for the case of worlds you would have to employ multiple representation relations.

5.5.3 Context Dependence

According to Lewis, the choice of the right representation relation is made by context. Of course we could say the same wrt. to the abovementioned variant of our theory that would be required in order to enable contingent identity. But there is also a second way in which judgements about essential properties might be said to vary with context. They could depend on tacit restrictions on what counts as possible in a given conversation.

Let us say that the content of a sentence φ is the set of all $\pi \in C$, such that φ is true at π. Given that permutations are our analysis of the logician's "possible worlds", this explication of content is ordinary possible worlds semantics.[6]

Since propositions are sets of permutations, they can be understood as restrictions on permutations. Since matters of essence are also a matter of restrictions on permutations, if sometimes essences vary with context, this can be treated in terms of some kind of propositional context. That would allow to apply, e.g., received models of context change, see Stalnaker (1978).

5.5.4 Haecceitist Differences

Two worlds differ *haecceitistically* if they are qualitatively the same but differ wrt. their representational properties. In Lewis (1983b, 1986a), Lewis

[6]Depending on whether we adopt the complications proposed in the last subsection, this notion of content would still not suffice for the purposes of semantic recursion. It does suffice for our purposes, here.

observes that a metaphysics of worldbound individuals should deny the existence of haecceitistic differences between possible worlds. If two worlds represent differently, this cannot be because of purely numerical differences between them; especially because, in a metaphysics of worldbound individuals, there are no identities across worlds.

But Lewis concedes that there are compelling examples of such differences. I might have lived in the 137th epoch of a world of one-way eternal recurrence, and I might have lived in the 17th epoch of the same world (or a world qualitatively just like the first). I might have been the prince, and I might have been the pauper (under completely the same circumstances). If I am the prince, I am there "but for the grace of god", see Feldman (1971). Now these examples can be used to provide an argument for the existence of haecceitist differences. E.g. from the first example it follows that there is a recurrence world according to which I live in the 17th epoch (and not in the 137th) and a second world, qualitatively identical to the first, according to which I live in the 137th epoch (and not in the 17th). These worlds can't be identical, so there are haecceitist differences between worlds after all.

The argument is sound. How should the anti-haecceitist reply? The anti-haecceitist need not accept the conclusion, he could also deny the hidden semantical premise that the possibilty operator quantifies precisely over the possible worlds of the metaphysician. He could

(a) give up on the idea that possibility is exactly quantification over possible worlds, or

(b) give up on the idea that the possible worlds of the semanticist and those of the metaphysician are the same.

In effect Lewis's counterpart-theoretic translation takes option (a) because it does not interpret possibility as quantification over possible worlds alone. And that is why a counterpart-theoretic treatment of the above example may come out exactly right. If I have counterparts in each of the epochs of the recurrence-world w, then it is both true that I might have lived in the 17th epoch, and that I might have lived in the 137th, because these statements are understood as "there is a world w and a counterpart of me in w that lives in the 17th (137th) epoch".[7]

I take option (b). The "possible worlds" of the semanticist are analysed in terms of permutations of individuals. Now it is easy to account for the

[7] Of course whether this analysis succeeds depends on whether I really have counterparts in these two epochs.

examples without assuming haecceitist differences between metaphysically possible worlds. Worlds do not represent at all, “worlds” do! And to admit that there are haecceitist differences between “worlds” is not metaphysically objectionable.

Suppose A_{17} and A_{137} are individuals in the 17th and the 137th epoch of a world of one-way eternal recurrence, resp., and there are π and ρ, which map me onto A_{17}, and A_{137}, respectively. Such a model is able to make it true that I might exist in the 17th epoch of that world and that I might exist in the 137th epoch of a world which is just like the first, however this will be expressed precisely.[8]

5.6 Comparisons

5.6.1 Hazen

My permutations resemble Allen Hazen’s *stipulations*, functions that specify how certain individuals are to be represented in a certain world.[9] For the simplest case (no nested modal operators) Hazen’s semantics can be formulated in terms of truth in a “stipulational world”, i.e. a world *plus* a stipulation; this is similar to “truth according to a representation”. My proposal has been inspired by Hazen’s, and so have the solutions to specific problems, like e.g. to the task of defining an actuality operator and the problem of haecceitist differences.

But there are also differences, both in detail (e.g. the quantifiers are actualist) as well as with respect to conceptual foundations. The main difference is that Hazen’s stipulations are limited to representation in a particular world. This limitation causes enormous technical complications in Hazen’s system. Hazen has to add a complicated apparatus of stipulational bookkeeping in order to be able to keep track of identifications made during the process of the stepwise evaluation of a modal formula with nested modal operators, see Hazen (1977).

[8]For this kind of treatment of the examples, see Hazen (1977, 1979), and also Lewis’s reconstruction of Hazen’s ideas in terms of indvidual possibilities, discussed in sec. 7 below.

[9]For details of Hazen’s proposal I have to refer to Hazen (1977). Hazen (1979) provides an accessible introduction to the main ideas.

5.6.2 Lewis on Possibilities

Finally let me compare representations with Lewis's individual possibilities. Although Lewis introduces this notion for a different purpose, namely the semantics of belief *de se* (see Lewis (1983a)) they will turn out to be intimately related to my representations.

This is how Lewis introduces individual possibilties in Lewis (1983b), p.395:

> *Possibilities are not always possible worlds.* There are possible worlds, sure enough, and there are possibilities, and possible worlds are some of the possibilities. But I say that *any* possible individual is a possibility, and not all possible individuals are possible worlds. Only the biggest ones are.
>
> The world is the totality of things. It is the actual individual that includes every actual individual as a part. Likewise a possible world is a possible individual big enough to include every possible individual that is compossible with it [...]. It is a way that an entire world might possibly be. But lesser possible individuals, inhabitants of worlds, proper parts of worlds, are possibilities too. They are ways that something less than an entire world might possibly be. A possible person, for instance, is a way that a person might possibly be.

Here, possibilities are equated with possible individuals and with ways things might possibly be. While at least the first identification seems to be peculiar (intuitively, possible individuals are particulars, while both ways and possibilities are universals, something various particulars could share), we will not challenge it here. So, let us simply accept Lewis's terminology for the following.

For which purposes are these possibilities employed by Lewis, and how they are related to what we have presented above?

Possibilities are employed as *possibilities for*; e.g. Lewis characterises the *epistemic* possibilities of a subject A as individuals A might have been, for all he knows. And, according to Lewis, the *metaphysical* possibilities for A (relative to a given counterpart-relation) are A's counterparts. [10]

[10] There are other kinds of possibilities that correspond to various kinds of accessibility relations between possible individuals, cf. Lewis (1983b), p.401.

This relation of being a possibility for ... and representations correspond closely. To say that B is a possibility for A in Lewis's sense seems to be the same thing as to say that B directly represents A. And since Lewis regards counterparts as possibilities, Lewis and I even agree that the semantics of possibilit*y* is to be defined in terms of possibilit*ies*.

Besides the above notion of an individual possibility there is also a notion corresponding to joint representations. For the job of *joint possibilities* Lewis proposes sequences of possible individuals Lewis (1983b, 1986a). Such a sequence is a possibility for another sequence *with the same index set*. (A sequence is a function from an index set (its domain) onto a value set (its range).) "A suitable pair of indiviudals in a given order, for instance (take this as a two-term sequence) is a way that a pair of individuals might possibly be." (Lewis (1986a), p.232). Lewis does not stop with two-place sequences. Other purposes are easily conceivable where one might need even bigger and bigger possibilities. For him, a possibility for the sequence of all individuals of a given world is the limiting case. But this still seems to be arbitrarily limited. Sequences of possible individuals are not limited to sequences of worldmates. Then why consider only possibilties for sequences of worldmates or sequences of worldmates as possibilities?[11] Consequently, we should allow sequences of individuals from different worlds. But then, a possibility for a sequence of all individuals of a certain world is no longer the limiting case. Instead the limiting case is a possibility for all possible individuals.

Now to say something is a joint possibility for simply seems to be a complicated way to specify a direct representation relation. If σ is a possibility for τ, then $\pi = \{\tau_i, \sigma_i | i \in I\}$ (with I being the joint index set of σ and τ) is a direct representation.

So representations are at least implicit in Lewis's own notion of a possibility *for*. Summing up, the difference between the approach here, and Lewis's semantics is not so much that modality is analysed in terms of possibili-

[11] To address the possibility of non-existence, Lewis proposes gappy sequences as possibilities, where some items of the index set do not receive a value. If σ is a gappy sequence that is a possibility for the sequence ρ, then there is an index n of ρ, such that σn is undefined. This is meant to symbolise that it is part of this possibility that ρn does not exist. To take a simple example, the empty sequence ( ) is a way the 1-place sequence $(X) = X$ might have been. I find gappy sequences strange. They do not seem to specify possibilities completely. If we say that individual Y is a possibility for X, we specify every qualitative property X has according to this possibility. But if we say that X does not exist, we only specify its absence from a particular world, namely the counterfactual world we are considering presently; we do not specify what properties X has.

ties, the differences concern (a) the nature of the constraints on admissible representations, and (b) how precisely possibilites are employed in semantics.

The latter difference is vast. Put in terms of Lewis's notion of possibilities I say that "Possibly, RAB" is true iff there is a possibility according to which RAB is true, while Lewis offers two analyses (and it is not quite clear which to apply to any occurrence of the above sentence). First that "Possibly, RAB" is true iff *there is a world* and, *in that world* both an individual C which is a possibility for A as well as an individual D which is a possibility for B and RCD is true; second that it is true iff there is a world and, in that world, a pair (C, D) which is a possibility for the pair (A, B) and RCD is true.

I dare not say which proposal is the most natural and simple one. At least to say that something is possibly true iff there is a possibility according to which it is true does not seem to me to be very much off the mark.

Chapter 6

Properties of Properties

In this chapter, I will introduce the notion of a perfectly natural property (see Lewis (1983d)). It allows to define the notion of a qualitative property and that of an intrinsic property. It also allows to predict differences between properties wrt. degree of naturalness, but I argue that it is not clear whether this does really present an advantage of the theory. Finally I ask whether the definitions could be had for less.

6.1 Introduction

There are lots of properties, and we want to be able to distinguish between different varieties. We want to be able to distinguish between properties which are *natural* to varying degrees, between intrinsic properties and *extrinsic* ones, and between *qualitative* and *non-qualitative* ones. The distinction between qualitative and non-qualitative (or pure and impure) properties and relations has already played some role, especially in Chap.1 and 3.

Qualitative properties are properties individuals have in virtue of how they are, not in virtue of which individuals they are. Take, for instance, 1KG := the property of having exactly one kg in mass; 1KG + Bonzo := the property of having 1kg in mass or being Bonzo, a monkey, which weighs more; or the set of red things. 1KG is an example of a property things have in virtue of how they are. Suppose there is a possible monkey, Bonzo', indiscernible from Bonzo (a monkey that differs only numerically from him). Then 1KG + Bonzo is an example of a non-qualitative property, a property which Bonzo has in virtue of which individual he is, not how he is. Bonzo' is exactly like Bonzo, but only the latter has the property 1KG + Bonzo.

The way an individual is (except for worlds) is not the way it is *by itself*. Qualitative properties may be constituted by relations to other individuals. Indeed, the whole world can matter; therefore, one can equally well characterize a qualitative property of an individual A as one had in virtue of the way A's *world* is, and of the place the individual has in the world. I.e. a qualitative property need not be *intrinsic*. A property is intrinsic if and only if, for any A, whether A has the property depends only on the thing itself. Intrinsic properties are properties things have "in virtue of the way they themselves are." (Lewis (1986a), p.61). Again, this is meant to be an elucidation, not a definition. Let's try to apply it. It is not easy to give examples of truly intrinsic properties; properties that are intrinsic according to intuition often turn out not to be after analysis. But a particular shape is a good candidate for being an intrinsic property.

1KG is a good example of a property which is *natural*, whose bearers have something in common. 1KG + Bonzo does not possess such a unity. More important is a comparative notion of naturalness:1KG is more natural than 1KG + Bonzo. Here is another example:

> Consider the difference between the class of crimson things and the class of colored things. Both are natural classes. But the latter is much more closely knit, has a much higher degree of unity, than has the class of colored things. Armstrong (1989b), p. 24

A theory of properties ought to give an account of intrinsicness and qualitativeness as well as their opposites, and it should contribute to a comparison of properties in terms of their naturalness. Armstrong's theory of universals (Armstrong (1978, 1989b)) promises to do this. In the following, I will sketch Lewis' set-nominalistic version of Armstrong's theory and evaluate its merits.

6.2 Universals *A Posteriori*

According to Armstrong (1997), the world consists of states of affairs which in turn consist of individuals and universals. If we could list all states of affairs, i.e. for every universal which (sequences of) individuals instantiate it, we would have a complete description of reality.

A theory of universals is developed in Armstrong (1978). Universals are wholly present wherever they occur, but they may occur in different places at the same time. Suppose having 1 kg in mass is a universal. Then, at every location where someone has 1 kg mass, this universal is wholly present. Thus, they are unlike individuals, which never occur at two distinct places at once. (Lewis comments that "by occurring repeatedly, universals defy intuitive principles. But this is no damaging objection, since plainly the intuitions were made for particulars." Lewis (1983d), p. 345.) Universals are also unlike sets. We wouldn't want to say that a set is wholly present wherever it is instantiated.

What are *the* universals? Armstrong claims that we can only rely on science for an answer. Science, he thinks, tries to discover the fundamental properties of things. So, for an answer to the question whether a purported universal is an actual one, we have to ask science. Who else could we ask? Armstrong thus terms his theory of universals *a posteriori realism*.—It is certainly true *in some sense* that science tries to discover the fundamental properties and relations because it tries to devise a set of laws, and the fundamental notions of these laws are (correlated with[1]) some set of properties and relations. But there is no guarantee that, even at the end of all enquiry, we end up with only one possible theory of the world. If you are a realist, you have to admit the possibility that, even at the end of all inquiry, science may fail to get at the universals. Of course, Armstrong is right in maintaining that there is nobody but science we could ask which are the universals. But it is not clear that a single correct answer is, in any sense, secured by scientific methodology and the facts alone. If this is not so, then a theory of universals is not so a posteriori after all. But that makes universals objectionable because it seems to imply that, among the best systems of scientific hypotheses, there is one that is better, but solely for metaphysical reasons. The same applies to Lewis' theory of perfectly natural properties (see the next section). I don't think I have devised a decisive argument against such a theory, but what I have said may be a reason to look for a less platonistic alternative. In the final section, I will briefly discuss such an alternative.

Although a theory of universals as the above is perfectly compatible with possible worlds realism, in Armstrong's work it is married to *combinatorialism*. There is an immediate consequence of this: if other worlds are built out of the universals of this world, there can't be *alien universals*; that is,

[1]While science usually formulates its laws in terms of functions ("determinables"), the theory of universals tends to break down such a function into a collection of properties ("determinates"). E.g. the determinable property mass will be replaced by the determinate properties having 0 kg in mass, . . . , having 1kg in mass and so on.

universals which are not instantiated in the actual world. Armstrong seems to be committed to saying the same against *alien individuals*, but doesn't (something we won't discuss here). In the following, I will rule out neither alien individuals nor alien universals (or alien perfectly natural properties, resp.)

6.3 Perfectly Natural Properties and Relations

According to our metaphysics, for every property there is a set of possible individuals which have the property. All along, we have interpreted predicates by assigning them sets of (possible) individuals. It is also natural to suppose that for every set of possible individuals, there is a corresponding property. If we combine the present metaphysics with set-nominalism, sets and properties are identified.

Lewis (1983d) proposes to replace universals by sets. First step: for every universal define the set of its instances. Every such set is called "perfectly natural". Second step: get rid of the universals. Instead of an ontology of universals, employ an ontology of sets and an additional second order property; namely, perfect naturalness. You need not make the second step. Lewis claims he remains neutral on the question whether primitive universals, primitive *tropes*, or primitive perfect naturalness are preferable. The following applications of PNRs do not force the second step, i.e. they are available for the theory of universals, too.

We will use the term "relations" to cover properties (one-place relations), too. We will abbreviate "perfectly natural properties or relation" as "PNR". Sometimes we will also name the set of PNRs by "PNR". Let us now inspect some applications. Before I go on to define "intrinsic" and "qualitative", let me first introduce the important notion of *supervenience*, which will prove helpful afterwards.

6.4 Supervenience

The notion of supervenience figures prominently in many discussions in the last century and continues to do so. Lewis gives the following illustration.

> A supervenience thesis is a denial of independent variation. Given an ontology of possibilia, we can formulate such theses in terms of differences between possible individuals or worlds. To say that so-and-so supervenes on such-and-such is to say that there can be no difference in respect of so-and-so without difference in respect of such-and-such. Beauty of statues supervenes on their shape, size, and colour, for instance, if no two statures, in the same or different worlds, ever differ in beauty without also differing in shape or size or colour. Lewis (1983b), p. 358

Conversely, no two statues can be the same wrt. shape, size and colour, without being the same wrt. their beauty. The example allows a simple formal treatment. We will identify the beauty of X with the set of possible individuals which are equally beautiful as X and the shape-size-colour of X as the set of possible individuals which have the same shape, the same size, and the same colour. We can then say that the set of beauties *supervenes* on the set of shape-size-colours, where the relation of supervenience is defined as follows:

Let M and N be sets of properties. M supervenes on N iff every two possible individuals which have the same M-properties also share their N-properties.

The definition is general enough to cover what we are going to discuss below, although there are many different notions of supervenience that have been discussed in the literature. One special case of the above definition is when the supervenient item is not a set (or a function of individuals into properties) but only a single property. For this purpose, we will simply identify singletons with their elements, such that, if P is a property then P supervenes on N iff $\{P\}$ supervenes on N. In this sense, e.g., a disjunctive property supervenes on the set of its disjuncts.

6.5 Defining “intrinsic”

We know that intrinsic properties are properties things have (or lack) in virtue of the way they themselves are. Now we need an explication of this intuitive notion. Lewis points out that such an explication is easily at hand, once we have an ontology of PNRs.

> it can plausibly be said that all perfectly natural properties are intrinsic. Then we can say that two things are duplicates iff (1) they have

> exactly the same perfectly natural properties, and (2) their parts can be put into correspondence in such a way that corresponding parts have exactly the same perfectly natural properties, and stand in the same perfectly natural relations.[...] Then we can go on to say that an intrinsic property is one that can never differ between two duplicates. Lewis (1986a), p. 61

This can be made slightly more precise. (1) above is redundant if we interpret "part" in (2) in a way such that A is always a part of itself (i.e. we don't use the relation of "proper part"). We will do so in the following. Suppose M is a set of relations, and $\leq$ is the part-whole relation conforming to the axioms of mereology.

Let an *M-isomorphism from A onto B* be a bijection f from the set of parts of A onto the set of parts of B, such that if and $R \in M$ or $R =\leq$ and $A_1, \ldots, A_n \leq A$, then

$$RA_1, \ldots, A_n \text{ iff } RfA_1, \ldots, f(An)$$

(an M-isomorphism from A onto B preserves the part-whole structure and all the relations in M).

Now we can say that A and B are *duplicates* iff there is a PNR-isomorphism from A onto B; and that a property P is intrinsic iff for every possible individual A and every duplicate A' of A, $A \in P$ iff $A' \in P$. Therefore, we can indeed say that an intrinsic property never differs between possible duplicates.

There is some useful alternative choice of terminology I want to mention here: the set of duplicates of A can justly be called A's nature (the way A is by itself)[2]. Then we can define that P is intrinsic iff no two possible individuals can differ wrt. P but agree wrt. their natures, i.e. P *supervenes* on the natures of its bearers.

Lewis contends that, in order for this to be an explication of "intrinsic", we need to assume that the perfectly natural properties are already intrinsic in the intuitive sense. But this is still not enough, witness the following example.

Let R be a relation, s.t. for every A and B RAB iff A and B have exactly the same distance from the (center of gravity of the) Eiffel tower. Suppose R were perfectly natural. Then no two sufficiently compound things, such

[2] In traditional terminology A's nature is often identified with A's essence.

as persons, would be duplicates unless we had placed them in exactly the same direction, relative to the Eiffel tower. Suppose Anton is a person facing the Eiffel tower. Is Anton' a duplicate of Anton? We need to check, among other things, whether he faces the Eiffel tower. Say he is alike Anton in a all other respects but not in respect of R. We can make Anton' an intrinsic duplicate just by turning him around! Or, suppose Anton' is living in a different world, where the Eiffel tower is not present. Then, Anton' can't be an intrinsic duplicate of Anton, for none of Anton's parts have any distance from the Eiffel tower, not even distance 0. Either we want to say, now, that no two parts of Anton' stand in the relation R, or all do. Either way, this will distinguish Anton' from Anton, since Anton has some pairs of parts which are equidistant from the tower, and others that are not. Thus, R distinguishes Anton and Anton' solely because they are living in different worlds, and nothing could be more extrinsic to them. If Anton and Anton' are exactly alike in virtue of how they themselves are and how R distinguishes them, then, by the definition of "intrinsic" every intrinsic property is likewise allowed to do so. But then, the definition doesn't explicate the intuitive criterion ("exactly alike in virtue of how they themselves are"). Therefore, in the following, we will replace the assumption that every perfectly natural property is already intrinsic by the assumption that every perfectly natural relation is already intrinsic to its pairs, which, of course, implies the former assumption.

Again, we can provide "intrinsic to its pairs" with a formal definition. Let us say that an n-place relation R is *intrinsic to it's n-tuples* iff for any $A_1, \ldots A_n$ and $B_1, \ldots, B_n$ if there is a PNR-isomorphism f from $A_1 + \ldots + A_n$ onto $B1 + \ldots + B_n$, such that $f(A_i) = B_i$, then $RA_1, \ldots A_n$ iff $RB_1, \ldots, B_n$.

The notions developed so far are helpful in defining two notions related to that of an intrinsic property: the traditional notion of an internal relation.[3] A relation R is internal iff it does not distinguish between duplicates; i.e. iff $RA_1, \ldots A_n$ and for every i A_i is a duplicate of $B_i (1 \leq i \leq n)$, then $RB_1, \ldots, B_n$. Obviously, every internal relation is intrinsic to it's pairs.

6.6 Dunn's Objections

The adequacy of Lewis' definition has been challenged in Dunn (1990). Dunn observes that

[3]see Lewis (1986a), p. 62

(3) “being a duplicate of B”

is an intrinsic property by Lewis’ criterion. But isn’t this a relational property?

> Consider the question whether a given object a is a perfect duplicate of an object b [...]. Clearly the answer to this question does not depend on a alone, it equally depends on b and its intrinsic properties. Dunn (1990), p. 185

So, Dunn claims to have an example for a property which is predicted to be intrinsic by Lewis’ criterion, but should come out extrinsic.

Dunn’s argument seems to be based on a criterion that allows him to discard some properties as being intrinsic. The criterion is roughly as follows: if you can refer to a property with a description in terms of a contingent relation to some particular objects, then the property is not intrinsic. E.g. if a candidate intrinisic property turns out to be the most remarkable property had by Ede, then the property is not intrinsic, because whether a given object A has the most remarkable property had by Ede, clearly does not depend on A alone, it equally depends on Ede and his properties.

The trouble with the criterion is that it is far too successful: if Dunn is right, then no property is intrinsic. We can refer to every property with descriptions that refer to some particular objects. Let P and Q be any properties. Suppose Ede is clever. The property of being P if Ede is clever and of being Q, else is identical to the property of being P, so it is intrinsic exactly if P is. But whether a given object A has the property of being P if Ede is clever and of being Q, else, clearly does not depend on A alone, it equally depends on Ede and his properties.

Theodore Sider (in Sider (1996)) defends Lewis’ account against this argument in considerable detail. He distinguishes two readings of (3):

> P1: the property had by a possible object x iff x and the counterpart of b in x’s world are duplicates
> P2: the property had by a possible object x iff x and b itself, in the actual world, are duplicates (Sider (1996), p. 13)

He then argues that P1 is not intrinsic according to both intuition and Lewis’ criterion, whereas P2 is indeed intrinsic according to Lewis’ criterion. Dunn is then portrayed to use P1 instead of P2, so Dunn’s argument seems to rest on an equivocation.

I agree with Sider that that the argument rests on an equivocation, but I don't think there is any ambiguity in (3). Only P2 is referred to by (3). P1 is the product of Sider's attempt to define the semantical value of an impure predicate as a set of possible individuals defined in terms of counterparts. We have seen in Sec.3.3 above that this is not possible. But this brings us finally on the right track: Dunn's criterion is not a criterion for being an intrinsic propery, it is a criterion for being an impure predicate. And Dunn's conflation is not between two possible meanings of (3). It is between the notion of an extrinsic property and the notion of an impure predicate! An impure predicate is a predicate such that whether it applies to a given object a in some circumstance c does not depend on a alone, it depends on how other objects are according to c. Below in Sec.7.2 I will make the notion of a pure predicate precise.

There is a second objection Dunn puts forward; namely, that the property expressed by

(4) "being such that Socrates is wise or not wise"

comes out intrinsic on Lewis' criterion. Now Dunn thinks this is odd since, although Reagan satisfies (4), this does not depend on intrinsic features of Reagan. The objection is fair and points out to a certain weakness of Lewis' explication of the intuitive notion of "intrinsic".

(5) A has P in virtue of the way A itself is

seems to be only insufficiently explicated by "P supervenes on the natures of things". If P is the universal property U, then (5) does not seem a correct thing to say: "A has U in virtue of the way A itself is." "Why, A has U anyway!" The problem seems to be that "in virtue of" states a sort of dependence and *supervenience* is in some cases an insufficient explication of *dependence*. The universal property supervenes on whatever set of properties you want; hence, which set you choose makes no difference; and hence, A's having U is not really *dependent* on A's having any other property.

But we shouldn't exaggerate the extent of the problem. First, this is a relatively harmless case of overkill of our explication. If P is not U, everything is fine, and we can indeed say that A has P in virtue of the way it itself is iff no possible object can lack P but have the same nature as A. Second, if we want to, we can strengthen the definition by requiring that there are individuals which have the complement of P in virtue of the way they themselves are. Anyway, since I do not see much harm done, I will stick to Lewis' definition above in the following.

Additional worries about (4) might arise again from the fact that (4) is about Socrates in some *syntactic* sense, but this isn't really a *dangerous* sense (c.f. Sider (1996).)

6.7 Qualitative relations

Qualitative properties are properties individuals have in virtue of how they are, not in virtue of the way which individuals they are: they don't make purely numerical distinctions. A qualitative property is not just a property an individual has in virtue of how it is *by itself* (we had these before), a qualitative property also depends on the whole way the world is, and on the location the individual has in the world. We can explicate this by the pattern of perfectly natural properties and relations the individual is involved in in two different ways: as closure under PNR-isomorphisms, or as closure under automorphisms.

The first sense explicates the dependence[4] on the way the world is and on the location of the bearer of the property. Let us assume for the moment that there are no perfectly natural trans-world relations (if $RA_1, \ldots, A_n$ and R is perfectly natural, then there is a world w, such that $A_1, \ldots, A_n$ are all in w). A and B are indiscernible iff there is a PNR-isomorphism from the world of A onto the world of B, such that B is the image of A under the isomorphism.[5] Finally,

> P is a qualitative property iff there are no possible individuals A and B, such that A is indiscernible from B, A has P and B lacks it (P is closed under indiscernibility).

For every individual A, we can define the qualitative character of A as the set of individuals indiscernible from A. Qualitative properties supervene on the qualitative character of their bearers; if P is a qualitative property, then

[4] the problem mentioned at the end of the last section put aside

[5] The definition applies also to cases where the world of A is the world of B, also. E.g., A may live in a world of two-way eternal recurrence, and B may be A's qualitative doppelgaengerin some other epoch.

there are no two possible individuals which differ wrt. P without differing in qualitative character.

The second sense explicates the absence of purely numerical distinctions. It is adapted from a definition in Fine (1977), p. 174. The idea is that if, for a field of application, we can freely permute possible individuals (while keeping the pattern of PNRs of their world and their place in them fixed), then the identity of the individuals seems to be of no importance. The BIG THING is the mereological sum of all individuals. Let a PNR-*automorphism* be a PNR-isomorphism π of the BIG THING onto itself, such that $\pi|W$ is a permutation of the worlds.[6]
I.e. a PNR-automorphism is a bijection π from U onto U, such that

(i) if R is an n-place PNR, then for any $A_1, \dots, A_n$
$RA_1, \dots, A_n$ iff $R\pi(A_1), \dots \pi(A_n)$,

(ii) for every $A, B : A \leq B$ iff $\pi(A) \leq \pi(B)$,

(iii) for every w: w is a world if and only if $\pi(B)$ is.

Under the assumption that there are no cross-world PNRs, we can say alternatively that a PNR-automorphism is a permutation of U which is the union of PNR-isomorphisms. We will make this assumption in the following; we will also drop the last condition (being a world seems to supervene on the part-whole relation and on PNR).

P is a *qualitative property* iff there is no PNR-automorphism π and no possible individual A, such that $A \in P$ and $\pi(A) \notin P$.

Note that, given the above assumption, every PNR-isomorphism g may be extended to a PNR-automorphism. If domain and range of g are different, take $g \cup g^{-} \cup Id_C$, where $C := U - \text{Field}(g)$, take $g \cup Id_C$, otherwise.[7] Therefore, preservation under every automorphism implies preservation under every isomorphism. On the other hand, every automorphism is the union of isomorphisms. Therefore, preservation under every isomorphism implies preservation under every automorphism.

[6]The last condition will be redundant iff worlds are definable as maximally PNR-interrelated things.

[7]This is only true if we stick to the assumption that there are no crossworld PNRs.

The definitions may be generalized to the case of relations, too. But notice a certain limitation of the isomorphism strategy: it will not work properly for trans-world relations. Suppose A, B, A' and B' all live in different worlds, A and A' are indiscernible, and so are B and B'. Let R be a trans-world relation which relates A to B, but not A' to B'. R is certainly not qualitative. But how can preservation under PNR-isomorphisms be of any help here, if these connect only single worlds, i.e. either the world of A and the world of A' or the world of B and the world of B'? Should we say that no trans-world relation is qualitative? This would be just too much, and in the following we want to say things like e.g. similarity relations between worlds are qualitative. Thus, in order to keep the definition sufficiently general, we will follow the second strategy. The criterion is simply that a relation is qualitative iff it is preserved under arbitrary automorphisms:

> R is qualitative iff
> if $(A_1, \ldots, A_n) \in R$ and π is a PNR-automorphism, then
> $(\pi(A_1), \ldots, \pi(A_n)) \in R$.

6.8 Naturalness

Armstrong (1989b) starts his introduction to the theory of universals with the familiar type/token distinction. E.g., the word "the" is a type, if I write it down on a blackboard, I produce a token of this type. You could collect arbitrary tokens and than ask whether such a collection "marks off a type". E.g. the class of tokens of the word "the" marks off a type. Now,

> What distinguishes the classes of tokens that mark off a type from those classes that do not?
> ibid., p. 13

This is what Armstrong calls "the problem of universals".

In this section, I will ask whether a theory of universals should solve the problem of universals. Whether and to what extent ought a theory of properties give an account of the apparent naturalness of some sets of individuals as opposed to others (or, what comes to the same thing, the apparent unity or coherence of some sets).

Armstrong goes on to say that he thinks the distinction in question is an objective one.

> [I]t is a distinction founded on a difference in the things themselves and not, say, in some different attitude we take up to the different classes. That a certain class of tokens marks off a type is not something we determine. (Who are we to determine the nature of things?) Hence I will call the good classes the natural classes. (*ibid.*, p. 14)

So for Armstrong, the solution for his problem what marks off a type is that it is an objective fact that some classes are natural.

He points out that naturalness permits of degree: there are maximally unnatural sets (just the non-qualitative ones) and maximally natural ones:

> the unity scale for natural classes appears to have a top. If there is a plurality of things all exactly alike, perfect twins, then the class has the highest possible degree of unity. Armstrong (1989b), p. 24

but there are also degrees of naturalness in between.

Often, if a class can be construed as the union of various natural classes, then the subclasses are more natural, e.g. the set of crimson things is more natural than the set of red things. Some of the distinctions are more substantial. Then the distinctions are based on objective resemblances. Take e.g. the set RED of red things, the set BLUE of blue things and the set GREEN of green things. Green is more similar to blue than red; therefore, BLUE ∪ GREEN is more *similarity-unified* than BLUE ∪ RED. Now, more similarity-unification leads to more naturalness; therefore, BLUE ∪ GREEN is more natural than BLUE ∪ RED.[8]

There is a another possible answer to the problem what marks off a type. It is that the classes that do are the ones *that look natural to us.* Then the distinction is, in one sense of the word, subjective. While Armstrong calls classes "natural" if they reflect the nature of things, Anthony Quinton, who introduced the term "natural class" into the debate about universals, (see, Quinton (1973)), actually took them to be the ones that look natural to us. Armstrong's objection against such a view "Who are we to determine the nature of things?" seems to be question-begging because adherents of the view that sameness of type is due to psychological naturalness may refrain from claiming that such classes determine the nature of things. So, instead of

[8]Resemblances obtain in virtue of some kind of partial identity between universals, see e.g. citeArmstrong[78b], p. 126. He admits that it is not clear whether such an analysis can be applied to the present example, but we could substitute the case of masses or lengths, here.

dismissing the idea that psychological naturalness is the key to the so-called problem of universals, let us ask which theoretical alternative is preferable.

Is the notion of naturalness addressed in the above problem of universals psychological of metaphysical? The problem is that either we can rely on our intuitions about which classes are natural, then the psychological notion seems to be appropriate, or the purported naturalness of classes is a purely theoretical feature of our theory of properties, then we can't praise the theories for giving an account of it. So suppose, the naturalness addressed in the problem of universals really is of a psychological nature.

Still, metaphysics could help to account for it. Does metaphysical naturalness explain psychological naturalness? Are the classes that look natural to us simply the classes that are metaphysically natural?

Unfortunately, even Armstrong's central example of differences in naturalness (various colours) provides no reason to think so. The problem is that it is an example of comparative psychological naturalness, but that it is less clear whether it is an example of comparative physical naturalness (see Hardin (1988)). First, if one inspects the class of objects of one colour more closely, it becomes clear that the class is physically diverse. Hardin cites Nassau (1983) with a list of 15 different physical causes of colour. The classes of coloured objects even seem to be very unnatural and "gerrymandered", in that very similar objects belong to different classes, while very different objects belong to the same class. Of course, to point out the low degree of naturalness of the colours does not refute Armstrong's contention that BLUE $\cup$ GREEN is *more* natural than BLUE $\cup$ RED. Isn't it the case that the similarities of the colours correspond to their distance in the spectrum, so that this is a case of physical naturalness after all? But spectra and colours are less closely related than the average philosopher tends to think, and the class of objects of a colour is also diverse with respect to their spectra. Now, this finally leads Hardin to think that the resemblances of colours are largely due to the perceiver, not to the perceived object.

> Objects with identical spectra will, to be sure, look to be the same color, but indefinitely large numbers of objects that are spectrally different will also look to be the same color [...]. How do we generalize from the class of spectrally identical objects to the very much larger class of chromatically identical objects, and to the still larger class of chromatically similar objects? Is there any way of doing this if we must confine our attention to the physical properties of the objects that are

> seen or the patterns of light which enable us to see them? We shall come to understand that there is no such way for the resemblances and differences of colors are grounded not just in the physical properties of objects but even more in the biological makeup of the animal that perceives them. (Hardin (1988), p. 7)

But then the relative naturalness of BLUE $\cup$ GREEN compared with BLUE $\cup$ RED is only of a psychological nature. A metaphysical theory, which locates them in the objects themselves, is unlikely to make correct predictions about psychological naturalness. If comparative metaphysical naturalness is thus stripped of the examples that served to motivate it, it is not clear why we should have such a notion at all.

Of course, metaphysics might provide *constraints* on psychological naturalness. E.g. the property of being psychological natural should only apply to qualitative properties, and the class of things indiscernible to a given thing should possess the highest degree of psychological naturalness.

6.9 Minimal Realism

Up to now, we have assumed that, in order to define "intrinsic" and "qualitative", we need some notion of fundamental property, such as "universal" or "perfectly natural property". In this part, we will see whether we can work on a weaker basis. It turns out that in order to define "intrinsic" and "qualitative", we only need mereology and duplication. We will adapt our definitions a bit:

The *nature* of a thing is the set of its duplicates. The set of duplicates of A is the property of being exactly like A is by virtue of itself, and therefore rightly called A's nature.

Obviously, we can equally well define an intrinsic property to be one that never distinguishes between things with the same nature. And, for the purpose of the definition of "qualitative", we could assume that the set of PNRs is the set of natures. But since the set of natures is not a very economical repertoire of fundamental properties, I will be more cautious and redefine "qualitative". A *proper match* from A onto B is an isomorphism from A onto B wrt. to the set of all natures. (If f is a proper match from A onto B, then A and $f(A)$ are duplicates.) An *automorphism* is a permutation of the set of possibilia which is the union of proper matches of worlds onto worlds. Finally, an n-place relation is *qualitative* iff it is preserved under arbitrary automorphisms.

You may define "nature" in terms of duplication, like we did above. But you need not do so. You could also try to define it in some other way. But you can also stop here and take "nature" to be a primitive. I call a theory that does so "minimal realism".

The weakness of minimal realism compared with a theory of PNRs displays itself in the fact that many different such systems could be used to define the same facts of duplication. Indeed, it is easy to take a given system of PNRs and define a different one which is equivalent wrt. duplication. Take e.g. the PNRs of the second system to be the complements of the PNRs of the first system. Now, from the standpoint of their contribution to the definitions of "intrinsic" and "qualitative", it is completely arbitrary which system to take .

Now, every task mentioned in the introduction can already be accomplished by minimal realism: the definitions of "intrinsic" and "qualitative" and the constraints on psychological naturalness mentioned in section 7.

Minimal realism no longer implies that there are additional metaphysical respects in which a scientific theory can be correct (or incorrect), and may thus seem to be preferable. I will not discuss whether minimal realism suffices for all purposes a theory of universals has been useful for. There is additional "work for a theory of universals" (see Lewis (1983d)), some of which is at least *prima facie* not reducible to the notion of nature. E.g. the use of perfectly natural properties or universals in combinatorialist accounts of the plenitude of worlds, see Armstrong (1989a); Lewis (1986a). Or take the notion of an *alien* property in many debates in metaphysics.

Chapter 7

Predicates and Propositions

In this chapter we will apply the notion of a qualitative property to our semantics. Every now and then we have already touched upon this notion in connection with counterpart semantics, now this connection will take center stage.

We will show how to apply the notion 'qualitative' beyond the realm of properties and relations, namely to predicates and propositions. We will define kinds of predicates and propositions in terms of this notion. Chap.5 contains a principle, in terms of the notion of a qualitative relation, about what kinds of propositions there are; we will show that certain natural assumptions about the counterpart relation and about the question what kinds of predicates there are allow us to show that, within our semantics, the principle holds.

7.1 Overview

Certain properties of things that represent are irrelevant for the content of the representation. E.g. whether p is true according to a book does not depend on whether the letters are roman or *sans serif*.

Our representations are *direct*, representees are represented to be exactly how their representatives themselves are. On p.89 above, I mention the following characteristic of a direct representation: "Whether a proposition is true according to a direct representation only depends on what things are represented in which ways. The identity of the representatives is never relevant for the question what is true according to a representation."

The last sentence states a general constraint on the notion of representation: it is never important for a representation who the representatives are. It is only important how they are or what they do. If they are to be representations at all, this applies to direct representations, too. It implies the following principle, which, therefore, should hold as well.

(1) Propositions do not distinguish between representations that represent the same things the same way.

In this book I have put forward a recursive definition of the notion of truth according to a (maximal complete) direct representation. The question is whether this definition conforms to (1). For the purpose of doing so we may restrict our attention to the question whether it holds of the semantics of QML. If it does this will also prove that it holds for the semantics of English^{-}, the part of English we interpret by translating it into QML.

Fortunately it does hold for our semantics of QML, for given the semantics in Chap.5 we can derive (1) from two assumptions (2) and (3) about our semantics we already made.

(2) Every proposition may be analysed in terms of names and qualitative predicates (Chap.1, p.12)

(3) The counterpart relation is qualitative (Chap.1, p.11)

My thesis is, in a nutshell, that whether a thing (or a sequence of things) falls under a predicate relative to a representation never depends on the identity of representatives, and that therefore, propositions never distinguish between representations that represent the same representees as being *in precisely the same way.*

We will sketch a proof. The main work consists in making (2) , (3) and (1) precise.

7.2 Properties of Predicates

In our semantics, whether a thing falls under a predicate is relative to arbitrary maximal representation relations. Thus, whether a thing (or a sequence of things) falls under the predicate relative to a representation, may be dependent (or independent) on a couple of things.

E.g. on how the thing is represented, i.e. on the qualitative character of the representative. Or on who is represented, i.e. on the identity of the thing itself. Or on how particular other things are represented. And there is, at least technically, a fourth possibility, namely dependence on the identity of the representative.

Depending on the predicate, various of the above features may be important. Below we will define a couple of *metaproperties*, properties of properties, that distinguish predicates in terms of the above list of features. We will define what it means for a predicate to be *qualitative*, to be *pure*, and to be *quasi-qualitative* (my term). I think the definitions are of independent interest (at least the distinctions between pure and impure, qualitative and non-qualitative play important roles in certain discussions in metaphysics).

But my real aim in providing these definitions is different. I want to be able to say generally what kind of predicates occur at all, and derive conclusions about what kind of propositions occur at all.

The metaproperties are now defined in terms of the following *supervenience* principles, adapted from similar things in Fine (1977).

Recall that a PNR-automorphism is a 1-1 function from *Posse* onto itself, such that the part-whole relation and perfectly natural properties and relations are preserved.

Def. qualitative P is qualitative iff it holds that if f is an automorphism and $\pi \circ \sigma = f \circ \rho \circ \sigma'$, then $\sigma \in V_\pi(P)$ iff $\sigma' \in V_\rho(P)$,

Def. pure P is pure iff it holds that if $\pi \circ \sigma = \rho \circ \sigma'$, then $\sigma \in V_\pi(P)$ iff $\sigma' \in V_\rho(P)$,

Def. quasi-qualitative P is quasiqualitative iff it holds that if f is an automorphism, then $\sigma \in V_\pi(P)$ iff $\sigma \in V_{f\circ\pi}(P)$.

An example of a qualitative predicate is "red"; $a \in V_\pi(red)$ iff $\pi(a)$ is red. Now if $\pi(b)$ and $\rho(b')$ are qualitatively indiscernible, both have the same colour; in such a case $b \in V_\pi(red)$ iff $b' \in V_\rho(red)$. "red " is also pure.

An example of an impure predicate is e.g.$= a$ ("being identical to a"). It holds that $b \in V_\pi(= a)$ iff $\pi(b) = \pi(a)$. Now it might be the case that $\pi(b) = \rho(c)$, but $b = a$, while $c \neq a$, and so $\pi(b) = \pi(a)$, but $\rho(c) \neq \rho(a)$, i.e. $b \in V_\pi(= a)$, while $c \notin V_\rho(= a)$. Another example is "being the father of Tom". These predicates are also not qualitative.

But all of the above predicates are quasi-qualitative. E.g. $= a$ is quasi-qualitative because for every automorphism f, $b \in V_\pi(= a)$ iff $\pi(b) = \pi(a)$ iff $f\pi(b) = f\pi(a)$.

The following things are easily seen.

1. If P is qualitative, then it is quasi-qualitative,
2. If P is qualitative, then it is pure,
3. If P is pure and quasi-qualitative, then it is qualitative.

7.3 Properties of Propositions

For the purpose of defining the relevant metaproperties, propositions (sets of permutations) may now be understood as 0-place predicates; functions from permutations into sets of 0-place sequences. This yields the following extensions of the above notions.

Def. qualitative p is qualitative iff it holds that $\pi \in p$ iff $\rho \in p$,

Def. pure p is pure iff it holds that $\pi \in p$ iff $\rho \in p$,

Def. quasi-qualitative p is quasi-qualitative iff it holds that if f is an automorphism, then $\pi \in p$ iff $f \circ \pi \in p$.

It turns out that

1. The pure propositions are exactly the qualitative ones.
2. The set of all permutations is pure (qualitative),
3. so is the empty set,
4. but no other set of permutations.

Again, pureness and qualitativeness are equivalent; in the case of propositions we do not even need the further assumption of quasi-qualitativeness. It may come as a suprise that only the necessarily true proposition and the necessarily false one turn out to be pure. There certainly seem to be propositions that are both pure and contingent, like the proposition that there are talking donkeys. But note that I construe this proposition as impure, namely as $\exists x(Ix@\&Tx)$ (there is a talking donkey in this world).

7.4 Qualitativeness and C

Let us finally explore how C fares wrt. qualitativeness. All of the following constraints on C should hold.

Anti-Haecceitism If $\pi \in C$ and f an automorphism, then $f \circ \pi \in C$

Supervenience of Essence If $\pi \in C$ and f is an automorphism, then $\pi \circ f \in C$

Representation is Qualitative If $\pi \in C$ and f an automorphism, ρ is that function, such that $\rho f(a) = f(b)$ iff $\pi a = b$, then $\rho \in C$

Anti-Haecceitism expresses, that, for every A, being a counterpart of A is a qualitative property. Supervenience of Essence expresses that being something that has B as its counterpart is qualitative property, likewise. And Representation is Qualitative bears its name, because the basic idea in our permutation semantics is that the sequence $\pi a, \pi b, \ldots$ represents the sequence $a, b, \ldots$ in a way the latter sequence might be. If we state, in terms of our general definition of a qualitative relation, that this relation is qualitative, then we arrive at the above principle. It follows from Anti-Haecceitism and Supervenience of Essence. I have argued for the ideas expressed by the above principles in Sec.1.3.4.

7.5 The Theorem

We may now make (1) precise:

(4) Every proposition expressible by the language is quasi-qualitative.

We can prove (4) if we assume Representation is Qualitative, and that every predicate is quasi-qualitative. With these assumptions we may prove first the following Lemma.

(5) Every open formula of the language defines a quasi-qualitative predicate, i.e. for every formula ϕ, every automorphism f and every assignment g: if $\pi \models \Diamond\phi[g]$, then $f \circ \pi \models \Diamond\phi[g]$

The proof proceeds by a straightforward induction. (4) follows immediately.

The significance of that theorem is the following. In Chap.1 we have argued that the counterpart relation is qualitative. In the same chapter we had to assume that impure predicates could be analysed away in terms of qualitative ones and names. So, every expressible proposition could be expressed in terms of qualitative predicates and names. It does not affect the range of expressible propositions if we assume that every predicate is qualitative. But then, it follows from an assumption we have already made, that we are justified in assuming that every predicate is quasi-qualitative. So, both our premises are justified, and hence the conclusion (4) above.

7.6 A Truthmaking Principle

(1), the principle that propositions do not distinguish between representation that represent the same things the same way bears a certain likeness to the following truthmaking principle, put forward by David Lewis.

> I hold, as an a priori principle, that every contingent truth must be made true, somehow, by the pattern of coinstantiation of fundamental properties and relations. The whole truth about the world [...] supervenes on this pattern. If two possible worlds were exactly isomorphic in their patterns of coinstantiation of fundamental properties and relations, they would thereby be exactly alike *simpliciter*. (citeLewis[94a], p. 412)

> To begin, we may be certain *a priori* that any contingent truth whatever is made true, somehow, by the pattern of instantiation of fundamental properties and relations by particular things. In Bigelow's phrase, truth is supervenient on being [...]. If two possible worlds are discernible in any way at all, it must be because they differ in what things are in them, or in how those things are. And "how things are" is fully given by the fundamental, perfectly natural, properties and relations that those things instantiate. (Lewis (1994a), p. 473f)

Lewis' principles are meant to provide a less contentious replacement of an idea made popular by David Armstrong, namely that for every truth there is something that makes it true. Despite its initial plausibility this truthmaker principle, Martin (1989); Armstrong (2004); Fox (1987) could at least be read to be ontologically inflationary, because it seems to demand that if A is P, then there is a third thing, the truthmaking fact that A is P, see Lewis (1999). Lewis tries to replace the truthmaker principle by the weaker principles quoted above. The second states that every contingent truth supervenes on what things there are and on how those things are.

This is at least very similar to our principle (1), which states that truth supervenes on how things are represented.

Appendix A

Counterpart Frames

I will now sketch a formal semantics for the language of quantified modal logic based on the suggestions in Chap.2. It employs counterparts of pairs (and of n-place sequences in general) and unrestricted quantifiers. For reasons of simplicity we will only consider a language without individual constants.

For every set B and $n > 0$, an n-place *sequence* over B is a function from the set $\{1, \ldots, n\}$ into B. $\emptyset$ is the 0-place sequence over B, and B^n is the set of n-place sequences over B. If σ is an n-place sequence and $1 \leq i \leq n$, we address $\sigma(i)$ as σ_i and, sometimes, σ as $(\sigma_1, \ldots, \sigma_n)$. If $n > 0$, σ is an n-place sequence over B, then τ is an m-place *transformation* of σ iff τ is an m-place sequence over $1, \ldots, n$. In that case $\sigma\tau = \sigma \circ \tau$ is an m-place sequence over B. E.g. if $\sigma = (a, b)$ and $\tau = (2, 1)$, then $\sigma\tau = (b, a)$. For every n-place sequence of variables $\sigma = (v_1, \ldots, v_n)$ and assignment f, $f\sigma = f(v_1, \ldots, v_n) = (f(v_1), \ldots, f(v_n))$, so $f\sigma$ is the sequence of values of the variables in σ wrt. to the assignment f.
A *(counterpart-)frame*[1] is a pair $(Posse, (C_n)_{n \in \mathbb{N}^+})$, such that $Posse$ is a non-empty set, for each $n \in \mathbb{N}^+$: $C_n \subseteq (Posse^n)^2$, and the following conditions hold for arbitrary $\sigma, \pi \in Posse^n$, $\rho \in Posse^m$, and m-place transformations τ of σ. We write $\pi C_n \sigma$ (π is a counterpart of σ) to express that $(\pi, \sigma) \in C_n$.

C.1 If $\pi C_n \sigma$, then $\pi\tau C_m \sigma\tau$

C.2 If $\rho C_m \sigma\tau$, then there is a π, s.t. $\rho = \pi\tau$ and $\pi C_n \sigma$

It follows from C.1 that if (c, d) is a counterpart of (a, b), then (d, c) is a counterpart of (a, b) (and *vice versa*). In terms of the correspondence

[1] counterpart frames are special cases of *metaframes* in the sense of Gabbay et al. (2007)

between counterparts and ways described in Chap.1: if (c, d) is in a maximally specific way in which (a, b) could be, then (d, c) is in a maximally specific way in which (b, a) could be. Another instance of C.1 is that if (c, d) is a counterpart of (a, b), then c is a counterpart of a and d is a counterpart of b, again this sounds correct in terms of the correspondence between possible counterparts and ways. The reverse, however, is not generally the case: if c is a counterpart of a and d is a counterpart of b we cannot conclude that (c, d) is a counterpart of (a, b). We may conclude from C.2, however that

(*) if c is a counterpart of a, then, for every b there is *some* d, such that (c, d) is a counterpart of (a, b).

(*) says that for every pair of possible individuals a and b and every maximally specific way W a might be in, there is some maximally specific way R, a and b might jointly be in, such that being W implies being R-related to some c. R does not require its relata to be worldmates. (*) simply ensures that the possibilities for a do not vanish, once we start talking about b; again this seems to be eminently plausible to me.

A *model* based on a frame $(Posse, (C_n)_{n \in \mathbb{N}^+})$ is a triple $(Posse, (C_n)_{n \in \mathbb{N}^+}, V)$ where V is a function s.t. for or every n-place predicate P: $V(P) \subseteq Posse^n$.

The following recursive semantics is just the semantics of first-order logic, with an additional clause for $\Diamond$. This clause uses the notion of a sequence of occurrences of free variables in a formula φ, $\mathrm{Fr}(\varphi)$.[2] Of course, $\mathrm{Fr}(\varphi)$ may be empty.

---

[2] We may define this more rigorously. First, we define, for any two sequences σ and τ, $\sigma_\sqcup\tau$ (τ appended to σ): $\sigma_\sqcup\tau := \sigma$, if $\tau = \emptyset$, else $\sigma_\sqcup\tau := \sigma + \tau_1, \ldots, \tau_n$, where $\sigma + u$ is that sequence that results from appending u at the end of σ.
Second, we define the deletion of all occurrences of v from a sequence σ: $\emptyset - v = \emptyset$, $(u) - v := (u)$, if $v \neq u$, $= \emptyset$, else; if $\sigma = \tau_\sqcup(u)$, then $\sigma - v := \tau - v_\sqcup(u) - v$.
Finally, $\mathrm{Fr}(\varphi)$ is defined in the following way:
$\mathrm{Fr}(Rv_1, \ldots, v_n) = (v_1, \ldots, v_n)$;
$\mathrm{Fr}(\neg\varphi) = \mathrm{Fr}(\varphi)$;
$\mathrm{Fr}(\varphi \wedge \psi) = \mathrm{Fr}(\varphi) \cup \mathrm{Fr}(\psi)$;
$\mathrm{Fr}(\exists v\varphi) = \mathrm{Fr}(\varphi) - v$;
$\mathrm{Fr}(\Diamond\varphi) = \mathrm{Fr}(\varphi)$.

Consider some arbitrary but fixed model $(Posse, (C_n)_{n\in\mathbb{N}^+}, V)$. If g is an assignment, then $g[a/v]$ is that assignment that is exactly like g, only that $g(v) = a$. Let us define $g \models \varphi$ (formula φ is true wrt. an assignment g) as follows.

1. $g \models P(v_1, \ldots, v_n)$ iff $(g(v_1), \ldots, g(v_n)) \in V(P)$,
2. $g \models \neg\varphi$ iff $g \not\models \varphi$,
3. $g \models \varphi \wedge \psi$ iff $g \models \varphi$ and $g \models \psi$,
4. $g \models \exists v\varphi$ iff $g[a/v] \models \varphi$ for some $a \in Posse$,
5. $g \models \Diamond\varphi$ iff

$$f \models \varphi \text{ for some } f\text{, s.t. if } \mathrm{Fr}(\varphi) = (v_1, \ldots, v_n), \text{ then } f(v_1, \ldots, v_n) C_n g.(v_1, \ldots, v_n)$$

(i.e. $\Diamond\varphi$ is true wrt. to assignment g iff φ is true wrt. some assignment f, s.t. if there are $n > 0$ occurrences of free variables in φ, then the sequence of values of these occurrences wrt. f is a counterpart of the sequence of their values wrt. g. This implies that if φ does not contain any free variables (φ is *closed*), then $g \models \Diamond\varphi$ iff $f \models \varphi$ for some f. Because φ is closed the latter holds exactly if $g \models \varphi$.)

The usual definitions of further connectives and operators apply, e.g.
$\varphi \rightarrow \psi$ is defined to be $\neg(\varphi \wedge \neg\psi)$, $\Diamond\varphi$ to be $\neg\Box\neg\varphi$.
A formula φ is *true* wrt. an assignment g iff $g \models \varphi$. φ is *valid* in a model $\mathcal{M}$ iff it is true wrt. every assignment in $\mathcal{M}$, it is valid on a class of frames $\mathcal{F}$ iff it is valid in every model based on some frame in $\mathcal{F}$, it is valid iff it is valid on the class of all frames. ($\mathbf{T}_c$) $\Diamond\varphi \rightarrow \varphi$ is valid for closed φ. (**K**) $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ and (**BFC**) $\Box\forall v\varphi \rightarrow \forall v\Box\varphi$ are valid. The proofs for the last two theorems use our conditions C.1 and C.2 above. (**T**)$\Box\varphi \rightarrow \varphi$, (**B**)$\varphi \rightarrow \Box\Diamond\varphi$, (**4**)$\Box\varphi \rightarrow \Box\Box\varphi$ are valid on the classes of reflexive, symmetrical, and transitive frames, resp.

A rule is validity-preserving in a model iff valid premises invariably yield valid conclusions, and validity-preserving on a class of frames iff it is validity-preserving in every model on some frame of that class; it is validity-preserving iff it is validity-preserving on the class of all frames. E.g. *modus ponens* and (**N**) $\varphi \vdash \Box\varphi$ are validity-preserving.

Appendix B

Internal models

I will now show that the language of QML, under the induced interpretation, cannot properly express trans-world relations. A model of CT is an *internal* model of QML precisely if it interprets no predicate of QML by a trans-world relation. For any model $\mathcal{M}$ of CT you might form its internal counterpart $\mathcal{M}^*$ as follows: $\mathcal{M}^*$ is precisely like $\mathcal{M}$, only that for any n-place predicate P of QML, $(a_1, \ldots, a_n) \in V^*(P)$ iff $(a_1, \ldots, a_n) \in V(P)$ *and* there is some world w, such that $(a_1, w) \in V^*(I), \ldots, (a_n, w) \in V^*(I)$. Now it turns out that as models of QML, $\mathcal{M}$ and $\mathcal{M}^*$ are elementarily equivalent, i.e.

Theorem For all models $\mathcal{M}$ of CT, assignments g, and all closed formulae ϕ: $\mathcal{M}, g \models \phi^w$ iff $\mathcal{M}^*, g \models \phi^w$.

This is a direct consequence of the Lemma below. If $\mathcal{M} = (D, V)$ is a model of CT, an assignment is *internal* (wrt. $\mathcal{M}$ and w) iff for all $v \neq w : g(v)$ is in $g(w)$ $((g(v), g(w)) \in V(I))$.

Lemma For all models $\mathcal{M}$ of CT, internal assignments g, and all formulae ϕ: $\mathcal{M}, g \models \phi^w$ iff $\mathcal{M}^*, g \models \phi^w$.

The proof is a straightforward induction. The atomic cases rest on the assumption that QML contains no individual constants. It would also do to assume that QML contains no *names of non-actuals*.

Appendix C

Counterparts and Permutations

Now I will argue, that the counterpart relation may be encoded as a family of permutations of the set of possible individuals C, such that quantification over counterparts can indeed be replaced by quantifation over members of C. I will argue that there is a set C of permutations of the set of possible individuals such that

(1) If $(B_1, \dots, B_n)$ is a counterpart of $(A_1, \dots, A_n)$, then there is a $\pi \in C$, such that $(\pi A_1, \dots, \pi A_n) = (B_1, \dots, B_n)$ (i.e. (8) on p.93) and

(2) if there is a $\pi \in C$, such that $(\pi A_1, \dots, \pi A_n) = (B_1, \dots, B_n)$, then $(B_1, \dots, B_n)$ is a counterpart of $(A_1, \dots, A_n)$ (i.e. (9) on p.93)

Counterpart Functions

First let me argue that a counterpart relation could be conceived of as a set of certain functions, which I will call a counterpart functions. These are partial functions from some set of possible individuals onto another.

First we will assume the following.

(3) If (C, D) is a counterpart of (A, B), then if $A = B$, then $C = D$

(3) expresses the necessity of identity: given (3) it implies that to be a non-identity pair is not a way an identity pair might be. For the finite case, (3) follows from constraint C.2 on counterpart relations assumed in Appendix A.

(3) implies that if If (C, D) is a counterpart of (A, B), then there is a function f, such that $C = fA$ and $D = fB$. (3) could and should be

generalised to counterparts of sequences of arbitrary length, then it implies that for every case where an n-place sequence σ is a counterpart of a sequence τ, there is a function f such that $\sigma_i = f\tau_i$ (for all $i \leq n$), i.e. every pair of a sequence and its counterpart gives rise to a function.

Now different such pairs might give rise to the same function. Fortunately, the counterpart relation does not distinguish between pairs of sequences that give rise to the same counterpart function.

First repetitions do not matter: if (A, A) is a counterpart of (B, B) exacly iff A is a counterpart of B (again this can be generalised). Second, order does not matter: If (C, D) is a counterpart of (A, B), then (D, C) is a counterpart of (B, A). Generally, if σ and τ are an n-place sequences, τ' is a reordering of τ, σ' is σ, reordered in the same way and σ is a counterpart of τ, then σ' is a counterpart of τ' as well. Counterpart relations are invariant under permutations of relata sequences, provided both relata a permuted in the same way.[1]

Summing up: instead of conceiving the counterpart relation as a relation between sequences, we may as well conceive of it as a set of functions. E.g. instead of saying that (C, D) is a counterpart of (A, B) and (D, C) is a counterpart of (B, A), we may say that that function from $\{A, B\}$ onto $\{C, D\}$ that maps A onto C and B onto D is a counterpart function.

We define a *counterpart function* to be a function f such that if σ is a sequence of all elements of f's domain, then $f\sigma$ is a counterpart of σ, where $f\sigma$ is that sequence of members of f's domain such that for all i in σ's index set, $(f\sigma)_i = f(\sigma_i)$.

Counterpart functions need not necessariliy be finite, just as counterpart relations may also hold between infinite sequences. (An infinite sequence over a set A is a function from some index set into A. Of course index sets may be infinite.) Of course our languages are finite, and so are observable cases of the multiple de re, in consequence. But we may also want to talk about de re modal properties of sets, where a counterpart of a set is defined in terms of a counterpart of (the sequence of) all of its members; and here we will soon need infinite counterpart relations.

[1]For every relation R there is a converse relation R_c, e.g. buying is the converse of selling and right-of is the converse of left-of. Sometimes relations are their own converse, like the relation of being a twin. Now it is very clear that if a pair (A, B) is possibly R-related, then (B, A) is possibly possibly R_c-related. For an illustration suppose that, that Dee is possibly right of Dum, but Dum is not possibly left of Dee.

Maximal Complete Counterpart Functions

Now I am able to show that the set of counterpart functions includes a set of permutations. This set can indeed be employed to do all the duties of the counterpart relation.

Above we have collected three important properties of the counterpart relation. There are some additional properties of the counterpart relation that will play a rôle in the following, all of which may easily be expressed in terms of counterpart functions. For every arbitrary counterpart function f with domain A and range B:

(4) f is 1-1

(5) for every a and b, if $fa = b$, then $f - \{(a, b)\}$ is a counterpart function.

(6) for every a: there is a counterpart function g, such that $g \subseteq f$ and ga is defined.

(7) $|\overline{A}| = |\overline{B}|$ (i.e. the complement set of the domain of f and the complement of the domain of f have the same cardinality.

(4) ensures the necessity of non-identity. Together with (3) this has the effect that a set could not have a different cardinality. Together with (3) it has the effect that two representatives are identical exactly if their representees are, a constraint we defended on p.43.

(5) is motivated by the fact that if we do not mention parts of a possibility, the possibility does not go away. E.g. if it is possible for A to be P and for B to be Q, then it is possible for A to be P. For the finite case, (5) follows from constraint C.1 in Appendix A above.

(6) accounts for the fact that if we simply introduce new individuals, what it is possible for the old individuals remains intact, e.g. if it is possible for A to be P, then it is possible for A and B that A is P. For the finite case, (6) follows from C.2 above.

(7) ensures also that a set could not have a complement set of a different cardinality. This is also something that holds wrt. counterpart frames as defined in Appendix A above.

Let C be the set of those counterpart functions which are permutations of the set of possible individuals. (5) implies that for every counterpart function f and all functions $g \subseteq f$, g is a counterpart function, and hence that

(2) if there is a $\pi \in C$, such that $(\pi A_1, \ldots, \pi A_n) = (B_1, \ldots, B_n)$, then $(B_1, \ldots, B_n)$ is a counterpart of $(A_1, \ldots, A_n)$

(6), (4), and (7), on the other hand, imply

(1) If $(B_1, \ldots, B_n)$ is a counterpart of $(A_1, \ldots, A_n)$, then there is a $\pi \in C$, such that $(\pi A_1, \ldots, \pi A_n) = (B_1, \ldots, B_n)$

qed.

Bibliography

Armstrong, D. (1978). *A Theory of Universals*. Cambridge: Cambridge University Press. Vol. 2 of Universals and Scientific Realism.

Armstrong, D. M. (1989a). *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.

Armstrong, D. M. (1989b). *Universals. An opinionated introduction*. Boulder, Colorado: Westview Press.

Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Armstrong, D. M. (2004). *Truth and Truthmakers*. Cambridge: Cambridge University Press.

Barwise, J. and R. Cooper (1981). Linguistics and philosophy; generalized quantifiers and natural language. *Linguistics and Philosophy 4*(2), 159–219.

Bauer, S. and H. Wansing (2002). Consequence, counterparts and substitution. *The Monist 85*, 483–497.

Chalmers, D. J. (2006). The Foundations of Two-Dimensional Semantics. In M. Garcia-Carpintero and J. Macia (Eds.), *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.

Chandler, H. S. (1975). Rigid designation. *Journal of Philosophy 72*, 363–369.

Chisholm, R. M. (1973). Parts as essential to their wholes. *The Review of Metaphysics 26*, 581–603.

Corsi, G. and S. Ghilardi (1992). Semantical aspects of quantified modal logic. In C. Bicchieri and M. Dalla Chiara (Eds.), *Knowledge, Belief, and Strategic Interaction*, pp. 167–195. Cambridge: Cambridge University Press.

Davies, M. and L. Humberstone (1980). Two notions of necessity. *Philosophical Studies 38*, 1–30.

Divers, J. (2002). *Possible Worlds*. London and New York: Routledge.

Divers, J. and J. Melia (2002). The analytic limit of genuine modal realism. *Mind 111*, 15–35.

Dunn, J. M. (1990). Relevant predication 2, intrinsic properties and internal relations. *Philosophical Studies 60*, 177–206.

Fara, M. and T. Williamson (2005). Counterparts and actuality. *Mind 114*(453), 1–30.

Feldman, F. (1971). Counterparts. *Journal of Philosophy 68*, 406–9.

Fine, K. (1977). Propositions, properties and sets. *Journal of Philosophical Logic 6*, 135–191.

Fine, K. (1978). Model theory for modal logic I. *Journal of Philosophical Logic 7*, 125–156.

Forbes, G. (1982). Canonical counterpart theory. *Analysis 42*, 33–37.

Forbes, G. (1984). Two solutions to Chisholm's paradox. *Philosophical Studies 46*(2), 171–187.

Forbes, G. (1987). Free and classical counterparts: response to Lewis. *Analysis 47*, 147–152.

Fox, J. F. (1987). Truthmaker. *Australasian Journal of Philosophy 65*, 188–207.

Gabbay, D., D. Skvortsov, and V. Shehtman (2007). *Quantification in Nonclassical Logic (Studies in Logic and the Foundations of Mathematics)*. New York, NY, USA: Elsevier Science Inc.

Ghilardi, S. (1991). Incompleteness results in Kripke semantics. *The Journal of Symbolic Logic 56*(2), 517–538.

Haas-Spohn, U. (1995). *Versteckte Indexikalität und subjektive Bedeutung.* Berlin: Akademie Verlag.

Hardin, C. L. (1988). *Color for Philosophers.* Indianapolis: Hackett Pub. Co.

Hazen, A. P. (1976). Expressive incompleteness in modal logic. *The Journal of Philosophical Logic 5*, 25–46.

Hazen, A. P. (1977). *The Foundations of Modal Logic.* Ph. D. thesis, University of Pittsburgh.

Hazen, A. P. (1979). Counterpart-theoretic semantics for modal logic. *Journal of Philosophy 76*, 319–338.

Heim, I. (1983). On the projection problem for presuppositions. In W. Barlow, Flickinger (Ed.), *Proceedings of WCCFL*, Stanford, pp. 114–125.

Hintikka, J. (1969). Semantics for propositional attitudes. In J. Davis, D. Hockney, and W. Wilson (Eds.), *Philosophical Logic*, pp. 21–45. Dordrecht: Reidel.

Hodes, H. T. (1984). Axioms for actuality. *Journal of Philosophical Logic 13*(1), 27–34.

Hughes, G. E. and M. J. Cresswell (1996). *A new introduction to modal logic.* London: Routledge.

Kratzer, A. (1977, 01). What 'must'and 'can'must and can mean. *Linguistics and Philosophy 1*(3), 337–355.

Kratzer, A. (1978). *Semantik Der Rede.* Königstein/Taunus: Scriptor.

Kratzer, A. (1981). The notional category of modality. In H. J. E. . H. Rieser (Ed.), *Words, Worlds, and Contexts.* Berlin: de Gruyter.

Kratzer, A. (1989). An investigation of the lumps of thought. *Linguistics and Philosophy 12*, 607–653.

Kratzer, A. (1991). Modality. In A. v. Stechow and D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research*, pp. 639–650. Berlin: de Gruyter.

Kripke, S. (1972). Naming and necessity. In D. Davidson and G. Harman (Eds.), *Semantics of Natural Language*. Boston: Reidel.

Kripke, S. A. (1963). Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 9*, 67–96.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Lewis, D. (1983a). Attitudes de dicto and de se. See Lewis (1983e), pp. 133–160.

Lewis, D. (1983b). Counterpart theory and quantified modal logic. See Lewis (1983e), pp. 26–46.

Lewis, D. (1983c). Counterparts of persons and their bodies. See Lewis (1983e).

Lewis, D. (1983d). New work for a theory of universals. *Australasian Journal of Philosophy 61*, 343–77.

Lewis, D. (1983e). *Philosophical Papers*, Volume 1. Oxford: Oxford University Press.

Lewis, D. (1986a). *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. (1986b). *Philosophical Papers*, Volume 2. Oxford University Press.

Lewis, D. (1994a). Humean supervenience debugged. *Mind 103*, 473–490.

Lewis, D. (1994b). *Reduction of Mind*, pp. 412–431. Oxford: Blackwell.

Lewis, D. K. (1968). Counterpart theory and quantified modal logic. *Journal of Philosophy 65*, 113–26.

Lewis, D. K. (1999). A world of truthmakers? In *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, D. K. (2003). Things *qua* truthmakers. In H. Lillehammer and G. R. Pereyra (Eds.), *Real Metaphysics, Essays in Honour of D.H. Mellor*. London: Routledge.

Linsky, B. and E. Zalta (1994). In defense of the simplest quantified modal logic. *Philosophical Perspectives 8*, 431–458.

Martin, C. (1989). Counterfactuals, causality, and conditionals. In J. Heil (Ed.), *Cause, Mind and Reality: Essays Honouring C.B. Martin*. Dordrecht: Kluwer.

Melia, J. (2003). *Modality*. Chesham: Acumen.

Montague, R. (1974). *Formal philosophy; selected papers of Richard Montague*. New Haven: Yale University Press.

Montanarelli, L. and T. J. O'Gorman (2005). *Strange but true: Chicago: tales of the Windy City* (1st ed ed.). Guilford, CT: Globe Pequot Press.

Nassau, K. (1983). *The physics and chemistry of color: the fifteen causes of color*. New York: Wiley.

Putnam, H. (1975). The meaning of "meaning". In K. Gunderson (Ed.), *Language, Mind and Knowledge*, pp. 215–71. Minneapolis: University of Minnesota Press.

Quine, W. V. (1976). *The ways of paradox, and other essays* (Rev. and enl. ed ed.). Cambridge, Mass.: Harvard University Press.

Quinton, A. (1973). *The nature of things*. London: Routledge and Kegan Paul.

Ramachandran, M. (1989). An alternative translation scheme for counterpart theory. *Analysis 49*, 131–41.

Ramachandran, M. (1990). Contingent identity in counterpart theory. *Analysis 50*, 163–66.

Ramachandran, M. (1998). Sortal modal logic and counterpart theory. *Australasian Journal of Philosophy 76*, 553–65.

Sider, T. (1996). Intrinsic properties. *Philosophical Studies 83*, 1–27.

Sider, T. (2008). Beyond the Humphrey objection. http://tedsider.org/papers/counterpart$_t$heory.pdf.

Skvortsov, D. and V. Shehtman (1993). Maximal Kripke-type semantics for modal and superintuitionisic predicate logics. *Annals of Pure and Applied Logic 63*, 69–101.

Stalnaker, R. (1978). Assertion. *Syntax and Semantics: Pragmatics 9*.

Stalnaker, R. (1981). A theory of conditionals. In *Ifs: Conditionals, belief, decision, chance, and time*. Dordrecht: Reidel.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Routledge.

Index