1

A Philosophers' Paradise

1.1 The Thesis of Plurality of Worlds

The world we live in is a very inclusive thing. Every stick and every stone you have ever seen is part of it. And so are you and I. And so are the planet Earth, the solar system, the entire Milky Way, the remote galaxies we see through telescopes, and (if there are such things) all the bits of empty space between the stars and galaxies. There is nothing so far away from us as not to be part of our world. Anything at any distance at all is to be included. Likewise the world is inclusive in time. No long-gone ancient Romans, no long-gone pterodactyls, no long-gone primordial clouds of plasma are too far in the past, nor are the dead dark stars too far in the future, to be part of this same world. Maybe, as I myself think, the world is a big physical object; or maybe some parts of it are entelechies or spirits or auras or deities or other things unknown to physics. But nothing is so alien in kind as not to be part of our world, provided only that it does exist at some distance and direction from here, or at some time before or after or simultaneous with now.

The way things are, at its most inclusive, means the way this entire world is. But things might have been different, in ever so many ways. This book of mine might have been finished on schedule. Or, had I not been such a commonsensical chap, I might be defending not only a plurality of possible worlds, but also a plurality of impossible worlds, whereof you speak truly by contradicting yourself. Or I might not have existed at all – neither I myself, nor any counterpart of me. Or there might never have been any people. Or the physical constants might have had somewhat different values, incompatible with the emergence of life. Or there might have been altogether different laws of nature; and instead of electrons and quarks, there might have been alien particles, without charge or mass or spin but with alien physical properties that nothing

in this world shares. There are ever so many ways that a world might be; and one of these many ways is the way that this world is.

Are there other worlds that are other ways? I say there are. I advocate a thesis of plurality of worlds, or *modal realism*,[1] which holds that our world is but one world among many. There are countless other worlds, other very inclusive things. Our world consists of us and all our surroundings, however remote in time and space; just as it is one big thing having lesser things as parts, so likewise do other worlds have lesser otherworldly things as parts. The worlds are something like remote planets; except that most of them are much bigger than mere planets, and they are not remote. Neither are they nearby. They are not at any spatial distance whatever from here. They are not far in the past or future, nor for that matter near; they are not at any temporal distance whatever from now. They are isolated: there are no spatiotemporal relations at all between things that belong to different worlds. Nor does anything that happens at one world cause anything to happen at another. Nor do they overlap; they have no parts in common, with the exception, perhaps, of immanent universals exercising their characteristic privilege of repeated occurrence.

The worlds are many and varied. There are enough of them to afford worlds where (roughly speaking) I finish on schedule, or I write on behalf of *impossibilia*, or I do not exist, or there are no people at all, or the physical constants do not permit life, or totally different laws govern the doings of alien particles with alien properties. There are so many other worlds, in fact, that absolutely *every* way that a world could possibly be is a way that some world *is*. And as with worlds, so it is with parts of worlds. There are ever so many ways that a part of a world could be; and so many and so varied are the other worlds that absolutely every way that a part of a world could possibly be is a way that some part of some world is.

The other worlds are of a kind with this world of ours. To be sure, there are differences of kind between things that are parts of different worlds – one world has electrons and another has none, one has spirits and another has none – but these differences of kind are no more than sometimes arise between things that are parts of one single world, for instance in a world where electrons coexist with spirits. The difference between this and the other worlds is not a categorial difference.

Nor does this world differ from the others in its manner of existing. I do not have the slightest idea what a difference in manner of existing is supposed to be. Some things exist here on earth, other things exist extraterrestrially, perhaps some exist no place in particular; but that is no difference in manner of existing, merely a difference in location or

[1] Or 'extreme' modal realism, as Stalnaker calls it – but in what dimension does its extremity lie?

lack of it between things that exist. Likewise some things exist here at our world, others exist at other worlds; again, I take this to be a difference between things that exist, not a difference in their existing. You might say that strictly speaking, only this-worldly things *really* exist; and I am ready enough to agree; but on my view this 'strict' speaking is *restricted* speaking, on a par with saying that all the beer is in the fridge and ignoring most of all the beer there is. When we quantify over less than all there is, we leave out things that (unrestrictedly speaking) exist *simpliciter*. If I am right, other-worldly things exist *simpliciter*, though often it is very sensible to ignore them and quantify restrictedly over our worldmates. And if I am wrong, other-worldly things fail *simpliciter* to exist. They exist, as the Russell set does, only according to a false theory. That is not to exist in some inferior manner – what exists only according to some false theory just does not exist at all.

The worlds are not of our own making. It may happen that one part of a world makes other parts, as we do; and as other-worldly gods and demiurges do on a grander scale. But if worlds are causally isolated, nothing outside a world ever makes a world; and nothing inside makes the whole of a world, for that would be an impossible kind of self-causation. We make languages and concepts and descriptions and imaginary representations that apply to worlds. We make stipulations that select some worlds rather than others for our attention. Some of us even make assertions to the effect that other worlds exist. But none of these things we make are the worlds themselves.

Why believe in a plurality of worlds? – Because the hypothesis is serviceable, and that is a reason to think that it is true. The familiar analysis of necessity as truth at all possible worlds was only the beginning. In the last two decades, philosophers have offered a great many more analyses that make reference to possible worlds, or to possible individuals that inhabit possible worlds. I find that record most impressive. I think it is clear that talk of *possibilia* has clarified questions in many parts of the philosophy of logic, of mind, of language, and of science – not to mention metaphysics itself. Even those who officially scoff often cannot resist the temptation to help themselves abashedly to this useful way of speaking.

Hilbert called the set-theoretical universe a paradise for mathematicians. And he was right (though perhaps it was not he who should have said it). We have only to believe in the vast hierarchy of sets, and there we find entities suited to meet the needs of all the branches of mathematics;[2] and we find that the very meagre primitive vocabulary of set theory, definitionally extended, suffices to meet our needs for mathematical

[2]With the alleged exception of category theory – but here I wonder if the unmet needs have more to do with the motivational talk than with the real mathematics.

predicates; and we find that the meagre axioms of set theory are first principles enough to yield the theorems that are the content of the subject. Set theory offers the mathematician great economy of primitives and premises, in return for accepting rather a lot of entities unknown to *Homo javanensis*. It offers an improvement in what Quine calls ideology, paid for in the coin of ontology. It's an offer you can't refuse. The price is right; the benefits in theoretical unity and economy are well worth the entities. Philosophers might like to see the subject reconstructed or reconstrued; but working mathematicians insist on pursuing their subject in paradise, and will not be driven out. Their thesis of plurality of sets is fruitful; that gives them good reason to believe that it is true.

Good reason; I do not say it is conclusive. Maybe the price is higher than it seems because set theory has unacceptable hidden implications – maybe the next round of set-theoretical paradoxes will soon be upon us. Maybe the very idea of accepting controversial ontology for the sake of theoretical benefits is misguided – so a sceptical epistemologist might say, to which I reply that mathematics is better known than any premise of sceptical epistemology. Or perhaps some better paradise might be found. Some say that mathematics might be pursued in a paradise of *possibilia*, full of unactualised idealisations of things around us, or of things we do – if so, the parallel with mathematics serves my purpose better than ever! Conceivably we might find some way to accept set theory, just as is and just as nice a home for mathematics, without any ontological commitment to sets. But even if such hopes come true, my point remains. It has been the judgement of mathematicians, which modest philosophers ought to respect, that *if* that is indeed the choice before us, then it is worth believing in vast realms of controversial entities for the sake of enough benefit in unity and economy of theory.

As the realm of sets is for mathematicians, so logical space is a paradise for philosophers. We have only to believe in the vast realm of *possibilia*, and there we find what we need to advance our endeavours. We find the wherewithal to reduce the diversity of notions we must accept as primitive, and thereby to improve the unity and economy of the theory that is our professional concern – total theory, the whole of what we take to be true. What price paradise? If we want the theoretical benefits that talk of *possibilia* brings, the most straightforward way to gain honest title to them is to accept such talk as the literal truth. It is my view that the price is right, if less spectacularly so than in the mathematical parallel. The benefits are worth their ontological cost. Modal realism is fruitful; that gives us good reason to believe that it is true.

Good reason; I do not say it is conclusive. Maybe the theoretical benefits to be gained are illusory, because the analyses that use *possibilia* do not succeed on their own terms. Maybe the price is higher than it seems, because modal realism has unacceptable hidden implications. Maybe the

price is *not* right; even if I am right about what theoretical benefits can be had for what ontological cost, maybe those benefits just are not worth those costs. Maybe the very idea of accepting controversial ontology for the sake of theoretical benefits is misguided. Maybe – and this is the doubt that most interests me – the benefits are not worth the cost, because they can be had more cheaply elsewhere. Some of these doubts are too complicated to address here, or too simple to address at all; others will come in for discussion in the course of this book.

1.2 Modal Realism at Work: Modality

In the next four sections, I consider what possible worlds and individuals are good for. Even a long discussion might be too short to convince all readers that the applications I have in mind are workable at all, still less that approaches employing *possibilia* are superior to all conceivable rivals. (Still less that *possibilia* are absolutely indispensable, something I don't believe myself.) Each application could have a book of its own. Here I shall settle for less.

The best known application is to modality. Presumably, whatever it may mean to call a world actual (see section 1.9), it had better turn out that the world we are part of is the actual world. What actually is the case, as we say, is what goes on here. That is one possible way for a world to be. Other worlds are other, that is *un*actualised, possibilities. 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 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.

But not just quantification: there is also the phrase 'at W' which appears within the scope of the quantifier, and which needs explaining. It works mainly by restricting the domains of quantifiers in its scope, in much the same way that the restricting modifier 'in Australia' does. In Australia, all swans are black – all swans are indeed black, if we ignore everything not in Australia; quantifying only over things in Australia, all swans are black. At some strange world W, all swans are blue – all swans are indeed blue, if we ignore everything not part of the world W; quantifying only over things that are part of W, all swans are blue.

Such modifiers have various other effects. For one thing, they influence the interpretation of expressions that are not explicitly quantificational,

but that reveal implicit quantification under analysis: definite descriptions and singular terms definable by them, class abstracts and plurals, superlatives, etc. An example: it is the case at world W that nine numbers the solar planets iff nine numbers those solar planets that are part of W. Another example: words like 'invent' and 'discover' are implicitly superlative, hence implicitly quantificational; they imply doing something *first*, before *anyone* else did. So the inventor of bifocals at W is the one who is part of W and thought of bifocals before anyone else who is part of W did. For another thing, besides restricting explicit or implicit quantifiers, our modifiers can restrict proper names. In Australia, and likewise at a possible world where the counterparts of British cities are strangely rearranged, Cardiff is a suburb of Newcastle – there are various places of those names, and we banish ambiguity by restricting our attention to the proper domain. Here I am supposing that the way we bestow names attaches them not only to this-worldly things, but also to other-worldly counterparts thereof. That is how the other-worldly Cardiffs and Newcastles bear those names in our this-worldly language. In the same way, the solar planets at W are those that orbit the star Sol of the world W, a counterpart of the Sol of this world. Natural language being complex, doubtless I have not listed all the effects of our modifiers. But I believe the principle will always stay the same: whatever they do, they do it by instructing us, within limits, to take account only of things that are part of a limited domain – the domain of things in Australia, or the domain of parts of a certain world.

Two qualifications concerning our restrictive modifiers. (1) I do not suppose that they must restrict all quantifiers in their scope, without exception. 'In Australia, there is a yacht faster than any other' would mean less than it does if the modifier restricted both quantifiers rather than just the first. 'Nowadays there are rulers more dangerous than any ancient Roman' would be trivialised if we ignored those ancient Romans who are not alive nowadays. 'At some small worlds, there is a natural number too big to measure any class of individuals' can be true even if the large number that makes it true is no part of the small world. (2) Of course there will usually be other restrictions as well; doubtless we are already ignoring various immigrant swans and their descendants, and also whatever freak or painted swans there may be in Australia or among the parts of world W, so our modifier 'in Australia' or 'at W' adds more restrictions to the ones already in force. In short, while our modifiers tend to impose restrictions on quantifiers, names, etc., a lot is left up to the pragmatic rule that what is said should be interpreted so as to be sensible. If that means adding extra tacit restrictions, or waiving some of the restrictions imposed by our modifiers, then – within limits – so be it.[3]

As possibility amounts to existential quantification over the worlds, with restricting modifiers inside the quantifiers, so necessity amounts to universal quantification. Necessarily all swans are birds iff, for any world W, quantifying only over parts of W, all swans are birds. More simply: iff all swans, no matter what world they are part of, are birds. The other modalities follow suit. What is impossible is the case at no worlds; what is contingent is the case at some but not at others.

More often than not, modality is *restricted* quantification; and restricted from the standpoint of a given world, perhaps ours, by means of so-called 'accessibility' relations. Thus it is nomologically necessary, though not unrestrictedly necessary, that friction produces heat: at every world that obeys the laws of our world, friction produces heat. It is contingent which world is ours; hence what are the laws of our world; hence which worlds are nomologically 'accessible' from ours; hence what is true throughout these worlds, i.e. what is nomologically necessary.

Likewise it is historically necessary, now as I write these words, that my book is at least partly written: at every world that perfectly matches ours up to now, and diverges only later if ever, the book is at least partly written.

[3]This discussion of restricting modifiers enables me to say why I have no use for impossible worlds, on a par with the possible worlds. For comparison, suppose travellers told of a place in this world – a marvellous mountain, far away in the bush – where contradictions are true. Allegedly we have truths of the form 'On the mountain both P and not P'. But if 'on the mountain' is a restricting modifier, which works by limiting domains of implicit and explicit quantification to a certain part of all that there is, then it has no effect on the truth-functional connectives. Then the order of modifier and connectives makes no difference. So 'On the mountain both P and Q' is equivalent to 'On the mountain P, and on the mountain Q'; likewise 'On the mountain not P' is equivalent to 'Not: on the mountain P'; putting these together, the alleged truth 'On the mountain both P and not P' is equivalent to the overt contradiction 'On the mountain P, and not: on the mountain P'. That is, there is no difference between a contradiction within the scope of the modifier and a plain contradiction that has the modifier within it. So to tell the alleged truth about the marvellously contradictory things that happen on the mountain is no different from contradicting yourself. But there is no subject matter, however marvellous, about which you can tell the truth by contradicting yourself. Therefore there is no mountain where contradictions are true. An impossible world where contradictions are true would be no better. The alleged truth about its contradictory goings-on would itself be contradictory. At least, that is so if I am right that 'at so-and-so world' is a restricting modifier. Other modifiers are another story. 'According to the Bible' or 'Fred says that' are *not* restricting modifiers; they do not pass through the truth-functional connectives. 'Fred says that not P' and 'Not: Fred says that P' are independent: both, either, or neither might be true. If worlds were like stories or story-tellers, there would indeed be room for worlds according to which contradictions are true. The sad truth about the prevarications of these worlds would not itself be contradictory. But worlds, as I understand them, are *not* like stories or story-tellers. They are like this world; and this world is no story, not even a true story. Nor should worlds be replaced by their stories, for reasons discussed in section 3.2.

Putting together nomological and historical accessibility restrictions, we get the proper treatment of predetermination – a definition free of red herrings about what can in principle be known and computed, or about the analysis of causation. It was predetermined at his creation that Adam would sin iff he does so at every world that both obeys the laws of our world and perfectly matches the history of our world up through the moment of Adam's creation.

As other worlds are alternative possibilities for an entire world, so the parts of other worlds are alternative possibilities for lesser individuals. Modality *de re*, the potentiality and essence of things, is quantification over possible individuals. 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. A nomologically or historically accessible world is similar to our world in the laws it obeys, or in its history up to some time. Likewise a counterpart of Oxford is similar to Oxford in its origins, or in its location *vis-à-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 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.

Sometimes one hears a short list of the restricted modalities: nomological, historical, epistemic, deontic, maybe one or two more. And sometimes one is expected to take a position, once and for all, about what is or isn't possible *de re* for an individual. I would suggest instead that the restricting of modalities by accessibility or counterpart relations, like the restricting of quantifiers generally, is a very fluid sort of affair: inconstant, somewhat indeterminate, and subject to instant change in response to contextual pressures. Not anything goes, but a great deal does. And to a substantial extent, saying so makes it so: if you say what would only be true under certain restrictions, and your conversational partners acquiesce, straightway those restrictions come into force.[4]

The standard language of modal logic provides just two modal expressions: the diamond, read as 'possibly', and the box, read as 'necessarily'. Both are sentential operators: they attach to sentences to make sentences, or

[4]See section 4.5; Kratzer, 'What "Must" and "Can" Must and Can Mean', and my 'Scorekeeping in a Language Game'.

to open formulas to make open formulas. So a modal logician will write

◇ for some x, x is a swan and x is blue

to mean that possibly some swan is blue, i.e. that there might be a blue swan; or

□ for all x, if x is a swan then x is a bird

to mean that necessarily all swans are birds. Likewise

◇ x is blue

is a formula satisfied by anything that could possibly be blue, and

□ x is a bird

is a formula satisfied by anything that must necessarily be a bird. When they attach to sentences we can take the diamond and the box as quantifiers, often restricted, over possible worlds. How to take them when they attach to open formulas – sentential expressions with unbound variables – is more questionable.

A simple account would be that in that case also they are just quantifiers over worlds. But that raises a question. Start with something that is part of this world: Hubert Humphrey, say. He might have won the presidency but didn't, so he satisfies the modal formula 'possibly x wins' but not the formula 'x wins'. Taking the diamond 'possibly' as a quantifier over worlds, (perhaps restricted, but let me ignore that), that means that there is some world W such that, at W, he satisfies 'x wins'. But how does he do that if he isn't even part of W?

You might reply that he *is* part of W as well as part of this world. If this means that the whole of him is part of W, I reject that for reasons to be given in section 4.2; if it means that part of him is part of W, I reject that for reasons to be given in section 4.3. Then to save the simple account, we have to say that Humphrey needn't be part of a world to satisfy formulas there; there is a world where somehow he satisfies 'x wins' *in absentia*.

We might prefer a more complex account of how modal operators work.[5] We might say that when 'possibly' is attached to open formulas, it is a quantifier not just over worlds but also over other-worldly counterparts of this-worldly individuals; so that Humphrey satisfies

[5]This is essentially the account I gave in 'Counterpart Theory and Quantified Modal Logic'.

'possibly x wins' iff, for some world W, for some counterpart of Humphrey in W, that counterpart satisfies 'x wins' at W. The satisfaction of 'x wins' by the counterpart is unproblematic. Now we need no satisfaction *in absentia*.

The simple and complex accounts are not in competition. Both do equally well, because there is a counterpart-theoretic account of satisfaction *in absentia* that makes them come out equivalent. Satisfaction *in absentia* is vicarious satisfaction: Humphrey satisfies 'x wins' vicariously at any world where he has a winning counterpart. Then according to both accounts alike, he satisfies 'possibly x wins' iff at some world he has a counterpart who wins.

The box and diamond are interdefinable: 'necessarily' means 'not possibly not'. So what I have said for one carries over to the other. According to the simple account, Humphrey satisfies the modal formula 'necessarily x is human' iff it is not the case that there is some world W such that, at W, he satisfies 'x is not human'; that is, iff at no world does he satisfy – *in absentia* or otherwise – x is not human'. According to the complex account, Humphrey satisfies 'necessarily x is human' iff it is not the case that for some world W, for some counterpart of Humphrey in W, that counterpart satisfies 'x is not human' at W; that is, iff there is no counterpart in any world of Humphrey who satisfies 'x is not human'. Taking satisfaction *in absentia* to be vicarious satisfaction through a counterpart, the simple and complex accounts again agree: Humphrey satisfies 'necessarily x is human' iff he has no non-human counterpart at any world.

(It is plausible enough that Humphrey has no non-human counterpart. Or, if I am right to say that counterpart relations are an inconstant and indeterminate affair, at any rate it is plausible enough that there is *some* reasonable counterpart relation under which Humphrey has no non-human counterpart – so let's fix on such a counterpart relation for the sake of the example.)

The alert or informed reader will know that if what I've said about how Humphrey satisfies modal formulas sounds right, that is only because I took care to pick the right examples. A famous problem arises if instead we consider whether Humphrey satisfies modal formulas having to do with the contingency of his existence. According to what I've said, be it in the simple or the complex formulation, Humphrey satisfies 'necessarily x exists' and fails to satisfy 'possibly x does not exist' iff he has no counterpart at any world W who does not exist at W. But what can it mean to say that the counterpart is 'at W' if not that, at W, the counterpart exists?[6] So it seems that Humphrey *does* satisfy 'necessarily

[6]We might just *say* it, and not mean anything by it. That is Forbes's solution to our present difficulty, in his so-called 'canonical counterpart theory' – my own version is

x exists' and *doesn't* satisfy 'possibly x does not exist'. That is wrong. For all his virtues, still it really will not do to elevate Humphrey to the ranks of the Necessary Beings.

What I want to say, of course, is that Humphrey exists necessarily iff at every world he has some counterpart, which he doesn't; he has the possibility of not existing iff at some world he lacks a counterpart, which he does. It's all very well to say this; but the problem is to square it with my general account of the satisfaction of modal formulas.

So shall we give a revised account of the satisfaction of modal formulas? Should we say that Humphrey satisfies 'necessarily ϕx' iff at every world he has some counterpart who satisfies 'ϕx'? Then, by the interdefinability of box and diamond, Humphrey satisfies 'possibly x is a cat' iff it is not the case that at every world he has some counterpart who satisfies 'not x is a cat'; and indeed that is not the case, since at some worlds he has no counterparts at all; so it seems that he *does* satisfy 'possibly x is a cat' even if he has not a single cat among his counterparts! This is no improvement. What next?

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 model 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?

---

hereby named 'official standard counterpart theory' – in which, if Humphrey has no ordinary counterpart among the things which exist at W, he does nevertheless have a counterpart at W. This extraordinary counterpart is none other than Humphrey himself – he then gets in as a sort of associate member of W's population, belonging to its 'outer domain' but not to the 'inner domain' of things that exist there fair and square. This isn't explained, but really it needn't be. It amounts to a stipulation that there are two different ways that Humphrey – he himself, safe at home in this world – can satisfy formulas *in absentia*. Where he has proper counterparts, he does it one way, namely the ordinary vicarious way. Where he doesn't, he does it another way – just by not being there he satisfies 'x does not exist'.

There might be a fallacy of equivocation. Maybe what it means for Humphrey to satisfy a formula *in absentia* is different in the case of different kinds of formulas, or in the case of different kinds of worlds. Maybe, for instance, he can satisfy 'x does not exist' at a world by not having a counterpart there; but to satisfy 'x is human' at a world he has to have a counterpart there who is human, and to satisfy 'x is human and x does not exist' he would have to have one who was human and yet did not exist. Or maybe the language is uniformly ambiguous, and different cases invite different disambiguations. Either way, that would disappoint anyone who hopes that the language of quantified modal logic will be a well-behaved formal language, free of ambiguity and free of devious semantic rules that work different ways in different cases.

Or maybe the satisfying of modal formulas does not always mean what we would intuitively take it to mean after we learn how to pronounce the box and diamond. Maybe, for instance, saying that Humphrey satisfies 'necessarily x is human' is *not* the right way to say that he is essentially human. That would disappoint anyone who hopes that the language of boxes and diamonds affords a good regimentation of our ordinary modal thought.

Whichever it is, the friend of boxes and diamonds is in for a disappointment. He can pick his disappointment to suit himself. He can lay down uniform and unambiguous semantic rules for a regimented formal language – and re-educate his intuitions about how to translate between that language and ordinary modal talk. He can discipline himself, for instance, never to say 'necessarily human' when he means 'essentially human'; but instead, always to say 'necessarily such that it is human if it exists'. Alternatively, he can build his language more on the pattern of what we ordinarily say – and equip it either with outright ambiguities, or else with devious rules that look at what a formula says before they know what it means to satisfy it.[7]

What is the correct counterpart-theoretic interpretation of the modal formulas of the standard language of quantified modal logic? – Who cares? We can make them mean whatever we like. We are their master. We needn't be faithful to the meanings we learned at mother's knee – because we didn't. If this language of boxes and diamonds proves to be a clumsy instrument for talking about matters of essence and potentiality,

[7]If he likes, he can give himself more than one of these disappointments. As I noted, Forbes's talk of non-existent counterparts in outer domains amounts to a stipulation that satisfaction *in absentia* works different ways in different cases; so I find it strange that he offers it in rejoinder to a proposal of Hunter and Seager that modal formulas of parallel form needn't always be given parallel counterpart-theoretic translations. But this divided treatment does not pay off by making the modal formulas mean what we would offhand expect them to – it is exactly the non-existent counterparts in the outer domains that keep Humphrey from satisfying 'necessarily x is human' even if he is essentially human.

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.

In any case, modality is not all diamonds and boxes. Ordinary language has modal idioms that outrun the resources of standard modal logic, though of course you will be able to propose extensions. Allen Hazen mentions several examples of this in his 'Expressive Completeness in Modal Languages'. But let me mention some more.

There is what I take to be numerical quantification: it might happen in three different ways that a donkey talks iff three possible individuals, very different from one another, are donkeys that talk. It scarcely seems possible to cover the entire infinite family of numerical modalities unless we resort to the pre-existing apparatus of numerical quantification. Then we need some entities to be the 'ways' that we quantify over. My candidates are the possible worlds and individuals themselves, or else sets of these.

There are modalised comparatives: a red thing could resemble an orange thing more closely than a red thing could resemble a blue thing. I analyse that as a quantified statement of comparative resemblance involving coloured things which may be parts of different worlds.

> For some x and y (x is red and y is orange and
> for all u and v (if u is red and v is blue, then
> x resembles y more than u resembles v))

Try saying that in standard modal logic. The problem is that formulas get evaluated relative to a world, which leaves no room for cross-world comparisons.

Maybe you can solve the problem if you replace the original comparative relation '. . . resembles . . . more than . . . resembles . . .' by some fancy analysis of it, say in terms of numerical measures of degrees of resemblance and numerical inequalities of these degrees. After that, you might be able to do the rest with boxes and diamonds. The fancy analysis might be correct. But still, I suggest that your solution is no fair. For that's not how the English does it. The English does not introduce degrees of resemblance. It sticks with the original comparative relation, and modalises it with the auxiliary 'could'. But this 'could' does not behave like the standard sentence-modifying diamond, making a sentence which is true if the modified sentence could be true. *I* think its effect is to unrestrict quantifiers which would normally range over this-worldly things. The moral for me is that we'd better have other-worldly things to quantify over. I suppose the moral for a friend of primitive modality is that he

has more on his plate than he thinks he has: other primitive modal idioms than just his boxes and diamonds.

Another modal notion which is badly served by diamonds and boxes is supervenience. The idea is simple and easy: we have supervenience when there could be no difference of one sort without differences of another sort. At least, this *seems* simple and easy enough; and yet in recent discussions[8] we get an unlovely proliferation of non-equivalent definitions. Some stick close to the original idea but seem too weak; others seem strong enough but out of touch with the original idea. A useful notion threatens to fade away into confusion. I offer this diagnosis of the trouble. There really *is* just one simple, easy, useful idea. However, it is unavailable to those who assume that all modality must come packaged in boxes and diamonds. Therefore we get a plethora of unsatisfactory approximations and substitutes.

To see why there is a problem about formulating supervenience theses, we need a few examples. First, a fairly uncontroversial one. A dot-matrix picture has global properties – it is symmetrical, it is cluttered, and whatnot – and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or isn't a dot.

A second example is more controversial and interesting. The world has its laws of nature, its chances and causal relationships; and yet – perhaps! – all there is to the world is its point-by-point distribution of local qualitative character. We have a spatiotemporal arrangement of points. At each point various local intrinsic properties may be present, instantiated perhaps by the point itself or perhaps by point-sized bits of matter or of fields that are located there. There may be properties of mass, charge, quark colour and flavour, field strength, and the like; and maybe others besides, if physics as we know it is inadequate to its descriptive task. Is that all? Are the laws, chances, and causal relationships nothing but patterns which supervene on this point-by-point distribution of properties? Could two worlds differ in their laws without differing, somehow, somewhere, in local qualitative character? (I discuss this question of 'Humean supervenience', inconclusively, in the Introduction to my *Philosophical Papers*, volume II.)

A third example. A person has a mental life of attitudes and experiences and yet – perhaps! – all there is to him is an arrangement of physical particles, interacting in accordance with physical laws. Does the mental supervene on the physical? We can distinguish two questions. (1) *Narrow* psychophysical supervenience: could two people differ mentally without

[8]Surveyed in Teller, 'A Poor Man's Guide to Supervenience and Determination'.

also themselves differing physically? (2) *Broad* psychophysical supervenience: could two people differ mentally without there being a physical difference somewhere, whether in the people themselves or somewhere in their surroundings? We can also distinguish questions in another way, cross-cutting the distinction of narrow and broad, depending on how restricted a range of possibilities we consider. If we restrict ourselves to worlds that obey the actual laws of nature, then even a dualist might accept some kind of psychophysical supervenience, if he believes in strict laws of psychophysical correlation. If we impose no restriction at all, then even a staunch materialist might reject all kinds of psychophysical supervenience, if he takes materialism to be a contingent truth. If we want to define materialism in terms of psychophysical supervenience, we will have to steer between these extremes.[9]

Supervenience means that there *could* be no difference of the one sort without difference of the other sort. Clearly, this 'could' indicates modality. Without the modality we have nothing of interest. No two dot-for-dot duplicate pictures differ in symmetry; they could not, and that is why symmetry is nothing but a pattern in the arrangement of dots. Maybe also it happens that no two dot-for-dot duplicate pictures differ in their origins. But if so, that just means that a certain sort of coincidence happens not to have occurred; it doesn't mean that the origin of a picture is nothing but a pattern in the arrangement of dots. Dot-for-dot duplicates perfectly well could come from different origins, whether or not they ever actually do.

So we might read the 'could' as a diamond – a modal operator 'possibly' which modifies sentences. 'There could be no difference of the one sort without difference of the other sort' – read this to mean that it is not the case that, possibly, there are two things which have a difference of the one sort without any difference of the other sort. That is: it is not the case that there is some world W such that, at W, two things have a difference of the one sort but not the other. That is, taking 'at W' as usual as a restricting modifier: there is no world wherein two things have a difference of the one sort but not the other. Is this an adequate way to formulate supervenience?

Sometimes it is. It will do well enough to state our supervenience theses about dot-matrix pictures. Symmetry (or whatnot) supervenes on the arrangement of the dots iff there is no world wherein two pictures differ in symmetry without differing in their arrangement of dots. It will do also to state narrow psychophysical supervenience: that thesis says that there is no world (or, none within a certain restriction) wherein two people differ mentally without themselves differing physically. So far, so good.

[9]See Kim, 'Psychophysical Supervenience', and my 'New Work for a Theory of Universals'.

But sometimes the formulation with a diamond is not adequate. We start to hit trouble with the thesis of broad psychophysical supervenience. The idea is that the mental supervenes on the physical; however, the physical pattern that is relevant to a given person's mental life might extend indefinitely far outside that person and into his surroundings. Then the thesis we want says that there could be no mental difference between two people without there being some physical difference, whether intrinsic or extrinsic, between them. Reading the 'could' as a diamond, the thesis becomes this: there is no world (or, none within a certain restriction) wherein two people differ mentally without there being some physical difference, intrinsic or extrinsic, between them. That is not quite right. We have gratuitously limited our attention to physical differences between two people in the same world, and that means ignoring those extrinsic physical differences that only ever arise between people in different worlds. For instance, we ignore the difference there is between two people if one inhabits a Riemannian and the other a Lobachevskian spacetime. So what we have said is not quite what we meant to say, but rather this: there could be no mental differences without some physical difference *of the sort that could arise between people in the same world.* The italicised part is a gratuitous addition. Perhaps it scarcely matters here. For it doesn't seem that the sort of very extrinsic physical difference that could never arise between people in the same world would make much difference to mental life. Nevertheless, insistence on reading the 'could' as a diamond has distorted the intended meaning.

For a case where the distortion is much more serious, take my second example: the supervenience of laws. We wanted to ask whether two worlds could differ in their laws without differing in their distribution of local qualitative character. But if we read the 'could' as a diamond, the thesis in question turns into this: it is not the case that, possibly, two worlds differ in their laws without differing in their distribution of local qualitative character. In other words: there is no world wherein two worlds differ in their laws without differing in their distribution of local qualitative character. That's trivial – there is no world wherein two worlds do anything. At any one world W, there is only the one single world W. The sentential modal operator disastrously restricts the quantification over worlds that lies within its scope. Better to leave it off. But we need *something* modal – the thesis is not just that the one actual world, with its one distribution of local qualitative character, has its one system of laws![10]

[10]One more example of the same sort of distortion. Let *naturalism* be the thesis that whether one's conduct is right supervenes on natural facts, so that one person could do right and another do wrong only if there were some difference in natural facts between the two – as it might be, a difference in their behaviour or their circumstances. Consider the theory that, necessarily, right conduct is conduct that conforms to divinely

What we want is modality, but not the sentential modal operator. The original simple statement of supervenience is the right one; in all cases: there *could* be no difference of the one sort without difference of the other sort. What got us into trouble was to insist on reading 'could' as a diamond. Just as in the case of modalised comparatives, the real effect of the 'could' seems to be to *un*restrict quantifiers which would normally range over this-worldly things. Among all the worlds, or among all the things in all the worlds (or less than all, in case there is some restriction), there is no difference of the one sort without difference of the other sort. Whether the things that differ are part of the same world is neither here nor there. Again the moral is that we'd better have other-worldly things to quantify over – not just a primitive modal modifier of sentences.

When I say that possible worlds help with the analysis of modality, I do not mean that they help with the metalogical 'semantical analysis of modal logic'. Recent interest in possible worlds began there, to be sure. But wrongly. For that job, we need no possible worlds. We need sets of entities which, for heuristic guidance, 'may be regarded as' possible worlds, but which in truth may be anything you please. We are doing mathematics, not metaphysics. Where we need possible worlds, rather, is in applying the results of these metalogical investigations. Metalogical results, by themselves, answer no questions about the logic of modality. They give us conditional answers only: if modal operators can be correctly analysed in so-and-so way, then they obey so-and-so system of modal logic. We must consider whether they may indeed be so analysed; and then we are doing metaphysics, not mathematics.

Once upon a time, there were a number of formal systems of sentential modal logic. (Also of quantified modal logic, but I shall not discuss those further.) Their modal operators, box and diamond, were said to mean 'necessarily' and 'possibly', but were not interpreted as quantifiers over

---

willed universal maxims. Suppose it is contingent what, if anything, is divinely willed. And suppose that facts about what is divinely willed are supernatural, not natural, facts. You might well expect that this divine-will theory of rightness would contradict naturalism; for if two people are alike so far as natural facts are concerned, but one of them lives in a world where prayer is divinely willed and the other lives in a world where blasphemy is divinely willed, then what is right for the first is not right for the second. But if we read the 'could' as a diamond, we get an unexpected answer. A difference in what universal maxims are divinely willed never could be a difference between two people in the same world. Within a single world, the only differences relevant to rightness are natural differences, such as the difference between one who prays and one who blasphemes. So indeed there is no world wherein one person does right and another does wrong without any difference in natural facts between the two. So either this divine-will theory of rightness is naturalistic after all; or else – more likely – something has gone amiss with our understanding of supervenience.

worlds. These systems differed from one another mostly by including or excluding various controversial axioms about iterated modality, most prominently these:

(B) If P, then necessarily possibly P.

(4) If necessarily P, then necessarily necessarily P.

(E) If possibly P, then necessarily possibly P.

It was possible to investigate the deductive interrelations and consequences of various modal principles. For instance, given the plausible further axiom

(T) If P, then possibly P.

and a fairly minimal (but not entirely uncontroversial) basic system K,[11] it turns out that (E) can be deduced from (B) and (4) together, and conversely. But what was not possible was to intuit clearly which of these principles were to be accepted, and why; or even to command a clear view of what was at issue.

At this point it was discovered, by several people at about the same time, that if you interpret the box and diamond as restricted quantifiers over a set of entities 'regarded as possible worlds', then (B), (4), (E), and (T) turn out to correspond to simple conditions on the relation whereby the box and diamond are restricted.[12] We spell this out as follows. A (*relational*) *frame* consists of a non-empty set – call it the set of *indices* – and a binary relation R on the indices. A *valuation* for the language of a

[11] K is given by rules of truth-functional implication; the rule that any substitution instance of a theorem is a theorem; the rule of interchange of equivalents, which says that if 'ϕ_1 iff ϕ_2' is a theorem, and $-\phi_2-$ comes from $-\phi_1-$ by substituting ϕ_2 for ϕ_1 at one or more places, then '$-\phi_1-$ iff $-\phi_2-$' is a theorem; and three axioms:

Possibly P iff not necessarily not P.

Necessarily (P and Q) iff (necessarily P and necessarily Q).

Necessarily (P iff P).

When a new system is made by adding further axioms to K, it is understood that the word 'theorem' in the rules of substitution and interchange applies to all theorems of the new system.

[12] The first discussions of this, some much more developed than others, are Hintikka, 'Quantifiers in Deontic Logic'; Kanger, *Provability in Logic*; Kripke, 'A Completeness Theorem in Modal Logic'; and Montague, 'Logical Necessity, Physical Necessity, Ethics, and Quantifiers'. There is also unpublished work of C. A. Meredith, reported in Prior, *Past, Present and Future*, page 42. A well known early discussion is Kripke, 'Semantical Considerations on Modal Logic'.

system of modal logic over a frame specifies a truth value for every sentence of the language at every index, and does so in conformity to the standard rules for the truth-functional connectives together with the following rules for modal operators:

'Necessarily ϕ' is true at i iff ϕ is true at all j such that iRj.

'Possibly ϕ' is true at i iff ϕ is true at some j such that iRj.

(Here is where we treat the modal operators as restricted quantifiers.) A frame *validates* a sentence iff every valuation over that frame makes that sentence true at every index; and validates a system of modal logic iff it validates every theorem of that system. Given the following correspondence between our axioms and conditions on frames –

(B) corresponds to being symmetric: if iRj, then jRi

(4) corresponds to being transitive: if iRj and jRk, then iRk

(E) corresponds to being 'euclidean': if iRj and iRk, then jRk

(T) corresponds to being reflexive: iRi

it is easy to see that by adding any combination of zero or more axioms to the basic system K, we get a system that is validated by all frames that satisfy the corresponding combination of conditions. Further, every such system is *complete* in the sense that if any sentence is validated by all frames that validate the system, then that sentence already is a theorem of the system. The same is true for a very much longer list of corresponding axioms and conditions. The results can be extended to quantified modal logic, and related results are available for systems weaker than K.

These metalogical investigations seemed to cast light on the status of the controversial axioms. Maybe we didn't yet know whether the axioms were to be accepted, but at least we now knew what was at issue. Old questions could give way to new. Instead of asking the baffling question whether whatever is actual is necessarily possible, we could try asking: is the relation R symmetric?

But in truth the metalogical results, just by themselves, cast no light at all. *If* the modal operators can be correctly interpreted as quantifiers over the indices of some or other frame, restricted by the relation of that frame, *then* we have found out where to look for illumination about controversial axioms. If not, not. To apply the results, you have to incur a commitment to some substantive analysis of modality. To be sure, you might not have to be a genuine modal realist like me. You might prefer an analysis on which the modal operators are quantifiers over some sort of abstract ersatz worlds – linguistic descriptions, maybe. (If you meant

that as a fully general analysis of modality, I would raise several objections; see section 3.2. If you meant it to apply only in certain limited cases, for instance to modal talk about how a chess game might have gone, I would not object at all.) But if the metalogical results are to be at all relevant to modality, *some* quantificational analysis has to be correct. If modal operators were quantifiers over towns restricted by the relation of being connected by rail, that would validate some system or other of modal logic. – So what, since modal operators are nothing of the sort? What good is it to know which *mis*interpretations would validate a system?

I myself, of course, do think that modal operators are quantifiers over possible worlds; that very often they are restricted; and that the applicable restriction may be different from the standpoint of different worlds, and so may be given by a relation of 'accessibility'. Therefore I do not just think that the indices of frames 'may be regarded as' possible worlds. I think that among all the frames, there are some whose indices *are* the possible worlds; and that among such frames there are some whose relations *do* give the correct restrictions on modal operators (correct for appropriate contexts). So for me, the metalogical results are applicable, because I believe that there exist frames which afford correct interpretations of the modal operators.

Return to an example I mentioned before: it is nomologically necessary that friction produces heat because at every world nomologically accessible from ours – every world that obeys the laws of ours – friction produces heat. Then, indeed, puzzling questions about the logic of iterated nomological necessity turn into more tractable questions about the relation of nomological accessibility. Is it symmetric? Transitive? Euclidean? Reflexive? In other words, is it so that whenever world W_1 obeys the laws of W_0, then also W_0 obeys the laws of W_1? Is it so that whenever W_2 obeys the laws of W_1 which in turn obeys the laws of W_0, then W_2 obeys the laws of W_0? Is it so that whenever W_1 and W_2 both obey the laws of W_0, then they obey each other's laws? Is it so that every world obeys its own laws? – A theory of lawhood can be expected to answer these questions, and we can see how different theories would answer them differently. (For instance, my own views on lawhood answer all but the last in the negative.) This transformation of questions is helpful indeed. But the help comes from a substantive theory of what nomological necessity is – not from metalogical investigations that keep silent about which frames, if any, afford correct interpretations. It is the substantive theory, not the metalogic, for which we need possible worlds.

1.3 Modal Realism at Work: Closeness

A counterfactual (or 'subjunctive') conditional is an invitation to consider what goes on in a selected 'counterfactual situation'; which is to say, at

some other possible world. Partly, the world in question is specified explicitly by the antecedent of the conditional: 'If kangaroos had no tails . . . .' Partly, it is specified by a permanent understanding that there is to be no gratuitous departure from the background of fact: ignore worlds where the kangaroos float around like balloons, since the kangaroos of our world are much too heavy for that. Partly, it is specified by temporary contextual influences that indicate what sorts of departures would be especially gratuitious; for instance, facts just mentioned may have a special claim to be held fixed.

Partly, it is not specified at all: no telling whether the kangaroos have stumps where the tails should be. So it is an idealisation to think that we have to do with a single world, rather than an ill-defined class. Under that idealisation, we can say that a counterfactual conditional 'If it were that A, then it would be that C' is true iff C is true at the selected A-world. More generally, the conditional is true at a world W iff C is true at the A-world selected from the standpoint of W.[13]

Within the approach to counterfactuals just sketched, there is room for debate on a number of questions.

(1) How might we best deal with the idealisation just noted? Should we write the analysis of conditionals so that it tolerates ties in the similarity relation? So that it tolerates incomparabilities? So that it tolerates a (somewhat far-fetched) situation in which there are no A-worlds *most* similar to W, but only more and more similar ones *ad infinitum*? How much should be done by complicating the analysis of counterfactuals, how much by joining a simple analysis of counterfactuals with a general treatment for phenomena of semantic indeterminacy?

(2) If one A-world is selected and another A-world is not, from the standpoint of W, that establishes a sense in which we may say that the first is *closer* to W. What are the formal properties of this 'closeness' ordering? Is it a well-ordering? Does it admit ties? Does it admit incomparabilities?

(3) Is it useful to describe it as a *similarity* ordering, saying that the selected A-worlds are the A-worlds most similar to W? We could mean too little or too much by that: too little if we meant only that the ordering had certain formal properties, too much if we meant that our immediate 'intuitions' of similarity could be relied on to follow the ordering. Is there an intermediate meaning that would be more satisfactory? To say that counterfactuals work by similarity is the skeleton of a theory. To flesh it out, we must say which are the important respects of comparison. How far can we answer that question once and for all? How far must we answer it differently for different sorts of counterfactuals in different sorts of contexts?

[13]See my *Counterfactuals* and Stalnaker, 'A Theory of Conditionals'.

(4) How do we connect the 'would' counterfactual with 'might' counterfactuals and probabilistic counterfactuals? Should we have a family of related connectives? Or should we have a single conditional connective, and apply modal or probabilistic modifiers either to the consequent or to the entire conditional?

(5) Is the indicative conditional something else altogether? Is it, for instance, the truth-functional conditional plus conventional or conversational implicatures? Or does it also work by truth of the consequent at a selected antecedent-world, with the difference between indicative and subjunctive being simply a difference in the principles of selection?

These questions have been much discussed, and I do not want to pursue them here.[14] I do want to point out that they are all within the family. They do nothing to threaten the core idea that counterfactuals have to do with what goes on at possible worlds given jointly by the antecedent, factual background, and contextual influences.

A challenge which goes deeper, and which does question the utility of bringing possible worlds into the story, goes as follows. Here is our world, which has a certain qualitative character. (In as broad a sense of 'qualitative' as may be required – include irreducible causal relations, laws, chances, and whatnot if you believe in them.) There are all the various A-worlds, with their various characters. Some of them are closer to our world than others. If some (A-and-C)-world is closer to our world than any (A-and-not-C)-world is, that's what makes the counterfactual true at our world. Now, whether or not this closeness ought to be called similarity, still somehow it's a matter of the character of the worlds in question. It's the character of our world that makes some A-worlds be closer to it than others. So, after all, it's the character of our world that makes the counterfactual true – in which case why bring the other worlds into the story at all?

To which I reply that is indeed the character of our world that makes the counterfactual true. But it is only by bringing the other worlds into the story that we can say in any concise way what character it takes to make what counterfactuals true. The other worlds provide a frame of reference whereby we can characterise our world. By placing our world within this frame, we can say just as much about its character as is relevant to the truth of a counterfactual: our world is such as to make an (A-and-C)-world closer to it than any (A-and-not-C)-world is.

If counterfactuals were no good for anything but idle fantasies about unfortunate kangaroos, then it might be faint praise to say that possible

[14]As well as the works cited in the previous footnote, see my 'Ordering Semantics and Premise Semantics for Counterfactuals'; my *Philosophical Papers*, volume II, chapter 17; and Stalnaker, *Inquiry*, chapters 6–8.

worlds can help us with counterfactuals. But, in fact, counterfactuals are by no means peripheral or dispensable to our serious thought. They are as central as causation itself. As I touch these keys, luminous green letters appear before my eyes, and afterward black printed letters will appear before yours; and if I had touched different keys – a counterfactual supposition – then correspondingly different letters would have appeared. That is how the letters depend causally upon the keystrokes, and that is how the keystrokes cause letters to appear.

Suppose that two wholly distinct events occur, C and E; and if C had not occurred, E would not have occurred either. I say that if one event depends counterfactually on another in this way (and if it's the right sort of counterfactual, governed by the right sort of closeness of worlds) then E depends causally on C, and C is a cause of E. To be sure, this is only the beginning of a counterfactual analysis of causation. Not all counterfactuals are of the right sort, and it is a good question how to distinguish the ones that are from the ones that aren't. We need an account of eventhood, and of distinctness of events. And not all effects depend counterfactually on their causes; for instance, we may have causation by a chain of stepwise dependence, in which E depends on D which depends on C, and thereby C causes E, yet E does not depend directly on C because of some alternate cause waiting in reserve.[15] You may or may not share my optimism about an analysis of causation in terms of counterfactual dependence of events. But even if you give up hope for an analysis, still you can scarcely deny that counterfactuals and causation are well and truly entangled.

Causal theories of this, that, and the other have been deservedly popular in recent years. These theories are motivated by imagining cases where normal patterns of counterfactual dependence fail. Normally, my perceptual experience depends on what is going on around me, in such a way as to make its content largely correct. Normally, my movements depend on my beliefs and desires, in such a way that they tend to serve my beliefs according to my desires. Normally, the way I am depends on the way I was just before, in such a way as to keep change gradual. What if these normal dependences were absent? If my perceptual experience would be the same no matter what was going on around me, I would not be perceiving the world. If the movements of my body would be the same no matter what I believed and desired, those movements would not be my actions. If the man who will wake up in my bed tomorrow would be exactly the same regardless of what befell me today, he would be an impostor.

If possible worlds help with counterfactuals, then, they help with many parts of our thought that we could scarcely imagine being without.

[15] I discuss these issues in my *Philosophical Papers*, volume II, part 6.

Closeness of worlds can also help us to say what it means for a false theory of nature to be close to the truth. False is false – and it takes only a trace of error to make a theory false – but false theories are not all on a par. We may reasonably think that present-day scientific theories, if not entirely free of error, are at any rate closer to the truth than previous theories were. We may hope that future theories will be closer still. How can we explain this?

Risto Hilpinen has proposed that we might explain this closeness to the truth (or 'truthlikeness' or 'verisimilitude') in terms of closeness of possible worlds. As in the case of counterfactuals, this closeness is a matter of some sort of similarity. A theory is close to the truth to the extent that our world resembles some world where that theory is exactly true. A true theory is closest to the truth, because our world *is* a world where the theory is true. As for false theories, the ones that can come true in ways that involve little dissimilarity to the world as it really is are thereby closer to the truth than those that cannot.

For instance, we have the simple, approximate gas laws; and then we have correction terms. But if the correction terms were all zero, things wouldn't be too different. (You couldn't tell the difference unless either the circumstances were extraordinary or you made a very careful measurement.) The closest of the approximate-gas-law worlds are pretty close to ours. That is why the approximate gas laws are close to the truth. Suppose we improve the gas laws by putting in the most important of the corrections. Then we get a theory that holds in some worlds that imitate ours still better, so the improved theory is still closer to the truth.

Just as in the case of counterfactuals, what we have here is the mere skeleton of an analysis. To put flesh on the bones, we need to say something about what an appropriate similarity ordering of worlds might be – what sort of respects of comparison are the ones that count. (It seems unlikely that we could use the same similarity ordering both for verisimilitude and for counterfactuals.) But even a skeleton is well worth having. It tells us what sort of flesh to look for – to explain what we mean by verisimilitude, pick out the appropriate respects of comparison of worlds.

Whether we must settle for a messy business of comparative similarity depends on whether we can hope for something cleaner. It would be nice to give equal weight to all agreements and disagreements between a theory and the truth, and never fuss about which ones matter most to verisimilitude. But the problem is harder than it may seem, and there seems to be little hope that egalitarian methods can ever deliver non-trivial comparisons of verisimilitude. Suppose we subject two rival theories to a true–false quiz covering all sentences in the appropriate language. When a theory declines to answer, that is better than a wrong answer and worse than a right answer. How do we translate the question-by-question

performance of rival theories into an overall comparison? Counting fails: all false theories alike give equal infinite numbers of right and wrong answers. Dominance fails: it cannot happen that one of two false theories sometimes does better than the other and never does worse.[16] If the quiz were better made, if questions were selected for their importance, if redundancy were avoided, and if there were less opportunity for errors to cancel out, then numerical score or dominance on the quiz could mean more. Of course, a selective quiz – unlike a quiz that includes all possible questions – calls for judgement on the part of the examiner. It is open to challenge by those who disagree about what are the most important things for a theory to get right. So what? *Any* standard for preferring one theory to another is open to challenge – if, *per impossibile*, the method of dominance had succeeded in ranking some false theories above others, it could still have been challenged by those who care little about truth. But there is a more serious difficulty with the selective quiz: our original problem returns for every question. When theories give the wrong answer to a question on the quiz, false is false – however, some mistakes are farther off the mark than others. Does anything go faster than light?' – 'No' says the truth (let us suppose). 'Yes' says the better theory, according to which a very few very rare particles do. 'Yes' says the worse theory, according to which most planes and some birds do. If the quiz were unselective, the difference between the better and worse theories would show up on some follow-up question. But if the quiz is selective, as it must be to give a meaningful comparison, maybe sometimes the revealing follow-up question will have been left out.

I don't deny that verisimilitude might be explained in terms of performance on a suitably selective quiz. However, the choice of which questions to include and how to weight them will be just as problematic, and will raise just the same issues about what it is important to get right, as the choice of a similarity relation of worlds on Hilpinen's proposal. In fact, I suggest that the best intuitive guide to what makes a quiz suitable is exactly that we want score on it to be a good measure of how closely our world resembles any of the worlds that conform to the theory under test. If so, there is no way to get out of judging which respects of comparison are the important ones – not unless, with absurd disdain for what we understand outside the philosophy room, we junk the very idea of closeness to the truth.

[16] *Ex hypothesi* both theories are false; so let F be the disjunction of a falsehood affirmed by one and a falsehood affirmed by the other; then F is a falsehood affirmed by both. Suppose one theory does better on one question: is it so that A? Then the other theory does better on another question: is it so that A iff F? Then neither theory dominates the other. The conjecture that dominance would give useful comparisons of verisimilitude is due to Popper, *Conjectures and Refutations*, page 233; the refutation is due to Miller and Tichý.

A merit of Hilpinen's proposal is that it distinguishes aspects of verisimilitude which comparison by means of quizzes tends to run together. A theory T defines a region in the space of possible worlds: namely, the class of all T-worlds. The whole truth defines another region: the unit class of our world. There are three relevant ways to compare these regions in terms of similarity distance. (1) Size: the smaller the region of T-worlds is, the more it resembles the point-sized region defined by the truth. (2) Shape: the more compact the region of T-worlds is, the less it consists of far-flung and scattered parts, the more it resembles the point-shaped truth.[17] (3) Separation: the distance, at closest approach, between the region of T-worlds and our world. It is the separation which most clearly deserves the name 'closeness to the truth'. But small size and compact shape also are merits of theories, and might be considered as aspects of verisimilitude or 'truthlikeness' in a broader sense. All three aspects are involved if we consider not only separation at closest approach, but also further questions of separation: how distant at most are the T-worlds from our world? How distant are they on average (with respect to some sort of measure)? As can be seen from the spatial analogy, these comparisons have to do with size and shape as well as separation at closest approach.

Verisimilitude, as such, has been discussed mostly in connection with scientific progress. We can credit the false theories of former times with some degree of closeness to the truth; and even those sceptics who are quite certain that science will never rid itself of all error may hope at least to approach the truth ever more closely.

But the verisimilitude of false theories is not limited to theories that are at some time accepted as true. It applies equally to deliberate falsifications: the theory of the frictionless plane, the massless test particle, the ideally rational belief system, and suchlike useful idealisations. These theories never were meant to be any better than truthlike. When we disregard friction in saying how things slide on a plane, that is fiction, truthlike but false. When we go on to say that the fiction about the frictionless plane is close to the truth about what really happens on slick black ice, *that* is physics and true. One handy way to tell the truth about complicated phenomena is to say how they resemble simpler idealisations. Maybe the same truth could in principle be told directly – it is hard to see why not – but there is no doubt that we do find it much easier to tell the truth if we sometimes drag in the truthlike fiction.[18]

[17]The variety – that is, dissimilarity – within a region reflects both its size and shape, just as a spatial region including points separated by at most 14 miles might be a long thin strip with very little area or might be a circular region of about 154 square miles. Bennett, in 'Killing and Letting Die', and Bigelow, in 'Possible Worlds Foundations for Probability', have discussed methods for disentangling variety due to size from variety due to shape.

When we do, we traffic in possible worlds. Idealisations are unactualised things to which it is useful to compare actual things. An idealised theory is a theory known to be false at our world, but true at worlds thought to be close to ours. The frictionless planes, the ideal gases, the ideally rational belief systems – one and all, these are things that exist as parts of other worlds than our own.[19] The scientific utility of talking of idealisations is among the theoretical benefits to be found in the paradise of *possibilia*.

1.4 Modal Realism at Work: Content

An inventory of the varieties of modality may include *epistemic* and *doxastic* necessity and possibility. Like other modalities, these may be explained as restricted quantification over possible worlds. To do so, we may use possible worlds to characterise the content of thought. The content of someone's knowledge of the world is given by his class of *epistemically accessible* worlds. These are the worlds that might, for all he knows, be his world; world W is one of them iff he knows nothing, either explicitly or implicitly, to rule out the hypothesis that W is the world where he lives. Likewise the content of someone's system of belief about the world (encompassing both belief that qualifies as knowledge and belief that fails to qualify) is given by his class of *doxastically accessible* worlds. World W is one of those iff he believes nothing, either explicitly or implicitly, to rule out the hypothesis that W is the world where he lives.

Whatever is true at some epistemically or doxastically accessible world is epistemically or doxastically possible for him. It might be true, for all he knows or for all he believes. He does not know or believe it to be false. Whatever is true throughout the epistemically or doxastically accessible worlds is epistemically or doxastically necessary; which is to say that he knows or believes it, perhaps explicitly or perhaps only implicitly.

Since only truths can be known, the knower's own world always must be among his epistemically accessible worlds. Not so for doxastic accessibility. If he is mistaken about anything, that is enough to prevent his own world from conforming perfectly to his system of belief.[20]

[18]See Scriven on the recognised inaccuracy – idealisation – of some so-called laws. See Glymour on the way we often credit superseded physical theories with being right in a limiting case. This connects our two applications: the verisimilitude of a superseded theory rests on the verisimilitude of an idealisation.

[19]Then it won't be much use trying to do without possible worlds and replacing them with ideally rational belief systems, as Ellis has proposed; for the ideal belief systems themselves are other-worldly. *I* can believe in Ellis's replacement for possible worlds. Can he?

[20]See Hintikka, *Knowledge and Belief*, and his subsequent discussions of knowledge and belief in *Models for Modalities* and *The Intentions of Intentionality*.

No matter how we might originally characterise the content of knowledge or belief, it ought to be possible afterward to introduce the distinction between worlds that do and worlds that do not conform to that content. That done, we could go on to introduce the epistemic and doxastic modalities. For instance if we began with a notion of belief as some sort of acceptance of interpreted sentences – perhaps of our language, perhaps of some public language the believer speaks, or perhaps of the believer's hypothetical 'language of thought' – then we could say that a doxastically accessible world is one where all the accepted sentences are true. I am quite sceptical about this order of proceeding, for reasons that need not be reviewed here.[21] A more promising plan, I think, is to characterise the content of knowledge or belief from the outset in terms of something rather like the epistemically or doxastically accessible worlds. (Let me concentrate simply on belief, passing over the added complications that arise when we distinguish someone's knowledge from the rest of his system of belief.) The class of doxastically accessible worlds is roughly what we want, but it isn't exactly right; some changes must be made.

For one thing, I said that the doxastically accessible worlds give the content of one's system of belief *about the world*; but not all belief is about the world. Some of it is egocentric belief; or, as I have called it elsewhere, 'irreducibly *de se*'.[22] Imagine someone who is completely opinionated, down to the last detail, about what sort of world he lives in and what goes on there. He lacks no belief about the world. For him, only one world is doxastically accessible. (Or, at most, one class of indiscernible worlds – let me ignore this complication.) And yet there may be questions on which he has no opinion. For instance he may think he lives in a world of one-way eternal recurrence, with a beginning but no end, with a certain course of history repeated exactly in every epoch; and he may have no idea which epoch he himself lives in. Every epoch of the world he takes to be his contains someone who might, for all he believes, be himself. He has no idea which one of them he is. If he did, for instance if he somehow became persuaded that he lived in the seventeenth epoch, he would believe more than he does. But he would not believe more about the world. The added belief would be not about the world, but about his own place therein.

So if we want to capture the entire content of someone's system of belief, we must include the egocentric part. We should characterise the content not by a class of possible worlds, but by a class of possible individuals – call them the believer's *doxastic alternatives* – who might,

[21]See Stalnaker, *Inquiry*, chapters 1 and 2.

[22]See my 'Attitudes *De Dicto* and *De Se*' and 'Individuation by Acquaintance and by Stipulation'; and see Chisholm, *The First Person*, for a parallel theory in a somewhat different framework.

for all he believes, be himself. Individual X is one of them iff nothing that the believer believes, either explicitly or implicitly, rules out the hypothesis that he himself is X. These individuals are the believer's doxastic possibilities. But they are not different possible ways for the world to be; rather, they are different possible ways for an individual to be, and many of them may coexist within a single world. (For further discussion of individual possibilities, in other words possible individuals, see section 4.4). Suppose that all of someone's doxastic alternatives have a certain property; then he believes, explicitly or implicitly, that he himself has that property.

One property that an inhabitant of a world may have is the property of inhabiting a world where a certain proposition holds. (Or, of inhabiting a world that falls in a certain set of worlds. In the next section, I shall suggest that these come to the same thing.) So if all of someone's doxastic alternatives inhabit worlds where a certain proposition A holds, then he believes that he himself inhabits an A-world. In other words, he believes that A holds at his world, whichever world that may be. We may say, simply, that he believes the proposition A. So belief about the world comes out as a special case of egocentric belief. And the original treatment of belief about the world in terms of doxastically accessible worlds still works, within its limits. The doxastic alternatives determine the doxastically accessible worlds, though not conversely: a world is accessible iff at least one of the alternatives inhabits it. If each alternative inhabits an A-world, then A holds at every accessible world, so it is doxastically necessary according to the original treatment that A holds.

The same person can have different systems of belief at different times. Suppose it is true, as I think it is, that a person persists through time by consisting of many different momentary stages located at different times. (This is a controversial view; for some discussion of it, see section 4.2.) Then we can say first that the various stages have various systems of belief; and then that the continuing person has a system of belief at a time by having a stage at that time which has that system of belief.

By treating the subjects of belief as momentary, we can subsume belief about what time it is as a special case of egocentric belief. You may last threescore years and ten; but the stage that does your believing at a given moment is a momentary stage. If that stage has as its doxastic alternatives various person-stages all of which are located at about noon on 11 March 1985, that is how you at that moment have a belief about what time it is. (On what it means to compare times across worlds, see section 1.6.) If, on the other hand, that stage has as its alternatives various stages on various hours of various days, that is how you, at that moment, are uncertain what time it is. Note that you can lose track of the time no matter how certain you are about what sort of world you live in, and about which continuing person in that world you are.

(Knowledge, as well as belief, may be egocentric: besides knowing what sort of world you live in, you can also know who in the world you are and what time it is. So again we don't get a complete characterisation of knowledge by taking a class of epistemically accessible worlds; rather, we need a class of possible individuals within worlds as the subject's *epistemic alternatives*. What the subject knows in the first place is that he is some or another one of these possible individuals. So if all of them have some property in common, then he knows that he has that property; and if all of them live in worlds where some proposition holds, then he knows that proposition.)

Besides providing for egocentric belief by switching from accessible worlds to alternative individuals, we must also provide for partial belief. Being a doxastic alternative is not an all-or-nothing matter, rather it must admit of degree. The simplest picture, idealised to be sure, replaces the sharp-edged class of doxastic alternatives by a subjective probability distribution. Thus you may give 90 per cent of your credence to the hypothesis that you are one or another of the possible individuals in *this* class, but reserve the remaining 10 per cent for the hypothesis that you are one of the members of *that* class instead. We can say that a doxastic alternative *simpliciter* is a possible individual who gets a non-zero (though perhaps infinitesimal) share of probability, but the non-zero shares are not all equal.

Precise numerical degrees of belief look artificial, so we might favour a coarser-grained system with some small number of distinct grades of belief. But whatever small number of grades we took, it is likely that our scale would seem sometimes too coarse to capture real distinctions and sometimes too fine to be realistic. A better response is to continue to treat a belief system as a precise numerical probability distribution, but then to say that normally there is no fully determinate fact of the matter about exactly which belief system someone has. There are a range of belief systems that fit him equally well, thought it may be that none fits perfectly; and there is no saying that his real belief system is one rather than another within this range. Then whatever coarse-graining is appropriate comes out as a spread of exact numerical values within the systems in the range. There may be more spread and there may be less; we needn't try to settle once and for all how coarse the grain should be.

We have another reason also to acknowledge that someone may have a multiplicity of belief systems. To a greater or lesser extent, we are all doublethinkers: we are disposed to think differently depending on what question is put, what choice comes before us, what topics we have been attending to. Belief is compartmentalised and fragmented.[23] Sometimes a doublethinking believer acts in a way that best fits one belief system,

[23]See Stalnaker, *Inquiry*, chapter 5; and my 'Logic for Equivocators'.

sometimes in a way that best fits another. And it should not be said just that his belief system changes rapidly; because, throughout, he remains simultaneously disposed toward both systems. In this way also, both systems may fit him equally well even if neither fits perfectly.

In such a case, there are two methods we might follow in saying what someone believes. There is no need to choose between the two once and for all, but it is useful to distinguish them. We might take an intersection, and concentrate on what is common to his many belief systems. Or we might instead take a union, and throw together the different things he believes under different systems.

To illustrate, suppose that hypochondria and good cheer are at war within you. You are simultaneously disposed toward both. Sometimes one is manifest, controlling your thought and conduct; sometimes the other. You have one belief system, the hypochondriac one, under which all your doxastic alternatives are in the early, invisible stages of a dread disease. You have another belief system, the cheery one, under which all your alternatives are healthy. Thus you have entirely different alternatives under the two systems. (Other cases of doublethink would be less extreme, and involve some overlap.) But though the two lots of alternatives differ in respect of health, they have much in common: for instance, all of them live in worlds where the disease in question is incurable. Under the method of intersection, you believe neither that you are diseased nor that you are healthy. Under the method of union, you believe that you are diseased (under one system) and also you believe that you are healthy (under the other). But though you believe that you are diseased and you believe that you are healthy, you do not believe that you are both diseased and healthy; because none of your alternatives under either system, and indeed no possible individual whatever, is both diseased and healthy.

In your state of doublethink, you have no whole-hearted belief about whether you are healthy; you are half-heartedly certain that you are diseased, half-heartedly certain that you are healthy. The two half-hearted certainties are not at all the same thing as partial belief. Your condition is not one of whole-hearted uncertainty about whether you are diseased or healthy, characterised by one unified belief system under which some of your alternatives are diseased, some are healthy, and your subjective probability is divided more or less evenly between the two subclasses. If you had the opportunity to bet on whether or not you were diseased, the difference between the two states would be plain. If you are whole-heartedly uncertain, you hedge your bets. If you are half-heartedly certain each way, you plunge one way or the other – but which way you go depends on exactly how the question is put to you, and on how you're feeling at the time. Indeed, in a more complicated case, belief could be both half-hearted and uncertain: you have one belief system in which your subjective probability is divided evenly between diseased and healthy alternatives,

another where it goes mostly or entirely to diseased alternatives, and still another where it goes mostly or entirely to healthy alternatives.

If content is given by a class of doxastic alternatives (or by a probability distribution), what is characterised is one whole system of belief, not several beliefs – the relevant notion of belief is singular, not plural. This built-in holism is one way in which the present approach contrasts with strategies in which there is a different belief for every different sentence of the languge of thought that is written in the 'belief box'. There is no sensible question whether something is one of your beliefs in its own right, or whether it is merely a consequence of some of your other beliefs. There is no sensible question whether your belief that you are hirsute is or isn't the same belief as your belief that you are hairy; your doxastic alternatives are all hairy, in other words they are all hirsute; and that's that. What is written in your 'belief box', if anything, or what word if any you might use to express yourself, is beside the point.

Of course, we can introduce a derivative notion whereby one belief system brings with it many different beliefs. We could do so in various ways. For instance we could say that each property common to all the believer's doxastic alternatives is one of his beliefs, namely his belief that he has that property. (As a special case, each proposition common to all his belief worlds is one of his beliefs, namely his belief that he inhabits a world where that proposition holds.)

A different way would be to say that he has one belief for every ordinary language belief-ascribing sentence (for short: belief sentence) that is true of him. That would be quite a different thing; because the connection between doxastic alternatives and the truth of belief sentences is far from uniform or straightforward. There are various ways for a system of belief to make a belief sentence true. I cannot propose any unified formula to cover all cases.

One way involves the doxastically accessible worlds. Each of Fred's doxastic alternatives inhabits a world where all things decay; and that is what makes it true to say that Fred believes that all things decay.

A second way involves not the worlds, but the doxastic alternatives themselves. Each of René's alternatives is immaterial; and that is what makes it true to say that René believes that he himself is immaterial. It isn't so, however, that each of René's alternatives inhabits a world where René is immaterial; for we may suppose that René is essentially material – he has no immaterial counterparts – in which case there are no such worlds. This means that René's alternatives are not among his counterparts.[24]

[24]At least, not under any ordinary counterpart relation. We could introduce a special 'counterpart-by-acquaintance' relation on which René's alternatives would be among his counterparts; see my 'Individuation by Acquaintance and by Stipulation'. This just moves the disunification. We get somewhat less variety of ways to make a belief sentence true in return for somewhat more variety of ways to have counterparts.

A third way involves the ascription of properties to things other than oneself via relations of acquaintance. Each of Ralph's doxastic alternatives is watching a spy at work, sneaking through the shadows; Ralph himself is watching Bernard, though he doesn't recognise him; thereby Ralph ascribes spyhood to Bernard; and that is how Ralph believes that Bernard is a spy.[25] It isn't so, however, that each of Ralph's alternatives inhabits a world where Bernard is a spy; for we may suppose that none of the other-worldly spies whom Ralph's alternatives watch is a counterpart of Bernard. Bernard gets into the act not through his other-worldly counterparts, but because he is the one Ralph is actually watching.

A relation of acquaintance needn't be so very direct and perceptual. Other relations will do, so long as they afford channels for the flow of information. For instance there is the relation which obtains when one has heard of something by name. Let us say that one is '*Londres*' – acquainted with something when one has heard of it under the name '*Londres*'. Each of Pierre's doxastic alternatives is '*Londres*'-acquainted with a pretty city; Pierre himself in *Londres*-acquainted with London; thereby Pierre ascribes prettiness to London; and that is how he believes that London is pretty. (See Kripke, 'A Puzzle About Belief'.) Likewise each of Fred's alternatives is 'arthritis'-acquainted with a disease that he has in his thigh; Fred himself is 'arthritis'-acquainted with arthritis; and that is how he believes that he has arthritis in his thigh. (See Burge, 'Individualism and the Mental'.) It isn't so, however, that each of Fred's alternatives has arthritis in his thigh; because arthritis is a disease of the joints which no possible individual has in his thigh. For the same reason, it isn't so that Fred has arthritis in his thigh at his doxastically accessible worlds.

A fourth way involves the acceptance of sentences. Each of Peter's doxastic alternatives is in a position to say truly 'Santa brings presents'; what is more, Peter and his alternatives more or less understand what this sentence means; and that is how Peter believes that Santa brings presents. It isn't so that Peter ascribes present-bringing to Santa under any relation of acquaintance, since there is no Santa for him to be related

[25]The so-called belief sentence 'Ralph believes that Bernard is a spy' has a mixed subject matter. It is not entirely about Ralph's system of belief. It is made true partly by Ralph's psychological state, and partly by his relationship to his surroundings. It is a matter of psychology that his system of belief has content given by a certain class of doxastic alternatives, all of whom watch spies. It is not a matter of psychology that the one he is in fact watching is none other than Bernard.

You might protest that belief is, by definition, that which belief sentences report; and psychology, by definition, covers such phenomena as belief; so if it turns out that relationships of the believer to external things get into the subject matter of belief sentences, then those relationships are *ipso facto* psychological! This may seem far-fetched; but after all it is a mere terminological proposal, and as such is harmless. However, it would compel us to introduce some new name for what hitherto has been called 'psychology', and there seems to be no good reason why we should have to do so.

to. Each of Peter's alternatives is 'Santa'-acquainted with a present-bringer, to be sure, but Peter himself is not 'Santa'-acquainted with anyone. Nor is it so, anyway not clearly, that each of Peter's alternatives inhabits a world where Santa brings presents. To be sure, each of them inhabits a world where someone with a red suit and a belly like jelly and so forth brings presents – but as any reader of *Naming and Necessity* should know, it is one thing to fit the Santa-stereotype, another thing to be Santa.

Four ways, so far, for a system of belief to make a belief sentence true; they cover a lot of the ground, but perhaps not quite all. Here is one further case. Each of Pierre's doxastic alternatives is '*Père Noel*'-acquainted with a present-bringer, although Pierre himself is not '*Père Noel*'-acquainted with anyone. Each of them is in a position to say truly '*Père Noel* brings presents'. (Pierre and his alternatives know English, and are not averse to mixing languages in their speech.) So Pierre believes that *Père Noel* brings presents. So far, it's just like the case of Peter. But also, Pierre believes that Father Christmas brings presents. Why so? Not because Pierre's doxastic alternatives are in a position to say truly 'Father Christmas brings presents' – we may suppose that they are not. Pierre has never heard the name 'Father Christmas', nor has it ever occurred to him to translate the name '*Père Noel*' into English. Presumably it's crucial that the two denotationless names '*Père Noel*' and 'Father Christmas' emerge from one tradition common to speakers of English and French. If there had been two fortuitously similar stories and if Pierre had been out of touch with the English story, then it would have been false to say that Pierre believes that Father Christmas brings presents. But how to work that fact into a general analysis of belief sentences? – Never mind; I have made my point that the connection of belief sentences with belief as characterised by doxastic alternatives is complicated and multifarious.

The use of classes of *possibilia* to specify content is supposed to be discredited by the way it imputes logical omniscience. Not so. We have seen several ways for someone to fall into inconsistency, either by holding impossible beliefs or by holding possible beliefs that conflict with one another.

(1) There is doublethink, as when our hypochondriac believes that he is healthy and also believes, but in a different compartment, that he is diseased. That is an extreme case. Often the walls of the compartments will be weaker and more temporary, due more to momentary inattention than to underlying confusion, and yet sufficient to produce lapses from logical perfection. Consider an everyday failure to draw a conclusion from several premises that one believes. Stalnaker (*Inquiry*, chapter 5) has shown how this can be explained as a case of compartmentalised thinking. Take the simplest way to believe something: a proposition holds throughout

your doxastically accessible worlds. Suppose that you believe that P, also you believe that Q, and P and Q jointly imply R in the sense that every world that is both a P-world and a Q-world is also an R-world; nevertheless, we may suppose that you fail to believe R. We may even suppose that *none* of your doxastically accessible worlds is an R-world. How can this be? – The answer is that you may be thinking double, with P and Q in different compartments. You believe that P by believing it in one system; that one gives you doxastically accessible worlds where P holds but Q and R do not. You believe that Q by believing it in the other system; that one gives you doxastically accessible worlds where Q holds but P and R do not. Thus you believe P and you believe Q, though in both cases half-heartedly; but you whole-heartedly disbelieve the conjunction of P and Q, and you whole-heartedly disbelieve R. You fail to believe the consequence of your two premises taken together so long as you fail to take them together.

(2) When René, an essentially material thinking thing, believes that he himself is immaterial, he self-ascribes a property contrary to his essence, and thereby believes the impossible. Likewise someone might ascribe to something else, via some relation of acquaintance, a property contrary to its essence.

(3) Someone might ascribe conflicting properties to the same thing via two different relations of acquaintance.[26] Pierre is both '*Londres*'-acquainted and 'London'-acquainted with London: each of his doxastic alternatives is '*Londres*'-acquainted with a pretty city and 'London'-acquainted with an ugly one; and that is how Pierre has inconsistent beliefs, believing that London is pretty and also believing that London is ugly. Of course none of his alternatives is in any way acquainted with anything that is both pretty and ugly, because there are no such things in any world to be acquainted with. It would not, I think, be true to say that Pierre believes that London is both pretty and ugly. (But if that were true, it would just go to show that belief sentences work in even more miscellaneous ways than I have given them discredit for – it would not be an objection to what I am saying.) This failure of beliefs to conjoin may suggest a case of doublethink; but it is not the same thing. I don't know whether leading philosophers and logicians like Pierre are less prone to doublethink than the rest of us, but at any rate Pierre is a paragon of mental unity. Far from keeping his '*Londres*'-thoughts and his 'London'-thoughts in separate compartments, he constantly bemoans his fate: 'Would that I had fetched up in *la belle Londres* instead of this dump London!' There is nothing in the least contradictory or impossible about

[26]Cresswell and von Stechow show how to account for arithmetical error along the lines of (2) and (3), provided that there is something akin to a relation of acquaintance that we can bear to numbers.

Pierre's alternatives or the worlds they are part of. Of course, that is because the alternatives – unlike Pierre himself, who is not one of them – are never '*Londres*'-acquainted and 'London'-acquainted with the same city.

(4) Someone could believe that a sentence is true when in fact it is subtly contradictory. Thus we may suppose that each of Duntz's doxastic alternatives is in a position to say truly 'There is a barber who shaves all and only those who do not shave themselves'; and that is how Duntz believes there is such a barber, and thereby believes the impossible. Of course, nobody could be in a position to say it truly and mean by it *exactly* what we (or Duntz) would mean; so none of the doxastic alternatives has the meaning exactly right. Note well that this is not the sort of case where Duntz has no idea what the sentence means, and only thinks that it means something or other true; in that case it would be wrong to describe his belief by indirect quotation. No; the indirect quotation is legitimate because he has a pretty good idea what the sentence means, even if his understanding is not quite good enough to enable him to notice the contradiction.[27] In summary: if we characterise content by means of *possibilia* we need not ignore the phenomenon of inconsistent belief. On the contrary, we are in a position to distinguish several varieties of it. *All* the varieties? – That question, no doubt, remains open.

If the content of belief, as given in terms of the subject's doxastic alternatives, is not tied in any uniform and straightforward way to the truth of ordinary language ascriptions of belief, and also is not tied to the subject's acceptance of inner sentences, how is it tied down at all? I would say that it is tied down mainly by belief–desire psychology. We suppose that people tend to behave in a way that serves their desires according their beliefs. We should take this principle of instrumental rationality to be neither descriptive nor normative but *constitutive* of belief. It enters into the implicit definition of what it is for someone to have a certain system of belief.

That is a rough approximation, and there is more to be said. The first thing is that what fits behaviour is not a system of belief alone but rather a combined system of belief and desire. Not only are the possible individuals divided into those which are and are not doxastic alternatives for the subject; also, there are some which he would rather be than others. In general, both belief and desire will admit of degree. Saying what it

[27]We may ask how it is that Duntz fails to notice the contradiction. He knows enough: we may suppose that he believes each of several premises, having to do with various aspects of the syntactic structure of the sentence and the meanings of the words, and from these premises taken together it follows that the sentence is contradictory. Then how can he fail to draw the conclusion? – We have addressed this question already. Duntz is no doubt a doublethinker, and never puts together all the things he knows. The different ways of falling into inconsistency interact, and Duntz combines our cases (1) and (4). See Stalnaker, *Inquiry*, chapters 4 and 5.

means for behaviour to fit a system of degrees of belief and desire is the business of decision theory. But here it will suffice to look at an absurdly simplified case, devoid of degrees or gradations: all black or white, no shades of grey. On the side of belief, some possible individuals are doxastic alternatives for the subject and others are not. On the side of desire, some individuals belong to the class in which the subject would prefer to be and others do not. (It is not assumed that the subject's preferences are selfish; maybe the preferred class consists of those individuals who inhabit possible worlds where mankind generally flourishes.) Now suppose that there is a certain bodily movement, which the subject is able to perform at will; and which is maximally specific with respect to his ability, so that he would not be able at will to perform it in one more specific way rather than another. Let it be waving the left hand in a certain way (for short: waving). Suppose further that each of the subject's doxastic alternatives is such that, if he were to wave, he would be in the preferred class. We understand this in terms of closeness of worlds and in terms of counterparts: each alternative is such that the closest world to his where his counterpart waves is one where his counterpart belongs to the preferred class. (We want the kind of closeness of worlds that's right for causal counterfactuals. We ignore complications about what happens if there are several counterparts in one world, or if several among the worlds where counterparts wave are tied for closest.) Then waving is a piece of behaviour that serves the subject's desires according to his beliefs. If he does wave, then to that extent the system of belief and desire in question is a system that fits his behaviour.

Besides the fit of belief and desire to behaviour at a moment, there is also fit over time. One way to think of this would be as fit between a succession of systems of belief and a stream of evidence: the changes in belief are as they should be, given the evidence. But it is easier to think of it as fit between the momentary system of belief and desire and present dispositions to follow contingency plans whereby future behaviour depends on what happens meanwhile. That way we can continue to concentrate on the present system of belief and desire of the momentary subject. Return to our simple case, all black and white, and elaborate it further. Suppose the rest of us are in a car parked near a restaurant; the subject is supposed to walk over and wave to us if the restaurant turns out to be open and not too crowded. What serves the subject's desires according to his beliefs is not waving now, and not waving unconditionally later, but rather following a certain contingency plan to wave or not depending on what he sees. He is able to follow this contingency plan at will, and it is maximally specific with respect to his ability. Each of the subject's doxastic alternatives is such that, if he were to follow the plan, he would be in the preferred class. That is, each of them is such that the closest world to his where his counterpart follows the plan is one where his

counterpart belongs to the preferred class. Then if the subject is now disposed to follow the plan in whatever way turns out to be right when he gets to the restaurant, to that extent the system of belief and desire in question is a system that fits his present behavioural dispositions.

(How does a momentary stage follow a plan that covers a period of time? – By being the first of a succession of suitably interrelated stages which together follow the plan. What makes a momentary stage able, in this sense, to follow a plan? – The fact that belief changes under the impact of evidence in such a way that, whatever may be observed, continuing to follow the plan will be the behaviour that fits the sytem of belief and desire of each subsequent stage. So the epistemic rationality of belief change has not, after all, been passed by; it is still there within the supposition that the subject is able to follow the contingency plan.)

What makes an assignment of a system of belief and desire to a subject correct cannot just be that his behaviour and behavioural dispositions fit it by serving the assigned desires according to the assigned beliefs. The problem is that fit is too easy. The same behaviour that fits a decent, reasonable system of belief and desire also will serve countless very peculiar systems. Start with a reasonable system, the one that is in fact correct; twist the system of belief so that the subject's alleged class of doxastic alternatives is some gruesome gerrymander; twist the system of desire in a countervailing way; and the subject's behaviour will fit the perverse and incorrect assignment exactly as well as it fits the reasonable and correct one.[28] Thus constitutive principles of fit which impute a measure of instrumental rationality leave the content of belief radically underdetermined.

However, instrumental rationality, though it is the department of rationality that has proved most tractable to systematic theory, remains only one department among others. We think that some sorts of belief and desire (or, of dispositions to believe and desire in response to evidence) would be unreasonable in a strong sense – not just unduly sceptical or rash or inequitable or dogmatic or wicked or one-sided or short-sighted, but utterly unintelligible and nonsensical. Think of the man who, for no special reason, expects unexamined emeralds to be grue. Think of Anscombe's example (in *Intention*, section 37) of someone with a basic desire for a saucer of mud. These beliefs and desires are unreasonable; though if twisted desire is combined with correspondingly twisted belief, then it may be that the failing lies entirely outside the purview of the department of instrumental rationality. So I say that other departments of rationality also may have a constitutive role. What makes the perversely twisted assignment of content incorrect, however well it fits the subject's behaviour, is exactly that it assigns ineligible, unreasonable content when

[28] I have shown how this can happen in my 'New Work for a Theory of Universals', pages 374–5, though only for a very simplified case.

a more eligible assignment would have fit behaviour equally well. The theory that implicitly defines the functional role of belief and desire, and so specifies *inter alia* what it is for a possible individual to be one of the subject's doxastic alternatives, is the constitutive theory not just of instrumental rationality but of rationality generally.[29]

I have objected to the radical indeterminacy, especially the indeterminacy between reasonable and perverse systems of belief and desire, that would result if we tried to get by with instrumental rationality as the only constitutive constraint. But I do not object at all to milder forms of indeterminacy. Far from being something forced upon us by the requirements of some theory, it seems independently plausible that there might be no straightforward and determinate fact of the matter about what a doublethinker does or doesn't believe. I said before that in cases of doublethink, or less remarkably in cases where the exact degrees of belief are indeterminate, someone might have multiple belief systems; none would fit him perfectly, all would fit him about equally well, and well enough. Now I have said what sort of fit I had in mind.

There is one further complication; doubtless not the last, but the last that I wish to consider here. I have been speaking as if the assignment of content were an assignment directly to a given subject. But I would rather say that the content belongs to some state – a brain state, perhaps – that recurs in many subjects. It recurs in many subjects in many worlds, the worlds being sufficiently similar in the anatomy of their inhabitants and in the relevant laws of nature; and it recurs in many subjects even in the same world, for instance if it is a world of eternal recurrence or if it is a world where the inhabitants' brains have a lot of hard-wiring in common. The recurrent state would tend to dispose anyone who had it to behaviour fitting a certain reasonable assignment of content. Therefore we can say that the state *is* a system of belief and desire with that content, and when a subject has that state he thereby has the content that belongs to the state. The reason why I prefer to attach content to the state, rather than directly to the subject, is that it leaves room for exceptional cases in which, despite the constitutive role of principles of fit, the subject's behaviour somehow fails to fit his system of belief and desire. I said that the state *tends* to dispose anyone who has it to behave in a certain way; but such a tendency might be defeated. Compare a state of a pocket calculator: that state tends, throughout all the calculators built to a certain plan, to cause '137' to be displayed when the 'recall' key is pressed, and so on; wherefore we call it the state of having the number 137 stored in memory. But there are a few calculators with defective 'recall' keys; they get into the very same state, but you press the key and nothing happens.

[29]See section 2.3; my 'New Work for a Theory of Universals', pages 373–7; and Grandy, 'Reference, Meaning and Belief', on 'principles of humanity'.

We can say of them along with the rest, by courtesy, that they have 137 stored in memory; and this is defined in terms of what the state tends to cause; but in the defective calculators the tendency is defeated. The state of the memory gets its numerical content in virtue of what it would generally, but not invariably, tend to cause; and so it might be, also, with a brain state which is assigned content as a system of belief and desire.[30]

Possible worlds and individuals are useful not only in connection with thought but also for the analysis of language. Suppose we want a systematic grammar, covering not only syntax but semantics, for a natural language or some reasonable imitation or fragment thereof. Such a grammar is meant to plug into an account of the social practice of using language. It encapsulates the part of the account that is different for different linguistic communities who are party to different conventions of language. What makes the grammar correct for a given population is that, when plugged into its socket, what results is a correct description of their linguistic practice – of the way they suit their words to their attitudes, of the way they suit their attitudes to others' words, and of their mutual expectations concerning these matters.

A principal way we use language is in conveying needed information. You know whereof you speak, and you want me to know something; so you tell me something true; I rely on you to know whereof you speak and be truthful; and that is how I come to have the knowledge you wanted me to have. But when you tell me the truth, and when I rely on you to be truthful, your words will not be true *simpliciter*. They will be true under some semantic interpretations and false under others. The right interpretation, for us, is the one that specifies truth conditions under which we are indeed truthful and do indeed rely on one another's truthfulness. So if a grammar is to plug into its socket in an account of the use of language, it has to specify truth conditions for (many or all) sentences of the language.

These may well depend on the circumstances of utterance. A sentence is said by some particular speaker, at some particular time, at some particular world. Further, it is said at a certain place; to a certain audience; accompanied perhaps by certain gestures of ostension; in the presence of certain conspicuous things; and in the context of previous discourse which influences what is to be presupposed, implicit restrictions of quantifiers, prevailing resolution of vagueness, and much more. All these things may be relevant to whether that sentence can be said truly. But speaker, time, and world determine the rest: the place is the place where that speaker is at that time, audience consists of those present whom the speaker intends to address, and so on.

[30]See my *Philosophical Papers*, volume I, pages 119–21.

I might even say that the *speaker* determines the rest. The appropriate world is the world that he is part of. As for time, of course it is not to be denied that we persist through time and speak at different times. But we do so by being composed of different temporal stages. The stages also may be called speakers; and if it is the momentary speaker we mean, then the appropriate time is the time at which the speaker is.

So the speaker, at a definite world and time, is one of those momentary subjects of attitudes just considered. His knowledge and belief are given by his epistemic and doxastic alternatives – those possible momentary individuals who might, for all he knows or believes, be himself. He can speak truly by luck if the sentence he says is true for him; but to exhibit the sort of truthfulness that members of a linguistic community expect from one another, the things he says will have to be true not only for him but also for all his alternatives. When language is used to convey information between truthful and trusting partners, the communication may take place all in this world; but nevertheless the truth conditions must involve other-worldly individuals. To plug into its socket in an account of the use of language, a semantically interpreted grammar has to specify which speakers at which times at which worlds are in a position to utter which sentences truly.

Then it must accomplish an infinite specification by finite means. Here is a way that can be done. First list a finite vocabulary of basic expressions – words, near enough – and assign each of them some sort of syntactic category and semantic value. Then list rules for building expressions from other expressions; and within each rule, specify the syntactic category and semantic value of the new expression as a function of the categories and values of the old expressions whence it was built. One syntactic category will be the sentences. Then specify truth conditions for sentences in terms of their semantic values.

The semantic values have two jobs. They are there to generate other semantic values; and they are there to generate truth conditions of sentences. The second job is what the whole system of semantic values is for; the first job is what gives us a whole system of semantic values.

I have said all this in a skeletal fashion, intending to say something that will be neutral between many conceptions of what the system of vocabulary, rules, categories, and semantic values might look like. For the same reason, I have chosen the colourless term 'semantic values' instead of some more familiar term that would convey some more definite idea of what the values might be and how they might do their job. The object is not that we should find entities capable of deserving names from the established jargon of semantics, but that we should find entities capable of doing the pair of jobs.[31]

[31]For instance, I don't think we should say that an ordinary proper name *refers* to a

We have a choice of strategies. What we want from our system of semantic values is a specification of which sentences are true for which of all the (momentary) speakers scattered through the worlds. We might put context-dependence outside the semantic values – call this the *external* strategy – by making the entire assignment of semantic values, from the words on up, be speaker-relative. Since different speakers are part of different worlds, this initial speaker-relativity brings *possibilia* into the picture, no matter what the semantic values themselves might look like. For a given speaker and sentence, we have first the semantic values for that speaker of each word of the sentence. In accordance with the rules of the grammar, these generate the semantic values for that speaker of expressions built up from these words. Among those expressions is the sentence itself; and the semantic value of the sentence for the speaker somehow determines whether it is true for him. We want a semantic value for a sentence, relative to a speaker, to deliver a truth value. We might even hope that it could just *be* a truth value – call this the *extreme* external strategy.

At the opposite extreme, we could assign semantic values once and for all, and put all the context-dependence inside them – call this the *internal* strategy. In that case *possibilia* may enter into the construction of the semantic values themselves. Else it will be hard for the fixed semantic value of a sentence to determine which of the speakers scattered over the worlds that sentence is true for.

In between, we might of course mix the two methods. We could put some of the context-dependence inside the semantic values, and some of it outside in the speaker-relativity of semantic values – call this the *moderate* external strategy.[32]

To illustrate this difference of strategies, and to illustrate various other choices and problems that arise, it will help to look at a miniature language. We shall have one kind of modification, namely modification of sentences; but that will do to illustrate phenomena that could take place also for modification of common nouns, verbs, quantifiers, and modifiers themselves in a more elaborate language. Our little language will have a categorial grammar with three categories altogether, one basic and two derived: *sentence*, *modifier*, *connective*. There are basic expressions in

---

bundle of properties. My name, for instance, refers to me – and I am not a bundle of properties. Property bundles might nevertheless be serviceable semantic values for proper names, along with other noun phrases. (See my 'General Semantics', section VII; and Montague, *Formal Philosophy*, chapter 8.) If so, it would be unwise to use 'refer' as our word for having a semantic value. There is, of course, no reason not to say both that my name has me as its referent and also that it has a certain property bundle as its semantic value.

[32]An example of a pure internal strategy is my treatment in 'General Semantics'. Moderate external strategies are to be found in Montague's papers on natural language in *Formal Philosophy*, and in Cresswell, *Logics and Languages*.

all three categories. What the semantic values for sentences are remains to be seen; a semantic value for a modifier is a function from semantic values for sentences to semantic values for sentences; a semantic value for a connective is a function from pairs of semantic values for sentences to semantic values for sentences. There are two grammatical rules.

Rule for Modifiers. If S is a sentence with semantic value s, and M is a modifier with semantic value m, then MS is a sentence with semantic value m(s).

Rule for Connectives. If S_1 and S_2 are sentences with semantic values s_1 and s_2 respectively, and C is a connective with semantic value c, then CS_1S_2 is a sentence with semantic value $c(s_1, s_2)$.

Given this much, all else depends on the basic expressions and their semantic values.

First let's try treating the language in an extreme external fashion: semantic values are assigned relative to a speaker, semantic values for sentences are mere truth values, semantic values for modifiers and connectives are made to fit and therefore are functions from and to truth values. For a little while all goes well. We have two basic sentences. They exhibit two kinds of context-dependence, both handled externally.

'Rains' is a basic sentence; its semantic value for any speaker is *truth* iff, at the world and time and the vicinity of the place where that speaker is, it is raining.

'Cold' is a basic sentence; its semantic value for any speaker is *truth* iff, at the world and time and the vicinity of the place where that speaker is, the temperature is below a certain level. This level is somewhat flexible, and depends on the previous course of the conversation in which the speaker is participating. If someone says something that requires a shift of the border to make it true for him, thereby the border shifts.

We also have one modifier and one connective, both truth-functional.

'Not' is a basic modifier; its semantic value for any speaker is the function that maps either truth value to the other.

'Iff' is a basic connective; its semantic value for any speaker is the function that maps a pair of truth values to *truth* iff both truth values in the pair are the same, and to *falsity* otherwise.

(We could have had a context-dependent modifier or connective; for some speakers its semantic value would be one truth-function, for others another. I omit an example.)

So far, so good. But suppose our little language also includes the modifier 'possibly'; and suppose that a sentence 'Possibly ϕ' is to be true for a speaker iff ϕ is true under some shift of world. (Let us postpone the important question of what happens to the speaker, and his time and place and so forth, when we shift worlds.) That frustrates the extreme external strategy. If semantic values for sentences are just truth values, there is, of course, no way we can derive the semantic value for a given speaker of 'Possibly ϕ' from the semantic values for that speaker of 'possibly' and of ϕ. The trouble is that we've discarded information about the truth value of ϕ for other worlds than the speaker's own. It would do us no good to reconstruct the grammatical rule for modifiers, abandon the function-and-argument method of generating semantic values for modified sentences, and devise some fancy semantic value for 'possibly'. Once the needed information is gone, we can't bring it back.

(But if the rule said that the semantic value of 'Possibly ϕ' for *this* speaker depends on the semantic value of ϕ for *other* speakers, then couldn't the semantic values be truth values? – There is a question, still postponed, of what happens if the world-shift takes us to a world with no speakers. But even setting that aside, the proposal rests on a misunderstanding. To be a semantic value is to be a big enough package of information. A semantic value worthy of the name must carry all the information that will be needed to generate other semantic values. Anything that we need to bundle together many of to get a big enough package is *ipso facto* not an adequate semantic value.)

Since the extreme external treatment fails, we have a choice between a moderate external and an internal treatment. The moderate external alternative could go as follows. Let our new semantic values for sentences be functions from worlds to truth values; then we get our truth conditions by saying that a sentence is true for a speaker iff its semantic value, for that speaker, assigns truth to that speaker's world. The rest gets adjusted to fit. Our new semantic values for modifiers and connectives are functions to and from the new semantic values for sentences. The rules for modifiers and connectives have the same form as before. As for the basic expressions:

> 'Rains' is a basic sentence; its semantic value for any speaker is the function that assigns *truth* to all and only those worlds W such that, for some counterpart X in W of the speaker, it is raining at W at the time and the vicinity of the place where X is. ('Cold' is similar.)
>
> 'Not' is a basic modifier; its semantic value for any speaker is the function that maps f to g iff both are functions from worlds to truth values and g(W) is *truth* when and only when f(W) is *falsity*.
>
> 'Iff' is a basic connective; its semantic value for any speaker is the function that maps e and f to g iff all three are functions from worlds

to truth values and g(W) is *truth* when and only when c(W) and f(W) are the same.

'Possibly' is a basic modifier; its semantic value for any speaker is the function that maps f to g iff both are functions from worlds to truth values and either g(W) is *truth* for all worlds and f(W) is *truth* for some world or else g(W) and f(W) both are *falsity* for all worlds.

Now we have accommodated the modifier 'possibly', thanks to the world-dependence within the semantic values. But there is still external context dependence; the semantic value for me of the basic sentence 'Rains' has to do with rain in the vicinity of my counterparts, the semantic value for you of 'Rains' has to do with rain in the vicinity of your counterparts.

I still haven't put in a context-dependent modifier or connective, but a true-to-life example could be now given: 'possibly' with accessibility restrictions, where the appropriate restrictions are somewhat flexible and depend, for a given speaker, on the previous course of the conversation in which the speaker is participating. Similarly, inconstancy in the counterpart relation (see section 4.5) could create another dimension of context-dependence, besides the sort already noted, in the semantic value of 'Rains'.

The present semantic values for sentences might look little different from the truth conditions that the whole system of semantic values is built to deliver. However, suppose our language turns out to contain another basic sentence.

'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.

'Am' has quite a simple truth condition: it is true for any speaker whatever. (Assuming, as I do, that anything is one of its own counterparts.) But its semantic values, for various speakers, are not so simple. In general, they will assign *truth* to the world where the speaker in question is and to some but not all other worlds. That's how 'Possibly not am' can come out true for a speaker, as, of course, it should. Call 'Am' a case of the 'contingent *a priori*' if you like – though it seems doubtful that there is any *one* thing to which both adjectives apply.

Given a speaker, his world is given; but when we shift worlds in connection with 'possibly', we don't necessarily shift speakers. What happens to the speaker when we shift worlds (our postponed question) may be that he completely disappears. We may shift to a world where there is no counterpart of a given speaker; that is how 'Possibly not am' comes out true. We might even shift to a world where there are no speakers

at all. Worlds started out fixed to speakers, but now they are varying independently.

So far, our moderate external strategy is working nicely; but now suppose it turns out that our little language contains some modifiers we haven't yet taken into account. Suppose there is 'past', and a sentence 'Past ϕ' is to be true for a (momentary) speaker iff ϕ is true, not with respect to the time when the speaker is, but with respect to some earlier time. Now we have to start over once more, taking semantic values for sentences as functions from world-time pairs (such that the time exists at the world) to truth values, and adjusting the rest to fit. Then we can say, for instance:

> 'Past' is a basic modifier; its semantic value for any speaker is the function that maps f to g iff both are functions from world-time pairs to truth values and g(W,t) is *truth* when and only when f(W,t′) is *truth* for some time t′ that exists at world W and is earlier than t.
>
> 'Rains' is a basic sentence; its semantic value for a given speaker is the function f from world-time pairs to truth values that assigns *truth* to all and only those pairs of a world W and time t such that, for some counterpart X in W of the speaker, it is raining at W at t in the vicinity of the place where X is.

As with worlds and 'possibly', so with times and 'past'. Given a speaker, his time is given; but when we shift times in connection with 'past', we never shift speakers. (For a speaker is momentary, and if present at one time he will never be found at an earlier time.) So when we speak of rain at t in the vicinity of the place where X is, that will not be his place at t – he has none – but his place when he exists.

And next suppose there is 'sorta', and a sentence 'Sorta ϕ' is to be true for a speaker iff ϕ is true for him under an adjustment of context-dependent flexible borders – such as the border for what counts as cold – that makes it easier for ϕ to be true. So 'Sorta cold' is true when it isn't quite cold enough to make 'Cold' true; 'Sorta not cold' is true when it isn't quite warm enough to make 'Not cold' true; 'Sorta sorta cold' is true when it isn't quite cold enough to make 'Sorta cold' true; and so on. We could make yet another new start, taking semantic values for sentences now as functions from world-time-border triples to truth values, and adjusting the rest yet again.

Is there no end to this? Maybe, maybe not. I'm making up the story of this little language as I go along, so let me make an end to it. Here is a conceivable phenomenon that turns out *not* to happen. There *isn't* a modifier 'reversedly' such that a sentence 'Reversedly ϕ' is true for a speaker iff ϕ is true for some hearer he is addressing. If there had been, we would have had to go back and take semantic values for sentences as functions from world-time-border-speaker quadruples; since it doesn't

happen (just as no such thing happens in English) perhaps we needn't. We can leave the speaker-relativity external to the semantic values.

By now the moderate external strategy has come to look cumbersome, and so we might wish we'd tried the internal alternative instead. The simplest method would be to say that a semantic value for a sentence, assigned once and for all, is a function from speakers to truth values. Again the semantic values for modifiers can be made to suit, and the rule for modifiers can prescribe a function-and-argument method of generating the semantic value of a modified sentence; and likewise for connectives. We read the truth conditions of a sentence directly off the semantic value.

Life cannot be that easy. Consider two sentences: 'Am' and 'Iff rains rains'. Both have the same truth condition: true for any speaker whatever. But they can't both have the same semantic value; because when we apply two more modifiers we get sentences 'Possibly not am' and 'Possibly not iff rains rains' which cannot have the same semantic value because they do not have the same truth conditions. The second is false for any speaker whatever; not so for the first.

So a better internal strategy would be to say that a semantic value for a sentence, assigned once and for all, is a function from speaker-world pairs to truth values. Adjust the rest to fit. A sentence is true for a speaker iff its semantic value assigns *truth* to the pair of that speaker and his own world. Now we can handle our problem about the two sentences, as follows.

> 'Rains' is a basic sentence; its semantic value is the function that assigns *truth* to all and only those pairs of a speaker Y and world W such that, for some counterpart X in W of Y, it is raining at W at the time and the vicinity of the place where X is.
>
> 'Am' is a basic sentence; its semantic value is the function that assigns *truth* to all and only those pairs of a speaker Y and world W such that W contains a counterpart of Y.
>
> 'Not' is a basic modifier; its semantic value is the function that maps f to g iff both are functions from speaker-world pairs to truth values and g(Y,W) is *truth* when and only when f(Y,W) is *falsity*.
>
> 'Iff' is a basic connective; its semantic value is the function that maps e and f to g iff all three are functions from speaker-world pairs to truth values and g(Y,W) is *truth* when and only when e(Y,W) and f(Y,W) are the same.
>
> 'Possibly' is a basic modifier; its semantic value is the function that maps f to g iff both are functions from speaker-world pairs to truth values and, for any Y, either g(Y,W) is *truth* for all worlds and f(Y,W)

is *truth* for some world or else g(Y,W) and f(Y,W) both are *falsity* for all worlds.

Now we can check that, because the embedded sentences 'Am' and 'Iff rains rains' have different semantic values despite their sameness of truth conditions, the sentences 'Possibly not am' and 'Possibly not iff rains rains' differ not only in semantic values but in truth conditions. As we would expect, the first is true for any speaker unless he has counterparts at all the worlds; the second is true for no speaker.

This is very like what we saw before under a moderate external strategy in considering the behaviour of 'possibly' and 'Am'. We needed to let world vary independently of speaker, despite the fact that a world is originally given as the world of a speaker. Taking speaker-world pairs is just another way to get independent variation. The pair delivers worlds twice over, not necessarily the same world both times, because there is the world of the speaker who is the first term of the pair and there is the world that is the second term of the pair.[33]

If we go on to consider the modifier 'past' under the internal strategy, we will find ourselves forced to say that the semantic values for sentences, assigned once and for all, are functions from speaker-world-time triples to truth values. And if we next consider 'sorta', we will have to say instead that they are functions from speaker-world-time-border quadruples. This begins to seem cumbersome. It's good luck that 'reversedly' is absent from the language, so that we may be spared functions from speaker-world-time-border-speaker quintuples. – Plainly, we are covering the same ground twice. There is no great divide between the moderate external and the internal strategies. There is a trivial translation between a speaker-relative assignment of semantic values that are functions from world-time-border triples and an assignment, once and for all, of semantic values that are functions from speaker-world-time-border quadruples. If pursued satisfactorily, the two strategies come to the same thing.[34]

It is clear from our little language that sameness of truth conditions – in the sense I gave to that phrase – does not imply sameness of meaning. Else 'Am' would mean the same as 'Iff rains rains', which surely it doesn't. It is less clear whether we should say that sameness of semantic values implies sameness of meaning. The semantic values are the same for 'Rains' and 'Not not rains'; or for 'Iff rains rains' and 'Iff am am'. Do these sentences mean the same or not?

[33]Either way, we have a form of 'double indexing'. See van Fraassen, 'The Only Necessity is Verbal Necessity', for discussion of the uses and origins of this device.

[34]For further discussion of this point, see my 'Index, Context, and Content'.

I think this is not a real question. Is there really anything in our theoretical or everyday use of the term 'mean' to suggest that we have settled the matter – settled it unequivocally, settled it the same way each time someone undertook to settle it? No, it is just a question of what to mean by 'mean'. Given a superfluity of more or less interchangeable semantic jargon, none of it very precisely pinned down, perhaps it might be convenient to reserve 'meaning' for the fine-grained notion of something that differs when – as in the examples just noted – we generate the same semantic value by different routes.

If this is what we want 'meanings' to be, we can let them encode the way a semantic value is generated. In view of the artificial simplicity of our illustrative language, it is an easy matter to let the generation of meanings go piggyback on the generation of semantic values, as follows. (For simplicity let's follow the internal strategy; if we preferred the external strategy, we could let meanings be speaker-relative along with the semantic values.) (1) The meaning of any basic expression is its semantic value. (2) If S is a sentence and M is a modifier, then the meaning of the sentence MS is the sequence of the meaning of M and the meaning of S. (3) If S_1 and S_2 are sentences and C is a connective, then the meaning of the sentence CS_1S_2 is the sequence of the meaning of C, the meaning of S_1, and the meaning of S_2. So a meaning amounts to a parsed expression with semantic values of words put in where the words themselves should be. Meanings determine semantic values; but not conversely, as witness the different meanings of 'Rains' and 'Not not rains' or the different meanings of 'Iff rains rains' and 'Iff am am'.[35]

Because meanings carry more information that semantic values (anyway, the semantic values so far considered) we can use them to make distinctions which would not show up in semantic values. Consider differences of triviality. Suppose that for every speaker, there is some world where he lacks a counterpart; that is a non-contingent matter, but it is far from trivial. It depends on just what the other worlds are like, on what sort of thing exactly can qualify as a 'speaker', and on the counterpart relation. If so, the semantic value of 'Possibly not am' is a constant function that always takes the value *truth*. So the sentence is a necessary truth, but it is not trivially so. The semantic value of 'Iff rains rains' is exactly the same; this sentence too is a necessary truth, but this time trivially so. This difference in triviality is captured by a difference of meanings; but not by a difference of semantic values, for there is no difference of semantic values.

[35]For further discussion of meanings, see my 'General Semantics'; for background, see Carnap on 'intensional isomorphism', *Meaning and Necessity*, section 14; and C. I. Lewis on 'analytic meaning' in 'The Modes of Meaning'.

(This raises a difficult problem.[36] Suppose it turns out that we have the modifier 'trivially' *within* our little language, and it works as we might expect. Then 'Trivially possibly not am' should be false for every speaker, but 'Trivially iff rains rains' should be true. This suggests that what we have been calling the 'semantic values' are not really quite big enough packages of information to do their jobs and deserve their names; and what we have been calling the 'meanings' are the things that really can do the job of the semantic values and deserve to be so called. Maybe something of the sort could and should be permitted, but it is not at all easy. The trouble comes when we ask what is the semantic value of 'trivially' itself? Our previous practice would lead us to think that it is a function which takes as argument the semantic value – hitherto called 'meaning' – of a sentence ϕ, and yields as value something whence we can retrieve the truth condition of 'Trivially ϕ'. Now let ϕ be the sentence 'Trivially iff rains rains'; and we have an argument of a function outranking the function itself in the set-theoretic hierarchy, which is impossible. What to do? Resort to queer set theory? Claim that it was illicit to stipulate that our little language contains the sentence 'Trivially trivially iff rains rains'? Allow the sentence, but insist that it can have no truth condition? Require the first 'trivially' and the second in the sentence to be homonymous words with different semantic values? No solution seems very nice.)

1.5 Modal Realism at Work: Properties

We have frequent need, in one connection or other, to quantify over properties. If we believe in possible worlds and individuals, and if we believe in set-theoretic constructions out of things we believe in, then we have entities suited to play the role of properties.

The simplest plan is to take a property just as the set of all its instances – *all* of them, this- and other-worldly alike. Thus the property of being a donkey comes out as the set of all donkeys, the donkeys of other worlds along with the donkeys of ours.[37]

[36]For discussion of it see Cresswell, 'Hyperintensional Logic', and Bigelow, 'Believing in Semantics'.

[37]I say 'set' not 'class'. The reason is that I do not want to restrict myself to properties of individuals alone; properties themselves have properties. Properties must therefore be sets so that they may be members of other sets.

When I use the term 'set' and 'class' in this book, the reader would not go far wrong to suppose that I am following the standard usage: 'class' is the more general term, and covers not only sets but also 'proper' classes'. Those are supposed to be set-like things which, by reason of the boundless rank of their members, are somehow disqualified from membership in any class or set. But in fact I use the terms to mark a somewhat different distinction, as follows. It is sometimes suggested that

The usual objection to taking properties as sets is that different properties may happen to be coextensive. All and only the creatures with hearts are creatures with kidneys; all and only the talking donkeys are flying pigs, since there are none of either. But the property of having a heart is different from the property of having a kidney, since there could have been an animal with a heart but no kidneys. Likewise the property of being a talking donkey is different from the property of being a flying pig. If we take properties as sets, so it is said, there is no distinguishing different but accidentally coextensive properties.

But according to modal realism, these 'accidentally coextensive' properties are not coextensive at all. They only appear so when we ignore their other-worldly instances. If we consider all the instances, then it never can happen that two properties are coextensive but might not have been. It is contingent whether two properties have the same this-worldly instances. But it is not contingent whether they have the same instances *simpliciter*.

It is a mistake to say that if a property were a set, then it would have its instances – its members – essentially, and therefore it never could be contingent whether something has or lacks it. Consider the property of being a talking donkey, which I say is the set of all talking donkeys throughout the worlds. The full membership of this set does not vary from world to world. What does vary from world to world is the subset we get by restricting ourselves to the world in question. That is how the number of instances is contingent; for instance, it is contingently true that the property has no instances. Further, it is a contingent matter whether any particular individual has the property. Take Brownie, an

---

there is an irreducibly plural way of referring to things, or quantifying over them. I say 'There are some critics such that they admire only one another' or 'There are all the non-self-members, and they do not comprise any sort of set or class', and I am not quantifying in the ordinary way over any set or class of critics or of non-self-members; rather I am quantifying over nothing but critics or non-self-members themselves, however I am quantifying over them in an irreducibly plural way. See Black; Stenius; Armstrong, *Universals and Scientific Realism*, volume I, pages 32–4; and especially Boolos. I find it very plausible that there is indeed such a thing as ontologically innocent plural quantification, and that it can indeed replace quantification over sets – sometimes. It would be delightful (except when I want to cite belief in sets as a precedent for my modal realism) if plural quantification could be iterated up the hierarchy, so that some fancy kind of plurally plural quantification over individuals could replace *all* quantification over sets or classes. But I think this project has very little hope of success. So I consider some apparent quantification over sets or classes of whatnots to carry genuine ontological commitment not only to the whatnots, but also to sets or classes of them; and then I use the word 'set'. But sometimes I think my quantification could be read as, or replaced by, innocent plural quantification that carries no commitment except to the whatnots themselves; and then I use the word 'class'.

An exception: since the phrase 'equivalence class' is standard, I use it whether or not I take there to be genuine ontological commitment.

other-worldly talking donkey. Brownie himself is, once and for all, a member of the set; hence, once and for all, an instance of the property. But it is contingent whether Brownie talks; Brownie has counterparts who do and counterparts who don't. In just the same way, it is contingent whether Brownie belongs to the set: Brownie has counterparts who do and counterparts who don't. That is how it is contingent whether Brownie has the property.

As it is for properties, so it is for relations. An instance of a dyadic relation is an ordered pair of related things; then we may take the relation again to be the set of its instances – all of them, this- and other-worldly alike. Again, it is no problem that different relations may happen to be coextensive; for this is only to say that the this-worldly parts of the sets are the same, and there is more to a set than its this-worldly part. Again, a pair may stand in a relation contingently, if it has counterpart pairs that do and counterpart pairs that don't.[38] In the same way, a triadic relation can be taken as a set of ordered triples, and so on. Also we can include relations of variable degree, since there is no reason why pairs and triples, for instance, cannot both belong to a single set.[39]

Often it is said that things have some of their properties relative to this or that. Thirst is not a property you have or lack *simpliciter*; you have it at some times and lack it at others. The road has different properties in different places; here it is surfaced, there it is mud. Nine has the property of numbering the planets at our world, but not at a possible world where a planet takes the place of our asteroid belt. (I mean the solar planets at present; and I mean to take another world where there are clear counterparts of the solar system and the present time.) Relative to Ted, Fred has the property of being a father, but relative to Ed, he has the property of being a son. Relative to the number 18, the number 6 has the property of being a divisor; but not relative to 17.

A property that is instantiated in this relative way could not be the set of its instances. For when something has it relative to this but not to

[38]Not just any pair of counterparts should count as a counterpart pair; it may be that pair ⟨X, Y⟩ counts as a counterpart of pair ⟨V, W⟩ partly because the relations between X and Y resemble those between V and W. See Hazen, 'Counterpart-Theoretic Semantics for Modal Logic'; my *Philosophical Papers*, volume I, pages 44–5; and the discussion of joint possibilities in section 4.4.

[39]There is a choice between various set-theoretic constructions of ordered pairs, triples, etc. I shall leave the choice unmade, since making it would serve no useful purpose. (At one point in section 4.4 I shall take them as sequences, consisting of terms indexed by numbers, because that makes it easy to leave gaps in them. But even that won't settle the matter. To make sequences by pairing the terms with their index numbers presupposes some different, prior construction of the term-index pairs, and I shall leave it open what that is to be.) So all that I say of pairs, triples, . . ., and relations is systematically ambiguous. No harm, unless I said something that would have different truth values on different disambiguations; which I have no intention of doing.

that, is the thing to be included in the set or not? Therefore we often see philosophers go to great lengths to provide for relative instantiation when they construct 'properties' in terms of possible worlds and individuals. A property is taken as a function from worlds to sets of things, giving for each world the things that have the property relative to that world. Or it is a function from world–time pairs to things, thus providing also for temporary properties like thirst. In the same way we could take the property of being surfaced as a function that assigns to each place the set of things surfaced there; or the property of being a son as a function that assigns to each person a set of sons; or the property of being a divisor as a function that assigns to each number the set of its divisors.

I find such constructions misguided: what is had by one thing relative to another might better be called a *relation*, not a *property*.[40] It may indeed turn out that one thing stands in a relation because another thing has a property, as when the part of the road that is at a certain place has the property of being surfaced, and that is how the whole road bears the 'surfaced at' relation to that place. Likewise it is by having temporal parts which are thirsty that a person is thirsty at various times. Of course, a disbeliever in temporal parts cannot agree; *he* thinks thirst is irreducibly relational. That is a central feature of his view and, for better or worse, it ought to be unhidden. That is why I do not approve of the terminology of 'properties' instantiated relative to this or that – it obfuscates and belittles the distinction between relations and genuine properties, and so puts us off guard against those theories that try to tell us that there are only relations where we might have thought there were genuine properties. (See section 4.2.) And that is why I offer a treatment of properties that requires things to have or to lack them *simpliciter*, together with a separate but parallel treatment of relations.

Likewise I have made no place for properties that admit of degree, so that things may have more or less of the same property. Instead, there are families of plain properties: the various lengths, the various masses. And there are relations to numbers, such as the mass-in-grams relation that (a recent temporal part of) Bruce bears to a number close to 4,500.

I identify propositions with certain properties – namely, with those that are instantiated only by entire possible worlds. Then if properties generally are the sets of their instances, a proposition is a set of possible worlds. A proposition is said to *hold* at a world, or to be *true at* a world. The proposition is the same thing as the property of being a world where that proposition holds; and that is the same thing as the set of worlds

[40]More precisely, what is had by X relative to Y is not a property of X. It *is* a property of the pair ⟨X, Y⟩ – on my account, any relation is a property of the pairs (or triples, or whatnot) that instantiate it.

where that propositions holds. A proposition holds at just those worlds that are members of it.[41]

Just as it is sometimes said that properties are had relative to this or that, so it is sometimes said that propositions hold relative to this or that. No harm in their holding at worlds, of course; but other relative holding requires a switch in what we mean by 'propositions'. For instance a *tensed* proposition, which is said to hold at some times but not others, can be taken as a set of world–time pairs; in other words a relation of worlds and times. If as I think (see section 1.6) no time is identically a common part of two different worlds, then this can be simplified: we can say that the tensed proposition is simply a property, that is a set, of times.

Likewise an *egocentric* proposition, which holds for some people but not others, could be taken as a property, that is a set, of people. And if we generalise, and countenance also egocentric propositions which hold for things other than people – such as the proposition that one is a poached egg – then we should say that the egocentric proposition is a property, that is a set, of possible individuals. But if we can already call it a 'property', what's the sense of also calling the same thing an 'egocentric proposition'?

There might be a good reason. The conception we associate with the word 'proposition' may be something of a jumble of conflicting *desiderata*. Part of the idea is that propositions are supposed to be true or false *simpliciter*. Or at any rate, their truth or falsity is not supposed to be relative to anything except the world – unlike a sentence, a proposition is not supposed to be true on one interpretation but false on another, true on one resolution of vagueness but false on another, true in Melbourne but false in Adelaide, true yesterday but false today, true for

[41]Distinguish my proposal from a different way of unifying propositions, properties, and relations. The idea is that relations properly speaking are two-place, three-place, and on up; properties are one-place relations; and propositions are zero-place relations. See, for instance Montague, *Formal Philosophy*, pages 122–3. This strikes me as misguided elegance. How can we make sense of it? – Only by giving *everything* one more place than meets the eye. The so-called n-place relations are instantiated not *simpliciter* but relative to a world. (Or for Montague, relative to an index that might or might not be a world.) I say that means they all have an extra, hidden place to them. Thus a proposition is supposed to be a zero-place relation, but it turns out to be a one-place relation – that is, a set of one-tuples of worlds. A so-called property is supposed to be a one-place relation, but it turns out to be a two-place relation of things to worlds; what is supposed to be a two-place relation turns out to be three-place; and so on up. The treatment of propositions is the only satisfactory part. If we identified a one-tuple of a world with the world itself, as we might but needn't, it is exactly my own treatment; if not, still sets of worlds and sets of their one-tuples would correspond so closely that we needn't care which ones get called the propositions. The rest of the unified treatment is not satisfactory because it relies on the obfuscatory notion of relative instantiation. Therefore the whole idea is best abandoned.

me but false for you. But another part of the idea is that propositions are supposed to be the objects of thought. They are supposed to be capable of giving the content of what we know, believe, and desire. But it is clear that some thought is egocentric, irreducibly *de se*, and then its content cannot be given by the propositions whose truth is relative to nothing but worlds; for those propositions do not discriminate between inhabitants of the same world. If you insist that propositions, rightly so called, must be true or false relative to worlds and nothing else, then you had better say that the objects of at least some thought turn out not to be propositions. Whereas if you insist that propositions, rightly so called, are the things that serve as objects of all thought, then you had better admit that some propositions are egocentric. The point is the same whichever way you say it: the objects of thought in general are not sets of possible worlds; they sometimes must be, and always can be, taken instead as sets of possible individuals.

Everyone agrees that it won't do to take a property as the sets of its this-worldly instances, because then two properties will be taken to be identical if they happen to be coextensive. Some will say that it is just as bad to take a property as the set of all its instances throughout the worlds, because then two properties will be taken to be identical if they are necessarily coextensive. The stock example concerns the properties of triangularity and of trilaterality. Necessarily, a planar figure bounded by line segments has the same number of angles as sides. So, throughout the worlds, all and only triangles are trilaterals. Yet don't we want to say that these are two different properties?

Sometimes we do, sometimes we don't. I don't see it as a matter for dispute. Here there is a rift in our talk of properties, and we simply have two different conceptions. It's not as if we have fixed once and for all, in some perfectly definite and unequivocal way, on the things we call 'the properties', so that now we are ready to enter into debate about such questions as, for instance, whether two of them ever are necessarily coextensive. Rather, we have the word 'property', introduced by way of a varied repertory of ordinary and philosophical uses. The word has thereby become associated with a role in our commonsensical thought and in a variety of philosophical theories. To deserve the name of 'property' is to be suited to play the right theoretical role; or better, to be one of a class of entities which together are suited to play the right role collectively. But it is wrong to speak of *the* role associated with the word 'property', as if it were fully and uncontroversially settled. The conception is in considerable disarray. It comes in many versions, differing in a number of ways. The question worth asking is: which entities, if any, among those we should believe in, can occupy which versions of the property role? My answer is, in part, that sets of *possibilia* are entities

we should believe in which are just right for *one* version of the property role.

There's no point in insisting that this one is the only rightful conception of the properties. Another version of the property role ties the properties more closely to the meanings of their standard names, and to the meanings of the predicates whereby they may be ascribed to things. 'Triangular' means having three angles, 'trilateral' means having three sides. These meanings differ. (Or do they? The conception of 'meaning' also is in disarray!) So on this conception of properties, we want to distinguish triangularity from trilaterality, though we never can distinguish their instances. We can put the distinction to use, for instance, in saying that one of the two properties is trivially coextensive with triangularity, whereas the other is non-trivially coextensive with triangularity.

This conception demands that properties should be *structured*. If we want to match up properties with the meanings of linguistic expressions that have syntactic structure, then we want to give the properties themselves some kind of quasi-syntactic structure. We can construct structured properties on the model of the structured 'meanings' considered in the previous section. We needn't build them from scratch; we can begin with the unstructured properties and relations we have already, the sets of this- and other-worldly instances. So these structured properties will require *possibilia* just as much as the unstructured ones did. We will need not only properties and relations of individuals; also we will make use of a higher-order unstructured relation that holds between properties and relations of individuals. It is a relation all the same – a set of pairs – and it is constructed out of *possibilia* just as much as first-order properties and relations of individuals are.

Let A be the relation of being an angle of; let S be the relation of being a side of. Suppose for simplicity that these can be left as unstructured relations; we could go to a deeper level of analysis if we like, but that would complicate the construction without showing anything new. Let T be the higher-order unstructured relation which holds between an unstructured property F of individuals and an unstructured relation G of individuals iff F is the property of being something which exactly three things bear relation G to. A certain unstructured property is the unique thing which bears T to A, and therefore it is the (unstructured) property of triangularity; it also is the unique thing which bears T to S, and therefore it is the (unstructured) property of trilaterality. Therefore let us take the structured property of triangularity as the pair ⟨T, A⟩, and the structured property of trilaterality as the pair ⟨T, S⟩. Since S and A differ, we have the desired difference between the two pairs that we took to be our two structured properties.

Likewise we can construct structured relations. And if at some deeper level of analysis, we had structured versions of the relation of being an

angle of and the relation of being a side – these might be pairs $\langle A_1, A_2 \rangle$ and $\langle S_1, S_2 \rangle$ respectively, or something still more complicated – then we could build these instead of the original A and S into our structured properties, getting structured triangularity as $\langle T, \langle A_1, A_2 \rangle \rangle$ and structured trilaterality as $\langle T, \langle S_1, S_2 \rangle \rangle$.

Likewise for propositions. If it is central to the role you associate with 'proposition' that there should be some sort of quasi-syntactic structure, so that it makes sense to speak of subject–predicate propositions or negative or conjunctive or quantified propositions, then sets of worlds will not do. But more complicated set-theoretic constructions out of *possibilia* can serve instead. In some cases, these might closely resemble the 'meanings' for sentences of our little language in the previous section. For instance we could associate the modifier 'not' with the unstructured relation N that holds between any unstructured proposition and its negation, that being the set of all worlds where the original proposition does not hold. Then a negative structured proposition could take the form $\langle N, P \rangle$, where P is a (structured or unstructured) proposition. Taking propositions as sets of worlds, it is nonsense to distinguish a proposition P from its double negation; the double negation of a proposition is the original proposition all over again. But the structured propositions P and $\langle N, \langle N, P \rangle \rangle$ do indeed differ; although they are equivalent, having the same truth value at every world.

Another kind of structured proposition corresponds to meanings we would have had if our illustrative language had been equipped for predication and if it had used individuals and properties as the semantic values, respectively, for individual constants and monadic predicates. Corresponding to an atomic predication in which the subject and predicate have as semantic values an individual A and a property P, we have as meaning the pair $\langle A, P \rangle$. This is a structured subject–predicate proposition; we might also call it a *singular* proposition or a *de re* proposition. It is true iff the individual A has the property P, otherwise false.

(Two elaborations. First, the properties so used might or might not themselves be structured. Second, we could just as well have a relation and several individuals: $\langle R, A, B \rangle$, a structured proposition which is true iff A and B stand in the relation R. It could be the meaning of a dyadic atomic predication in which R is the semantic value of the predicate, and A and B the semantic values of two individual constants that appear as arguments.)

These singular propositions have been much discussed, under a variety of names, but mostly in connection with inappropriate questions. Should we believe that they exist? – Of course we should. We *must*, if we believe in properties and we believe in individuals and we believe in ordered pairs of things we believe in. You don't even have to believe in the sets of *possibilia* that I call properties, just in entities suited to occupy some or other version of the property role.

Are they, rightly speaking, *propositions*? – Certainly they (and their more complicated relatives) occupy a version of the proposition role. They do not occupy the one and only rightful version, because nothing in our tangled and variable usage suffices to settle which version that would be.

Are they objects of thought? – This much is true: somehow, by our thought, we do ascribe properties to individuals. (Not by thought *alone*, of course, special cases aside; rather, by thought plus the relations of subject to environment.) Whenever you ascribe a property to an individual, there is the pair of that property and that individual. So your accomplishments in property ascribing can be characterised in terms of property-individual pairs, in other words in terms of the singular propositions that are true according to your ascribing. That much ought to be uncontroversial, and that much is enough to provide a good sense in which we can say that singular propositions are objects of thought.

Are they *the* objects of thought? That is, are they the entities that serve best for characterising the subject? – Surely that cannot be answered once and for all. It all depends on what purpose the characterisation is meant to serve.

When it is meant to serve a narrowly psychological purpose, revealing how the subject's actions serve his desire according to his belief and how his belief evolves under the impact of his experience, then the use of singular propositions to characterise his thought will be rather unsatisfactory. It will tend to suppress relevant information about how exactly the subject does his ascribing; it will drag in psychologically irrelevant information about which individuals exactly stand at the far end of the various relationships that connect him to other parts of his world.

(To illustrate, remember Pierre. Let us present him with an unlimited bus pass, and take him to an international bus station. First he comes to an English bus with its destination shown in English, and he shuns it. Why? Because he believes it goes to London; that is, because he ascribes the relation of *going to* to the pair of the bus and London; that is, because the singular proposition ⟨*going to*, the bus, London⟩ is, in the appropriate sense, an object of his belief. Next he comes to another bus, a French bus with its destination shown in French, and he hops on with glee. Why? Because he believes it goes to London, that is, because the singular proposition ⟨*going to*, the other bus, London⟩ is an object of his belief. Evidently something relevant has been left out. I don't say that we *cannot* tell the whole story if we insist on characterising Pierre's thought by means of singular propositions. We could, for instance, mention the singular propositions which pair the two buses with the two properties of *going to an ugly city* and *going to a pretty city*. The characterisation by means of singular propositions is badly matched to the needs of belief–desire psychology, but I don't deny that with sufficient effort we can overcome the mismatch and pull out all the information we need.)

But when we are interested less in the subject's psychology, and more in his dealings with the things around him, as happens if we are interested in him as a partner in cooperative work and as a link in channels for information, then it is otherwise. The more he and we ascribe the same properties to the same individuals, the better we fare in trying to coordinate our efforts to influence those individuals. We learn from him by trying to ascribe the same properties to the same things that he does. We teach him by trying to get him to ascribe the same properties to things that we do. What matters is *agreement* about how things are; and we agree not when we think alike, but when we ascribe the same properties to the same things. To characterise him and ourselves in the sense that is relevant to our agreement, singular propositions are just right. When the same singular propositions are true according to him and according to us, that is when we ascribe the same properties to the same things.

In short, there is no contest between structured and unstructured versions of the properties, relations and propositions. Given the combined resources of set theory and modal realism, we have both versions. (That is: we have suitable candidates to fill both versions of the roles associated with the terms 'property', 'relation', and 'proposition'.) Both versions require *possibilia*. We needn't worry about which versions better deserve the names, since previous use of the names has not been uniform enough to settle the matter. I shall reserve the names 'property', 'relation', and 'proposition', when used without an adjective, for the unstructured versions: the sets of instances, or of worlds. Likewise I shall reserve them for properties and relations of the kind that do not admit of degree and that are instantiated *simpliciter*, not relative to anything, and for propositions of the kind that hold or not relative only to a world. But all this is terminology, not doctrine.[42]

There is another great rift in our talk of properties. Sometimes we conceive of properties as *abundant*, sometimes as *sparse*. The abundant properties may be as extrinsic, as gruesomely gerrymandered, as miscellaneously disjunctive, as you please. They pay no heed to the qualitative joints, but carve things up every which way. Sharing of them has nothing to do with similarity. Perfect duplicates share countless properties and fail to share countless others; things as different as can be imagined do exactly the same. The abundant properties far outrun the predicates of any language we could possibly possess. There is one of them for any condition we could write down, even if we could write at infinite length and even if we could name all those things that must remain nameless because they

[42]But sometimes, especially when considering alternatives to modal realism in chapter 3, I shall use the names 'property', 'relation', and 'proposition' in a vague and neutral way, to apply to whatever the most satisfactory occupants of the appropriate roles might be.

fall outside our acquaintance. In fact, the properties are as abundant as the sets themselves, because for any set whatever, there is the property of belonging to that set. It is these abundant properties, of course, that I have identified with the sets.

The sparse properties are another story. Sharing of them makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific, the sets of their instances are *ipso facto* not entirely miscellaneous, there are only just enough of them to characterise things completely and without redundancy.

Physics has its short list of 'fundamental physical properties': the charges and masses of particles, also their so-called 'spins' and 'colours' and 'flavours', and maybe a few more that have yet to be discovered. In other worlds where physics is different, there will be instances of different fundamental physical properties, alien to this world. (See section 3.2, where these alien properties get in the way of the project of building ersatz possible worlds out of this-worldly constituents.) And in unphysicalistic worlds, the distribution of fundamental physical properties won't give a complete qualitative characterisation of things, because some of the 'fundamental' properties of things will not be in any sense physical. What physics has undertaken, whether or not ours is a world where the undertaking will succeed, is an inventory of the *sparse* properties of this-worldly things. Else the project makes no sense. It would be quixotic to take inventory of the *abundant* properties – the list would not be short, nor would we discover it by experimental and theoretical investigation.

I would not recommend that we enter into debate over whether the properties really are abundant or whether they really are sparse. We needn't choose up sides. Rather we should acknowledge that we have both conceptions, and an adequate account of what there is ought to accommodate both.[43]

If we have the abundant properties (as we do, given set theory and *possibilia*) then we have one of them for each of the sparse properties. So we may as well say that the sparse properties are just some – a very small minority – of the abundant properties. We need no other entities, just an inegalitarian distinction among the ones we've already got. When a property belongs to the small minority, I call it a *natural* property.[44]

[43]Here I am in partial agreement with Bealer, who advocates a twofold scheme of abundant 'concepts' and sparse 'qualities'. However, he brings the abundant-versus-sparse division into line with the structured-versus-unstructured division, whereas I take the two divisions as cutting across each other.

[44]The name is borrowed from the familiar term 'natural kind'; the contrast is meant to be with unnatural, gerrymandered, gruesome properties. The name has proved to have a drawback: it suggests to some people that it is supposed to be *nature* that distinguishes the natural properties from the rest; and therefore that the distinction is a contingent matter,

Probably it would be best to say that the distinction between natural properties and others admits of degree. Some few properties are *perfectly* natural. Others, even though they may be somewhat disjunctive or extrinsic, are at least somewhat natural in a derivative way, to the extent that they can be reached by not-too-complicated chains of definability from the perfectly natural properties. The colours, as we now know, are inferior in naturalness to such perfectly natural properties as mass or charge; grue and bleen are inferior to the colours; yet even grue does not plumb the real depths of gruesomeness. If it did, we would not have been able to name it.

Relations, like properties, can be conceived as abundant or as sparse: a relation for any set of pairs (or triples, or . . .) whatever, or else a minimum basis of relations sufficient to characterise the relational aspects of likeness and difference. Again we may say that some relations are natural, or that some are more natural than others; and that the natural relations are the same sort of thing as other relations, just a distinguished minority among the sets of pairs, triples, and so on. Also propositions can be conceived as abundant or sparse, and sets of worlds may accordingly be divided into the more and less natural. This is automatic, given the division of properties plus the identification of propositions with properties of worlds.

In systematic philosophy we constantly need the distinction between the more and the less natural properties. It is out of the question to be without it. I have discussed some of its uses in 'New Work for a Theory of Universals' and in 'Putnam's Paradox'. Here I shall mention only one.

We distinguish *intrinsic* properties, which things have in virtue of the way they themselves are, from *extrinsic* properties, which they have in virtue of their relations or lack of relations to other things. How to draw this distinction? Some approaches fail, some fall into circularity. (See my 'Extrinsic Properties'.) But if we start by distinguishing natural from unnatural properties, then the distinction between intrinsic and extrinsic properties is not far away. It cannot be said that all intrinsic properties are perfectly natural – a property can be unnatural by reason of disjunctiveness, as the property of being tripartite-or-liquid-or-cubical is, and still it is intrinsic if its disjuncts are. But 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. (Maybe the

---

so that a property might be natural at one world but not at another. I do not mean to suggest any such thing. A property is natural or unnatural *simpliciter*, not relative to one or another world.

second clause is redundant. That depends on whether we acknowledge some *structural* properties – properties having to do with the way a thing is composed of parts with their own properties and relations – as perfectly natural.) Then we can go on to say that an *intrinsic* property is one that can never differ between two duplicates.

There is a corresponding distinction among relations. An *internal* relation is one that supervenes on the intrinsic natures of its *relata*: if X_1 and Y_1 stand in the relation but X_2 and Y_2 do not, then there must be a difference in intrinsic nature either between the Xs or else between the Ys. If X_1 and X_2 are duplicates (or identical), and so are Y_1 and Y_2, then the pairs $\langle X_1, Y_1 \rangle$ and $\langle X_2, Y_2 \rangle$ stand in exactly the same internal relations. Relations of similarity or difference in intrinsic respects are internal; for instance, the relations of closeness of worlds that figured in my accounts of counterfactuals and of verisimilitude. (See section 1.3.)

Some other relations, notably relations of spatiotemporal distance, are not internal; they do not supervene on the natures of the *relata*. If X_1 and X_2 are duplicates (or identical), and so are Y_1 and Y_2, it may yet happen that the pairs $\langle X_1, Y_1 \rangle$ and $\langle X_2, Y_2 \rangle$ stand in different relations of distance. Consider a (classical) hydrogen atom, which consists of an electron orbiting a proton at a certain distance. If we take a duplicate of the electron and a duplicate of the proton, then they needn't exhibit the same distance – they may not comprise an atom, they may be in different galaxies or different worlds.

However there is a different way in which relations of distance do supervene on intrinsic character. If, instead of taking a duplicate of the electron and a duplicate of the proton, we take a duplicate of the whole atom, then it will exhibit the same electron–proton distance as the original atom. Although distance fails to supervene on the intrinsic natures of the *relata* taken separately, it does supervene on the intrinsic nature of the composite of the *relata* taken together – in this case, the composite hydrogen atom.

There are other relations for which not even that much is true, for instance the relation of having the same owner. It involves more than the *relata* taken either separately or together, since it also drags in the owner and however much of the rest of the world it takes for there to be the institution of ownership. Thus we don't just have the internal relations versus all the rest; we have a three-way classification. I shall say that a relation is *external* iff it does not supervene on the natures of the *relata* taken separately, but it does supervene on the nature of the composite of the *relata* taken together. A relation of intrinsic similarity is internal; a relation of distance is external; but the relation of having the same owner is neither internal nor external.

I distinguish *duplication* from *indiscernibility*. Two things are duplicates iff they have the same intrinsic qualitative character; and that is a matter of the perfectly natural (hence *ex officio* intrinsic) properties of those

things and their parts, and of the perfectly natural external relations of their parts. Two things are *indiscernible* iff they have the same intrinsic and extrinsic qualitative character. Extrinsic qualitative character, wherein duplicates may differ, consists of extrinsic properties that are, though not perfectly natural, still somewhat natural in virtue of their definability from perfectly natural properties and relations. Indiscernibles share all their somewhat natural properties. They do not, of course, share all their properties without exception – not if we admit, for any set, a property of belonging to that set, as we automatically do if we identify properties with sets.

To illustrate, contrast two kinds of eternal recurrence. Some worlds exhibit *one-way* eternal recurrence: there is a beginning of time and then there is a first epoch, a second epoch just like the first, a third, and so *ad infinitum*. Then corresponding inhabitants of the different epochs are duplicates – they differ in no intrinsic respect – but they are not indiscernible. They differ in their extrinsic qualitative character in that one inhabits the first epoch, another inhabits the seventeenth, and so on. Other worlds exhibit *two-way* eternal recurrence: there is no last epoch and no first, the epochs are ordered like the integers rather than the natural numbers. Then the corresponding inhabitants of different epochs are not only duplicates but indiscernibles. But still they don't share all their properties, because for any two of them there are sets which contain one without the other.

Many philosophers are sceptical about the distinction between natural and gruesome properties. They think it illegitimate, unless it can somehow be drawn in terms that do not presuppose it. It is impossible to do that, I think, because we presuppose it constantly. Shall we say that natural properties are the ones that figure in laws of nature? – Not if we are going to use naturalness of properties when we draw the line between laws of nature and accidental regularities. Shall we say that they are the ones that figure in the content of thought? – Not if we are going to say that avoidance of gratuitous gruesomeness is part of what constitutes the correctness of an ascription of content. Shall we say that they are the ones whose instances are united by resemblance? – Not if we are going to say that resemblance is the sharing of natural properties. Unless we are prepared to forgo some of the uses of the distinction between natural and unnatural properties, we shall have no easy way to define it without circularity. That is no reason to reject the distinction. Rather, that is a reason to accept it – as primitive, if need be.

I would willingly accept the distinction as primitive, if that were the only way to gain the use of it elsewhere in our analyses. The contribution to unity and economy of theory would be well worth the cost. But I think there are two attractive alternatives: theories which, for some price both

in ontology and in primitives, give us resources to analyse the distinction without forgoing any of its applications. I have two such theories in mind. One is a sparse theory of immanent universals, more or less as presented in D. M. Armstrong's *Universals and Scientific Realism*. The other is a theory of tropes, more or less as in D. C. Williams's 'On the Elements of Being', but made sparse in a way that imitates Armstrong's theory.[45] In the contest between these three alternatives – primitive naturalness, universals, or tropes – I think the honours are roughly even, and I remain undecided.

The two theories go as follows. To each perfectly natural property there corresponds a universal, or else a set of tropes. Wherever the property is instantiated, there the corresponding universal, or one of the corresponding tropes, is present. Let us assume that unit positive charge is a perfectly natural property, which is instantiated by momentary stages of various particles. For short: charge is instantiated by particles. Wherever there is a charged particle, there the universal of charge, or else one of the tropes of charge, is present. It is located there, just as the particle itself is. Indeed, it is part of the particle. It is not a spatiotemporal part: the universal or trope occupies the whole of the spatiotemporal region, point-sized or larger, that the particle itself occupies. Besides the universal or trope of charge, other universals or tropes also will be present as further non-spatiotemporal parts of the same particle. For instance, there will be a universal or trope of mass.

The difference between universals and tropes comes when we consider two instances of the same perfectly natural property – for instance, two particles each having unit positive charge. Each one contains a non-spatiotemporal part corresponding to charge. But if this non-spatiotemporal part is a universal, then it is the same universal for both particles. One and the same universal recurs; it is multiply located; it is wholly present in both particles, a shared common part whereby the two particles overlap. Being alike by sharing a universal is 'having something in common' in an absolutely literal sense. If the non-spatiotemporal part whereby a charged particle is charged is a trope, on the other hand, then there are different tropes for different charged particles. There is no recurrence, no sharing of a multiply located non-spatiotemporal part. Instead, we say that the charge-trope of one particle and the charge-trope of another are *duplicate* tropes, in a way that a charge-trope and a mass-trope, say, are not.

[45] A somewhat similar theory of universals is the principal system of Goodman, *The Structure of Appearance*, provided we take it to apply not only to appearance but to things generally. Other advocates of trope theory – under a variety of names, and with various differences of doctrine – include Stout, Campbell, and Johnston.

If there are universals, we can say that the particle is composed partly of its several universals. But not entirely; because another particle exactly like it would have the very same universals, and yet the two particles would not be the same. We can say that the particle consists of its universals together with something else, something non-recurrent, that gives it its particularity. Then we need a primitive notion to say how that something gets united with the universals. I shall call this union 'instantiation'. (I trust there will be no confusion with the 'instantiation' of a property-taken-as-a-set by its members.) We can either say that the universal is instantiated by the whole of a particular; or that it is instantiated by the part that gives the particularity, the residue which is left if we take an ordinary particular and subtract its universals.

(It cannot be said, unfortunately, that a universal is instantiated by just anything that has it as a part. For one thing, the relation of part to whole is transitive; so if a universal of charge is part of a particle which is part of an atom, then the universal in turn is part of the atom; but it is the particle, not the atom, which instantiates the universal. And so on up; the universal is part of everything, however big, that the particle is part of. Further, suppose there are disunified wholes composed of miscellaneous parts, as indeed I believe (see section 4.3). These might include universals which they do not instantiate.)

If there are tropes, we might say that the particle is composed entirely of its tropes; there is no problem with a second particle exactly like it, since that second particle is composed not of the same tropes but of duplicate tropes. Then we need a primitive notion – 'instantion' in yet another sense – to say how the tropes that comprise the particle are united. It is an advantage of tropes over universals that we need no special thing to confer particularly – that is, non-recurrence – since the tropes are particular already. The companion drawback is that we need the primitive notion of duplicate tropes, whereas with universals we just say that it is one and the same universal throughout the charged particles.[46]

A theory of universals might attempt to analyse all similarity in terms of shared universals. (Whether it can succeed depends on what can be said about similarity between universals themselves; see Armstrong, *Universals and Scientific Realism*, chapters 22 and 23.) A theory of tropes must be less ambitious. It cannot analyse all similarity, because duplication of tropes is itself a primitive relation of similarity. But duplication of tropes is much better behaved than other relations of similarity that we

[46]A universal recurs; a trope has duplicates. We could also imagine an intermediate thing that sometimes recurs and sometimes has duplicates. A trope theorist who also believes in strict identity over time might say that charge recurs along the world-line of one persisting particle, but is duplicated between one persisting particle and another. Campbell and Johnston favour this sort of theory.

might contemplate taking as primitive. The similarity of particles is a messy business: particles can be alike in one respect and not in another, for instance when they are alike in mass but opposite in charge. A theory that starts with similarity-in-some-respect and attempts to recover the respects of comparisons by analysis will run into serious trouble. (See Armstrong, chapter 5; and Goodman, *The Structure of Appearance*, chapter V.) It's simpler with tropes: two charge tropes are alike or not, and that's that. If you will not countenance primitive similarity in any form, then trope theory is not for you. But if you will, then duplication of tropes is an especially satisfactory form for primitive similarity to take.

A universal unifies the set of all and only those particulars that instantiate it. A maximal set of duplicate tropes – that is, a set of tropes that are duplicates of one another but not of any other trope not included in the set – likewise unifies the set of all and only those particulars which instantiate some trope in the set. If we accept a theory of universals or of tropes, we can define a perfectly natural property (of particulars) as any set that is thus unified.

This may seem roundabout. If indeed we accept a theory of universals, why not give up the plan of identifying properties with sets of their instances, and say that the universal itself is the property? Or if we accept a theory of tropes, why not say that the set of duplicate tropes is the property? Surely these things, if they exist, are fine candidates for the role of properties – and no *possibilia* are needed.

Yes and no. In the first place, we would still need *possibilia* if we wanted to acknowledge uninstantiated properties alien to this world. Universals and tropes are present in their instances, and so must have instances if they are to be present at all. If uninstantiated properties are universals, they are other-worldly universals. If they are sets of tropes, they are sets of other-worldly tropes.

In the second place, universals or sets of duplicate tropes would be fine for the role of *sparse* properties, but the sparse properties are not enough. There may be no urgent need to quantify over all of the very abundant and very gruesome properties that modal realism has on offer as sets of *possibilia*. But certainly we want to go well beyond the perfectly natural properties. When we speak of the various properties that a believer ascribes to himself and the things around him, or when we say that Fred hasn't many virtues, or when we say that sound taxonomy will take account of the biochemical as well as the anatomical properties of organisms, then we quantify over properties that are neither flagrantly gruesome nor perfectly natural. We would not wish to repudiate all properties that are in any way disjunctive or negative or extrinsic. However, universals or tropes are credible only if they are sparse. It is quite easy to believe that a point particle divides into a few non-spatiotemporal parts in such a way that one of them gives the particle

its charge, another gives it its mass, and so on. But it is just absurd to think that a thing has (recurring or non-recurring) non-spatiotemporal parts for *all* its countless abundant properties! And it is little better to think that a thing has a different non-spatiotemporal part for each one of its properties that we might ever mention or quantify over. The most noteworthy property of this bed is that George Washington slept in it – surely this is true on some legitimate conception of properties – but it is quite unbelievable that *this* property corresponds to some special non-spatiotemporal part of the bed! This is not one of the perfectly natural properties that might correspond to a universal or a trope; rather, it is a property that gains a degree of derivative naturalness, because it is definable in a not-too-complicated way from the perfectly natural properties. The universals or sets of duplicate tropes would not be good candidates to serve as the abundant properties, or even the not-too-abundant-and-not-too-sparse properties. They make a useful adjunct to a broader theory of properties, not a replacement for it.

(A note on terminology. Sometimes 'universal' becomes just another rough synonym for 'property'. The two words are used loosely and interchangeably, equally infected with indecision between rival versions of the definitive theoretical role. On that usage, any candidates whatever for the role of properties, abundant or sparse, could equally deserve the name of universals. But I do not use the two words loosely and interchangeably. (I regret to say that I once did, in 'An Argument for the Identity Theory'.) Instead, I reserve the word 'universal' strictly for the things, if such there be, that are wholly present as non-spatiotemporal parts in each of the things that instantiate some perfectly natural property.)

Just as monadic universals or tropes might serve to single out the perfectly natural properties, so polyadic universals or tropes might serve to single out the perfectly natural relations. Indeed, if we buy into universals or tropes just in order to avoid taking naturalness as primitive, it seems that we had better be able to cover the relations as well as the properties.[47] Suppose we have a dyadic universal or trope corresponding

[47]There just might be another way to define naturalness of relations: by a very short list, fixed once and for all. It seems a little strange to discuss naturalness of relations in a general way when we have only one really clear example: the spatiotemporal relations. Maybe a few more: maybe part–whole and identity. Maybe set membership. Maybe, if we're unlucky, an irreducible relation of causal or lawful connection. But it's still a short list. If we tried to define the natural properties once and for all by a short list – there are the mass properties, the charge properties, the quark colours and flavours, . . . – we ought to suspect that we had left off not only the this-worldly natural properties we have yet to discover, but also the nameless alien natural properties that are found only at other worlds. It seems a bit less clear that we need to leave room for alien natural relations. What if the few natural relations of this-worldly things are the only ones to be found at any world? I regard this hypothesis as far-fetched, but not altogether absurd.

to the relation of being a certain minute distance apart; and suppose a proton and an electron are that distance apart, and together comprise an atom. Then the dyadic universal or trope is present as a non-spatiotemporal part of the atom. It has the same divided location that the atom itself has. But in a different way; unlike the atom, the universal or trope is not itself divided. It doesn't have one part in the proton and another in the electron. If we accept this theory, we just have to accept that an undivided thing can have a divided location. It is part of the atom; but no part of it is part of the proton or part of the electron. If we accept this theory, we must say that the proton and the electron do not exhaust the atom.[48] All this is disturbingly peculiar, much more so that the monadic case, but if the price is right we could learn to tolerate it.

The atom has the structural property of consisting of a proton and an electron a certain distance apart. Is there a structural universal or a structural trope to correspond to this property? If so, that too is present as a non-spatiotemporal part of the atom. We might think that if sparseness is wanted, then this extra thing is superfluous. We already have the monadic universals or tropes of the two particles, and the dyadic universal or trope of distance between them. The presence of these already settles the atom's structure – so what would a structural universal or trope add? But just as the atom itself is not some extra thing over and above its proton and its electron and their distance, so we might say that the atom's structural universal or trope is no extra thing. It is somehow composed of the simpler universals or tropes, and so is nothing over and above them; so we needn't complain of its redundancy. It is not entirely clear how the composition of structural universals would work and so I think it doubtful whether a theory of universals ought to admit them.[49] Structural tropes, on the other hand, seem unproblematic.

The question of primitive naturalness versus universals or tropes is

[48]I said that an external relation, although it does not supervene on the intrinsic natures of its *relata* taken separately, does supervene on the intrinsic nature of the composite of its *relata* – for instance, the electron–proton distance supervenes on the intrinsic nature of the whole atom. To make this work under a theory of universals or tropes, 'composite' has to be understood in a special way. The *relata* are just the electron and the proton, but their composite has to be augmented to include also their distance-universal or distance-trope, and any other dyadic universals or tropes that may connect the electron and the proton. (See Williams, 'Necessary Facts', pages 603–5.) Might we throw in too much, and falsely certify the relation of having the same owner as external because we had thrown in a corresponding universal or trope? No fear! – The alleged universal or trope would be superfluous, so a sparse theory will deny its existence. Just as we can safely say that all perfectly natural properties are *ex officio* intrinsic, so we can say that all perfectly natural relations are external, and those will be the only relations to which there correspond dyadic universals or tropes.

[49]See my 'Against Structural Universals'; and Armstrong, *Universals and Scientific Realism*, volume II, pages 69–70.

peripheral to the defence of modal realism which is the main business of this book; I have nevertheless taken it up here for several reasons. First, because of a question I have already considered: whether the benefits of modal realism are diminished if we believe in universals or tropes and therefore have less need of properties taken as sets of *possibilia*. Second, because of the question how satisfactory it might be to replace my genuine possible worlds and individuals by ersatz ones constructed out of this-worldly universals or tropes; I take up this question in section 3.2. And third, because universals or tropes turn out to complicate many of our discussions of the tenets of modal realism and the difference between varieties of it. To take one example: I noted above that a universal is part of anything that a particular that instantiates it is part of. That makes it a common part of all worlds wherein it is instantiated; which means that so long as I remain neutral about the existence of universals, I need to qualify my denial of trans-world identity. (See section 4.2.)

1.6 Isolation

I hope that by saying what theoretical purposes it is meant to serve, I have helped to make clear what my thesis of plurality of worlds is. Now I shall address some further questions of formulation and state some further tenets of my position.

A possible world has parts, namely possible individuals. If two things are parts of the same world, I call them *worldmates*.[50] A world is the mereological sum[51] of all the possible individuals that are parts of it, and so are worldmates of one another. It is a maximal sum: anything that is a worldmate of any part of it is itself a part. (This is just a consequence of my denial that worlds overlap.) But not just any sum of parts of worlds is itself a world. It might, of course, be only part of a world. Or it might consist of parts of two or more different worlds; thus it might be spread

[50]Worldmates are compossible in the strongest sense of the word. Two things are compossible in another sense if they are vicariously worldmates, in virtue of their counterparts; that is, iff some one world contains counterparts of both of them. Two things are compossible in yet another sense iff some one world contains intrinsic duplicates of both. In this third sense, any two possible individuals are compossible (except, perhaps, when one is too big to leave room for the other); see section 1.8.

[51]The *mereological sum*, or *fusion*, of several things is the least inclusive thing that includes all of them as parts. It is composed of them and of nothing more; any part of it overlaps one or more of them; it is a proper part of anything else that has all of them as parts. Equivalently: the mereological sum of several things is that thing such that, for any X, X overlaps it iff X overlaps one of them. For background on the mereology that I shall be using extensively in this book, see Leonard and Goodman; or Goodman, *Structure of Appearance*, section II.4.

over logical space, not wholly within any one world, and its parts might not all be worldmates of one another.

What, then, is the difference between a sum of possible individuals that is a possible world, and one that is not? What makes two things worldmates? How are the worlds demarcated one from another? Why don't all the *possibilia* comprise one big world? Or, at the other extreme, why isn't each possible neutrino a little world of its own? In Perry's terminology: what is the unity relation for possible worlds?[52]

I gave part of the answer in my opening section, when I said that nothing is so far away from us in space, or so far in the past or the future, as not to be part of the same world as ourselves. The point seems uncontroversial, and it seems open to generalisation: whenever two possible individuals are spatiotemporally related, they are worldmates. If there is any distance between them – be it great or small, spatial or temporal – they are parts of one single world.

(Better: for any two possible individuals, if every particular part of one is spatiotemporally related to every particular part of the other that is wholly distinct from it, then the two are worldmates. This formulation avoids difficults that might be raised concerning partial spatiotemporal relatedness of trans-world mereological sums; difficulties about multiply located universals; and difficulties about whether we ought to say that overlapping things are spatiotemporally related.)

This is perhaps more controversial than it seems. Didn't I speak, in connection with predetermination, of worlds that diverge? That is, of worlds that are exactly alike up to some time, and differ thereafter? Doesn't that presuppose trans-world comparison of times, simultaneity or succession between events of different worlds? Trans-world spatiotemporal relations between the participants in those events, or the spacetime regions in which they happen?

I think not. Trans-world comparisons, yes; trans-world spatiotemporal relations, no.

Suppose two worlds are exactly alike up to a certain time, and diverge thereafter. I explain it thus. There is an initial segment of one world, and there is an initial segment of the other, which are perfect duplicates. They are maximal such segments: they are not respectively included in two larger initial segments which are also duplicates. There is a correspondence between the parts of these two segments under which the corresponding parts also are duplicates; and under which corresponding parts are related alike spatiotemporally, and as whole to part. Therefore the corresponding parts are excellent counterparts. They are so whether you take a counterpart relation that stresses similarity of intrinsic

[52]The question is raised by Richards. I am grateful to him, and to David Johnson, for helpful discussion of it.

character, or one that stresses extrinsic match of origins, or even one that stresses historical role. (Except insofar as something that is part of the duplicated region has a historical role lying partly outside that region.) Temporal cross-sections of the worlds, for instance, are excellent counterparts: there are counterpart centuries, or weeks, or seconds. Likewise there are counterpart places: galaxies, planets, towns. So things that are parts of the two worlds may be simultaneous or not, they may be in the same or different towns, they may be near or far from one another, in very natural counterpart-theoretic senses. But these are not genuine spatiotemporal relations across worlds. The only trans-world relations involved are internal relations of similarity; not indeed between the very individuals that are quasi-simultaneous (or whatever) but between larger duplicate parts of the two worlds wherein those individuals are situated.

Suppose you discovered – say, from a well-accredited oracle – that large parts of human history were re-enacted, with interesting variations, in remote galaxies at times in the distant past and future. In speaking of these re-enactments, you would surely introduce counterpart-theoretic comparisons of place and time. You might say that a remarkable event in one of them took place last year in Headington; when you would also say, without any conflict, that it will take place about 6.4×10^{12} years hence, 3.8×10^{9} light years away in the general direction of the constellation Centaurus. You should have no greater difficulty in squaring talk about other-worldly goings-on last week in Didcot with my denial that there are any spatiotemporal relations between parts of different worlds.

So we have a sufficient condition: if two things are spatiotemporally related, then they are worldmates. The converse is much more problematic. Yet that is more or less the doctrine that I propose. Putting the two halves together: things are worldmates iff they are spatiotemporally related. A world is unified, then, by the spatiotemporal interrelation of its parts. There are no spatiotemporal relations across the boundary between one world and another; but no matter how we draw a boundary within a world, there will be spatiotemporal relations across it.

A first, and simplest, objection is that a world might possibly consist of two or more completely disconnected spacetimes. (Maybe *our* world does, if indeed such disconnection is possible.) But whatever way a world might be is a way that some world is; and one world with two disconnected spacetimes is a counterexample against my proposal. Against this objection, I must simply deny the premise. I would rather not; I admit some inclination to agree with it. But it seems to me that it is no central part of our modal thinking, and not a consequence of any interesting general principle about what is possible. So it is negotiable. Given a choice between

rejecting the alleged possibility of disconnected spacetimes within a single world and (what I take to be the alternative) resorting to a primitive worldmate relation, I take the former to be more credible.

I cannot give you disconnected spacetimes within a single world; but I can give you some passable substitutes. One big world, spatiotemporally interrelated, might have many different world-like parts. *Ex hypothesi* these are not complete worlds, but they could seem to be. They might be four-dimensional; they might have no boundaries; there might be little or no causal interaction between them. Indeed, each of these world-like parts of one big world might be a duplicate of some genuinely complete world. There are at least four ways for one big world to contain many world-like parts. Each is a way that a world could be; and so, say I, each is a way that some world is.

(1) The spacetime of the big world might have an extra dimension. The world-like parts might then be spread out along this extra dimension, like a stack of flatlands in three-space.

(2) The world-like parts might share a common spacetime. There might be several populations, interpenetrating without interaction in the single spacetime where all of them live. If so, of course the inhabitants had better not interact with the shape of their spacetime as we do with the shape of ours; else this interaction enables the different populations to interact indirectly with one another.

(3) Time might have the metric structure not of the real line, but rather of many copies of the real line laid end to end. We would have many different epochs, one after another. Yet each epoch would have infinite duration, no beginning, and no end. Inhabitants of different epochs would be spatiotemporally related, but their separation would be infinite. Or instead there might be infinitely many infinite regions laid out side by side in space; then there would have to be infinite spatial distances between points in different world-like regions.

(4) Or time might have the metric structure of the real line, as we normally suppose. And yet there might be infinitely many world-like epochs one after the other. Each might be of finite duration; but their finitude might be hidden from their inhabitants because, as the end of an epoch approaches, everything speeds up. Suppose that one generation lives and dies in twelve months, the next in six, the next in three, . . . so that infinitely many generations fit into the last two years of their epoch. Similarly, world-like regions of finite diameter might be packed spatially, with shrinkage as things approach the edge.

If you thought, as I did too, that a single world might consist of many more or less isolated world-like parts, how sure can you be that you really had in mind the supposed possibility that I reject? Are you sure that it was an essential part of your thought that the world-like parts were in no way spatiotemporally related? Or might you not have had in mind,

rather, one of these substitutes I offer? Or might your thought have been sufficiently lacking in specificity that the substitutes would do it justice?

A second objection concerns spirits, and episodes in the mental lives of spirits, which are traditionally supposed to be outside of space. However sure we are that no such deficient things are worldmates of ours, is it not at least possible that the traditional story might be true? If so, then some world is populated by such spirits. But that is no objection. I do not say that all worlds are unified by spatiotemporal interrelatedness in just the same way. So the interrelation of a world of spirits might be looser than that of a decent world like ours. If the spirits and their doings are located in time alone, that is good enough. (To make sense of that, maybe time and space would have to be more separable at the world of the spirits than they are at our world; but that is surely possible.) I can even allow marvellous Spirits who are spatiotemporally related to other things by being omnipresent – for that is one way among others to stand in spatiotemporal relations. I am not sure why I need to defend the possibility of spirit tales – after all, people have been known to accept impossible theories, as witness naive set theory – but in fact I think I give them at least as much room in logical space as they deserve.

A third objection concerns the possibility that there might be nothing, and not rather something. If a world is a maximal mereological sum of spatiotemporally interrelated things, that makes no provision for an absolutely empty world. A world is not like a bottle that might hold no beer. The world *is* the totality of things it contains, so even if there's no beer, there's still the bottle. And if there isn't even the bottle, there's nothing there at all. And nothing isn't a very minimal something. Minimal worlds there can indeed be. There can be nothing much: just some homogeneous unoccupied spacetime, or maybe only one single point of it. But nothing much is still something, and there isn't any world where there's nothing at all. That makes it necessary that there is something. For it's true at all worlds that there is something: it's true whenever we restrict our quantifiers to the domain of parts of a single world, even if the only part of some world is one indivisible nondescript point. Of course, if we don't restrict quantifiers from the standpoint of one world or another, then all the more is it true that there is something rather than nothing: there is logical space, the totality of the worlds in all their glory.

How bad is this? I think the worst of it is the fear that I might offer to *explain* why there is something rather than nothing, just by saying that this is a necessary truth. But don't fear; I do not think that would be an explanation. For an explanation, I think, is an account of etiology: it tells us something about how an event was caused. Or it tells us something general about how some, or many, or all events of a certain kind are caused. Or it explains an existential fact by telling us something about how several events jointly make that fact true, and then perhaps

something about how those truthmaker events were caused. So I think there is nothing I might say that could count as explaining why there is something rather than nothing; and that includes saying, truly, that there is no world where there is nothing.[53]

So far I am stonewalling. I accept the unwelcome consequences of my thesis, and claim they are not as bad as you might think. But there is one more objection to consider, and this one really does seem to me to call for a retreat. The last resort would be a primitive worldmate relation, but I think it won't be necessary to fall back that far.

Imagine a theory of spacetime that is built for Newtonian mechanics, or for common sense. (Old-fashioned Newtonian mechanics, as opposed to recent reformulations that are still in a sense Newtonian, but do away with absolute rest.) This theory will say that any two spacetime points are related by a spatial distance and a temporal distance: two different distances. One but not both of these distances may be zero, thus absolute simultaneity and absolute rest both are well defined. I suppose this is a way the world might have been, therefore it is a way that some world is. But we have good reason to think that our world is different. In our relativistic world, any two spacetime points have only one distance between them; it may be a spatial distance, it may be a temporal distance, or it may be a zero distance which is neither spatial nor temporal ('space-like' interval coded by a positive real, 'time-like' interval coded by a positive imaginary, or 'light-like' interval). Of course there are other differences between Newtonian and relativistic spacetime, but this difference of two distances versus one is the difference that matters to ontology.

We name the properties and relations that figure in our world; so what we call 'spatiotemporal relations' are relations that behave in the relativistic way, with spatial or temporal distance but not both. Now when we talk about the Newtonian world, are we talking about the possibility of different behaviour on the part of those same relations? Is it that those very relations might double up to give us two distances, one of each kind, between the same two points? Or are we talking instead about some different relations that might take the place of the spatiotemporal relations of our world?[54]

[53]I find it pleasing that another view, the one I like second best after my own, also seems to make it come out necessary that there is something rather than nothing. This is the 'combinatorial' view: in place of other worlds, we have constructions in which the elements of this world – elementary particulars and universals, perhaps – are put together in different combinations. (See section 3.2, in which I present this as a form of 'linguistic ersatzism'.) But as D. M. Armstrong has noted in discussion, there is no way to combine elements and make nothing at all. So there is no combinatorial possibility that there might be nothing.

[54]What does this question mean? Maybe one thing, maybe another, depending on our underlying theory of natural properties and relations; and on that question I am staying

If it is the former, no worries. The Newtonian world is just as much spatiotemporally interrelated as ours is, even if the spatiotemporal relations behave differently there. But if it is the latter, then strictly speaking I cannot say that the Newtonian world is *spatiotemporally* interrelated. It has its system of external relations, whereby its parts are arranged, which are analogous to the spatiotemporal relations whereby the parts of our world are arranged. But these Newtonian impostors are not to be called the 'spatiotemporal relations', because that is the name we gave to the different relations that hold between the parts of our world. (It is beside the point that when we named the relations of our world, we may have thought they behaved in the Newtonian rather than the relativistic way. However much we intended to name relations that conformed to some theory, doubtless we intended much more to name relations that are pervasive in our world.) Similarly, *mutatis mutandis*, if the inhabitants of a Newtonian world talk about the possibility of a world like ours. Suppose they did pretty much what we did in naming what they call 'spatiotemporal relations'; and suppose it is not so that the very same relations behave in the Newtonian way at one world and in the relativistic way at the other. Then they should not say, strictly speaking, that our world is 'spatiotemporally interrelated'.

I do not know how to answer the question whether we have the same relations in the different worlds. It might even have different answers in different cases: some pairs of a Newtonian and a relativistic world use the same relations (doubled up for the Newtonian world), other pairs don't. Also, I suppose some worlds are interrelated by systems of external relations that differ more, at least in their behaviour, than Newtonian doubled-up distances differ from relativistic distances. It would be nice to suppose that all worlds are interrelated by the very same relations, namely the ones that we call 'spatiotemporal', despite whatever behavioural differences there may be. I do not reject this supposition. But I am unwilling to rely on it.

What I need to say is that each world is interrelated (and is maximal with respect to such interrelation) by a system of relations which, if they are not the spatiotemporal relations rightly so called, are at any rate analogous to them. Then my task is to spell out the analogy. At least some of the points of analogy should go as follows. (1) The relations are

---

neutral between three alternatives. (See section 1.5.) (1) Maybe naturalness is a primitive, applied to properties or relations understood as sets. Then we have families of relations that can serve as the common spatiotemporal relations of all the worlds, and we have other families of less inclusive relations that can serve as the different special spatiotemporal relations for different kinds of worlds, and the question is which relations are more natural. (2) Maybe a relation is natural when its instances share a relational universal; then the question is what universals there are. (3) Maybe a relation is natural when its instances contain duplicate tropes; then the question is what tropes there are.

natural; they are not gruesome gerrymanders, not even mildly disjunctive. (2) They are *pervasive*: mostly, or perhaps without exception, when there is a chain of relations in the system running from one thing to another, then also there is a direct relation. (3) They are *discriminating*: it is at least possible, whether or not it happens at every world where the relations are present, that there be a great many interrelated things, no two of which are exactly alike with respect to their place in the structure of relations. (4) They are *external*: they do not supervene on the intrinsic natures of the *relata* taken separately, but only on the intrinsic character of the composite of the *relata*. (See section 1.5. The definition of what it is for a relation to be external involved *possibilia* but not yet possible worlds, and so is available at this point without circularity.) When a system of relations is analogous to the spatiotemporal relations, strictly so called, let me call them *analogically* spatiotemporal.[55]

I have some hope that it might be possible to bypass the messy idea of analogically spatiotemporal relations. A much simpler alternative would be that worlds are unified by *external* interrelatedness, of whatever sort. On this suggestion, *any* natural external relations will do to unify a world. Every part of a world bears some such relation to every other part; but no part of one world ever bears any such relation to any part of another.

[55]There are three different conceptions of what the spatiotemporal relations might be. There is the dualist conception: there are the parts of spacetime itself, and there are the pieces of matter or fields or whatnot that occupy some of the parts of spacetime. Then the spatiotemporal relations (strict or analogical) consist of distance relations that hold between parts of spacetime; relations of occupancy that hold between occupants and the parts of spacetime they occupy; and, derivatively from these, further distance relations between the occupants, or between occupants and parts of spacetime.

There are two simpler monistic conceptions. One of them does away with the occupants as separate things: we have the parts of spacetime, and their distance relations are the only spatiotemporal relations. The properties that we usually ascribe to occupants of spacetime – for instance, properties of mass, charge, field strength – belong in fact to parts of spacetime themselves. When a part of spacetime has a suitable distribution of local properties, then it is a particle, or a piece of a field, or a donkey, or what have you.

The other monistic conception does the opposite: it does away with the parts of spacetime in favour of the occupants (now not properly so called), so that the only spatiotemporal relations are the distance relations between some of these. I tend to oppose the third conception, at least as applied to our world, for much the reasons given in Nerlich, *The Shape of Space*. I tend, more weakly, to oppose the dualist conception as uneconomical. I suppose it may be, however, that there are worlds of all three sorts; if so, that would give more reason than ever to doubt that the same system of spatiotemporal relations serves to unify all the worlds. Throughout this book, I shall presuppose that there are such things as spatiotemporal regions, whether or not there also are distinct things that occupy those regions. But I believe this presupposition plays no important role, and I could have been more neutral at the cost of clumsier writing. I certainly don't mean to suggest that the existence of spacetime and its parts is an essential tenet of modal realism.

Never mind whether the relations in question are spatiotemporal, either strictly or analogically.

If the simplification is to have a hope, the restriction to natural relations must bear a good deal of weight. It will have to exclude more than just the gruesome gerrymanders. For what about the relation of *non-identity*? (Here I am indebted to discussion with James Grieve.) It qualifies on my definition as an external relation, and it obtains invariably between the particular parts of different worlds. However, we may fairly deny it a place in our select inventory of the natural relations. It would be superfluous to include it if we have the resources to introduce it by definition; and we do, since X and Y are non-identical iff there is a class that one of X and Y belongs to and the other does not. (If you think there is need to cover non-identity of 'proper classes' you should add a clause: '. . . or there is something that belongs to one of X and Y but not the other'.)

I find it hard to say whether this simplification could succeed. My problem is a lack of test cases. What natural external relations could there be besides the (strictly or analogically) spatiotemporal relations? I would reject some candidates for further external relations that might be offered: for instance, primitive genidentity relations, non-qualitative counterpart relations (see section 4.4), or a primitive worldmate relation.

Perhaps the following will do as a test case. If so, it looks unfavourable for the simplification. We tend to think that positive and negative charge are natural intrinsic properties of particles; but suppose not. Suppose instead there are natural external relations of like-chargedness and opposite-chargedness. (Then we can introduce extrinsic versions of the charge properties. To be neutral is to be like-charged to some particles and opposite-charged to none; to be negative is to be like-charged to most of the lightweight particles that orbit much heavier clumps of particles hereabouts; and to be positive is to be opposite-charged to a negative particle.) On this view, as opposed to the standard view, the relations of like- and opposite-chargedness do *not* supervene on the intrinsic natures of two particles taken separately; an electron and a positron may be perfect intrinsic duplicates. That is the point of calling the relations external. They are natural *ex hypothesi*. They are pervasive (at least, given the appropriate laws) in what whenever two particles are connected by a chain of such relations, they are connected directly. But they are very far from discriminating (again, given the appropriate laws): if there are as few as three particles, there must be two of them that are alike so far as these relations are concerned. If this story, or something like it, could be true, then here we have external relations that are not strictly or analogically spatiotemporal.

Could two particles in different worlds stand in these external relations of like- or opposite-chargedness? So it seems, offhand; and if so, then

the simplification fails. I would welcome a reason why particles at different worlds cannot stand in these relations – other than a verificationist reason that I would find unpersuasive – but failing such a reason, I am inclined to reject the simplification. Then I must stick instead with my underdeveloped suggestion that the unifying external relations have to be, if not strictly spatiotemporal, at least analogically spatiotemporal.

There is a second way in which the worlds are isolated: there is no causation from one to another. If need be, I would put this causal isolation alongside spatiotemporal isolation as a principle of demarcation for worlds. But there is no need. Under a counterfactual analysis of causation, the causal isolation of worlds follows automatically. Therefore it contributes nothing to the demarcation of one world from another. No matter how we solve the demarcation problem, trans-world causation comes out as nonsense.

When we have causation within a world, what happens is roughly as follows. (For simplicity I ignore complications having to do with causal pre-emption and overdetermination, and with the idealisation of supposing that we always have closest antecedent-worlds. Taking these matters into account would do nothing in favour of trans-world causation.) We have a world W where event C causes event E. Both these events occur at W, and they are distinct events, and it is the case at W that if C had not occurred, E would not have occurred either. The counterfactual means that at the closest worlds to W at which C does not occur, E does not occur either.

Try to adapt this to a case of trans-world causation, in which the events of one world supposedly influence those of another. Event C occurs at world W_C, event E occurs at world E_E, they are distinct events, and if C had not occurred, E would not have occurred either. This counterfactual is supposed to hold – where? It means that at the closest worlds to – where? – at which C does not occur, E does not occur – where? – either.

Normally the counterfactual is supposed to hold at the world where the one event causes the other; so maybe if the causation goes between two worlds, the counterfactual ought to hold at both. So we have:

(1) at the closest worlds to W_C at which C does not occur, E does not occur either, and

(2) at the closest worlds to W_E at which C does not occur, E does not occur either.

But (1) looks wrong: since we are looking at a supposed case of trans-world causation, it is irrelevant to ask whether we get E at worlds close to W_C; we ought to be looking at worlds close to W_E, the world where the supposed effect did take place. And (2) looks even worse: we ought

to be hypothesising the removal of C from a world like W_C: removing it from a world like W_E is irrelevant. In fact, the closest world to W_E at which C does not occur might very well be W_E itself!

So should we make sure that we make our revisions to the right worlds by specifying explicitly which worlds the events are to be removed from? Like this:

(1′) at the closest worlds to W_C at which C does not occur at W_C, E does not occur at W_E, and

(2′) at the closest worlds to W_E at which C does not occur at W_C, E does not occur at W_E.

But this is worse than ever. What can these double modifications mean: at *this* world, an event does not occur at *that* world? C just *does* occur at W_C, E just *does* occur at W_E; there is no world at which these facts are otherwise. You might as well say that in Auckland it rains in Melbourne, but in Wellington it doesn't rain in Melbourne. There is no way to make literal sense of this, unless by taking the outer modifier as vacuous. (That is why you instantly thought of two ways to make *non*-literal sense of it: in Auckland they say it rains in Melbourne, but they don't say so in Wellington; it rains a lot in Melbourne compared to Auckland, but not compared to Wellington.)

Try this. As the one world is to ordinary causation, so the pair of worlds is to trans-world causation. So put pairs for single worlds throughout:

(3) at the closest world-pairs to the pair $\langle W_C, W_E \rangle$ such that C does not occur at the first world of the pair, E does not occur at the second world of the pair.

This makes sense, but not I think in a way that could make it true. For I suppose that the closeness of one world-pair to another consists of the closeness of the first worlds of the pairs together with the closeness of the second worlds of the pairs. We have to depart from W_C for the first world of a closest pair, since we have to get rid of C. But we are not likewise forced to depart from W_E for the second world of a closest pair, and what is so close to a world as that world itself? So the second world of any closest pair will just be W_E, at which E does occur, so (3) is false.

(If there were significant external relations between worlds, that might provide another respect of comparison for world-pairs. But to this I say, first, that even if trans-world external relations are not absolutely forbidden by our solution to the problem of demarcation, the permitted ones would be such things as our imagined relations of like- and opposite-chargedness, which don't seem to do anything to help (3) to come true;

and second, that if our special world-pair counterfactuals are supposed to make for causal dependence, they had better be governed by the same sort of closeness that governs ordinary causal counterfactuals, but ordinary closeness of worlds does not involve any trans-world external relations that might make world-pairs close.)

When it seems to us as if we can understand trans-world causation, I think that what must be happening is as follows. We think of the totality of all the possible worlds as if it were one grand world, and that starts us thinking that there are other ways the grand world might have been. So perhaps what we really have in mind is:

> (4) at the closest alternative grand worlds to ours where C does not occur in the part corresponding to W_C, E does not occur in the part corresponding to W_E.

But this is thoroughly misguided. If I am right, the many worlds already provide for contingency, and there is no sense in providing for it all over again. Or else I am wrong, and the many worlds do not provide for genuine contingency. (As some think; see section 2.1.) But then it makes no sense to repeat the very method you think has failed, only on a grander scale. The worlds are all of the maximal things that are suitably unified. If they fall into grand clusters, and yet grander clusters of clusters, and so on, that is neither here nor there. By 'worlds' I still mean *all* the worlds. (And how could they fall into clusters – what sort of relation could unify a cluster without also merging the worlds within it?) There is but one totality of worlds; it is not a world; it could not have been different. Therefore (4) is nonsense, intelligible only if taken as vacuous.

So there isn't any trans-world causation. And not because I so stipulate as a principle of demarcation, but as a consequence of my analyses of causation and of counterfactuals. This is the real reason why there couldn't be a very powerful telescope for viewing other worlds. The obstacle isn't that other worlds are too far away, as Kripke jokingly says; and it isn't that they're somehow 'abstract', as of course he really thinks. (See *Naming and Necessity*, pages 44 and 19.) Telescopic viewing, like other methods of gathering information, is a *causal* process: a 'telescope' which produced images that were causally independent of the condition of the thing 'viewed' would be a bogus telescope. No trans-world causation, no trans-world telescopes.

Likewise, if there is no trans-world causation, there is no trans-world travel. You can't get into a 'logical-space ship' and visit another possible world. You could get into what you confusedly think is a logical-space ship, turn the knob, and disappear. And a perfect duplicate of you at your disappearance, surrounded by a perfect duplicate of your ship, could appear *ex nihilo* at some other world. Indeed, there are plenty of worlds

where aspiring logical-space travellers disappear, and plenty of worlds where they appear, and plenty of qualitative duplications between ones that disappear and ones that appear. But none of this is travel unless there is one surviving traveller who both departs and arrives. And causal continuity is required for survival; it is a principal part of what unifies a persisting person. It is so within a single world: if there is a demon who destroys people at random, and another who creates people at random, and by a very improbable coincidence the creating demon replaces a victim of the destroying demon, the qualitative continuity could be perfect, but the lack of causal dependence would still make it not be a genuine case of survival. Likewise across worlds. No trans-world causation, no trans-world causal continuity; no causal continuity, no survival; no survival, no travel. All those people in various worlds who meet their ends in 'logical-space ships', as well as the more fortunate ones who appear *ex nihilo* in such ships, are sadly deluded.

But if you'd like to see a world where Napoleon conquered all, don't give up hope. Maybe ours is one of those big worlds with many world-like parts, spatiotemporally related in some peculiar way. Then you might get your wish, near enough, by means of a special telescope or a special spaceship that operates entirely within our single world. You won't see the world-like part where Napoleon himself is, of course; you're there already, and he didn't conquer all. But I presume you'd be content with a world-like part where the conqueror was an excellent counterpart of Napoleon. I would be the last to denounce decent science fiction as philosophically unsound. No; tales of viewing or visiting 'other worlds' are perfectly consistent. They come true at countless possible worlds. It's just that the 'other worlds' that are viewed or visited never can be what *I* call 'other worlds'.

1.7 Concreteness

Because I said that other worlds are of a kind with this world of ours, doubtless you will expect me to say that possible worlds and individuals are concrete, not abstract. But I am reluctant to say that outright. Not because I hold the opposite view; but because it is not at all clear to me what philosophers mean when they speak of 'concrete' and 'abstract' in this connection. Perhaps I would agree with it, whatever it means, but still I do not find it a useful way of explaining myself.

I can say this much, even without knowing what 'concrete' is supposed to mean. I take it, at least, that donkeys and protons and puddles and stars are supposed to be paradigmatically concrete. I take it also that the division between abstract and concrete is meant to divide entities into fundamentally different kinds. If so, then it is out of the question that

an abstract entity and a concrete entity should be exactly alike, perfect duplicates. According to my modal realism, the donkeys and protons and puddles and stars that are parts of this world have perfect duplicates that are parts of other worlds. This suffices to settle, whatever exactly it may mean, that at least some possible individuals are 'concrete'. And if so, then at least some possible worlds are at least partly 'concrete'.

A spectator might well assume that the distinction between 'concrete' and 'abstract' entities is common ground among contemporary philosophers, too well understood and uncontroversial to need any explaining. But if someone does try to explain it, most likely he will resort to one (or more) of four ways.[56]

First, the Way of Example: concrete entities are things like donkeys and puddles and protons and stars, whereas abstract entities are things like numbers. That gives us very little guidance. First, because we have no uncontroversial account of what numbers are. Are the paradigms of abstractness meant to be the von Neumann ordinals – certain pure sets? Are they meant to be structural universals, instantiated here and there within our world, like the tripartiteness that is instantiated wherever there is a proton composed of quarks (if quarks themselves are mereological atoms)? Are they 'irreducible *sui generis* abstract entities'? And even given a useful account of the nature of numbers, there are just too many ways that numbers differ from donkeys *et al.* and we still are none the wiser about where to put a border between donkey-like and number-like.

At least the Way of Example has something to say about some parts of other worlds. As noted above, some parts of other worlds are *exactly* like donkeys, because they are donkeys, so those at any rate are paradigmatically concrete. Likewise for other-worldly puddles and protons and stars. So far, so good. But other parts of other worlds are, for instance, chunks of other-worldly spacetime – are those paradigmatically concrete? And if ordinary particulars contain universals or tropes as (non-spatiotemporal) parts, then worlds composed of ordinary particulars will in turn have universals or tropes as parts; in which case not all the parts of worlds are paradigmatically concrete. Indeed we might contemplate a theory of numbers – one which says, for instance, that the number three is the structural universal of tripartiteness – according to which some parts of worlds would turn out to be paradigmatically abstract.

And what of a whole world? Is it sufficiently donkey-like, despite its

[56] I shall pass over a fifth way, offered by Dummett in chapter 14 of *Frege: Philosophy of Language*, in which the distinction between abstract and concrete entities is drawn in terms of how we could understand their names. Even if this fifth way succeeds in drawing a border, as for all I know it may, it tells us nothing directly about how the entities on opposite sides of that border differ in their nature. It is like saying that snakes are the animals we instinctively most fear – maybe so, but it tells us nothing about the nature of snakes.

size? And perhaps despite the fact that it consists mostly of empty spacetime? I am inclined to say that, according to the Way of Example, a world is concrete rather than abstract – more donkey-like than number-like. I am also inclined to say that a world is more like a raven than a writing-desk; and that it is ping rather than pong. But I know not why.

Second, the Way of Conflation: the distinction between concrete and abstract entities is just the distinction between individuals and sets, or between particulars and universals, or perhaps between particular individuals and everything else. That accords well enough with our examples. It is safe to say that donkeys and the like are particular individuals, not universals or sets. It is a defensible, if not trouble-free, view that numbers are sets; alternatively, it is arguable that they are universals. So far, so good. I say that worlds are individuals, not sets. I say that worlds are particulars, not universals. So according to the Way of Conflation in either version, I say that worlds are concrete.

Third, the Negative Way: abstract entities have no spatiotemporal location; they do not enter into causal interaction; they are never indiscernible one from another.

The Negative Way and the Way of Conflation seem to disagree rather badly. As for the first part, the denial that abstract entities are located, I object that by this test some sets and universals come out concrete. Sets are supposed to be abstract. But a set of located things *does* seem to have a location, though perhaps a divided location: it *is* where its members *are*. Thus my unit set is right here, exactly where I am; the set of you and me is partly here where I am, partly yonder where you are; and so on. And universals are supposed to be abstract. But if a universal is wholly present in each of many located particulars, as by definition it is, that means that it is where its instances are. It is multiply located, not unlocated. You could just declare that an abstract entity is located only in the special way that a set or a universal is located – but then you might as well just say that to be abstract is to be a set or universal. Your talk of unlocatedness adds nothing. Maybe a *pure* set, or an *uninstantiated* universal, has no location. However these are the most dispensable and suspect of sets and universals. If it is said that sets or universals generally are unlocated, perhaps we have a hasty generalisation. Or perhaps we have an inference: they're unlocated because they're abstract. If so, we had better not also say that they're abstract because they're unlocated.

As for the second part, the denial that abstract entities enter into causal interaction, this too seems to disagree with the Way of Conflation. Is it true that sets or universals cannot enter into causal interaction? Why shouldn't we say that something causes a set of effects? Or that a set of causes, acting jointly, causes something? Or that positive charge causes effects of a characteristic kind whenever it is instantiated? Many authors have proposed to identify an event – the very thing that most surely can

cause and be caused – with one or another sort of set. (For instance, in 'Events' I propose to identify an event with the set of spacetime regions where it occurs.) Must any such identification be rejected, regardless of the economies it may afford, just because sets are supposed to be 'abstract'?

As for the third part, the denial that abstract entities can be indiscernible, indeed I do not see what could be said in favour of indiscernible universals. But as for sets, I should think that if two individuals are indiscernible, then so are their unit sets; and likewise whenever sets differ only by a substitution of indiscernible individuals. So, *pace* the Way of Conflation, it seems that the Negative Way does not classify universals, or sets in general, as abstract.

What does it say about worlds? Other worlds and their parts certainly stand in no spatiotemporal or causal relations to *us*. Worlds are spatiotemporally and causally isolated from one another; else they would be not whole worlds, but parts of a greater world. But by the same token, we stand in no spatiotemporal or causal relation to *them*. That doesn't make us abstract. It's no good saying that, for us, we are concrete and an other-worldly being is abstract; whereas, for that other-worldly being, he is concrete and we are abstract. For one thing is certain: whatever the abstract–concrete distinction is, at least it's supposed to be a very fundamental difference between two kinds of entities. It has no business being a symmetrical and relative affair.

So the right question is: do other worlds and their parts stand in spatiotemporal and causal relations to anything? Parts of worlds do: they stand in (strictly or analogically) spatiotemporal relations, and in causal relations, to other parts of their own worlds. (With exceptions. Maybe a tiny world might have only one part. A chaotic and lawless world might have no causation. But I presume we don't want to say that parts of worlds are abstract in these special cases, concrete otherwise.) Whole worlds, however, cannot stand in spatiotemporal and causal relations to anything outside themselves, and it seems that nothing can stand in such relations to its own parts. Should we conclude that worlds – including the one we are part of – are abstract wholes made of concrete parts? Perhaps, indeed, divisible exhaustively into concrete parts? That seems unduly literalistic – presumably the Negative Way should be construed charitably, so that wholes can inherit concreteness from their parts. As for indiscernibility, I have no idea whether there are indiscernible worlds; but certainly there are indiscernible parts of worlds, for instance indiscernible epochs of a world of two-way eternal recurrence. So according to the Negative Way, charitably read, I say that worlds and their parts – including the universals, if such there be! – are concrete.

Fourth, the Way of Abstraction: abstract entities are abstractions from concrete entities. They result from somehow subtracting specificity, so

that an incomplete description of the original concrete entity would be a complete description of the abstraction. This, I take it, is the historically and etymologically correct thing to mean if we talk of 'abstract entities'. But it is by no means the dominant meaning in contemporary philosophy.

A theory of non-spatiotemporal parts of things, whether these be recurring universals or non-recurring tropes, makes good sense of some abstractions. We can say that unit negative charge is a universal common to many particles, and is an abstraction from these particles just by being part of each of them. Or we can say that the particular negative charge of this particular particle is part of it, but a proper part and in that sense an abstraction from the whole of it. But we cannot just identify abstractions with universals or tropes. For why can we not abstract some highly extrinsic aspect of something – say, the surname it bears? Or its spatiotemporal location? Or its role in some causal network? Or its role in some body of theory? All these are unsuitable candidates for genuine universals or tropes, being no part of the intrinsic nature of the thing whence they are abstracted.

We can also make good sense of abstractions, or an adequate imitation thereof, by the stratagem of taking equivalence classes. For instance, we abstract the direction of a line from the line itself by taking the direction to be the class of that line and all other lines parallel to it. There is no genuine subtraction of specific detail, rather there is multiplication of it; but by swamping if not by removal, the specifics of the original line get lost. For instance, the direction comprises many located lines; it is located where its members are, namely everywhere; so it is not located more one place than another, and that is the next best thing to not being anywhere. But sets in general cannot be regarded thus as abstractions: most sets are equivalence classes only under thoroughly artificial equivalences. (And the empty set is not an equivalence class at all.) Further, if we abstract by taking equivalence classes, we need not start with paradigmatically concrete things. Thus directions may be abstracted from lines, but the lines themselves may be taken as certain sets of quadruples of real numbers.

So even if universals and equivalence classes are abstractions, it remains that the Way of Abstraction accords badly with the Ways of Example and of Conflation. It accords no better with the Negative Way: if we can abstract the spatiotemporal location of something, that abstraction will not be unlocated; rather, there will be nothing to it except location. Likewise if we can abstract the causal role of something, then the one thing the abstraction will do is enter into causal interactions.

Unless understood as universals or tropes or equivalence classes, abstractions are obviously suspect. The inevitable hypothesis is that they are verbal fictions: we say 'in the material mode' that we are speaking about the abstraction when what's true is that we are speaking abstractly

about the original thing. We are ignoring some of its features, not introducing some new thing from which those features are absent. We purport to speak of the abstraction 'economic man'; but really we are speaking of ordinary men in an abstract way, confining ourselves to their economic activities.

According to the Way of Abstraction, I say that worlds are concrete. They lack no specificity, and there is nothing for them to be abstractions from. As for the parts of worlds, certainly some of them are concrete, such as the other-worldly donkeys and protons and puddles and stars. But if universals or tropes are non-spatiotemporal parts of ordinary particulars that in turn are parts of worlds, then here we have abstractions that are parts of worlds.

So, by and large, and with some doubts in connection with the Way of Example and the Negative Way, it seems that indeed I should say that worlds as I take them to be are concrete; and so are many of their parts, but perhaps not all. But it also seems that to say that is to say something very ambiguous indeed. It's just by luck that all its disambiguations make it true.

1.8 Plenitude

At the outset, I mentioned several ways that a world might be; and then I made it part of my modal realism that

> (1) absolutely every way that a world could possibly be is a way that some world is, and
>
> (2) absolutely every way that a part of a world could possibly be is a way that some part of some world is.

But what does that mean? It *seems* to mean that the worlds are abundant, and logical space is somehow complete. There are no gaps in logical space; no vacancies where a world might have been, but isn't. It seems to be a principle of plenitude. But is it really?

Given modal realism, it becomes advantageous to identify 'ways a world could possibly be' with worlds themselves. Why distinguish two closely corresponding entities: a world, and also the maximally specific way that world is? Economy dictates identifying the 'ways' with the worlds.

But as Peter van Inwagen pointed out to me, that makes (1) contentless. It says only that every world is identical to some world. That would be true even if there were only seventeen worlds, or one, or none. It says nothing at all about abundance or completeness. Likewise for (2).

Suppose we thought a maximally specific 'way' should be the same kind of things as a less specific 'way': namely a property, taken as a set. Then a maximally specific 'way' would be a unit set. Now indeed the 'ways' are distinct from the worlds. Further, they are abstract in whatever sense sets are. But this does nothing to restore content to (1). A 'possible way' is a *non-empty* set, and (1) now says trivially that each of the unit sets has a member.[57]

Or perhaps a 'way' should be not a unit set, but an equivalence class under indiscernibility. I am agnostic about whether there are indiscernible worlds. If there are, I myself would wish to say that there are indiscernible ways a world could be, just as I would say that a world of two-way eternal recurrence affords countless indiscernible ways – one per epoch – for a person to be. But others might not like the idea of indiscernible 'ways'. They might therefore welcome a guarantee that, whether or not worlds ever are indiscernible, 'ways' never will be. Now (1) says trivially that each of the equivalence classes has a member.

Or suppose we thought a 'way' should be the intrinsic nature of a world, a highly complex structural universal (as in Forrest, 'Ways Worlds Could Be'.) Given that thesis, a 'possible way' is an *instantiated* universal. Now (1) says trivially that each of these has a world to instantiate it.

We might read (1) as saying that every way we *think* a world could possibly be is a way that some world is; that is, every seemingly possible description or conception of a world does fit some world. Now we have made (1) into a genuine principle of plenitude. But an unacceptable one. So understood, (1) indiscriminately endorses offhand opinion about what is possible.

I conclude that (1), and likewise (2), cannot be salvaged as principles of plenitude. Let them go trivial. Then we need a new way to say what (1) and (2) seemed to say: that there are possibilities enough, and no gaps in logical space.

To which end, I suggest that we look to the Humean denial of necessary connections between distinct existences. To express the plenitude of possible worlds, I require a *principle of recombination* according to which patching together parts of different possible worlds yields another possible

[57]Some critics have thought it very important that the 'ways' should be 'abstract' entities and distinct from the worlds. For instance, see Stalnaker, 'Possible Worlds'; and van Inwagen, who writes 'the cosmos, being concrete, is not a way things could have been. . . . And surely the cosmos cannot itself be identical with any way the cosmos could have been: to say this would be like saying that Socrates is identical with the way Socrates is, which is plain bad grammar.' ('Indexicality and Actuality', page 406.) But to me, the choice whether to take a 'way' as a unit set or as its sole member seems to be of the utmost unimportance, on a par with the arbitrary choice between speaking of a set or of its characteristic function.

world. Roughly speaking, the principle is that anything can coexist with anything else, at least provided they occupy distinct spatiotemporal positions. Likewise, anything can fail to coexist with anything else. Thus if there could be a dragon, and there could be a unicorn, but there couldn't be a dragon and a unicorn side by side, that would be an unacceptable gap in logical space, a failure of plenitude. And if there could be a talking head contiguous to the rest of a living human body, but there couldn't be a talking head separate from the rest of a human body, that too would be a failure of plenitude.

(I mean that plenitude requires that there could be a separate thing *exactly like* a talking head contiguous to a human body. Perhaps you would not wish to call that thing a 'head', or you would not wish to call what it does 'talking'. I am somewhat inclined to disagree, and somewhat inclined to doubt that usage establishes a settled answer to such a far-fetched question; but never mind. What the thing may be called is entirely beside the point. Likewise when I speak of possible dragons or unicorns, I mean animals that fit the stereotypes we associate with those names. I am not here concerned with Kripke's problem of whether such animals are rightly called by those names.)

I cannot altogether accept the formulation: anything can coexist with anything. For I think the worlds do not overlap, hence each thing is part of only one of them. A dragon from one world and a unicorn from a second world do not themselves coexist either in the dragon's world, or in the unicorn's world, or in a third world. An attached head does not reappear as a separated head in some other world, because it does not reappear at all in any other world.

Ordinarily I would replace trans-world identity by counterpart relations, but not here. I cannot accept the principle: a counterpart of anything can coexist with a counterpart of anything else. Counterparts are united by similarity, but often the relevant similarity is mostly extrinsic. In particular, match of origins often has decisive weight. Had my early years gone differently, I might be different now in ever so many important ways – here I envisage an other-worldly person who is my counterpart mainly by match of origins, and very little by intrinsic similarity in later life. It might happen (at least under some resolutions of the vagueness of counterpart relations) that nothing could be a counterpart of the dragon unless a large part of its surrounding world fairly well matched the dragon's world; and likewise that nothing could be a counterpart of the unicorn unless a large part of its surrounding world fairly well matched the unicorn's world; and that no one world matches both the dragon's world and the unicorn's world well enough; and therefore that there is no world where a counterpart of the dragon coexists with a counterpart of the unicorn. Considered by themselves, the dragon and the unicorn

are compossible. But if we use the method of counterparts, we do not consider them by themselves; to the extent that the counterpart relation heeds extrinsic similarities, we take them together with their surroundings.

It is right to formulate our principle of recombination in terms of similarity. It should say, for instance, that there is a world where something like the dragon coexists with something like the unicorn. But extrinsic similarity is irrelevant here, so I should not speak of coexisting *counterparts*. Instead, I should say that a *duplicate* of the dragon and a *duplicate* of the unicorn coexist at some world, and that the attached talking head has at some world a separated duplicate.

Duplication is a matter of shared properties, but differently situated duplicates do not share all their properties. In section 1.5, I defined duplication in terms of the sharing of perfectly natural properties, then defined intrinsic properties as those that never differ between duplicates. That left it open that duplicates might differ extrinsically in their relation to their surroundings. Duplicate molecules in this world may differ in that one is and another isn't part of a cat. Duplicate dragons in different worlds may differ in that one coexists with a unicorn and the other doesn't. Duplicate heads may differ in that one is attached to the rest of a human body and the other isn't.

Not only two possible individuals, but any number should admit of combination by means of coexisting duplicates. Indeed, the number might be infinite. Further, any possible individual should admit of combination with itself: if there could be a dragon, then equally there could be two duplicate copies of that dragon side by side, or seventeen or infinitely many.

But now there is trouble. Only a limited number of distinct things can coexist in a spacetime continuum. It cannot exceed the infinite cardinal number of the points in a continuum. So if we have more than continuum many possible individuals to be copied, or if we want more than continuum many copies of any single individual, then a continuum will be too small to hold all the coexisting things that our principle seems to require.

Should we keep the principle of recombination simple and unqualified, follow where it leads, and conclude that the possible size of spacetime is greater than we might have expected? That is tempting, I agree. And I see no compelling reason why a possible spacetime can never exceed the size of a continuum. But it seems very fishy if we begin with a principle that is meant to express plenitude about how spacetime might be occupied, and we find our principle transforming itself unexpectedly so as to yield consequences about the possible size of spacetime itself.

Our principle therefore requires a proviso: 'size and shape permitting'. The only limit on the extent to which a world can be filled with duplicates of possible individuals is that the parts of a world must be able to fit

together within some possible size and shape of spacetime. Apart from that, anything can coexist with anything, and anything can fail to coexist with anything.

This leaves a residual problem of plenitude: what are the possible sizes and shapes of spacetime? Spacetimes have mathematical representations, and an appropriate way to state plenitude would be to say that for every representation in some salient class, there is a world whose spacetime is thus represented. It is up to mathematics to offer us candidates for the 'salient class'. (See section 2.2 for further discussion.)

We sometimes persuade ourselves that things are possible by experiments in imagination. We imagine a horse, imagine a horn on it, and thereby we are persuaded that a unicorn is possible. But imaginability is a poor criterion of possibility. We can imagine the impossible, provided we do not imagine it in perfect detail and all at once. We cannot imagine the possible in perfect detail and all at once, not if it is at all complicated. It is impossible to construct a regular polygon of nineteen sides with ruler and compass; it is possible but very complicated to construct one of seventeen sides. In whatever sense I can imagine the possible construction, I can imagine the impossible construction just as well. In both cases, I imagine a texture of arcs and lines with the polygon in the middle. I do not imagine it arc by arc and line by line, just as I don't imagine the speckled hen speckle by speckle – which is how I fail to notice the impossibility.

We get enough of a link between imagination and possibility, but not too much, if we regard imaginative experiments as a way of reasoning informally from the principle of recombination. To imagine a unicorn and infer its possibility is to reason that a unicorn is possible because a horse and a horn, which are possible because actual, might be juxtaposed in the imagined way.

In 'Propositional Objects' Quine suggested that we might take a possible world as a mathematical representation: perhaps a set of quadruples of real numbers, regarded as giving the coordinates of the spacetime points that are occupied by matter. His method could be extended to allow for various sizes and shapes of spacetime, for occupancy by different kinds of matter and by point-sized bits of fields, and perhaps even for occupancy of times by non-spatial things. In section 3.2, I shall argue that we should not *identify* the worlds with any such mathematical representations. However we should accept a *correspondence*: for every Quinean ersatz world, there is a genuine world with the represented pattern of occupancy and vacancy. This is just an appeal to recombination. But we are no longer applying it to smallish numbers of middle-sized things, horses or horns of heads. Instead, we are applying it to point-sized things, spacetime points

themselves or perhaps point-sized bits of matter or of fields. Starting with point-sized things that are uncontroversially possible, perhaps because actual, we patch together duplicates of them in great number (continuum many, or more) to make an entire world. The mathematical representations are a book-keeping device, to make sure that the 'size and shape permitting' proviso is satisfied.

Another use of my principle is to settle – or as opponents might say, to beg – the question whether laws of nature are strictly necessary. They are not; or at least laws that constrain what can coexist in different positions are not. Episodes of bread-eating are possible because actual; as are episodes of starvation. Juxtapose duplicates of the two, on the grounds that anything can follow anything; here is a possible world to violate the law that bread nourishes. So likewise against the necessity of more serious candidates for fundamental laws of nature – perhaps with the exception of laws constraining what can coexist at a single position, for instance the law (if such it be) that nothing is both positive and negative in charge. It is no surprise that my principle prohibits strictly necessary connections between distinct existences. What I have done is to take a Humean view about laws and causation, and use it instead as a thesis about possibility. Same thesis, different emphasis.

Among all the possible individuals there are, some are parts of this world; some are not, but are duplicates of parts of this world; some, taken whole, are not duplicates of any part of this world, but are divisible into parts each of which is a duplicate of some part of this world. Still other possible individuals are not thus divisible: they have parts, no part of which is a duplicate of any part of this world. These I call *alien* individuals. (That is, they are alien *to* this world; similarly, individuals could be alien to another world. For instance, many individuals in this world are alien to more impoverished worlds.) A world that contains alien individuals – equivalently, that is itself an alien individual – I call an alien world.

In 'New Work for a Theory of Universals', I defined an alien natural property as one that is not instantiated by any part of this world, and that is not definable as a conjunctive or structural property build up from constituents that are all instantiated by parts of this world.[58] Anything that instantiates an alien property is an alien individual; any world within which an alien property is instantiated is an alien world.

But not conversely: we could have an alien individual that did not instantiate any alien properties, but instead combined non-alien properties

[58]Perhaps, as Armstrong has suggested in discussion, I should have added a third clause: '. . . and that is not obtainable by interpolation or extrapolation from a spectrum of properties that are instantiated by parts of this world'.

in an alien way. Suppose that positive and negative charge are not, strictly speaking, incompatible; but suppose it happens by accident or by contingent law that no this-worldly particle has both these properties. Then an other-worldly particle that does have both is an alien individual but needn't have any alien properties.

A world to which no individuals, worlds, or properties are alien would be an especially rich world. There is no reason to think we are privileged to inhabit such a world. Therefore any acceptable account of possibility must make provision for alien possibilities.

So it won't do to say that all worlds are generated by recombination from parts of this world, individuals which are possible because they are actual. We can't get the alien possibilities just by rearranging non-alien ones. Thus our principle of recombination falls short of capturing all the plenitude of possibilities.

A principle which allowed not only recombination of spatiotemporal parts of the world but also recombination of non-spatiotemporal parts – universals or tropes – would do a bit more. It would generate those alien individuals that do not instantiate alien properties. But I say (1) that such a principle, unlike mine, would sacrifice neutrality about whether there exist universals or tropes, and (2) that it still wouldn't go far enough, since we also need the possibility of alien properties.

Although recombination will not generate alien worlds out of the parts of this world, it nevertheless applies to alien worlds. It rules out that there should be only a few alien worlds. If there are some, there are many more. Anything alien can coexist, or fail to coexist, with anything else alien, or with anything else not alien, in any arrangement permitted by shape and size.

1.9 Actuality

I say that ours is one of many worlds. Ours is the actual world; the rest are not actual. Why so? – I take it to be a trivial matter of meaning. I use the word 'actual' to mean the same as 'this-worldly'. When I use it, it applies to my world and my worldmates; to this world we are part of, and to all parts of this world. And if someone else uses it, whether he be a worldmate of ours or whether he be unactualised, then (provided he means by it what we do) it applies likewise to his world and his worldmates. Elsewhere I have called this the 'indexical analysis' of actuality and stated it as follows.

> I suggest that 'actual' and its cognates should be analyzed as *indexical* terms: terms whose reference varies, depending on relevant features of the context of utterance. The relevant feature of context, for the term 'actual', is the

> world at which a given utterance occurs. According to the indexical analysis I propose, 'actual' (in its primary sense) refers at any world w to the world w. 'Actual' is analogous to 'present', an indexical term whose reference varies depending on a different feature of context: 'present' refers at any time t to the time t. 'Actual' is analogous also to 'here', 'I', 'you', and 'aforementioned' – indexical terms depending for their reference respectively on the place, the speaker, the intended audience, the speaker's acts of pointing, and the foregoing discourse. ('Anselm and Actuality', pages 184–5.)

This makes actuality a relative matter: every world is *actual at* itself, and thereby all worlds are on a par. This is *not* to say that all worlds are actual – there's no world at which that is true, any more than there's ever a time when all times are present. The 'actual at' relation between worlds is simply identity.

Given my acceptance of the plurality of worlds, the relativity is unavoidable. I have no tenable alternative. For suppose instead that one world alone is *absolutely* actual. There is some special distinction which that one world alone posseses, not relative to its inhabitants or to anything else but *simpliciter*. I have no idea how this supposed absolute distinction might be understood, but let us go on as if we did understand it. I raise two objections.

The first objection concerns our knowledge that we are actual. Note that the supposed absolute distinction, even if it exists, doesn't make the relative distinction go away. It is still true that one world alone is ours, is this one, is the one we are part of. What a remarkable bit of luck for us if the very world we are part of is the one that is absolutely actual! Out of all the people there are in all the worlds, the great majority are doomed to live in worlds that lack absolute actuality, but we are the select few. What reason could we ever have to think it was so? How could we ever know? Unactualised dollars buy no less unactualised bread, and so forth. And yet we *do* know for certain that the world we are part of is the actual world – just as certainly as we know that the world we are part of is the very world we are part of. How could this be knowledge that we are the select few?

D. C. Williams asks the same question. Not about 'actuality' but about 'existence'; but it comes to the same thing, since he is discussing various doctrines on which so-called 'existence' turns out to be a special status that distinguishes some of the things there are from others. He complains that Leibniz 'never intimates, for example, how he can tell that *he* is a member of the existent world and not a mere possible monad on the shelf of essence' ('Dispensing with Existence', page 752).

Robert M. Adams, in 'Theories of Actuality', dismisses this objection. He says that a simple-property theory of absolute actuality can account for the certainty of our knowledge of our own actuality by maintaining

that we are as immediately acquainted with our own absolute actuality as we are with our thoughts, feelings, and sensations. But I reply that if Adams and I and all the other actual people really have this immediate acquaintance with absolute actuality, wouldn't my elder sister have had it too, if only I'd had an elder sister? So there she is, unactualised, off in some other world getting fooled by the very same evidence that is supposed to be giving me my knowledge.

This second objection concerns contingency. (It is due to Adams, and this time he and I agree.) Surely it is a contingent matter which world is actual. A contingent matter is one that varies from world to world. At one world, the contingent matter goes one way; at another, another. So at one world, one world is actual; and at another, another. How can this be *absolute* actuality? – The relativity is manifest!

The indexical analysis raises a question. If 'actual' is an indexical, is it or is it not a rigidified indexical? In a context where other worlds are under consideration, does it still refer to the world of utterance, or does it shift its reference? Compare 'now', which is normally rigidified, with 'present', which may or may not be. So you say 'Yesterday it was colder than it is now', and even in the scope of the time-shifting adverb, 'now' still refers to the time of utterance. Likewise you say 'Yesterday it was colder than it is at present', and the reference of 'present' is unshifted. But if you say 'Every past event was once present', then the time-shifting tensed verb shifts the reference of 'present'. I suggest that 'actual' and its cognates are like 'present': sometimes rigidified, sometimes not. What if I'd had an elder sister? Then there would have been someone who doesn't actually exist. (Rigidified.) Then she would have been actual, though in fact she is not. (Unrigidified.) Then someone would have been actual who actually isn't actual. (Both together.) In the passage just quoted I called the unrigidified sense 'primary'; but not for any good reason.[59]

I said that when I use it, 'actual' applies to my world and my worldmates; that is, to the world I am part of and to other parts of that world. Likewise, *mutatis mutandis*, when some other-worldly being uses the word with the same meaning. But that left out the sets. I would not wish to say that any sets are *parts of* this or other worlds,[60] but nevertheless I would like to say that sets of actual things are actual. Sometimes we hear it said that sets are one and all unlocated; but I don't know any reason to believe this, and a more plausible view is that a set

[59]For various examples that require or forbid rigidification if they are to make sense, see my *Philosophical Papers*, volume I, page 22; for further discussion, see Hazen, 'One of the Truths about Actuality' and van Inwagen, 'Indexicality and Actuality'.

[60]But not because I take it that the part–whole relation applies only to individuals and not sets, as I said in *Philosophical Papers*, volume I, page 40; rather, because I now take it that a set is never part of an individual.

is where its members are. It is scattered to the extent that its members are scattered; it is unlocated if, but only if, its members are unlocated. That applies as much to location among the worlds as it does to location within a single world. Just as a set of stay-at-home Australians is in Australia, so likewise a set of this-wordly things is this-worldly, in other words actual. In the same way, a set of sets that are all in Australia is itself in Australia, and likewise a set of actual sets is itself actual; and so on up the iterative hierarchy.

I might sometimes prefer to use the word 'actual' a bit more broadly still. There is no need to decide, once and for all and inflexibly, what is to be called actual. After all, that is not the grand question: what is there? It is only the question which of all the things there are stand in some special relation to us, but there are special relations and there are special relations. Suppose there are things that are not our world, and not parts of our world, and not sets built up entirely from things that are parts of our world – but that I might nevertheless wish to quantify over even when my quantification is otherwise restricted to this-worldly things. If so, no harm done if I sometimes call them 'actual' by courtesy. No harm done, in fact, if I decline to adopt any official position on the question whether they are actual or whether they are not! It is no genuine issue.

The numbers, for instance, might well be candidates to be called 'actual' by courtesy. But it depends on what the numbers are. If they are universals, and some or all of them are non-spatiotemporal parts of their this-worldly instances which in turn are parts of this world, then those numbers, at least, are actual not by courtesy but because they are parts of this world. Likewise for other mathematical entities.

Properties, taken as sets of all their this- and other-worldly instances, are another candidate. By what I said above about actuality of sets, only those properties are actual whose instances are confined to the actual world. But most of the properties we take an interest in have instances both in and out of this world. Those ones might be called 'partly actual'; or they might as well just be called 'actual', since very often we will want to include them in our otherwise this-worldly quantifications.

Events fall in with the properties; for I see no reason to distinguish between an event and the property of being a spatiotemporal region, of this or another world, wherein that event occurs. (See my 'Events'.) An event that actually occurs, then, is a set that includes exactly one this-worldly region. That makes it partly actual, and we may as well just call it 'actual'.

Propositions, being sets of worlds, also fall in with the properties taken as sets. A proposition is partly actual at just those worlds where it is true, for it has just those worlds as its members. So we might call at least the true propositions 'actual'; or we might just call all propositions 'actual', distinguishing however between those that are and are not actually true.

Not only sets but individuals may be partly actual – big individuals, composed of parts from more worlds than one, and so partly in each of several worlds. If there are any such trans-world individuals that are partly in this world, hence partly actual, should we call them 'actual' *simpliciter*? – That depends. We needn't, if we think of them just as oddities that we can mostly ignore. I think they are exactly that. (See section 4.3.) But if we were reluctant to ignore them in our quantifying, perhaps because we thought that we ourselves were among them, then we might appropriately call them 'actual'.[61]

[6]In *Philosophical Papers*, volume I, pages 39–40, I distinguished three ways of 'being in a world': (1) being *wholly* in it, that is, being part of it; (2) being *partly* in it, that is, having a part that is wholly in it; and (3) existing *from the standpoint of* it, that is, 'belonging to the least restricted domain that is normally – modal metaphysics being deemed abnormal – appropriate in evaluating the truth at that world of quantifications'. If the world in question is actual, that is almost my present distinction between being actual, being partly actual, and being actual by courtesy; the only difference in the terminologies being that I would not now throw all sets into the lower grade. I distinguish all of the above from (4) existing *according to* a world: I claim that something exists according to a world – for instance, Humphrey both exists and wins the presidency according to certain worlds other than ours – by having a counterpart that is part of that world. On being part of versus existing according to, see section 4.1.