The Simplicity of Everything

Cian Seán Dorr

A dissertation
presented to the faculty of
Princeton University
in candidacy for the degree of
Doctor of Philosophy

Recommended for acceptance
by the Department of Philosophy

January 2002

Abstract

Part One of my dissertation is about composite objects: things with proper parts, like plates, planets, plants and people. I begin chapter 1 by pointing out that if one were to judge by the way we normally speak about composite objects, one would suppose that we were all completely certain of a theory I call *folk mereology*. For instance, we seem to be completely convinced that whenever some things are piled up, there is an object—a pile—which they compose. I point out that folk mereology is neither an analytic truth nor a theory for which we have conclusive empirical evidence. So what are we to make of the feeling that it *makes no sense* to deny folk mereology? What this shows, I claim, is that the standard which an assertion about composite objects has to meet in order to be *correct* is not strict and literal truth, but something less demanding.

In the last part of chapter 1 and in chapter 2, I argue that we have no evidence for the existence of composite objects, and that without such evidence we ought to believe that there aren't any composite objects. Chapter 3 is devoted to working out some details of a broadly "fictionalist" explanation of the fact that claims about composite objects can be correct without being strictly and literally true.

In Part Two I defend a parallel view about complex attributes—properties, relations and propositions. In our ordinary talk about attributes, we seem to presuppose

a theory according to which there are very many of these entities; but, as I argue in chapter 4, we cannot justifiably be certain that this theory is true. I conclude that in this case too, correct assertion diverges from strict and literal truth. Chapter 5 is devoted to an argument that there are in fact no complex attributes. Finally, in chapter 6, I develop a fictionalist account of the conditions under which ordinary assertions about attributes are correct.

The whole globe of the earth, nay the whole universe *may be consider'd as an unite*. That term of unity is merely a fictitious denomination, which the mind may apply to any quantity of objects it collects together; nor can such an unity any more exist alone than number can, as being in reality a true number. But the unity, which can exist alone, and whose existence is necessary to that of all number, is of another kind, and must be perfectly indivisible, and incapable of being resolved into any lesser unity. (Hume 1739, I.II.ii)

One very important heuristic maxim which Dr. Whitehead and I found, by experience, to be applicable in mathematical logic, and have since applied to various other fields, is a form of Occam's Razor. When some set of supposed entities has neat logical properties, it turns out, in a great many instances, that the supposed entities can be replaced by purely logical structures composed of entities which have not such neat properties. In that case, in interpreting a body of propositions hitherto believed to be about the supposed entities, we can substitute the logical structures without altering any of the detail of the body of propositions in question. This is an economy, because entities with neat logical properties are always inferred, and if the propositions in which they occur can be interpreted without making this inference, the ground for the inference fails, and our body of propositions is secured against the need of a doubtful step. The principle may be stated in the form: 'Whenever possible, substitute constructions out of known entities for inferences to unknown entities.' (Russell 1924, pp. 160–161)

Contents

Acknowledgements

I thank all those among the faculty and students of the philosophy departments at Princeton and New York Universities with whom I have had the opportunity to discuss these ideas. Particular thanks are due to Hartry Field, Benj Hellie, Mark Johnston, Gideon Rosen and Kieran Setiya; and especially to Jessica Moss, whom I have far more to thank for than I could possibly hope to express here.

My most important debt of gratitude is due to my adviser, David Lewis: for his invaluable advice, without which this dissertation would have been so much worse than it is, and taken so much longer than it did; for many illuminating conversations down the years; and for his writings, from which I have learnt so much of what I think I understand about philosophy. Sadly, I will never be able to thank David in person. This work is dedicated to his memory.

Preface

This dissertation is about two sorts of complex entities: composite particular things, in Part One; and complex attributes, in Part Two. My view about these entities is easy to state: there aren't any. They exist only according to certain false theories.

I don't hold this view because of any complicated metaphysical argument. Rather, I hold it because I don't see that we have any *evidence* for the existence of these sorts of entities, just as we have no evidence for the existence of angels or elves or aliens in our midst. I do give arguments for the claim that we don't have any evidence for the existence of complex entities, and for the claim that in the absence of such evidence, the reasonable thing to do is to believe that there aren't any of them; some of these arguments are rather complicated. But the basic thought is as simple as can be. And for me at least, no arguments were required to shake my faith in the existence of complex things: as soon as it occurred to me to ask myself what evidence I had for believing in them, it just seemed obvious that the answer was 'little or none'. What could I point to? The evidence of the senses? This can be explained equally well on the assumption that there aren't any complex things. The testimony of those who taught me to believe in complex things? They were no better-equipped than I to resolve the question. The belief in complex entities soon came to seem a mere superstition, surviving only because it generally doesn't occur to us to question it.

If complex entities don't really exist, how are we to make sense of the things we say that appear to presuppose that there *are* complex entities? According to one central tradition in analytic metaphysics, this sort of question is to be answered by showing that when we say things that, taken at face value, entail the existence of entities of the disputed sort, the propositions which we are really judging to be true, and expressing, don't entail the existence of those entities. The philosopher's task is to explain the rules whereby these more complicated propositions come to be associated with sentences which seem superficially like simple existential claims. Russell seems to be endorsing something like this approach when he recommends substituting 'logical structures' for certain supposed entities with 'neat logical properties' when we are interpreting propositions 'hitherto believed to be about the supposed entities' (see the quotation on p. v).

The most important objection to this programme, I think, is that it carries the commendably charitable policy of refusing to attribute false metaphysical beliefs to ordinary people to an excess. We must admit that there are some people, namely our metaphysician colleagues, who do hold false metaphysical beliefs. Surely they sometimes have these beliefs in mind when they use ordinary sentences. If they also believe (rightly or wrongly) that others can be relied on to share these beliefs, we must moreover admit that they sometimes use the ordinary sentences to *express* the false metaphysical beliefs. But given that this sometimes happens, it seems unduly optimistic to assume that philosophically uncontaminated people never hold or express the same beliefs. Perhaps we should not go so far as Hume went in holding the beliefs of the vulgar to be a tissue of philosophical absurdities; but nor do we want to fall into Berkeley's error of declaring that our metaphysical system is just an articulation of what everyone who hasn't been confused by bad philosophy has

always believed.

Is it really necessary for us to embark on this patronising business of telling other people what they think in the first place? I hope not: I think that the demand to "make sense" of ordinary talk doesn't have to be understood as an invitation to engage in psychological speculation. Instead, we should think of the datum that needs to be explained as the fact that there is clearly *some* sense in which it's *correct* to say that there are tables, in which it isn't correct to say that there are golden mountains or flying pigs. There's a salient line here, and we need to explain what makes some sentences fall on one side of it while other sentences fall on the other side. The distinction is there even if both sentences are used by ordinary people to convey false beliefs: we might put it by saying that whereas 'there are tables' merely conveys a false *metaphysical* belief, 'there are flying pigs' is more thoroughly infected with falsity.

Russell thought that the task of making sense of ordinary talk about entities that don't really exist would be a difficult and technical one. To me, by contrast, the task seems quite easy: we can just say that 'there are tables', unlike 'there are flying pigs', *would* be true if there were composite objects. There are some interesting questions about how exactly this sort of thing is to be understood—I answer some of them in chapters 3 and 6—but they are questions of detail. This proposal enjoys all 'the advantages of theft over honest toil' (Russell 1919, p. 71), but does not labour under the one big disadvantage of the 'method of postulation': it does not require us to believe in entities for whose existence we have no good evidence. What more could we want?

Part I

Composition

Chapter 1

The existence of composite objects

1.1 Folk mereology and its justification

Let us say that two things are *disjoint* iff they have no part in common, and that some things *compose* something iff each of them is part of it, and no part of it is disjoint from all of them. What do some things have to be like for there to be something that they compose? This is the central question of mereology, the theory of parts and wholes. Let us follow van Inwagen (1990, section 2) in calling it the *special composition question*.[1]

Here's a little map of the enormous logical space of answers to the special composition question. At one extreme, there is *Universalism* about composition, the view that whenever there are some things—no matter how disparate and spread out they might be—there is something that they compose. At the other extreme, there is *Nihilism* about composition, the view that it never happens that several things

[1]I should note that I am using 'the things compose something' to express what van Inwagen would express by 'the things have a sum'; for van Inwagen, some things compose something just in case it is a sum of them, and any two of them are disjoint. So van Inwagen's "special composition question" is slightly different from mine.

compose something, no matter how tightly bound together they might be.[2] These labels, again, are due to van Inwagen (1990, section 8).[3] All the other possible answers lie between Universalism and Nihilism: they are more demanding than the former, but less demanding than the latter. Of these views, the ones that have come in for most notice are those according to which some things compose something just in case they are *sufficiently closely associated*. Views of this sort can be roughly ordered according to how demandingly they interpret 'closely associated': from those which merely require spatial proximity, through views which require some sort of "bonding" or functional interrelations, all the way to van Inwagen's notorious theory that several things compose something just in case 'their activity constitutes a life' (1990, section 9). But we should not forget that most of the possible answers don't belong anywhere on this axis: for instance, the view that some things compose something just in case they are sufficiently disconnected and far apart.

Judging by the way we talk, you'd swear we knew a lot about mereology, and in particular about the answer to the special composition question. For example, we

[2]However, it is normally allowed that when "some things" are one in number, "they" compose the one and only thing that is one of them. As the scare quotes indicate, there is something odd about using plural expressions when you know you are talking about just one thing. I take it that this oddness is just a pragmatic matter: if there is only one bottle of beer on the wall, 'there are some bottles of beer on the wall' is literally true. At any rate, that's how I'll always understand plural expressions.

Given our definitions, the claim that the things that are identical to x compose x is equivalent to the claim that x is part of itself. It is customary among philosophers to say that everything is part of itself, although not a *proper* part of itself. This is generally thought of as a stipulation rather than a claim that needs to be argued for—perhaps because 'is part of' is just *defined* as 'is identical to, or is a proper part of'.

[3]Except that van Inwagen uses 'Universalism' for the view that it is *necessary* that whenever there are some things, there is something they compose, and 'Nihilism' for the view that it is *necessary* that several things never compose one thing.

seem to be incredibly confident that whenever some plates are piled up, each on top of the next, there is something which they compose. If I were to tell you that there was a plate in the sink with another plate on top of it, and another plate on top of that, and another plate on top of that, and no other plates nearby, and you believed me, you would immediately start to give every sign of believing that there was a *pile of plates* in the sink. You would say 'Yes' if someone asked you whether there was a pile of plates in the sink. Indeed, you might even say that I had *told* you that there was a pile of plates in the sink. Saying that there is a pile of four plates in the sink seems to be a merely a less awkward and plodding way to say the very same thing that I said. But a pile of plates is something composed of plates: if the sink doesn't contain any object composed of plates, it certainly doesn't contain a pile of plates.

This isn't an isolated example. Many descriptions are like the description 'are four plates, each on top of the next' in this respect: we seem to be utterly confident that whenever some things satisfy the description, there is something that they compose. Rows of books, heaps of sand, rocks, solar systems: for each of these kinds of thing, there is some description of the properties and relations of some things which we take to license the immediate inference that they compose something which is a row of books, heap of sand, rock, or solar system.[4]

Let *folk mereology* be the theory which comprises all the general claims about

[4]Here's an even clearer example: we seem to take it for granted if some things are exactly similar to some other things as regards their properties and relations to one another and to their surroundings, and the latter things compose something, then the former things compose something as well (see Markosian 1998). In other words, it couldn't happen that some things differed from some other things *only* as regards whether they composed something—for instance, it couldn't happen that on the other side of the sun there are some particles exactly like the particles which make up the Earth and everything on it, without there also being some composite things there.

composition and parthood which we ordinarily treat as if we were certain that they were true. Unless the people whose ordinary talk we are considering are all philosophers with a certain background, folk mereology so defined will be silent on some questions. It won't say whether there is anything composed of exactly one hawk and exactly one handsaw, since in general people don't talk as if they were certain one way or the other.[5] So folk mereology does not entail Universalism. Let us agree to count as honorary members of the folk certain philosophers who believe Mereology—the conjunction of Universalism with the principle that it never happens that some things compose more than one thing, and the principle that parthood is transitive—and in fact treat it it as if they were certain it was true. That will ensure that folk mereology is weak enough to be entailed by Mereology.

In ordinary life, we speak as if we were utterly certain that folk mereology is true. In some cases, this appearance is misleading: certain philosophers have informed us that they don't really believe parts of the theory, even though they speak as if they did.[6] But I suppose many people really are certain.[7] Is this certainty justified? Of course some people have an *excuse* for their certainty: it has never occurred to them that the theory might be false. But such innocence once lost is never regained. Now

[5]Indeed, we often talk as if there were no such things, for instance when we are asked to count the number of physical objects in a room. But that isn't decisive, since such counts often also exclude a great many things which definitely do exist according to folk mereology—if you were asked to count the number of things in the sink, you wouldn't count the pile of plates along with the plates themselves.

[6]Foremost among them is van Inwagen (1990). The quotation from Hume on page (v) suggests that Hume may be another such philosopher.

[7]Well, actually I'm not entirely confident about this: the attribution of beliefs about metaphysics to people who have never thought about metaphysics is a tricky affair, and prone to indeterminacy. I imagine that if you asked the right questions, you could get a typical physicist—even a typical reader of *Scientific American*—to say all sorts of things which contradict folk mereology.

that it has occurred to *us* that the theory might be false, can *we* reasonably continue to be so certain?

Worries about the epistemological status of our belief in folk mereology don't come naturally. Whereas everyone who hears about Cartesian scepticism immediately sees that there is a serious question to be answered about the justification of our belief in an external world, one is initially *baffled* by the challenge to justify one's beliefs about folk mereology. One wants to respond that it *makes no sense* to suppose that there are plates piled up in the sink without there being a pile of plates in the sink, just as it makes no sense to suppose that there are plates piled up in the sink without there being any plates in the sink. What we feel, in other words, is that the denial of the theory is *inconsistent*, and the theory itself *analytic*. We are justified in being certain of it in the same distinctive and absolute way that we are justified in being certain that all plates are plates, or that no plates are spherical. Can this be right?

Certainly there are *some* principles about the circumstances under which some things compose something which really are analytic. An example is the principle 'Whenever some things compose something, they compose something'. We are completely within our epistemic rights to take this as certain. Not all analytic truths are as easily recognised as this one: sometimes one needs to know facts about the meanings of non-logical words to recognise a sentence as analytic. So it is possible that, upon proper analysis, various other more interesting-looking principles about composition will turn out to be analytic. If 'are piled up' just meant 'are such as to compose a pile', for example, then 'Whenever some plates are piled up, they compose something' would be analytic after all.

But most of the principles of folk mereology are not analytic. 'Whenever there are four plates such that the first is on top of the second, which is on top of the third,

which is on top of the fourth, they compose something' is an example. Try hard as I might to find a hidden contradiction in the following story, I can find none:

C There are four plates such that the first is on top of the second, which is on top of the third, which is on top of the fourth, and there is nothing which has any of these plates as a proper part.

This isn't inconsistent in the narrowly logical sense. Nor is there any plausible view about the meanings of any of its constituent expressions which would make it turn out to be analytically false.

It is sometimes suggested that it is true in virtue of the meaning of the phrase 'pile of plates' that whenever some plates are spatially arranged in that way, they compose a pile of plates. But the words 'pile of plates' don't even occur in C, so how could any special facts about their meaning make a difference to whether C is inconsistent?

I said that there was no *plausible* analysis of the words contained in C which would make it turn out to be inconsistent; on the other hand, there are plenty of completely implausible analyses which would have this effect. If 'x is a plate' meant 'x is a vessel and x is shallow and ... and whenever there are some things that are piled up, they compose something...', C would indeed be inconsistent. If such an analysis were correct, we would have to know that folk mereology or some fragment of it was true before we could decide whether any given thing was a plate or not. So the very same questions that I am asking about our epistemic right to the principles of folk mereology would then have to be asked about our assumption that "plate-candidates" really are plates. Wherever the bump in the carpet is moved, our ordinary practices are still going to end up committed to something epistemologically problematic.

I've been assuming that the claim that C is inconsistent is the sort of claim that needs to be defended by giving an analysis of some of the words it contains. But of course you don't always have to work like this to uncover the inconsistency in a sentence: some sentences, like 'Some bachelors are not bachelors' and 'There is something that isn't identical to anything', wear their inconsistency right on the surface. We call these sentences 'logical falsehoods'. Could C belong in this category? Could the principles of folk mereology be logical truths?

It seems manifest to me that they are not. There is a great chasm between the principles of folk mereology and genuine logical truths: unlike genuine logical truths, folk mereology makes existence claims. It doesn't make *absolute* existence claims, since it is consistent with the claim that there is nothing at all. But it makes *conditional* existence claims, of the form 'if these things do such-and-such, then there is this *other* thing, not identical to any of them, which they compose'; and that is as much a disqualification from being a truth of logic as any absolute existence claim would be. Thus, in the relevant way, folk mereology is like the theory that there is a God, and the theory that if there is anything at all there is a God, and the theory that there are black holes, and the theory that every star is orbited by at least one planet. All these theories concern what Hume called 'connections between distinct existences'; and hence they cannot be analytic truths.[8]

[8]By 'distinct' I mean 'not identical'. Some philosophers (Armstrong 1978b; Lewis 1991, 1993) use 'distinct' to mean 'disjoint': in *this* sense, folk mereology does not concern connections between distinct existences. Likewise, if you used 'x is distinct from y' to mean 'either x is not identical to y, or one of x and y is a God', you would be able to say that the theory that if there is anything at all there is a God did not concern connections between distinct existences.

The use of 'distinct' for 'disjoint' is associated with the mystifying thesis that *composition is identity*, originally defended by Don Baxter (1988). As explicated by David Lewis (1991, section 3.6), this amounts to the claim that composition is

I admit that I don't know how to *argue* for the relevance of this distinction. To do that, I would need to have a theory which could explain why we can be justified in being certain of the things *I* count as logical truths, like the claim that everything is identical to something. But I have no theory about that to offer; and even if I had one, I doubt that it would be possible to argue for it without begging the question. All I can really do is encourage you to imagine, as vividly as you can, what it is like to doubt or disbelieve folk mereology, and hope that, as you get accustomed to doing this, your initial feeling that this view makes no sense will evaporate. You will just find yourself doing what you supposed couldn't be done, namely *imagining* or *supposing* folk mereology to be false, in that more-than-nominal sense in which you simply can't imagine or suppose, for example, that there is something that isn't identical to anything.[9]

---

analogous to identity in various ways. Could such analogies explain our justification for being certain a priori of the truth of folk mereology? One might try to argue for folk mereology by means of an argument for Mereology (which all the aforementioned authors endorse) along the following lines: 'Composition and identity are similar in this, that and the other respect. Therefore composition and identity are also similar in respect of being governed by extremely simple and general principles: Mereology in the case of composition, the axioms of the logic of identity in the case of identity.' But this argument obviously wouldn't justify one in being *certain* of Mereology: analogies very often go only so far. The idea that composition might be in any important way analogous to identity is further criticised in van Inwagen 1994.

[9]I am generally loath to hold that philosophers who seem to be disagreeing with one another really mean different things by their words, so that the disagreement is illusory. However, I find the claim that C isn't inconsistent so *very* obvious that when I meet philosophers who flatly insist that it *is* inconsistent, my general faith in face-value interpretations begins to waver. I wonder whether these philosophers might not mean something different from me by the quantifier 'there are'.

There is an attractive hypothesis that can make sense of everything they say: what they mean by 'there are' is what I mean by 'doubts about the truth of folk mereology aside, there are' or 'according to the mereological fiction, there are' (see section 1.2 below). I don't think it's *easy* for the differences in two peoples' speech habits to lead to them assigning different meanings to a logical expression as basic as 'there

So folk mereology is not analytic: everything I will say in the rest of this chapter, and in chapters 2 and 3, will be founded on this assumption. Could it be that we are nevertheless justified in being certain that folk mereology is true? I can't see how. Surely if empirical evidence ever justifies us in being certain of something, it justifies us in being certain about *local* matters of fact. But folk mereology makes claims of sweeping generality, about how things are everywhere and at all times, unobserved as well as observed—plates that are piled up in distant galaxies compose piles just as nearby ones do. It is absurd to think that our empirical evidence could justify us in being certain about this.

---

are': the thing *I* mean by 'there are'—which I will tendentiously label 'standard unrestricted quantification'—is, I should think, a "reference magnet": that is, the fact that one means *that* by 'there are' is robust under a wide range of variations in one's use of 'there are'. (See Lewis 1984; Horwich 1998; and the introduction to Sider 2002.) Nevertheless, given *enough* of a difference of use, 'there are' can come to mean something else: anything can mean anything, after all. And differences in which claims one *treats as analytic* are very significant, as differences of use go: we can happily disagree about how many bachelors there are and what they are like, and maybe we can even disagree about whether the Pope is a bachelor (see Harman 1996); but if you start treating the inference 'x is an unhappy adult male, therefore x is a bachelor' as analytic, I will start to wonder if we are talking past each other.

If the hypothesis is correct, then I face the task of explaining what *I* mean by 'there are' to these philosophers who apparently disagree with me about C. This might be easy: they might have some word or expression in their language which seems to me to be properly translated into my language as 'there are'. This has generally been the case in real life when I have come across philosophers with whom I am inclined to take this line. And I don't think this is just good luck: I can hardly imagine how someone could do without any way of expressing *standard unrestricted quantification*. Still, it would be nice to be able to say something which could introduce standard unrestricted quantification to someone who was so benighted as to lack any word for this wonderful concept, or at least enable someone to figure out which of the various concepts they already have is the one I have in mind when I speak of standard unrestricted quantification. Sometimes I imagine that this could be done by pointing out that unrestricted quantification is involved in *especially few* analytic truths, compared to other notions in the same grammatical category. (See footnote 17 on p. 17 below.)

Some say that empirical evidence isn't the only way to get to be justified in being certain of a synthetic truth. We also have a faculty of *a priori insight*, whereby certain things are revealed to us. Maybe you have had, or think you have had, an a priori insight that justifies you in being certain of folk mereology. But are you really *certain* that that experience really was an a priori insight? If not, I can't see how it could justify you in being certain of anything: if you grant the *possibility* that the experience was not an a priori insight, you should grant the possibility that the experience was not veridical, i.e. that folk mereology is in fact false. And when you consider the analogy between your situation and the situation of others who have supposed themselves to have a priori insight into synthetic matters of fact—Euclidean physical geometry, say, or Descartes' laws of motion—surely you have no choice but to concede that your experience *might* be merely the illusion of an a priori insight.[10]

1.2 Fictionalism about folk mereology

And yet, there is this lingering feeling that it would make no sense to say that some plates were piled up but composed no pile. When I try to imagine a course of experience that would convince me that I was right when I said that there were four plates in the sink, each on top of the next, but wrong when I said that there was a pile of plates in the sink, I get stuck. Even if a god came and told me that this was so, I would want to say to him: 'You must have misunderstood. You are treating the claim that there is a pile of plates in the sink as a controversial metaphysical claim, and I am prepared to believe whatever you tell me about controversial metaphysical

[10]I address the question whether "a priori insight" or "intuition" might give us justification for believing folk mereology with something less than certainty in section 1.4.4.

claims. But when I said that there was a pile of plates in the sink, I didn't *mean* to be making a controversial metaphysical claim—I meant to be saying something *uncontroversially* entailed by the claim that there were some plates in the sink, each on top of the next. If you think you can show me that I was wrong by telling me some controversial metaphysics, you must have misunderstood what I meant to say.'[11]

To see how this could make sense, we need to draw a distinction between the question whether a sentence is *strictly and literally true*—true "taken at face value"—and the question whether it is *correct* to assert the sentence in an ordinary, non-philosophical context. In effect, the response I imagined making to the god is that the sentence 'there is a pile of plates in the sink' can correctly be asserted even though it is not strictly and literally true. Given the standards that govern our ordinary talk about material objects, it is correct to assert 'there is a pile of plates in the sink' just in case it is correct to assert 'there are some piled-up plates in the sink'. And this is a general phenomenon: if it is correct to assert a sentence S, and S' follows from S together with the principles of folk mereology, then it is ordinarily correct to assert S'.

It would be good to have a theory of correctness, which could explain why inferences made in accordance with the principles of folk mereology preserve it. One very general form such a theory could take is this: we find an operator, $\bigcirc$, of which we make the following claim: for any sentence φ, φ is correctly asserted iff $\ulcorner\bigcirc\varphi\urcorner$ is strictly and literally true. There is no special reason to expect this operator should already exist in the language: we might have to introduce it, by explaining how it features in valid inferences. But there are various operators that we already understand tolerably

[11]This thought experiment is borrowed from Burgess and Rosen (1997, p. 3) and Yablo (2000).

well which seem like they might do the job. For example, one could analyse $\ulcorner\bigcirc\varphi\urcorner$ as $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$, for some specific ψ. If we choose a sentence ψ which logically entails folk mereology, we will be able to explain why it is correct to assert, and to reason in accordance with, the principles of folk mereology: for in general, when φ is a logical consequence of ψ, $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$ is strictly and literally true.[12]

Of course, it is an oversimplification to pretend that there is just one standard of correctness that governs our practice always and everywhere. There may be many relevant operators $\bigcirc_1 \ldots \bigcirc_n$, and it may be a vague and context-dependent matter which of them (if any) is in play on any given occasion. So even if every word in a sentence is precise and univocal, it can be a vague or context-dependent matter whether the assertion of the sentence is correct. This gives us the option of adopting a deflationary approach to many philosophical disputes about material objects. For example, some say that every statue is co-located with, but not identical to, a lump of matter; others deny this. *Maybe* the right thing to say about this dispute is that the parties are talking past one another: in some contexts it is correct to say that statues are lumps of matter, in other contexts it is correct to deny this. Of course, the disputants could insist that their utterances should be interpreted strictly and literally; but on this interpretation, it is not clear that the appeals to intuition which feature in the argument have any epistemological authority.[13]

A suggestive way to pronounce the operator $\bigcirc$ is as 'According to the mereological

[12]Maybe we don't have to work so hard to find an operator fit to play the role of $\bigcirc$: following a suggestion of Yablo's (2000), we might try something like 'Doubts about the existence of composite objects aside...'. But the need for an informative explanation of this operator is hardly less urgent than the need for an informative theory of correctness which drove us to look for an operator in the first place.

[13]I'll have more to say about statues and lumps of matter in section 3.3 below.

fiction...'. One can then think of the task of filling in the placeholder 'ψ' in the analysis of $\bigcirc$ as 'If it had been the case that ψ...' as the task of spelling out the "explicit content" or "principles of generation" of this fiction. The analogy between ordinary talk of composite objects and talk about the content of works of fiction is in some ways a compelling one. But this way of putting things is also liable to mislead, in several ways. First, to call something a "fiction" is to imply that it is not strictly and literally true; but I do not (yet) mean to take any stand on the question whether folk mereology is strictly and literally as well as fictionally true. (Perhaps 'story' or 'legend' would be a better word.) Second, it suggests a commitment to the existence of an entity, "the mereological fiction", which is the subject matter of our ordinary talk about composite objects. But if it could be correct to say 'there is a pile of plates in the sink' even if there are really no such things as piles, it is even more obvious that it could be correct to say 'there is a pile of plates in the sink' even if there are really no such things as *fictions*! Third, it suggests that the explicit content of the fiction might be a contingent matter, dependent on people's beliefs, intentions and actions in the same way that what is the case according to a novel depends on its author's beliefs, intentions and actions. Fourth, it suggests that the role of the fiction in constituting the correctness-conditions of our utterances depends on our *intending* that our utterances should be judged according to that standard. The view of ordinary talk about composite objects that I have been advocating counts as a form of "fictionalism" only if one can be a fictionalist without accepting any of these claims. However, now that I have disavowed them, I will make free use of the metaphor of fiction.

1.3 Pragmatic and semantic versions of fictionalism

The divergence between correctness and strict and literal truth could be treated as a pragmatic phenomenon, or as a semantic one. One who regards it as a pragmatic phenomenon will think of truth *simpliciter* as being equivalent to strict and literal truth, so that the gap between correctness and strict and literal truth is equally a gap between correctness and truth *simpliciter*.[14] One who regards it as a semantic phenomenon will instead think of truth *simpliciter* as diverging from strict and literal truth, and as coinciding, at least in typical cases, with what I have been calling correctness. The dispute between the pragmatic and semantic approaches can be made to sound momentous: the proponent of the pragmatic view accusing the proponent of the semantic view of mutilating the logical structures of sentences about composite objects; the proponent of the semantic view gasping in outrage at the suggestion that the principles of folk mereology might be *false*. There may, in fact, be some way to attach genuine significance to this dispute. But I think the default view should be that the line between the domains of semantics and pragmatics is so unclear that there is no determinate fact of the matter as regards which of these approaches is the right one.

The price of adopting the semantic approach is that one is obliged to hold that some very simple and apparently straightforward sentences don't have the sort of se-

[14]Even though 'S is true' and 'S is strictly and literally true' are *semantically* equivalent, the proponent of the pragmatic approach will want to recognise a *pragmatic* difference between them: 'S is true' can correctly be asserted whenever S can, whereas 'S is strictly and literally true' can only be correctly asserted when S really is strictly and literally true. 'Strictly and literally' is semantically inert, but it has the pragmatic function of neutralising the forces which make correctness diverge from truth.

mantics one would most naturally assume them to have. This sort of deviousness has been out of fashion ever since certain phenomenalists attempted to analyse 'There is a table in the next room' as some big conjunction of counterfactual conditionals about experience. However, there is no need to subject the apparent logical form of the sentences in question to anything like this sort of mutilation. A less revisionary proposal is that ordinary sentences about material objects contain a single unpronounced "according to the fiction" operator, of the sort discussed in the previous section (see Lewis 1978, p. 262). The same sentence used first in the street and then in the metaphysics seminar room has different meanings; the former utterance does, while the latter utterance does not, contain the unpronounced operator.[15] 'Strictly and literally speaking' is another operator, whose function is to undo the effects of any unpronounced 'according to the mereological fiction' operator that might be lurking in the sentence to which it is applied. A sentence is true strictly and literally speaking iff it is true when all unpronounced operators are deleted. A hardline defender of the semantic account might make the same move for *strict and literal truth* as for truth itself; then we would need some other word for the status I am using 'strictly and literally true' to express: 'true taken at face value' or 'true when all unpronounced operators are deleted' or something like that. Those who want to adopt the semantic approach but balk at 'true but not strictly and literally true' should feel free to mentally substitute one of these other expressions whenever I use the expression 'strictly and literally'.

[15]I suppose it's conceivable that a proponent of the semantic approach might hold that the move from the street to the seminar room affects only the pragmatics, and not the semantics. This is the mirror image of what I have been calling the pragmatic approach: in the street, correct assertability goes hand in hand with truth; but in the metaphysics room, one can correctly assert false sentences, provided that they *would* be true if only they didn't contain any unpronounced operators.

The proponent of the semantic approach is not bound to attribute the change of meaning that sentences about composite objects undergo between the seminar room and the street to an unpronounced operator. It might instead be due to a change of meaning of some of the words that are pronounced as part of the sentence. For example, *quantifiers* could change their meanings: the meaning of 'there is' in the street could be the same as the meaning of 'according to the mereological fiction there is' in the seminar room, and similarly for other quantifiers.[16] (Something more than this is required to account for the change of meaning of quantifier-free sentences, if there are such things in English as quantifier-free sentences.) This strategy achieves the same effect as the unpronounced-operator proposal, without requiring any modification whatsoever in our views about the "logical forms" of ordinary sentences.[17]

[16]Thus, the ordinary meaning of 'someone loves someone else' is the strict and literal meaning of 'according to the fiction, there is someone such that according to the fiction, there is someone else such that the former loves the latter'. This will be equivalent to the more familiar 'according to the fiction, someone loves someone else', provided that the logic of 'according to the fiction' makes all but the outermost occurrence of 'according to the fiction' in any sentence redundant.

[17]Even though the fictionalist view can in this way be formulated as a view according to which 'there are' can mean different things, and means something different in the mouth of a typical person in the street from what it means in the mouth of a philosopher inclined to doubt the existence of composite things, I still think there is a big difference between the view I am advocating and the popular view—presented, for instance, by Putnam (1987)—that debates about the metaphysics of composition are mere pseudo-debates, in which the appearance of disagreement arises from the participants' assigning different meanings to the quantifiers. What the issue ultimately comes down to, I think, is the question whether of the various different concepts that might be expressed by a concept in the same grammatical category as the quantifier 'there are', one concept is especially *important* and *fundamental*. Important and fundamental in what sense? Perhaps this: whereas the other concepts that are candidates to be the meaning of 'there are' can be analysed in terms of it, it can't be analysed in terms of them. I *suspect* that many philosophers who are inclined to the view that debates about composition are pseudo-debates actually agree that of the many senses of 'there are' which they recognise, one is important

If the claim that it is correct to assert a sentence just in case it is fictionally true is interpreted pragmatically, it must be distinguished from the plainly false claim that it is always *pragmatically appropriate* to utter sentences that are fictionally true. It is generally recognised that a host of phenomena can make pragmatic appropriateness come apart from truth: we should allow that all the same factors can make pragmatic appropriateness come apart from fictional truth. Someone who thinks the sentence 'There is a pile of plates in the sink' is a *true* answer to the question 'What is in the sink?' may nevertheless regard it as a pragmatically inappropriate one—supposing that there is also a vial of plutonium in the sink, someone who knowingly neglected to mention the fact would be guilty of deception. The very same phenomenon can arise even if 'There is a pile of plates in the sink' is only fictionally true. The right pragmatic interpretation of the claim that a sentence is "correctly asserted" iff it is fictionally true is something like this: fictional truth plays the distinctive role in constraining and determining pragmatic appropriateness which has normally been ascribed to *truth*.

Indeed, there are many circumstances in which further pragmatic factors would make it inappropriate simply to utter some sentence that is fictionally but not strictly and literally true. Surely there are many people who really do wholeheartedly believe folk mereology; doubtless some of these people falsely believe that everyone believes

---

and fundamental in this sense, and furthermore, that when 'there are' is understood in this way, 'there are no composite things' is true. The only remaining issue between me and such a philosopher is the question whether, when metaphysicians who apparently disagree about composition, like Lewis and van Inwagen, use the words 'there are', they should be understood as expressing the important and fundamental concept. But given that they clearly intend to be talking about something important and fundamental—something categorically different from restricted quantification and quantification-according-to-a-fiction—surely my interpretation of these philosophers has to be the correct one.

folk mereology, and indeed that everyone believes that everyone believes folk mereology, and so on. If I say 'There is a pile of plates in the sink' to such a person despite believing it to be strictly and literally false, I will be guilty of a certain sort of dishonesty. I know that my audience will end up with a belief that I take to be false; moreover, she will mistakenly attribute this belief, and the intention to convey it to her, to me. This is a venial kind of dishonesty—after all, the same thing happens whenever your audience has a background belief which you take to be false, but which they take to be common knowledge, and from which you know they will draw false inferences based on what you say. But if I wanted to be *perfectly* honest, or *perfectly* pragmatically correct, I should have revealed my metaphysical scruples.

This brings up an important problem for the pragmatic view. Within communities of people who are all completely confident that folk mereology is strictly and literally true, and who know this about each other, and who know *this* about each other, and so on, sentences about ordinary material objects are used to *convey* beliefs which entail the strict and literal truth of parts of folk mereology. If, before the interview with the god which I described in section 1.2, I had believed myself to be a member of this kind of community, my protest that it was not *incorrect* for me to assert sentences like 'There is a pile of plates in the sink' would have a certain hollowness. I admit, after all, that I used to believe that there really was such an entity as the pile of plates in the sink; that it was part of my aim in uttering the sentence 'There is a pile of plates in the sink' to induce this belief in others; that I expected the others to form the belief because of their recognition of my intention. . . . According to Grice's (1957) analysis of speaker-meaning, it follows that I *meant* (speaker-meant) that there really was such an entity as the pile of plates in the sink. And this seems like the right thing to say. So what I have learned from the god doesn't merely contradict the

proposition *semantically expressed* by 'there is a pile of plates in the sink': it also contradicts at least one of the propositions that I *speaker*-meant by the sentence. In what meaningful sense, then, can I still claim that what the god told me doesn't really contradict what I meant when I said 'There is a pile of plates in the sink'? How can this sentence have been *correct*, given that it both semantically expresses a falsehood, and is used to communicate a falsehood?

The semantic approach faces a similar challenge. If 'There is a pile of plates in the sink' is not only strictly and literally false, but also used to communicate a falsehood, what justification could there be for taking it to have anything other than the obvious truth-conditions?

This objection depends on the premise that the belief that folk mereology is strictly and literally true is a "mutual belief" in our society. I'm not sure that this is right. As I've already said, the task of determining what people who have never given any thought to the matter believe about metaphysical questions is a difficult one; given the right leading questions, people can be made to say all sorts of things. So the task of determining what people who have never given any thought to the matter believe about the metaphysical beliefs of others is even more difficult; the pressure to conclude that people don't really have an opinion, or that it's not determinate what they think, is even greater.

Instead of pursuing this argument further, let me briefly sketch a different reply. What we do when we associate a certain proposition with a sentence as the proposition which has to be true for an utterance of that sentence to be *correct* is closely analogous to what we do when we engage in the more familiar semantic project of associating a proposition with a sentence as the one which has to be true for that sentence to be *true*. Statements about the correctness-conditions of sentences in a linguistic

community, like statements about truth-conditions, while they may be ultimately grounded in facts about the mental states of the members of the community, are not simply generalisations about what people in general speaker-mean by their uses of the sentences. Other conditions are in play as well. In the present case, an important consideration counting against making the (strict and literal) existence of a pile of plates in the sink a condition for the correctness of the sentence 'There is a pile of plates in the sink', even in the community of wholehearted believers in folk mereology, is the fact that they treat the inference from 'There are some plates piled up' to 'There is a pile of plates' as if it were completely unproblematic, indeed valid. It is indeed a desideratum on the assignment of correctness-conditions that if a sentence is generally used to convey a certain belief (and this is common knowledge, etc.), that belief should be entailed by the sentence's correctness condition. But it is another desideratum that inferences that are generally treated as valid should come out really being "valid", in the sense that the correctness of the premises should entail the correctness of the conclusion. In this case these desiderata are in conflict. My claim is that even if the first wins out when the assignment of *truth*-conditions is concerned, the second prevails for the assignment of correctness-conditions.

You might disagree with this judgment. You might think that an ordinary assertion of 'there is a pile of plates in the sink' is flatly *wrong* if there is strictly and literally speaking no such thing. If you agree with my claim that even if we found out that these metaphysical commitments were false, we would continue to talk in the same way, you must think that that would be because finding this out would lead us to *abandon* our old, metaphysically committed practice, and replace it with a new, fictionalistic one. Perhaps I can still convince you to adopt a "revolutionary" counterpart to the "hermeneutic" version of fictionalism I have been defending up to

now.[18] Even we, who have not had the benefit of a conversation with the god, would do well do make the switch to a practice whose standard of correctness is given by truth in the mereological fiction. Why should we do this? Well, it would be very useful if we could be completely confident that the substitution of a description like 'compose a pile' for a description like 'are piled up' would always preserve correctness. It is *easier* to use singular expressions than plural ones: 'there are three piles of plates on the table' is much more succinct than any sentence which has the same consequences as regards the arrangement of plates but makes no mention of piles or other objects composed of plates. But given the way our current practice works, the long-winded claim about the plates taken individually has the advantage that it can be correct even if there really aren't any such things as piles of plates. Caution pulls in one direction; the desire to communicate efficiently pulls in another direction. An unfortunate dilemma to be faced with all the time; a good reason to adopt a practice in which it does not arise.[19]

The distinction between the revolutionary and hermeneutic forms of fictionalism may not really be such a significant one. After all, the new practice that the revolutionary advocates is not completely unrelated to the old one—it is not as if someone advocated using the words 'gnomon' and 'deodand' and 'palimpsest' to refer to newly discovered insect species, on the grounds that their old meanings had fallen into desuetude. No: what is supposed to be good about the new practice is that it achieves some purpose which the old practice somehow set out to achieve, but failed

[18]These terms are from Burgess and Rosen 1997, p. 6.

[19]In section 1.4 and chapter 2 I will argue that there is in fact no good reason to believe that folk mereology is strictly and literally true: this will provide further support for "revolutionary" fictionalism. These considerations in favour of adopting revolutionary fictionalism are spelled out in more detail in Dorr and Rosen 2001.

at. (We wanted to be able to talk about piles of plates as a less cumbersome way to express claims we could equally well have expressed by talking about individual plates: unfortunately for us, we failed, and ended up saying things which presupposed a substantial metaphysical theory of composition.) Thus, the revolutionary is committed to *some* view about the nature of our current practice. The propositions which the revolutionary thinks we *ought* to assign to sentences as correctness-conditions are *already* assigned to them by a scheme of assignment which gives a high importance to fulfilling the relevant purposes, and a low importance to speakers' unreflectively held metaphysical opinions. But the notion of 'correctness conditions' isn't so very well-understood that we can be confident that there are determinate and generally-agreed facts about the weights to be given to the various considerations which count in making a given assignments of correctness conditions the true one. Thus, it is possible that the hermeneut and the revolutionary are really saying the same thing, but appear to disagree because they use the word 'correct' in different ways.

1.4 Why believe folk mereology?

I will postpone further discussion of the question what it takes for an ordinary assertion about composite objects to be correct until chapter 3. The rest of this chapter and all of the next one will be devoted to the question which theory of composition is strictly and literally true. There is an enormous range of consistent theories to choose from.[20] None of these theories can be conclusively ruled out by conceptual considerations alone. But how should we distribute our credence among them? Should most of

[20]In section 2.5 I will give a highly tendentious argument that the appearance that theories other than Nihilism are consistent is misleading. But for the moment let's assume that this is wrong.

our credence go to folk mereology? Or to Nihilism? Or to van Inwagen's "organicist" view? Or should it be distributed more evenly among the different possible theories, including those that are too random and complicated to have ever been written down or defended?

In this section I will take an look at some strategies that might be used to argue for one or another distribution of credence among the theories. The discussion will be tentative and exploratory.[21] While I will be making some critical points, especially about the prospects that any of the strategies might yield a justification of folk mereology, my main aim in mapping out the epistemological territory is to put us in a position to appreciate the epistemological significance of the conclusions for which I will be arguing in chapter 2.

1.4.1 Direct empirical evidence

Perception is a good way to find out about all sorts of things. Might it also be a good way to find out the truth about composition? It is natural to think that we can *just see* that there are composite objects. It sounds right, for example, to say that right now I can see that there is a sandwich on the table before me, and see that it has some slices of bread and cheese as parts. From this I can infer, by the definition of 'composite', that the sandwich is composite; and hence infer that there is at least one composite object, i.e. that Nihilism is false. Of course, perception alone won't be enough to establish any interesting generalisations about composition: for all perception tells me, the facts about composition could be quite different outside that small region of the universe that I am at present able to observe. But if we can find out some immediate and local facts about composition by perception, we might

[21] Parts of this discussion will be drawn from Dorr and Rosen 2001.

be able to find out a lot more by relying on memory, testimony and induction. For example, if on each of the many occasions when I have had perceptual justification for believing that some plates were piled up, I have had equally good perceptual justification for believing that they composed a pile, it would probably be reasonable for me to infer that *whenever* some plates are piled up, they compose a pile.

This strategy faces a serious problem. If your visual system is functioning normally, your visual experience depends only on the motions of the photons in the vicinity of your eyes. And it seems to be a consequence of any remotely acceptable theory about photons that the behaviour of photons depends entirely on the arrangement of small things: what sorts of elementary particles there are, and how they stand to one another; what the curvature is at each point in spacetime; that sort of thing. Nowhere in physics do we come across a law that predicts that a photon will behave one way when it comes across some particles that compose something, and behave a different way when it comes across some particles of the same sorts, arranged in the same ways, that don't compose anything. Similar claims seem to be true for the other sense modalities. Hence, whenever some things in your environment compose something, things would have looked, sounded, felt, smelt and tasted to you exactly the same way that they actually do if those things had been arranged exactly as they actually are, without composing anything.

So clearly there is some sense in which we can't *resolve* the question whether there are composite objects like sandwiches, chairs, piles of plates, etc., by appealing to the way things perceptually appear. However, this isn't quite the end of the story for the appeal to direct empirical evidence. The dispute between the commonsense belief in external objects and the theory that all of your experience is caused by a demon, no less powerful than malicious, devoted to the goal of ensuring that your experience is

exactly as it would be if things were the way we normally take them to be, is also not resoluble in the most straightforward way by appeal to direct empirical evidence, since both parties agree on the facts about perceptual experience. Nevertheless, it is sometimes claimed that the dispute can in another sense be settled by empirical methods. It *looks to be the case that* there are external bodies; it does not *look to be the case that* the demon hypothesis is true. Other things being equal, when it looks to be the case that p, it is reasonable to believe to a high degree that p.[22] The situation with regard to folk mereology might be thought to be analogous. Although things wouldn't look any different if the sandwich didn't exist (provided that those particles continued to be arranged as they actually are), nevertheless it looks to be the case that the sandwich exists and is a composite object; so I can reasonably, though of course defeasibly, rule out theories like Organicism and Nihilism according to which there aren't any sandwiches.

This view raises some interesting questions. Suppose that there were a community of people who took, say, Organicism for granted in the same way that we take folk mereology for granted. They would doubtless not think of their perceptual states as being systematically misleading. Instead, they would say things like 'it looks to be the case that there are some things arranged sandwichwise' in the circumstances in which we say 'it looks to be the case that there is a sandwich'. Does this mean that it doesn't look to be the case to them that there are sandwiches? If so, this appeal to direct empirical evidence starts to look like it should better be described as an appeal to "common sense" or "what we find it natural to believe" (see section 1.4.4 below). If not, how are we to decide whether they or we are right about the contents

[22]See Pryor 2000 for a recent exposition of this sort of view.

of perception?[23]

1.4.2 Indirect empirical evidence

You can't tell that there are electrons just by opening your eyes and looking out the window. If there is empirical evidence for the existence of electrons, it is evidence of an indirect sort, gathered painstakingly by science. Could we have evidence of this sort for believing some theory of composition? Scientists are no less prone than anyone else to talk as if they had conclusive a priori grounds for being certain of folk mereology, so most of the scientific theories we actually have could not be strictly and literally true unless there were a wide range of composite objects. The policy of believing the propositions strictly and literally expressed by our best scientific theories has a lot to be said for it under most circumstances. But it is not just because they are produced by those excellent folks with the white coats that the theories merit this attitude; it is because the theories have emerged from a distinctive process in which empirical testing plays a central role. If a scientific theory makes some assumption which has never been tested—which has been taken so much for granted that alternative theories which do not make the assumption have never even been formulated—then the parts of the theory that rely on that assumption do not deserve the sort of deference that the rest of the theory may have earned. This seems precisely to be status of the assumptions about mereology that are built into our

[23]Perhaps someone will claim that the debate would be resolved by psychological evidence indicating that a propensity to think in folk-mereological terms was somehow hardwired into our brains, including our perceptual systems. I am not sure what it would be for there to be evidence of which this was the unambiguously correct description. But even if we had such evidence, its effect would just be to make the appeal to direct perception turn into a biologically-flavoured variant of the appeal to "natural belief" discussed below in section 1.4.4.

scientific theories.

To see whether there really is evidence of a scientific sort for believing folk mereology or some other theory of composition, we should compare the scientific theories we actually have with other theories which make different claims about composition, or which are completely neutral about composition. Only if our actual theories come out ahead by scientific standards will we have identified a reason to believe folk mereology. And when we apply these scientific standards, we had better not count it as a significant point in favour of our theories that they fit better with existing theories, or are easier to understand: this would be no more than another version of the appeal to epistemic conservativism or natural belief which I will discuss in section 1.4.4. Likewise, we should not count it as a point in favour of our theories that they can explain the "data", where the "data" are described in such a way as to presuppose the existence of composite things; this would be no more than another version of the appeal to direct empirical evidence.

One obvious way to come up with mereologically neutral alternatives to our current scientific theories is by using some version of the 'according to the mereological fiction' operator described in section 1.2. This gives us an easy way of shedding unwarranted mereological assumptions while keeping the rest of the theory unchanged. Unfortunately, theories of this form have a deservedly bad reputation in science. Using "hedging" operators of this kind, the most dreadful theories can be made to *look* nice and simple. The theory 'As far as observable matters are concerned, it is just as if T' might look to be roughly as simple as T. But in the sense that matters to the epistemology of science, this theory is *extremely* complex. The bad precedent set by the phenomenalist has made many philosophers of science look askance at all theories formulated in an "as if" way. I think that this is too hasty; for one thing, it

gives the indispensability argument for the existence of mathematical entities more credit than it deserves (see Yablo 2000, MS). Nevertheless, it gives us a good reason to see whether we can do without 'according to the fiction' operators in formulating alternatives to mereologically committed scientific theories.

Given a typical scientific theory, it is in fact a trivial matter to come up with an alternative theory in which the mereological assumptions are eliminated, by systematically replacing singular terms and singular predicates with plural ones.[24] The process can be formalised as an algorithm. Given an ordinary scientific theory which makes the usual sorts of assumptions about mereology, we generate a new theory which doesn't by

(i) replacing every occurrence of 'there is something which' ($\exists x$) with 'there are some things which' ($\exists\!\!\exists xx$).[25]

(ii) replacing every occurrence of 'for every thing' ($\forall x$) with 'whenever there are some things' ($\forall\forall xx$)[26]

[24]This method is very similar to van Inwagen's recipe for giving "paraphrases" of sentences apparently committed to composite non-living things like chairs and tables (van Inwagen 1990, sections 10–11). Another recent expression of more or less the same idea is in Morton (1975).

[25]This notation for plural quantification and variables is due to Burgess and Rosen 1997, section 1.c.1.b.

[26]What do we do if the old theory contained *plural* quantifiers? It seems that in that case we will need to introduce "perplural" quantifiers in the new theory, quantifiers which stand to plural quantifiers just as plural quantifiers stand to singular ones (see Hazen 1997). Ordinary language doesn't contain any clear examples of such quantifiers, though I think it does contain perplural referring expressions—'the Beatles and the Stones' seems to be an example, on one of its disambiguations; so does 'Bob and Carol and Ted and Alice'.

Some philosophers have claimed not to understand perplural quantifiers, or to be able to understand them only as disguised quantifiers over sets. This sort of view would have implications similar to the implications of the view that ordinary *plural*

(iii) replacing every occurrence of 'is part of' with 'are among' (some things *xx are among* some things *yy* iff whenever something is one of *xx*, it is one of *yy*)

(iv) replacing every occurrence of 'is identical to' with 'are the same things as' (*xx are the same things as yy* iff *xx* are among *yy* and *yy* are among *xx*)

(v) replacing every singular predicate in the theory with a new plural predicate. Thus 'is a molecule' is replaced by 'are arranged molecule-wise', 'has mass m' is replaced by 'have mass m', 'is located one nanometre away from' is replaced by 'are located one nanometre away from', and so on.[27]

The new theory constructed using this method will not be a *logical consequence* of the old theory upon which it was based, since it will contain predicates that did not occur in the old theory ('are arranged molecule-wise', 'have mass M', and so forth). But it is very plausible that the meanings of these new predicates are systematically related to the meanings of the predicates of the old theory in such a way that the new theory follows *analytically* from the old one. Things could be arranged molecule-wise without there being any molecules; but if there are molecules, then there are things

quantification is disguised quantification over sets, discussed below. In any case, few extant theories in science seem to make any essential use of both plural quantifiers and assumptions about mereology.

[27]Remember that these plural expressions should be understood in such a way that there is nothing contradictory in the claim 'there are some things such that there is only one of them'. If the old theory was compatible with the claim that atoms are simple, the new theory we get by applying the algorithm will be compatible with the claim that whenever there are some things arranged atomwise, there is only one of them.

arranged molecule-wise.[28] This doesn't depend on the assumption that the predicates of the new theory orthographically contain the corresponding predicates of the old theory. Even if we had written the new theory using nothing but arbitrary predicate-letters, the meanings of these predicate-letters would still have been fixed in such a way that if the old theory was in fact true, the new theory could not have failed to be true.[29]

(The singulars-to-plurals algorithm is designed for dealing with theories in which mereological vocabulary is used tenselessly. It will break down if it is applied in the obvious way to a theory in which the basic mereological concept is 'x is part of y at t': the consistent sentence 'x is part of y at t but not at t'' will go over into the nonsensical 'xx are among yy at t but not at t''.[30] What are we to do with

[28] *Which* things are arranged molecule-wise, if the old theory, which entails the existence of molecules, is true? We can't say that some things are arranged molecule-wise iff there is exactly one of them, and it is a molecule: since the old theory said that some molecules have proper parts, the new theory says that some things arranged molecule-wise are more than one in number. One option would be to say that some things are arranged molecule-wise iff they are all simple, and they compose a molecule. But this only works if everything is composed of simple things; whereas the old theory may have been compatible with the existence of *gunk*, i.e. composite objects which lack simple parts. An option which will work even if there is gunk is to say that if there is a molecule, its parts—*all* of its parts—are arranged molecule-wise. In general, for some things to be arranged F-wise, where 'F-wise' is one of the new predicates introduced by the algorithm, it must be the case that everything composed of some of them is itself one of them.

[29] It is worth noting that on one possible analysis of the 'according to the mereological fiction' operator, the proposal to get our new scientific theories by writing 'according to the mereological fiction' in front of the old ones is really just a notational variant of the present proposal. See footnote 14 of chapter 3.

[30] The fact that 'Cian and Roger are among the members of the Department in 2001 but not in 1999' is consistent no more shows that 'xx are among yy at in 2001 but not in 1999' is consistent than the fact that 'Bush is the President in 2001 but not in 2005' is consistent shows that 'x is identical to y in 2001 but not in 2005' is consistent.

scientific theories that are formulated in this way, and which make essential reference to mereologically variable composite things? One approach would be to interpret all this talk of time-relative parthood along "four-dimensionalist" lines: 'x is part of y at t' can first be rendered as 'the timeslice of x at t is part of the timeslice of y at t', and then the algorithm can be applied as before.[31] It might be objected that a new theory generated by this method will be inferior to the original theory because it entails the existence of things that last only for an instant, while the original theory does not. This wouldn't be a problem if there was independent reason for believing in instantaneous things, as I am inclined to think there is.[32] But in any case, there are other possible strategies that could allow us to retain the "three-dimensionalist" character of a scientific theory. We could, for example, introduce a new four-place predicate, taking two plural expressions and two times as arguments, to do the work done by the predicate 'the thing that xx compose at t is identical to the thing that yy compose at t'' in the old theory. 'x is F at t' will now be transformed into 'there exist a t' and some yy such that $R(xx, t, yy, t')$ and yy are arranged F-wise at t'.[33])

The method I gave for eliminating commitment to composite things from scientific

[31]This account of time-relative parthood in terms of temporal parts is described in detail by Sider (2002, chapter 3).

[32]I am persuaded that I must believe in spacetime points by the argument in Field 1985. And I don't see any good reason to believe in anything else: in this I go along with the argument of Sider 2002, section 4.9—except that Sider thinks that there is reason to believe in mereological sums of spacetime points as well as the spacetime points themselves.

[33]Something even more complicated than this would be required for theories which allow for some things to compose two different things at one time. If—as seems very likely—the composite things discussed by the theory fall into finitely many kinds, such that some things never compose more than one thing of any kind at any time, we could accommodate this by having a different four-place predicate for each kind of thing.

theories relied essentially on the use of plural quantification. My assumption—now widely accepted, thanks to Boolos (1984)—is that the only things which all sentences of the form 'There are some xx such that $\varphi(xx)$' are committed to are the individual ones of xx. 'There are some particles such that there are more than one of them, and there are no sets' is logically consistent. So is 'There are some particles such that there are more than one of them, and there are no complex or composite things of any sort'. Some philosophers deny this assumption. They think that plural existentially quantified sentences, if they are intelligible at all, must be taken to entail the existence of sets or some other sort of special abstract entities.[34] If this was your view, you might not be happy with the theory generated by the singular-to-plural algorithm. You might complain that the theory you began with was thankfully free of commitment to *sets*, whereas the new theory entails the existence of sets thanks to its use of plural quantifiers. It seems reasonable to think that if one had to believe in one or the other, one should believe in composite things rather than sets.[35]

Even if this were a good reason to reject the competitors to our current scientific theories generated by the singular-to-plural algorithm, there are many other competitors to be investigated. Perhaps the laws of physics cannot be formulated adequately without dragging in *something* other than ordinary microphysical things (particles, spacetime points, etc.). Still, we might be able to get by with many fewer of these "extra" things than are contemplated in folk mereology, and we might do better

[34]This view is defended explicitly by Hazen (1993). Oliver and Smiley (2001) give a thorough critical survey of attempts to analyse sentences involving plurals in terms of sentences in which the only terms and quantifiers are singular.

[35]It's not so easy to say why this might be. Composite things are supposed to be *concrete* and therefore unmysterious; but in chapter 2 I will argue that if there are any composite things, they are not causally efficacious, and therefore lack one of the main hallmarks of concreteness.

to adopt a very different theory about their nature and their relations to ordinary microphysical things.[36]

1.4.3 Default reasonableness

Suppose that empirical evidence does little or nothing to resolve the debate between the different theories of composition. What then? Some empiricists have thought that one can justifiably believe a synthetic claim only if one has empirical evidence that counts in favour of that claim. It's an appealing thought, but on reflection it can't be right. There are many propositions which, in the absence of relevant empirical evidence, should not merely be disbelieved, but assigned an extremely low degree of belief. For example, if one had no empirical evidence that bore either way on the question, one ought to believe to a very low degree that there are exactly 123,456,789 things. So one ought to believe to a very high degree—and hence ought to believe, *simpliciter*—that it is not the case that there are exactly 123,456,789 things, even in the absence of supporting empirical evidence. The belief that there are not exactly

[36]The most convincing example I know of of a physical theory in which assumptions about composite things play an essential role is Hartry Field's nominalistic version of Newtonian gravitational theory (Field 1980). The basic predicates of Fieldian physics all apply only to spacetime points; however, Field makes extensive use of the expressive resources of Mereology to express the laws which platonistic theories express using references to mathematical entities. I think that the work done by mereological sums in Field's theory could just as well be done by plural quantification. But if plural quantification is deemed unacceptable, it would be wrong to assume that the only alternatives are to supplement the ontology of spacetime points with mereological sums of them, or with mathematical entities. The fact that mereological sums and mathematical entities, though very different in structure, can do the same "expressive work" in formulating the laws is a hopeful sign that the work might be done in yet other ways. Perhaps some system of entities that no-one has yet thought of will allow a formulation of the laws that is simpler than either Field's or the platonists'.

123,456,789 things is a *default reasonable belief*.[37] It is a belief one is automatically justified in having, unless one has some very peculiar and specific defeating evidence.[38]

Surely there are more interesting examples of default reasonable beliefs than this one. For example, there is a good case to be made for the view that for our ordinary practice of induction to be reasonable, some belief roughly to the effect that the world is a uniform place, governed by simple laws, in which better explanations are true more often than worse ones, will have to turn out to be default reasonable. Could folk mereology, or Nihilism, or some other theory of composition, be justified in the same way?

My own view is that Nihilism can be justified in this way. Just as, in the absence of positive empirical evidence for the existence of dogs or electrons or goblins or phlogiston, one should believe that there aren't any such things, so, in the absence of positive empirical evidence for the existence of things with proper parts, one should believe that there aren't any of *those* things.[39] And I don't think there is any positive evidence for the existence of things with proper parts. The presumption that there

[37]This useful expression is due to Field (2000).

[38]Instead of saying that one can be justified in believing that there are not exactly 123,456,789 things even when one has no empirical evidence that counts in favour of this belief, you could instead say that the empirical evidence all of us have—in fact, all possible states of empirical evidence, except for those which involve appearing to have been told that there are exactly 123,456,789 things by some apparently omniscient being, and the like—does count in favour of the belief. But if we wanted to say this, I suppose we would have to say that the evidence one would have if one had just been told that there were between 123,456,000 and 123,457,000 things by some apparently omniscient being still "counted in favour of" the belief that there are not exactly 123,456,789 things, which seems like an odd way to use the expression 'count in favour of'.

[39]Contrariwise, in the absence of relevant empirical evidence, one *should* believe that there are non-dogs, non-electrons, non-goblins, non-phlogiston, and things that *don't* have proper parts.

aren't any composite things rests, I think, on a more general principle: other things being equal, one ought to believe that the world is a simpler place, rather than a complicated one. But for there to be interesting mereological relations among things is for the world to be complicated, in a certain respect.[40]

You either find this sort of thing convincing or you don't. Perhaps you hold that the sort of simplicity which I hold so dear is simply not a "theoretical virtue" at all; or that I have overrated its importance at the expense of some other virtue which is had to a high degree by the theory that there *are* composite things. These disagreements aren't easily resolved by argument; one finds oneself constantly bumping up against rock bottom.

However, I do want to say something against the idea that folk mereology itself might be a default reasonable belief. Folk mereology attributes a specific and distinctive sort of structure to the world. In that respect it is similar to general relativity, and Aristotle's theory of the heavens, and unlike, say, Nihilism, and the denial of Aristotle's theory of the heavens. It is very implausible to think that it could be reasonable to believe *that* sort of theory in the absence of relevant evidence. If we know anything at all about epistemology, we know that you need empirical evidence to be justified in believing far-reaching and specific hypotheses about the structure of the world. If epistemology didn't work like this—if the correct story about rationality in induction were sensitive not just to schematic, abstract features of theories such as simplicity and strength, but to the fine distinctions between the contents of dif-

[40]When I talk about "the world" I do not mean to be understood as talking about a concrete thing that has every other actual concrete thing as a part: I don't believe in any such thing. For now, it will do to understand me as talking about some kind of abstract thing, such as the conjunction of all true propositions. I really don't believe in those either, but let's not worry about that until part II.

ferent theories—it would lose much of its point. It would be useless to try to resolve our object-level debates about the nature of the world by appealing to facts about evidence, reasonableness, and so forth, since we would all adopt fine-tuned epistemological theories according to which our own beliefs were reasonable and those of our opponents unreasonable, given the evidence.

To see that folk mereology really does make a very far-reaching and specific claim about the structure of the world, it helps if we eliminate specifically mereological vocabulary from the theory, replacing it with quantification over properties. Suppose for the sake of definiteness that folk mereology is formulated in such a way that 'is part of' is the single primitive mereological expression; the other mereological vocabulary is defined in terms of this predicate in the usual way. For the theory to be true, then, there must be a binary relation which obeys certain laws. For example, the following must hold: whenever there are some things that are piled up, there is something x to which all of them bear that relation, and which is such that for everything y that bears that relation to x, there is something that bears that relation both to y and to one of them. (This is what the claim 'when things are piled up, they compose something' looks like when 'compose' is replaced by its definition in terms of 'is part of'.) There will also be generalisations about the interaction between the special relation and other notions: for example, whenever the special relation holds between x and y, x must weigh no more than y. If there is no binary relation that obeys all those laws, folk mereology is not true.

1.4.4 Epistemic conservativism, natural belief and intuition

A default reasonable belief is a belief that *anyone* ought to have, in the absence of relevant empirical evidence. Some philosophers hold that it can happen that someone

ought to believe that p when their empirical evidence is neutral as regards whether p even when p is not a default reasonable belief. In other words, it can happen that *some* people ought to believe that p in the absence of relevant evidence, while others ought not to believe that p. What makes the difference, if not empirical evidence? Different answers are possible. For example, some say that one should believe that p in such circumstances just in case one *already* believes that p: this is a version of *epistemic conservativism* (Harman 1989). Others say that one ought to believe that p just in case one *finds it natural* to believe that p. A third view of this sort is that whether one ought to believe that p depends on whether one *intuits* that p (Bealer 1999).

The apparent conflict between these views and the view that in the absence of relevant empirical evidence one ought to believe that φ iff φ is a default reasonable belief may be more apparent than real. Perhaps 'ought to believe' and other terms of epistemic evaluation are ambiguous. For example, everyone should agree that it is true in *some* sense that one ought to believe in accordance with one's intuitions. Provided one follows one's intuitions where they lead, one is immune to a certain sort of epistemic criticism—a sort of criticism that plays an important role in discussions of hiring decisions in philosophy departments, for example. This is compatible with its being true in a different sense that one ought to believe not what one *does* intuit, but what one *ought* to intuit, and with its being true that one ought to intuit that p iff p is a default reasonable belief. Likewise, everyone should agree that there are sorts of epistemic criticism to which one is immune provided that one changes one's views only when this is required in order to maintain coherence, or provided one believes what comes naturally.

The proponents of this sort of approach to epistemology have a relatively easy

time defending the claim that most people ought to believe folk mereology. It is quite plausible that most people do start out believing folk mereology, and that they find it natural and intuitive to do so. (Although this isn't *entirely* clear: see footnote 7.) Nevertheless, there is a consideration which may defeat the prima facie justification this allegedly confers. If we start out believing folk mereology, we probably also start out believing that folk mereology is analytic. (We certainly *treat* it as if it were analytic!) Moreover, we believe folk mereology in large part *because* we take it to be analytic. When we make inferences such as the inference from 'some piled-up plates are in the sink' to 'there is a pile of plates in the sink', we do so because these inferences strike as as unproblematic in the same way as other analytically valid inferences. But folk mereology isn't analytic. So our belief in folk mereology is based on a "lemma" which, on reflection, we find to be false. But if it is to be remotely plausible that we ought to believe what we start out believing, an exception must be made for beliefs which are based on lemmas which we discover to be false. It's just obvious that when you change your mind about some important belief upon which many other beliefs are based, you often ought to change your mind about many of those beliefs as well. The same sort of argument works for natural belief and intuition: if we find it natural to believe, and intuit, folk mereology, we do so largely because we find it natural to take folk mereology to be analytic, and because the inferences licensed by folk mereology strike us intuitively as analytically valid. Again, if the prescriptions to believe what comes naturally or what you intuit are to have any credibility, exceptions must be made for propositions which we find it natural to believe, or intuit, on the basis of some lemma which we discover to be false.

Even if this argument doesn't undermine the prima facie justification of folk mereology, we still can't conclude that belief in folk mereology is justified *simpliciter*. We

must first scrutinise other beliefs that enjoy the same sort of *prima facie* justification, and make sure that these beliefs do not conflict with folk mereology. If there is a conflict, some further method will have to be used to resolve it. In the next chapter, I will present an argument which, I hope, will convince those who favour this sort of epistemology that there is such a conflict, and that it should be resolved by giving up our belief in folk mereology.

Chapter 2

Against composite objects

2.1 An argument from counterfactuals

Let us return once again to those plates in the sink. Suppose that they compose something. In section 1.4.1, I pointed out that if the plates had been arranged the same way but composed nothing, the behaviour of photons would have been exactly as it actually is. I concluded from this that if the plates had been arranged in the same way but composed nothing, the facts about peoples' visual experiences would have been exactly as they actually are. This argument can be generalised. What goes for photons also goes for electrons, and quarks, and spacetime points, and so forth: in fact *all* the microphysical facts would have been the same if the plates had been arranged in that way and composed nothing. And this seems to be a good reason to think that *all* other facts—even facts about non-microphysical things like chairs, computers, cats, and people, assuming that there are such things—would have been the same if the plates had been arranged in that way and composed nothing. Given the way the world works, there couldn't be a device that did a better job of detecting

the presence of composite objects than peoples' visual systems do. Any device you could invent would have to depend on the microphysical facts in more or less the same way that the facts about peoples' visual experiences depend on the facts about the arrangement of particles in their brains. So if the plates had been arranged in that way but composed nothing, *everything* would have been exactly as it actually is, but for the absence of that single composite thing.

If this is right, then if the plates do compose something, the thing they compose has a very curious role in the causal economy of the world. It is a *pointless* thing; it is something we should ignore, if we are interested in prediction and explanation. It doesn't really *do* anything—it just sits there inertly. In short, it is an epiphenomenon, an object with no causal power to affect what goes on at the world. It has exactly the same sort of status that *souls* have according to traditional epiphenomenalist substance dualism in the philosophy of mind. The only difference is that it is a lot more boring than souls are supposed to be—whereas a soul is supposed to have many properties which are logically (though not nomologically) independent of the facts about the body to which it belongs, there doesn't seem to be anything left to say about the pile once one has said what its parts are and how they are arranged.

If this argument is any good, it can be adapted to show that *plates*, if they exist, are epiphenomena. If there is a plate in a certain region of space, then there are also some molecules there. But given that those molecules are there, and are related to one another as they actually are, it would make no difference to anything if they failed to compose anything. Some molecules arranged in that way but composing nothing would be just as good as a plate for any ordinary purpose. If all the plates in my kitchen dresser were to cease to exist, but all the molecules in my kitchen dresser were to stay arranged exactly as they are, I wouldn't care very much. My guests

would have no new reason to worry about their food getting all over the tablecloth. In fact they would never know, unless I told them—but come to think of it, I would never know either. Likewise if there were no molecules there, but all the atoms stayed arranged in the same way. And I wouldn't miss the atoms much if I got to keep the protons, neutrons and electrons. . . .

What we have, then, is an argument for the conclusion that *all* composite objects are epiphenomena. Let's take a closer look at the steps of the argument, to see how it might be resisted. Assume that *xx* are some things, more than one in number, which compose something. Then:

P1 If *xx* had been arranged as they actually are but composed nothing, all the microphysical facts would have been just as they actually are.

P2 If all the microphysical facts had been just as they actually are, and *xx* had been arranged as they actually are but composed nothing, then everything else would have been just the same.[1]

C1 So, if *xx* had been arranged as they actually are but composed nothing, everything else would have been just the same.

[1]What does it mean, exactly, to say that "everything else" would have been the same? We might interpret this to mean that all facts *about things other than the object composed by xx* would have been just the same. But this is a bit too strong. Suppose that *xx* compose an object, *a*, which is itself a proper part of some larger object *b*. It seems somewhat plausible that if *xx* had been arranged just as they are but composed nothing, neither *a* nor *b* would have existed. This isn't the only possible thing to say: maybe we should instead say that *b* would still have existed in that case, and would still have had all of *xx* as parts, though of course it wouldn't have had *a* as a part. It seems best to remain neutral between these options. So let's understand 'everything else would have been just the same' to mean 'all the facts about objects that don't overlap any of *xx* would have been just the same'.

P3 If C1 is true, then the thing *xx* compose is an epiphenomenon.

C2 So the thing *xx* compose is an epiphenomenon.

The argument is valid, so all that remains is to consider the plausibility of the premises.

For P1 to be false, there would have to be some microphysical things other than *xx* themselves, which would (or at least might) have been arranged differently if *xx* had been arranged as they actually are without composing anything. It is easy to see how this could be the case if the distinction between particles which are proper parts of bigger things and particles which aren't proper parts of anything played some role in the fundamental laws of microphysics; but evidently it does not. Microphysics is *nomologically self-contained*: microphysical goings-on can in principle be given complete explanations in microphysical terms, to the extent that they can be explained at all. A Laplacean demon who wanted to know how particles were going to be arranged in the future would have everything he needed once he knew all about the properties and relations of particles in the past, and about the microphysical laws.[2] If determinism is true, he would be able to infer everything about the properties and relations of particles in the future. If it isn't, then he wouldn't; but we couldn't help him by telling him about non-microphysical goings-on, such as the thoughts and feelings of any Cartesian Egos that happened to be hanging around, or the facts about composition.

P1 would also be false if the object composed by *xx* itself itself counted as a

[2]The word 'particles' here is standing for whatever things are mentioned by the true microphysical theory. So "particles" might be real particles, or particle-time-slices, or points of spacetime, or points of a field, or strings, or something else I know not what.

microphysical object: for if it is a microphysical object, the fact that it exists is a microphysical fact which would have been different if it hadn't existed. So our argument will only establish that composite objects *that are not microphysical* are epiphenomenal. This is no real limitation, however. There are many acceptable ways of understanding 'microphysical': provided that the things we count as microphysical are things of which everything else is composed, we can fix the threshold as low as we want. It seems likely to me that if P1 and P2 are true on *any* reading of 'microphysical', they are true on all of them. If so, it follows that *all* composite objects are epiphenomena, since for every composite object, there is *some* sense of 'microphysical' on which it doesn't count as microphysical.

For P2 to be false, there would have to be some facts about composite things disjoint from *xx* which would (or might) have been different if *xx* had composed nothing but all the microphysical facts were the same. Perhaps some of those things wouldn't have existed if that had been the case; or perhaps they would still have existed, but would have had different parts; or perhaps they would have existed and had the same parts but been different in some other way. Someone might think, for example, that if the microphysical facts had been as they actually are but the particles which actually compose the chair I'm sitting on had failed to compose anything, then from this time forth, no particles arranged chairwise would ever have composed anything. The ability of particles arranged chairwise to compose things is fragile: if any particles arranged chairwise should ever fail to compose something, the spell will be broken and there will be no more chairs. Even though the facts about chairs are causally irrelevant to the microphysical facts, and causally irrelevant to the fact that in such-and-such circumstances people don't fall onto the ground, they are causally relevant to the existence of future chairs.

Even if something like this is true, the sort of causal efficacy that it would allow composite objects to have is so very unlike the sort we normally suppose them to have that it seems hardly worth having. We normally suppose that the chair's causal influence has something to do with its physical interactions with nearby material objects, not with its maintenance of the ability of particles to compose chairs. But in any case, there seems to be no good reason to think that anything like this *is* true. In evaluating a counterfactual, one tries as much as possible to respect the laws of nature—even when the antecedent of the counterfactual can only be true if there is some exception to the laws, one does not gratuitously introduce new exceptions (see Lewis 1979). But surely the fact that whenever some particles are arranged chairwise they compose a chair, if it is a fact, is a nomologically necessary truth: a law of nature, or if you prefer, a consequence of some more general law of nature. So in considering a counterfactual whose antecedent requires one to suppose that this law has an exception, we should hold the law fixed in all other cases: even if these particles had been arranged chairwise but failed to compose anything, all other particles arranged chairwise would still have composed something. The same is true of other principles about composition.[3]

[3]—But if it is a law of nature that all particles arranged chairwise compose chairs, it is a law of nature that if any particles arranged chairwise ever fail to compose a chair, no particles arranged chairwise will ever again compose anything. Why not hold this law fixed when evaluating the counterfactual?

—The same sort of reasoning will allow you to conclude that if any collision between two particles had ever violated the law of the conservation of momentum, all collisions thenceforth would also have violated the law. The task of deciding what to hold fixed and what not to hold fixed in evaluating a counterfactual is, famously, a difficult one—even if one knows how to distinguish (consequences of the) laws of nature from other propositions, still the task of deciding what it is for a violation of the laws to be "minimal" is not trivial, and cannot be reduced to any sort of syntactic test. Nevertheless, we seem to be able to tell the difference between sensible ways

If *microphysical reductionism* is true, then any composite object can be described completely by listing the microphysical things that are parts of it, and stating the microphysical facts about them. There are no *emergent* properties of or relations among composite things, logically independent of the microphysical facts and the facts about parthood. If microphysical reductionism is false, then there are emergent properties or relations, and hence more possible ways for P2 to fail. Even if the same composite objects (disjoint from xx) would have existed, with the same parts, had the microphysical facts been the same although xx didn't compose anything, it might be held that the distribution of emergent properties and relations among those composite objects would have been different in that case. But this is utterly incredible. Even if there are emergent properties or relations, surely whether a composite thing has a given emergent property, or whether some composite things stand in a given emergent relation, is *nomologically* determined by the microphysical facts about the parts of the thing in question. Microphysics is *nomologically complete*: all the facts about what the world is like, including the facts about the distribution of emergent properties and relations, are nomologically determined by the microphysical facts. Take mental properties for example, which are the most plausible candidates to be emergent properties. It obviously isn't just a coincidence that various kinds of rearrangements of the particles in a person's brain coincide with certain kinds of changes in the person's mental state: the facts about people's mental properties are at least nomologically *constrained* by the microphysical facts. But when we consider all we have learnt about how the brain works, the only reasonable conclusion is that the microphysical facts leave *no* room for variation in mental properties. Unlike Descartes, we couldn't

of performing the task and frivolous, arbitrary ones—and surely the one suggested above falls into the latter category.

nowadays believe that while the facts about peoples' experiences are nomologically determined by the microphysical facts about their brains, the facts about their rational thoughts are not. So, for much the same reasons that we should think that the facts about the *existence* of other composite objects wouldn't have been any different if the microphysical facts had been the same but xx composed nothing, we should also think that the facts about the distribution of any emergent properties and relations there might happen to be wouldn't have been any different in that case.

P3 is the hardest premise to argue for. 'x is an epiphenomenon' certainly doesn't follow from just any old claim of the form 'if x hadn't existed even though the A-facts had been just as they actually are, nothing else would have been any different'. It might be true that if everything had ceased to exist throughout 1999 but things in the year 2000 had been just as they actually are, everything else would have been just the same. That wouldn't show that the things that exist only in 1999 are epiphenomena. Again, it might be true that if a given particle p had not existed, but the particles that in fact bump into p had had the exact same spatiotemporal trajectories that they actually have, nothing else would have been any different—and that wouldn't show that p was an epiphenomenon. Why should the thing composed by xx be any different? It is not hard to find a relevant difference between these cases and the case described in P3. The things that inhabit 1999 get to be causally efficacious *by* influencing how things are in 2000: things are the way they are in 2000 partly *because* of how those things are. p gets to be causally efficacious by influencing the trajectories of the particles it bumps into; they have the trajectories they have partly because of p and its properties. By contrast, given what we know about the way microphysics works, it seems to be false to say that the particles that compose a certain thing are arranged as they are *because* they compose it. Their being arranged in that way

already has a perfectly good microphysical explanation; there is no need to appeal to the fact that they compose something.[4]

Some people have objected to P3 on the grounds that C1, while true, is *vacuously* true, and hence should not be taken to have any significance for questions about causation. (A counterfactual ⌜If it had been the case that φ, it would have been the case that ψ⌝ is vacuously true if it remains true when any other sentence is substituted for ψ.) That C1 is vacuously true follows from two widely-held views: (i) that when some things arranged a certain way compose something, it is *metaphysically necessary* that if those things are arranged that way, they compose something, and (ii) that counterfactuals with metaphysically impossible antecedents are vacuously true. But both of these claims are subject to doubt. (ii) in particular faces many counterexamples that are quite independent of the present discussion. Many philosophers would agree that it is metaphysically necessary that I am not a bird; but while it may be inde-

[4]On the interpretation which I gave in footnote 1, C1 is compatible with the claim that, if *xx* had been arranged as they actually are without composing anything, certain bigger things that have all of *xx* as parts wouldn't have existed. Suppose this is true: then there is some pressure, I think, to say that the object composed by *xx* is a *cause* of the existence of these larger objects that have it as a part. Does this mean that the object composed by *xx* isn't an epiphenomenon after all? On an unusually strict sense of 'epiphenomenon', I suppose it does. But in the sense that matters—the sense that is relevant to epistemology—being an epiphenomenon is compatible with having causal influence on other entities, provided that they too are epiphenomenal. Suppose we take a world where epiphenomenalist substance dualism is true—there are *souls* in addition to physical things, and the state of each soul is nomologically determined by the state of its corresponding body—and add a new category of *spirits*, whose states are nomologically determined by the states of their corresponding souls. In this new world, we must admit that the souls aren't *entirely* epiphenomenal. Each soul is a necessary intermediary in the causal chain whereby the state of a certain body determines the state of a certain spirit. Nevertheless, they are clearly still epiphenomenal in the relevant sense. Nothing about the physical world, or about any other soul, or about any spirit other than the one associated with it, is at it is because of a given soul.

terminate what sort of bird I would be, if I were a bird, it doesn't seem to be the case that I would be a bird who was also not a bird. Nolan (1997) assembles many other arguments against (ii) which should, I think, convince even the most hardened anti-essentialist.[5]

So P3, although clearly the weakest link in the argument, is nevertheless quite defensible. Nevertheless, it would be nice to have an argument for the conclusion that composite things are epiphenomena that didn't rely on this premise, or on any other contentious claim about the relevance of counterfactuals to questions about causation.[6] I will give such an argument in the next section. The new argument will depend entirely on direct intuitions about the causal facts at certain possible worlds, and on intuitions about the relevance of certain structural analogies between possible worlds.

[5]I will discuss some reasons for doubting (i) in section 2.3.2 below.

[6]We would be on firmer ground if we replaced C1 with a simpler counterfactual: if *xx* had not composed anything, nothing else would have been any different. But this claim is harder to justify than C1: to defend it, we would have to argue that *xx* would still have existed and been arranged in the same way even if they had composed nothing. Why shouldn't we say instead that if they hadn't composed anything, they wouldn't have existed at all, or would have been arranged differently, e.g. by being more scattered than they actually are? One way to respond to this question would be to distinguish different senses of the counterfactual, and then argue that in the sense that's relevant to causal matters—the sense on which the argument from 'if *xx* had not composed anything, nothing else would have been any different' to 'the thing *xx* compose is an epiphenomenon' is valid—it is true that if *xx* had not composed anything, they would still have been arranged just as they actually are (see Lewis 1973a). But these are delicate and difficult questions which we should avoid if we can.

2.2 An argument by analogy

To begin with, consider a (conceptually) possible world which I call 'Mirror World'. I'm going to describe how things are at Mirror World without specifying what the causal facts are there; I want you to have a certain intuition about what the causal facts would have to be like, given that the other facts are as I describe them to be. Microphysically, Mirror World is just like the actual world: every particle that actually exists exists at Mirror World, with the same microphysical properties and relations. No actual thing that is not a particle exists at Mirror World. But there are some things that aren't particles at Mirror World: there are exactly as many of these new things as there are particles, and the new things are related one-one to the particles by a new relation R. I'll speak of things that bear R to one another as "reflections of" one another, and of the new things at Mirror World that bear R to the particles as "reflections". Thus, each particle at Mirror World has a unique reflection, and each reflection is the reflection of a unique particle. And that is a complete description of the reflections. The only respect in which any two reflections differ is by being reflections of different things.

Surely, given that the facts at Mirror World are as I said, the reflections at Mirror World must be epiphenomena. If it is ever possible, just by looking at the structure of properties and relations at a possible world, to tell that certain things at that world are epiphenomena, then Mirror World is an example of such a world. What is it about Mirror World that makes this be so clear?

One relevant fact is that the reflections are characterless. They have no distinguishing intrinsic properties, and hence they never undergo real change—as opposed to mere "Cambridge change", i.e. change in how they are related to other things. But

things typically don't get to be causes by undergoing Cambridge changes: if I go in my sleep from being five miles south of an intact barn to being five miles south of a burning barn, I don't thereby cause anything.

But it's not quite right that the only way to be causally efficacious is by having a changeable intrinsic character. One way to imagine a Newtonian world of point-particles is to imagine the particles to be intrinsically alike, locating all change in the spatial relations between the particles. The particles at such a world would surely not have to be epiphenomena: one particle could for example *attract* or *repel* another nearby particle.

The relation R at Mirror World seems however to be quite different in relevant respects from the spatial relations at a world of intrinsically characterless point-particles. The extension of R is monolithic and unchanging; the laws have said all there is to say about it once they have said that every particle bears it to something, and that two particles never bear it to the same thing. There are no laws which say how particles behave differently depending on whether they bear R to something, because this isn't a respect in which particles differ from one another. In general, if one was forbidden from ever mentioning R, one wouldn't lose the ability to make any interesting distinctions, of the sort that might have a role in the explanation of anything.[7]

[7]If you thought that being located in time was a matter of bearing a certain fundamental relation, *location*, to times, or points of spacetime, then it would follow from the description of the reflections at Mirror World that they are not located in time. This might be cited as a further reason for taking them to be epiphenomenal: only temporal things can be causes. But it would be rash for us to build such a specific presupposition about what it is to be temporal into our concept of causation, on pain of having to concede that empirical investigation of the nature of time might well reveal that nothing causes anything. So we should leave open the possibility that being the reflection of something that bears *location* to a time or a spacetime point

Since the reflections at Mirror World are epiphenomena, there must be something wrong with the following argument. 'If a certain reflection had not existed, then since it is a law of nature that everything has a reflection, the particle it is a reflection of wouldn't have existed either. But many other facts about particles would have been different if neither the particle nor its reflection had existed. Therefore, many other facts about particles would have been different if the reflection had not existed. So things are as they are in part because the reflection exists; so the reflection is not an epiphenomenon.' Even if we aren't sure where the argument goes wrong, we know it goes wrong somewhere. Either the fact that facts about particles counterfactually depend on the existence of the reflection isn't sufficient for it to have causal power, or else it isn't true at Mirror World that if a reflection hadn't existed, the thing it reflects wouldn't have existed either.[8]

Now consider a world, "World One", that is very like Mirror World. It too contains all the actual particles, and then some other things. The only difference is that whereas the relation R at Mirror World is one-one, the analogous relation P at World One is many-one. Whenever there are some particles (more than one), there is exactly one non-particle that they, and only they, bear P to. So if n is the number of particles, the number of things at Mirror World is $2n$, and the number of things at World One is $2^n - 1$.

---

is a close enough relation to the time to enable one to be a cause.

[8] Which is it? I say that it depends how the counterfactual is being understood. There is a sense in which would be sufficient for the reflection to have causal power if the facts about particles counterfactually depended on whether it exists; there is a different sense in which it is true at Mirror World that if a reflection hadn't existed, the thing it is the reflection of wouldn't have existed either. One example of a sense that works in the latter way is the "backtracking" sense in which it's generally true that if an effect hadn't occurred, its cause wouldn't have occurred either (see Lewis 1973a).

If the non-particles at Mirror World are epiphenomena, then so are the non-particles at World One. What difference could it make whether the relation is one-one as at Mirror World, or many-one as at World One? The non-particles are still characterless and changeless; the structural features which made R causally irrelevant apply in equal measure to P.

World Two is just like World One, except for two minor changes:

> At World Two, P is reflexive: everything, particles included, bears P to itself.
>
> At World Two, non-particles sometimes bear P to each other: if every particle that bears P to x also bears P to y, then x bears P to y.

These changes ensure that Mereology comes out true at World Two, if we interpret 'is part of' as expressing the relation P. (Hence, folk mereology too is true at World Two under this interpretation, since folk mereology is entailed by Mereology.) But the non-particles at World Two are surely epiphenomenal if those at World One are.

Let us suppose, first, that microphysical reductionism is true. That is, there are no emergent properties; there are no respects in which two composite things could be dissimilar if they were alike as far as the microphysical properties of their minute parts were concerned. If, moreover, Mereology is true, then the actual world fits the description of World Two exactly. Composite things *are* all alike, except that some have parts of one sort, and others have parts of another sort. So our conclusion about World Two applies to the actual world: all non-microphysical things are epiphenomena.

Could the causal efficacy of composite objects perhaps be rescued if we deny that Mereology is the true theory of composition? To see that the answer is no, consider

some variants of Mirror World in which the law governing the relation R is more complicated. Perhaps there are some kinds of particle that have no reflections, and others that have two reflections apiece. I can't see how this could make any difference to the conclusion that reflections are epiphenomena. What is there for the reflections to make a difference to at the new worlds that they couldn't make a difference to at the original Mirror World? Likewise, then, if we make corresponding alterations in World Two. Let some particles—perhaps those that are scattered far apart—compose nothing at all. Let other particles—perhaps those that are arranged statue-wise—compose more than one thing. Provided that there is a law of nature which allows one to predict whether some particles compose something just by knowing their properties and relations, to one another and to other particles, the status of the remaining non-particles as epiphenomena is not affected. And surely, whatever the true theory of composition might be, the number of objects composed by any given particles is determined with at least nomological necessity by the microphysical facts.

What happens if we drop the assumption that microphysical reductionism is true? Suppose that there are some "emergent" properties of composite things that are logically independent of the microphysical and mereological facts, although their distribution is *nomologically* determined by the microphysical facts. Could such properties confer causal power on the things that have them? Again, we can see that the answer is no by considering a corresponding variation of Mirror World. At the variant world, reflections are not all exactly alike, but come in several different kinds; there is a law of nature which enables one to predict what kind of reflection a given particle will have, given its intrinsic properties and its relations to other particles. This change does nothing to diminish the force of the intuition that reflections are epiphenomena.[9]

[9]The intuition that reflections at Mirror World are epiphenomena does begin to

So we do not need to assume microphysical reductionism for the argument to work: emergent properties of complex things pose no problem provided that their distribution is nomologically determined by the microphysical facts. The premise we do need to rely on is the nomological completeness of microphysics. If we allowed the reflections at Mirror World not merely to have distinguishing properties and relations of their own, but to have properties whose distribution was not nomologically determined by the facts about particles, there would be no particular reason to expect them to be epiphenomena. If the laws governing the new properties and relations took the right form, reflections could easily prove to be causally efficacious—they might be involved in a parallel system of causal relations having nothing to do with the particles, or they might actually be able to make a difference to the motions of the particles. Likewise, there could be all sorts of magical goings-on whereby the emergent

---

fade if we give the reflections so many "emergent" properties and relations of their own that the asymmetry between particles and reflections disappears. With enough new properties and relations, one would be able to read off the properties and relations of the particles from the properties and relations of the reflections as easily as vice versa. It's not clear what to say about the causal or counterfactual facts at a world with this sort of redundancy. One possible view is that at such a world, everything that happens is *overdetermined*, having logically independent sufficient causes both in the realm of reflections and in the realm of particles.

Maybe we should say the same thing about World Two: if we added so many emergent properties and relations that one could read off the microphysical facts from the facts about the emergent properties and relations (and the laws), the world might properly be described as a world of massive overdetermination rather than a world of epiphenomena. (I'm not sure if this is right: even with all those extra properties and relations, World Two is still not symmetric, unlike Mirror World: it might be relevant that there are many fewer particles than composite things.) But even the most ardent proponents of emergent properties have typically imagined a much more fragmentary correspondence than this. Even if every controversial claim of identity between properties is false—so that all our mental predicates, and all the distinctive predicates of biology, chemistry, etc., correspond to unique emergent properties or relations—still the facts about the distribution of all of these properties and relations are not remotely rich enough to determine the microphysical facts.

properties of composite things caused particles to swerve from their paths, or could cause other composite things to have other emergent properties without making any difference to the microphysical facts. But we know very well that that's not what the world is like.

Suppose I'm wrong about this: suppose that microphysics isn't nomologically complete, and that because of this composite things are sometimes causally efficacious. Still, the sort of causal efficacy they have is very different from what we might have thought it would be like: a composite thing gets to make a difference to the world only in virtue of its its *radically* emergent properties and relations: emergent properties and relations whose distribution isn't nomically determined by the microphysical facts. If there are some composite things that don't instantiate any such properties or relations, then they at least are epiphenomena. So if the only radically emergent properties and relations are mental ones, for example, we will have to conclude that the only composite things that aren't epiphenomena are the ones with mental lives. But as far as I can see, mental properties and relations are the only even remotely plausible candidates to be *radically* emergent. So we are still in for a big surprise about the causal efficacy of composite things![10]

In the next section, I will consider a few objections to this argument.

[10]In section 2.5 I will return to the question whether sceptical consequences can be drawn from my argument even without the premise that microphysics is nomologically complete.

2.3 Objections

2.3.1 Is structure all that matters?

—The description of World Two is not enough to enable us to tell whether the non-particles there are epiphenomena. For it leaves open two possibilities: the relation P either is parthood, or it isn't. If it isn't parthood, then the non-particles really are epiphenomena; but if it is parthood, the causal facts are quite different. If P isn't parthood, then it is right to say that "change" from bearing P to things of one kind to bearing P to things of another kind isn't real change, but only "Cambridge" change. But if P is parthood, then it is real change. If the energy of one of the atoms that is part of a molecule increases, this is a change in the molecule as well as the atom; whatever the atom causes by changing in that way, the molecule causes as well.

—I find the idea that we might need to specify *that* sort of fact about a world in order to be able to read off the causal facts quite incredible. Think how absurd it would be if someone maintained that the *identities* of particular objects at the actual world needed to be fixed in order to determine the causal facts, so that the causal facts would have been different if things had been just as they actually are except that certain actual things didn't exist, their roles being played instead by things that don't actually exist. This just isn't the sort of difference that matters as far as the causal facts are concerned. Or again, consider a world just like the actual world, but with different properties playing the roles that are actually played by the properties *being an electron* and *being a positron*. This too seems not to be the sort of difference that could conceivably matter as far as the causal facts are concerned. But the objector is

in effect claiming that a difference of this sort does matter. Here is a world, matching the description of World Two, full of epiphenomena; here is another world, perfectly isomorphic to the first one, except that *this* relation (P) has been replaced by *that* one (parthood); at the latter world, nothing is epiphenomenal.

If the objector is right that this sort of difference *does* matter, the concept of causation is a strangely parochial one. We live at a world full of things with parts: we introduce the word 'causation' in such a way that the facts about the application of this word in an arbitrary possible world are sensitive to the facts about which things are parts of which at that world. Now consider the inhabitants of a world structurally like ours, except that nothing is *part* of anything else; instead, there is a relation of "schmarthood" which behaves just as parthood behaves at the actual world. They introduce a word 'schmausation' that works just like our word 'causation', except that the facts about schmausation are sensitive to the facts about schmarthood instead of parthood. We say that their world is a world at which almost all properties are not involved in causation; they say that our world is a world at which almost all properties are not involved in schmausation. But why should we care so much about causation, and not at all about schmausation? Why should we regard it as a weighty consideration against believing some theory that it postulates lots of entities that play no role in causation, but no problem at all when a theory postulates lots of entities that play no role in schmausation? Why should we regard a causal link to something as a possible source of justified belief about it, when a merely schmausal link is of no use from the point of view of justification? If the concept of causation really is parochial in this way, it doesn't seem to be suited to playing the epistemological role normally associated to it. Perhaps we should introduce a new and more general concept to play that role.

2.3.2 The metaphysical necessity of the laws governing composition

—The laws of composition—laws of the form 'if some particles are arranged in such-and-such way, they compose something'—are not merely nomologically necessary, but *metaphysically* necessary as well. So any object composed entirely of causally efficacious things is itself causally efficacious. The claim that *xx* are arranged in a certain way is *metaphysically necessarily equivalent* to the claim that the thing that *xx* compose has a certain property. So if something happens because *xx* are arranged a certain way, it also happens because the thing that *xx* compose has that property. So the thing *xx* compose is causally efficacious, provided that *xx* themselves are causally efficacious.

—If laws of composition are metaphysically necessary, they aren't very similar to the canonical examples of truths that are metaphysically necessary without being knowable for certain a priori. These canonical examples fall into two categories. The first category includes such sentences as 'Hesperus is Phosphorus' and 'Whatever is made of water is made of H_2O molecules'. These are claims that would be counted as *analytic* on a fairly coarse-grained conception of meaning according to which 'Hesperus' and 'Phosphorus', and 'is made of water' and 'is made of H_2O molecules', counted as synonyms. But laws of mereology don't have this status. Even in the coarse-grained sense of 'analysis', there is no plausible way to find a hidden inconsistency in the conjunction of a microphysical description of the arrangement of some particles with the claim that they don't compose anything. The second category includes sentences like 'If Socrates exists, Socrates is human': the metaphysical necessity of these claims is supposed to follow from facts about the essences of the entities mentioned in the

sentence. But laws of composition aren't claims about what anything has to be like in order to exist.[11]

But suppose that the status of metaphysical necessity is not subject to these constraints of logical form, so that it could for all we know be had by laws of composition. In that case, metaphysically necessary equivalence is too weak a relation between propositions to guarantee the sharing of causal powers by the things the propositions are about. For if it makes sense to suppose that laws of composition are metaphysically necessary, then it also makes sense to suppose that the law of Mirror World (that every particle has a reflection) is metaphysically necessary. But in describing Mirror World I said nothing about whether the law that every particle has a unique reflection was metaphysically necessary or contingent. Nevertheless, the description I gave was enough to entail that the reflections were epiphenomena. So they would be epiphenomena, even if the law was metaphysically necessary.

If the facts about metaphysical necessity were relevant to causation in the way assumed in the objection, and metaphysical necessity was not subject to constraints on logical form, the charge that a theory posits epiphenomena would be an easy one to escape. The proponents of the theory need only specify that the laws connecting the alleged epiphenomena to causally efficacious things are metaphysically as well

[11]It might be suggested that a law of composition is metaphysically necessary in virtue of the essences of fundamental microphysical properties and relations. For example, it might be essential to *electronhood* to be such that, if anything instantiates it and is among some things arranged tablewise, those things compose a table. But I can't see how anyone could find this sort of view remotely credible unless they held the "dispositional essentialist" view that laws of nature generally are metaphysically necessary, being built into the essences of the fundamental properties (Shoemaker 1980, 1998; Swoyer 1982; Ellis and Lierse 1994). And no-one who held *this* view could endorse the objection we are considering, unless they were prepared to deny that the reflections at Mirror World are epiphenomena, and to hold that epiphenomenalist substance dualism of the sort we have been considering is inconsistent.

as nomically necessary. But in fact, the version of substance dualism in which the behaviour of minds is determined with metaphysical necessity by the behaviour of physical things is no easier to believe than the version in which the determination is merely a law of nature. If anything, the mental things are even more thoroughly under the sway of the physical ones if the former theory is true.[12]

2.3.3 Gunk

—If all composite things are epiphenomenal, it follows that if everything is composite, everything is an epiphenomenon. But this is false. It is not unlikely that the actual world is a "gunk world", where everything is composite. We used to think that atoms were simple, but then we discovered that they were made up of electrons, protons and neutrons. We thought these were simple until we discovered quarks. Why shouldn't this sequence go on for ever?[13] But it is very unlikely that everything is epiphenomenal: *something* must have causal efficacy, else how does anything ever happen?

—It does indeed follow from my view that at a world where there is a relation that obeys the laws of Universalism, such that everything bears the relation to something else, and such that all the facts about the a thing are nomically determined by the facts about the things that "compose" it, everything is an epiphenomenon. This verdict seems correct: it is confirmed by considering other sorts of worlds with 'infinite

[12]In my scepticism about the power of metaphysical necessity to put theories beyond the reach of epistemological worries, I am influenced by Hartry Field—see especially the response to to Lewis (1986b, pp. 111–12) in Field 1988, pp. 233ff..

[13]Sider 1993 gives an argument that has some similarity to this one, although he is arguing only for the *possibility* of gunk, which I am not denying. (But see footnote 19 below.)

descending sequences' of levels. For example, we could consider a variant of Mirror World with infinitely many planes, such that the facts about plane $n+1$ nomically determine the facts about plane n, while the facts about plane n don't even come close to nomically determining the facts about plane $n+1$. Or imagine a world with infinitely many Malebranchean gods; all the doings of god n are directly due to the intervention of the much wiser and more powerful god $n+1$. These are bizarre possibilities, but it seems to me that the right thing to say about them is that they contain no objects that are causally efficacious in the ordinary sense. Things happen, but they happen without being caused by the doings of causally efficacious things.

The hypothesis that the actual world is anything like this does indeed seem very implausible. Positing a whole lot of epiphenomenal things is bad enough; positing infinite hierarchies of them, and leaving out all the causally efficacious things, is certainly worse. I am happy to draw the conclusion that it is very unlikely that the world contains only composite things.

But what about the inductive argument for the view that it is not unlikely that everything is composite? It is without force. We did not—or so I say—"discover" that atoms were made of protons, or that protons were made of quarks. What really happened was that we got empirical reason to believe first that there were no atoms, and then that there were no protons. We shouldn't be confident that we have reached the end of this sequence of discoveries; that's why we shouldn't be confident that there are any quarks, as opposed to things arranged quarkwise. But it would obviously be absurd to argue inductively that since atoms turned out not to exist, and protons turned out not to exist, perhaps nothing at all exists. It's perfectly obvious that all sorts of things do exist. Things are interesting and various, and they can't be

interesting and various without being numerous.[14]

Before leaving this topic, I should point out that my argument does nothing to reduce the likelihood that we live at an amazing world where—loosely speaking, i.e. speaking according to the mereological fiction—each electron turns out to be a miniature galaxy, with its own little electrons, each with its own galaxy.... 'Whenever there are some things arranged galaxy-wise, there are some things which are among them, and whose volume is only a tiny fraction of theirs, which are also arranged galaxy-wise' is perfectly consistent, and consistent with the claim that there no composite things. It could be true, for example, if the only things there were were spacetime points.

2.4 Why there are no composite things

I conclude that composite things, if there are any, are epiphenomenal. But there is no good reason to believe that there are epiphenomenal things. So there is no good reason to believe that there are composite things.

In section 1.4 I considered various epistemological standpoints from which it might

[14]If physics were reformulated in such a way as to eliminate assumptions about mereology in the manner described in section 1.4.2, it would contain claims like 'whenever there are some things arranged atomwise, there are some things among them arranged protonwise, and some other things among them arranged electronwise', and 'whenever there are some things arranged protonwise, some of them are arranged quarkwise, and some others of them are also arranged quarkwise, and the rest of them are also arranged quarkwise.' From these claims it follows that whenever some things are arranged atomwise or protonwise, there are more than one of them. The conclusion 'whenever there are some things, there are more than one of them' is obviously not supported by these discoveries, since it is logically inconsistent. (Proof: take any thing; the things identical to it are one in number, since whenever x is one of them and y is one of them, $x = y$.)

appear that there *was* good reason to believe in composite things. If all the composite things there are are epiphenomena, all these attempts to justify belief in them face severe difficulties.

(i) There seems to be a deep incoherence in the idea that we might have *direct perceptual evidence* for the existence of epiphenomenal things. Surely for a person to *see* an object, there must be some sort of causal connection between the object and the person. So if composite things are epiphenomena, we can't see them or have any other sort of perceptual access to them. Moreover, once we find out that composite things are epiphenomena, we can no longer justifiably believe in them on perceptual grounds. Even if it still *looks to be the case* that there are composite objects, this gives us at best a defeasible reason to believe that they exist; and surely a sound philosophical argument showing that there was no causal connection between our visual experiences and the composite objects they purport to represent would be an example of a defeater.[15]

(ii) It isn't incoherent, in the same way, to suppose that we might have good scientific reason for believing in epiphenomenal things. Some philosophers think that we have good scientific reason for believing in numbers and sets, even though these entities are epiphenomenal. Nevertheless, it ought to be uncontroversial that the postulation of realms of epiphenomenal things is a major theoretical vice: other things being roughly equal, theories that do not posit them should be preferred. And as we

[15]According to many popular theories of mental representation, causal relations to the environment are of fundamental importance in fixing the contents of mental states, especially perceptual experiences. So if composite objects are epiphenomena, the contents of our perceptual experiences don't really entail that composite objects exist. It doesn't really *look to be the case* that there are composite objects after all: if we come to believe in composite objects on the basis of visual experience, we are extrapolating, going beyond what is really represented in the experience.

have seen in section 1.4.2, it is not particularly hard to construct scientific theories that are at least roughly as good in other respects as our current best theories, but do without composite things. I can't see how any of the respects in which these theories might be thought to be inferior to our present theories could be significant enough to outweigh the consideration that they do not, whereas our present theories (taken at face value) do, entail the existence of lots of epiphenomenal entities.

(iii) It's surely a good maxim of inductive reasoning that one ought not believe in epiphenomenal things unless one has positive empirical reason to do so. Thus, the belief that there are no epiphenomenal things (or at any rate very few of them) is a default reasonable belief. So once we realise that composite things would be epiphenomenal, if there were any, and realise that we have no empirical evidence for their existence, we should come to believe that there are no (or at any rate very few) composite things.[16]

Even if you don't accept that the belief that there are no epiphenomena is a default reasonable belief, surely you should agree that in the absence of relevant evidence, one should not have any firm opinions as regards the question how many epiphenomenal things there are and what they are like. But among all the different ways in which a realm of epiphenomena might be structured, those in which the structure looks

[16]Someone might hold that while the belief that there are no epiphenomena was a default reasonable belief, folk mereology (or some other theory contradicting Nihilism) was *also* a default reasonable belief. Then from these two default reasonable beliefs, it could be inferred that microphysics is not nomologically complete: so this would also be a default reasonable belief. But surely this isn't something we have any reason to believe in the absence of relevant evidence. And even if it were, we seem to have very good evidence that microphysics *is* nomologically complete, so our evidence *defeats* our default reasonable belief. In principle, I suppose one could respond by holding onto the belief in folk mereology and giving up the belief that there are no epiphenomena—but that would be perverse.

anything like a mereology are in the minority. So even if one insists on maintaining an open mind about the existence of epiphenomena, still, if one is guided by default reasonable belief, one ought to believe to a low degree that there are composite things.

(iv) If the argument that composite things are epiphenomena is sound, it would be feeble to try to justify one's belief in composite things on the grounds that it is the view of common sense, the view one started out with, the view one finds natural, the view one intuits to be true. All these authorities also say that there are no epiphenomena, or at least none that we can know anything about, and that many composite things can be found out about by normal methods. So something has to be given up. But surely we shouldn't give up the claim that we can't know much about epiphenomena: it's just obvious that the methods we actually use for finding out about material objects are no good for finding out about epiphenomena. So on reflection, it's the claim that there are composite things that has to go.

None of the strategies works: the reasonable thing to do is to embrace the natural conclusions that there are probably no epiphenomena at all, and if there are any epiphenomena, they probably don't have the sort of structure described in folk mereology or any other theory of composition. If that's what the world is like, then there is no relation that does what a relation would have to do to be expressed by the predicate 'is a proper part of'. The predicate 'is a proper part of' should be regarded as a *fictional predicate*, tied to the mereological fiction that governs our ordinary practice of talking about material objects in the same way that the predicate 'is a snark' is tied to the poem *The Hunting of the Snark*, and in a similar way to the way fictional names like 'Sherlock Holmes' are tied to the fictions in which they are introduced. Like 'is a snark', 'is a proper part of' is *semantically defective*: these predicates stand to ordinary predicates as empty names like 'Sherlock Holmes' stand to names that

have referents. We can put this by saying that 'is a snark' fails to express a property; 'is a proper part of' fails to express a relation.[17]

Strictly and literally speaking, then, nothing is a proper part of anything, for the same reason that there are no snarks.[18] If a simple thing is defined as a thing that has no proper parts, then everything is simple.

Kripke (1972, pp. 156–158) argues that when a sentence S is untrue because some of the predicates it contains fail to express properties, the sentence ⌜It is possible that S⌝ is untrue for the same reason. It isn't even *possible* for there to be unicorns, or snarks, or phlogiston, or people who have the same hat size as Sherlock Holmes. If this view is correct—and it seems plausible enough to me—then it will follow from the failure of 'is a proper part of' to express any relation that it is *impossible* for there to be anything that is a proper part of anything else, and hence that it is *necessary* that everything is simple.[19]

[17]I am concentrating on the predicate 'is a *proper* part of' in order to accommodate the view that 'is part of' (as used by philosophers) is *defined* as 'is a proper part of, or is identical to'. If this view is correct, 'is part of' isn't a *wholly* semantically defective predicate. Just as, if we defined "snelephants" as things that are either snarks or elephants, 'all elephants are snelephants' would be strictly and literally true, so 'everything is part of itself' is strictly and literally true, if it's definitionally equivalent to 'everything is a proper part of, or identical to, itself'.

[18]That's what I say; but there is a dispute amongst philosophers of language over what should be said about the (strict and literal) truth-value of sentences containing semantically defective predicates (see section 3.1 below). One radical thought is that every sentence containing such predicates must be truth-valueless. But even if the property shared by 'Nothing is part of anything', 'There are no snarks' and 'There is no phlogiston' is not truth, it is clearly a property that makes it appropriate to assert the sentences that have it even in otherwise strict philosophical contexts.

[19]If it is impossible for there to be something which is part of something else, then it is impossible for there to be gunk. Some people (e.g. Lewis 1991, p. 70; Sider 1993) have objected to this claim, on the grounds that it is obvious that gunk is possible. (I'm not sure I understand why this objection should be considered better than the simpler argument from the premise 'It is obvious that composite things are possible';

2.5 Is the claim that there are composite things inconsistent?

I don't see any reason to deny the *possibility* of epiphenomena. It is possible that there are epiphenomenal things whose pattern of relations to one another and to particles is exactly the sort of pattern described by folk mereology, or any other mereological theory. But this doesn't just seem *unlikely*—it seems deeply irrelevant. If there happened to be such a realm of epiphenomena, this wouldn't in any significant way be a vindication of the strict and literal truth of our folk theory of the world.

Suppose that, unlikely though it may be, there really are a lot of epiphenomenal things, related to microphysical things by a relation obeying laws structurally isomorphic to the principles of folk mereology. Would those things really be the ordinary middle-sized objects of common sense? Could an epiphenomenon deserve to be called a plate, a rock, a planet or a plant? I think not. A central part of our idea of any of these kinds of thing is that they are involved in certain patterns of causal relations—in particular causal relations with us and our uses of the words 'plate', 'rock', 'planet'

---

but let that pass.) The intuition that gunk is possible can be broken up into two steps. First, there could be a binary relation that obeyed some appropriate mereological laws—such as the laws of Mereology—and such that everything bore it to something else. Second, this relation could be parthood. The first step does indeed seem to me to be obvious. In the intended sense of 'possible', the mere logical consistency of the claim that there is gunk should be enough to establish that it is possible that there is *some* relation that satisfies it. The second step is the one I disagree with, at least when I am in a mood to accept the Kripkean view of semantically defective predicates. Since (assuming the Kripkean view) 'gunk is possible' can be true only if there is *actually* such a relation as parthood, it doesn't seem to be the sort of thing that could be established just by "conceivability" evidence. The claim that gunk is possible is on a par with the claim that it is possible for there to be an infinite series of snarks: it is certainly possible for there to be an infinite series of *somethings*, but that they could be *snarks* is just not the sort of thing that one can reasonably insist on on grounds of conceivability.

and 'plant'. So (strictly and literally speaking) there are no such things, even in the unlikely event that the world happens to be structured as folk mereology says it is. Thus, we should be *more* confident that there are no plates, rocks, planets or plants than that there are no epiphenomenal things. If there are no epiphenomenal things, there are no composite things and hence no plates, etc.;[20] but if there are epiphenomenal things, there are *still* no plates, etc.

A similar claim can plausibly be made about the predicate 'is part of'. A relation governed by laws structurally isomorphic to parthood as represented in folk mereology, but such that anything that bore it to anything else was an epiphenomenon, would not, I think, be parthood. This is clearest on a causal theory of reference, according to which 'is part of' expresses the relation, if there is one, that has the right sort of causal relation to our use of mereological vocabulary. Whatever exactly the "right" sort of causal relation is, surely it isn't one that a relation which never held between causally efficacious things could stand in. Things aren't quite so clear if we adopt a more "descriptivist" approach to the fixing of the semantic value of 'is part of'. But even on such an approach, it seems plausible that a central component of the role which a relation would have to play to be expressed by 'is a proper part of' would involve certain causal relations. For example, we might think that no admissible interpretation could make the sentence 'There is an epiphenomenon that is composed

[20]It might be suggested that plates and so forth are all *simples*—the things we normally think of as the proper parts of plates are not really parts of them at all, but merely spatially contained within them. But this is of no help: the argument that all composite things are epiphenomena is really just a special case of an argument for the conclusion that all non-microphysical things are epiphenomena. Certainly plates aren't microphysical things: none of the entities that need to be mentioned in explanations of microphysical phenomena could with any plausibility at all be thought to be a plate, since they're all much too small.

entirely of things that are not epiphenomena' come out true.

So the conclusion that everything is simple cannot be avoided just by embracing the view that there are lots of epiphenomenal things, and an interesting relation that each of these things bears to some microphysical things. Even if such a relation existed, it wouldn't be parthood—'parthood' should be regarded in more or less the same way that we regard the names of fictional characters who bear a coincidental resemblance to certain actual people.

The only way that's left open for the believer in composite objects is to give up the premise that microphysics is nomologically complete. If it isn't, then the arguments of sections 2.1 and 2.2 fail; we can have non-microphysical things that are not epiphenomena. Would *that*, at least, be a situation in which 'parthood' expressed a relation that things sometimes bear to other things? I'm not so sure. As I have had occasion to point out (see p. 57 above), a world in which some composite things were causally efficacious, thanks to the failure of microphysics to be nomologically complete, would be a very strange world indeed. The only way for a composite thing to be involved in causal relations is by having some *radically emergent* property or relation for the non-microphysical laws to get a grip on. For example, a person might be causally relevant to the movement of some particles in her pineal gland, if she had the radically emergent property of *deciding to move one's arm*, and if there was a special law correlating the possession of this property by a composite object with abnormal movements of the particles in their pineal glands. So even if composite things are sometimes causally efficacious thanks to their radically emergent properties, it is much harder for them to be causally efficacious than we normally suppose. We normally assume that it is enough for a composite thing to be causally relevant in a situation if it is composed of some parts all of which are causally relevant in the

situation. We don't think we need to do anything more to justify the claim that the stone is causally efficacious, apart from telling the microphysical story about the motions of the stone's component particles and their combined effect on the motions of the particles that compose the glass.

I am tempted to conclude from this that the entities at such a world would not really be *composite objects* at all. A *genuine* composite object would automatically inherit the causal powers of its parts; it wouldn't have to depend on radically emergent properties and relations for its causal efficacy. So *no matter what the world is like*, there are no genuine composite objects. The very concept of a composite object imposes inconsistent demands that nothing could satisfy. Nihilism is analytically true!

I put this argument forth in a tentative spirit; nothing else I have to say depends on its being sound. And even if it were sound, the most it could show is that it is inconsistent for there to be composite objects *in the full sense of the term*; even if nothing could *perfectly* fulfil the "composite object" role, it is clearly conceivable that there might be some entities that behaved enough like composite objects to make it natural to classify them as composite objects in a looser sense. That said, it is worth trying to state the argument a bit more rigorously. Let us introduce the expression 'yy r-compose x' to abbreviate 'each of yy bears r to x, and whenever z bears r to x, there is some z' such that z' bears r both to z and to one of yy'. (This is isomorphic to the definition of 'compose' in terms of 'part': so 'parthood-compose' is equivalent to 'compose'.) Then we can set up the argument as follows:

(1) It is analytically true that for any relation r, if r = parthood, then (it is necessary that for any x and yy, if yy r-compose x, and none of yy is an

epiphenomenon, then x is not an epiphenomenon).

(2) It is analytically true that there is no relation r such that it is necessary that for any x and yy, if yy r-compose x, and none of yy is an epiphenomenon, then x is not an epiphenomenon.

(3) Therefore, it is analytically true that there is no such relation as parthood.

(4) Therefore, it is analytically true that nothing has any proper parts.

The real question about this argument is whether there is any sense of 'necessary' on which both (1) and (2) come out true. If 'necessary' just means 'nomologically necessary' then (2) seems to be false. If we can imagine a world where composite things are causally efficacious thanks to radically emergent properties and relations, we can imagine that the laws of this world might be set up in such a way that it is nomologically impossible for there to be any epiphenomena. (2) looks much better, and (1) looks equally good, if 'necessary' means 'metaphysically necessary'. I suppose someone *might* claim that it could be *essential* to a relation r, and hence metaphysically necessary, that if r exists, the laws are such as to rule out epiphenomena.[21] If this strange doctrine were correct, we would have to find some even stronger sense of 'necessary' on which both (1) and (2) were true. I am inclined to think that this can be done.

[21] According to "dispositional essentialists" like Shoemaker (1980, 1998), it is *always* essential to properties and relations that they stand in the nomological or dispositional relationships that they in fact stand in. But something even stronger than this would be required for it to be essential to a relation r that if r exists, no r-composite things are epiphenomena. If 'no r-composite things are epiphenomena' is a law of nature, it doesn't seem to be the sort of law that can be captured as a set of nomological or dispositional relationships involving r.

2.6 People

One thing we know about people is that if there are any of them, they are not microphysical things. So by the argument of section 2.2, if there are any people, they are all epiphenomena. But if I exist, I am a person. So if I exist, then my impression that certain facts about me—for instance, my choices—make a difference to events outside myself is an illusion.

Those who have followed me thus far will differ in their reactions to this point in a way that seems beyond the power of any philosophical argument to reconcile. For my part, I am convinced that if I existed, I would have to be causally efficacious. If there are epiphenomena, then they are of no importance to me; whereas if I exist, then I am of great importance to myself. But if I exist, then I am an epiphenomenon. Hence I do not (strictly and literally speaking) exist. The activities I used to ascribe to myself—thinking these thoughts, feeling these feelings, and so forth—are not performed by any single thing: rather, they are jointly performed by those things I used to call 'the particles that compose me'. If I want to say things that are strictly and literally true, I had better stop using this word 'I' all the time, since it doesn't refer to anything. Instead we—we particles, that is—should say 'we'.[22]

[22]The question whether we are justified in having the belief we express using 'I exist' seems more or less irrelevant as far as the epistemological status of folk mereology (or for that matter, the belief that there are composite objects) is concerned. Suppose that from my own case I can be certain that there is at least one person, who is not a microphysical thing. Suppose that the problem of other minds is solved somehow, so that I can come reasonably to believe in other people as well. Still, the theory that anything remotely like folk mereology is true seems to go far beyond what I could have any reason to believe about the structure of the world. Even if there is a distinctive relation between people and certain particles, there are many ways in which such a relation could be governed by basic laws quite different from those described in folk mereology. Nor is there any reason that I can see to think that the distinctive relation

Van Inwagen (1990, section 12) claims that this is incoherent, on the grounds that it is impossible for several things jointly to think, feel, or have experiences. I don't understand why he says this. When I consider the hypothesis that the particles that I normally think of as "the particles that compose David Lewis", say, fail to compose anything—and surely van Inwagen must recognise that this hypothesis is a coherent one—I find descriptions like 'the particles believe in a plurality of worlds' extremely apt for describing various distinctive facts about the particles' arrangement. One is free to insist that this is just ungrammatical, or that it conflicts with the meaning of verbs like 'believe'. But how could this be a truth of deep metaphysical or epistemological significance, as opposed to a shallow fact about the way our language works, to which one might respond by advocating a minor linguistic reform?

If it *is* coherent to think that some things jointly think, feel or experience, the idea that you can be justified immediately and beyond all doubt in believing in your own existence seems to be in trouble. If you were informed that most of the world's thinking and feeling was done jointly by particles which didn't compose anything, wouldn't it then be reasonable to have doubts? What if you were informed that there

---

must be *parthood*.

If it seems harder to believe that you are a simple, non-microphysical thing than that you are a composite thing with microphysical things as parts, this is presumably because it seems that we *already* have good reason to believe in composite things, whereas we don't have any reason to believe in non-microphysical simples. But if I'm right, then we don't have any more reason to believe in the former than in the latter; both are equally mysterious; both go beyond anything we have reason to believe on ordinary empirical grounds. In fact the question whether you are complex is just the question whether any relation that connects you to microphysical things does whatever a relation would have to do to merit the name 'parthood'. Since it's presumably more likely that an arbitrary relation about which you know nothing else fails to do these things than that it does do them, even if you are sure that you exist, you should still regard it as more likely that you're simple than that you're complex.

were some particles who were currently having exactly the same thoughts and feelings (qualitatively speaking) as you, right down to wondering whether they composed something? At some point, I think, you would have to give up the belief that you exist. But if so, then shouldn't you also give up the belief when you learn that if it is true, you are an epiphenomenon?[23]

[23]Much of the material in this section parallels section 6 of Dorr and Rosen 2001.

Chapter 3

Questions for the mereological fictionalist

3.1 A Kripkean challenge

In this chapter, I will try to give a more fleshed-out version of the fictionalist theory of ordinary talk about composite objects which I began sketching in chapter 1. Recall that in section 1.2 I suggested that a theory of this practice might take the form of an analysis—or some other sort of explanation—of an operator $\bigcirc$, concerning which one makes the following claim: a sentence φ is correctly asserted (in ordinary contexts) iff the sentence $\ulcorner\bigcirc\varphi\urcorner$ is strictly and literally true. I suggested, furthermore, that $\ulcorner\bigcirc\varphi\urcorner$ might be analysed as $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$, for some specific sentence ψ which logically entailed folk mereology. Let's call this theory (or theory-schema) the *counterfactual theory of correctness*. In this section I will discuss an important challenge to the counterfactual theory of correctness, arising from some of Kripke's doctrines in the philosophy of language.

I have been arguing for the claim that the only material objects there are are microphysical ones. If this is even close to being correct, then our language must

have a lot of proper names that don't have referents—*empty* names. For example, the proper name 'Mars' (as used in discussions of astronomy) is an empty name, since it surely refers to a non-microphysical material object if it refers to anything at all.[1] It belongs in the same category as 'Vulcan' (a name introduced by nineteenth-century astronomers for a postulated planet orbiting between the Sun and Mercury) and the name 'Mars' as used in discussions of Roman mythology.

Furthermore, in section 2.4, I argued that the predicate 'is a proper part of' is "semantically defective" in much the same way as the predicate 'is a unicorn'. If this is true, then probably many other common predicates are semantically defective as well.[2]

There is much debate among philosophers of language about the truth-values of sentences involving empty names and semantically defective predicates. I take it to be uncontroversial is that *simple, positive* sentences involving empty names and semantically defective predicates are never strictly and literally true. I won't bother

[1]This isn't wholly uncontroversial. Salmon (1998) holds that many proper names which are generally assumed to be empty—names like 'Sherlock Holmes' and 'Vulcan'—really refer to abstract objects. It would be convenient for me if this were true; unfortunately, I find the view completely incredible.

[2]Obviously other mereological predicates, like 'is part of', 'overlaps', and 'compose', are liable to inherit the defectiveness of 'is a proper part of'—although, if we want to say that it is analytic that everything is part of, and overlaps, itself, and that the things identical to any thing compose it, we will want to say that these predicates aren't *fully* semantically defective in the same way that 'is a proper part of' is (see footnote 17 of chapter 2). But perhaps the phenomenon of defectiveness goes further still, extending to predicates like 'is a cow', 'is red', 'is an oxygen atom', 'is a person'.... I have already argued that there aren't any cows, red things, oxygen atoms, or people: everything there is is much too small. It is fairly plausible that predicates with empty extensions are always semantically defective, unless they are explicitly defined in terms of, or syntactically built up from, predicates whose extensions are not empty. If this is even close to being true, the vast majority of English singular predicates are semantically defective.

making this notion of a simple, positive sentence precise: I have in mind sentences like 'Vulcan is a planet', 'Vulcan orbits the Sun', 'Vulcan exists', 'there is a unicorn in the room', 'there are unicorns'. The problem for the counterfactual theory comes from a further thesis, suggested by some claims of Kripke's:

(1) When φ is a simple, positive sentence involving an empty proper name or a semantically defective predicate, no sentence of the form $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$ is non-vacuously true.

If (1) were true, we would have to give up the counterfactual theory of correctness. For there are many simple, positive sentences involving empty names and semantically defective predicates—sentences like 'Mars is a planet', 'Mars orbits the Sun', 'Mars exists', 'Mars has some proper parts', 'Something has some proper parts'—such that it is correct to assert them, and incorrect to assert their negations.

Why believe (1)? One possible argument is based on the claim

(2) When φ is a simple, positive sentence involving an empty proper name or a semantically defective predicate, the sentence $\ulcorner$It is metaphysically possible that $\varphi\urcorner$ is not true.

(2) is defended—at least as far as 'Sherlock Holmes' and 'is a unicorn' are concerned—in the addenda to *Naming and Necessity*. Although Kripke characterised this as a 'surprising' claim and said that 'it doesn't ever convince anyone' (Kripke 1972, p. 23), by now it is the orthodox view. Unfortunately, to get from (2) to (1), one would have to rely on the following principle:

(3) For the counterfactual $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$ to be non-vacuously true, $\ulcorner$It is metaphysically possible that $\varphi\urcorner$ must also be true.

But this claim must be rejected, as I have already argued (see p. 49 above). At least, there is *one* good sense of the counterfactual for which (3) fails: and that's all that we need for the purposes of the counterfactual theory of correctness.

But perhaps there is some better reason to believe (1). For one thing, there may be some sense of possibility, weaker than metaphysical possibility, such that (2) and (3) are both true when 'it is metaphysically possible that' is replaced by an operator expressing the weaker sort of possibility.[3] Even if there is no such sense of possibility, there is another way to argue for (1) that is worth considering. A common diagnosis of the problem with simple sentences involving empty names or semantically defective predicates is that they *fail to express propositions*.[4] It is because of this that they are not true: for a sentence to be true, it must express a true proposition. But if these sentences fail to express propositions, there is some pressure to say that sentences which embed these sentences as the consequents of counterfactual conditionals also fail to express propositions. Much of the point of talking about propositions in semantics would be lost if we couldn't specify the semantic values of a wide range of operators and sentential connectives as functions from (sequences of) propositions to propositions.

If you follows this argument where it leads, you end up having to say some very strange things. For whatever reason there is to think that the semantics of the counterfactual conditional is given by a function from pairs of propositions to propositions, there is at least as much reason to think that the semantics of the operator 'It is not

[3]Nolan (1997) argues quite convincingly that there is no such sense of possibility: every sentence whatsoever features in the consequent of some true counterfactuals and some false ones.

[4]This might be motivated, in the case of empty names, by the Millian view that the semantic contribution of a proper name is exhausted by its referent. But many neo-Fregeans, such as Evans (1982), hold what amounts to the same view.

the case that' is given by a function from propositions to propositions. So you will end up saying that sentences like 'It is not the case that Vulcan is a planet' and 'it is not the case that there are unicorns' fail to express propositions, and hence are not true. This conclusion is so bizarre that it seems reasonable to take it as a *reductio* of the assumptions that lead to it. However, some philosophers with Millian sympathies have proposed strategies for blocking the bizarre conclusion which do not block the corresponding argument in the case of the counterfactual conditional.[5] These strategies have the common feature that they treat all empty names as equivalent, so that the substitution of one empty name for another can make no difference to the truth value of a sentence. Likewise, all semantically defective predicates are equivalent, provided they are in the same syntactic category (e.g. they have the same number of argument places). If this were true, we would still have to drop the counterfactual theory of correctness. For clearly it is correct to say 'Mars is a planet' but not correct to say 'The Eiffel Tower is a planet'; it is correct to say 'my leg is a proper part of me', and not correct to say 'my leg is more phlogisticated than me' or 'my leg has me as a proper part'.

I'm not convinced that these objections show that the counterfactual theory needs to be given up. It seems to me that we can make perfectly good, non-trivial sense of counterfactuals involving empty names and semantically defective predicates. For example, sentences like 'if Vulcan had been a planet, Vulcan would have been a heavenly body', 'if Vulcan had existed, Vulcan would have been a planet' and 'if there had been a unicorn in my living room, there would have been a unicorn in my house' strike me as non-vacuously true. The project of coming up with a comprehensive semantic theory for the language which explains how these sentences get to be true

[5]See Braun 1993, Fitch 1993, and the last section of Salmon (1998).

does indeed face formidable difficulties: what kind of thing shall we assign to 'Vulcan' as a "semantic value", given that we can't assign it a planet? But these difficulties apply to every interesting sentence involving empty names or semantically defective predicates. Only if one is prepared to embark on a wholesale revision of one's intuitions about all these sentences—including propositional attitude sentences ('Leverrier believed that Vulcan was a planet'), "intensional transitives" ('Astronomers sought Vulcan for years'), and indicative conditionals ('If Vulcan exists, astronomers must be conspiring to hide its existence from the general public')—is there any reason not to stand by one's intuitions about counterfactual conditionals.

If I'm wrong that counterfactuals can be strictly and literally true even when they contain empty names or semantically defective predicates, we need to look for a successor to the counterfactual theory of correctness. A good place to start looking is among those sorts of sentences for which it's most intuitively obvious that they can be strictly and literally true even if they contain empty names and semantically defective predicates. Sentences like 'According to the Sherlock Holmes stories, Sherlock Holmes lives in London' and 'According to *The Hunting of the Snark*, some snarks are Boojums' are among the clearest examples of all.[6] Inspired by these precedents,

[6]Many philosophers follow Kripke (MS) and van Inwagen (1977) in holding that the name 'Sherlock Holmes' is not *always* an empty name: it has one sense on which it refers to a special sort of abstract object, a so-called "fictional character". The name is supposed to occur in this sense in sentences like 'Sherlock Holmes is among the most vividly realised fictional detectives.' Very few philosophers, however, have been willing to agree with Salmon's claim that the name refers to an abstract object even in the sentence 'According to the Sherlock Holmes stories, Sherlock Holmes lives in London', so that the stories falsely say of a certain abstract object that it lives in London (Salmon 1998).

It's not entirely clear to me how these claims about the metaphysics of fictional characters are supposed to carry over to the case of fictional predicates. Is there some sense of 'snark' on which it is true that there are snarks—a sense which occurs,

we might try taking the pronunciation of $\bigcirc$ as 'according to the mereological fiction', which I allowed in section 1.2 only as a manner of speaking, at face value. We would then face the task of saying just what sort of entity the mereological fiction is supposed to be. We would get strange results if we took it to be a contingent entity, whose existence and properties depend on us and our practices, in the same way that the existence and properties of the Holmes stories and *The Hunting of the Snark* are often held to depend on contingent facts about Conan Doyle and Carroll. Surely we want to say that even if everyone had always been a convinced and literal-minded Nihilist, it would still have been true according to the mereological fiction that there were composite objects. We will do a bit better if we take the mereological fiction to be something whose existence and content are necessary. But even then, it seems to make sense to ask how things would have been if it hadn't existed, or had had different properties: and it should be true that for any one entity, if that entity hadn't existed, it would still have been the case that $\bigcirc$ there are composite things.

Thus, the 'according to' construction that English provides with is not quite adequate for our purposes. What we want is something that is like the counterfactual conditional in that it makes a sentence out of two sentences, rather than making a sentence out of a singular term and a sentence like 'according to'; but which shares the ability of 'according to' to generate true sentences containing empty names and semantically defective predicates. I can't find any conditional in the language that satisfies these desiderata more clearly than the counterfactual conditional itself does; but I don't see why we shouldn't be able to introduce one. We could stipulate that $\ulcorner \psi \rightarrowtail \varphi \urcorner$ is to be understood as a sentence just like $\ulcorner$According to the fiction that

perhaps, in the sentence 'Snarks are among the least vividly realised of Carroll's fictional creatures'? If so, how many of them are there?

ψ, $\varphi\urcorner$, except that unlike the latter sentence it does not entail $\ulcorner$The fiction that ψ exists$\urcorner$. It can hardly be claimed to be beyond our conceptual powers to understand a conditional that works like this: all it does is combine some features which are already had separately by other expressions which we already do understand.

None of this will satisfy hard-liners who insist that no sentences involving empty names can ever be true, or (more plausibly) that the substitution of one empty name for another, or of one semantically defective predicate for another of the same grammatical category, can never make the difference between a true sentence and a false one. But for this view to deserve to be taken seriously at all, it must be combined with the claim that interesting sentences involving empty names, even if they are never *true*, have some other property which makes it correct to assert them in ordinary contexts. So even if the hard-line view is right, we really have nothing to worry about: whatever account the hard-liner gives of this property, we can build into our theory about the correctness of ordinary assertions about material objects. Instead of saying that φ is correctly asserted iff $\ulcorner\bigcirc\varphi\urcorner$ is *strictly and literally true*, we will now say that φ is correctly asserted iff $\ulcorner\bigcirc\varphi\urcorner$ is F, where 'F' stands for this other property.

Thus, the counterfactual theory of correctness has little to fear from the objection based on (1): the case for (1) is not very strong, and even if (1) were true, the problem could easily be overcome by making minor modifications to the theory. However, the following weaker and more intuitive version of (1) poses a more powerful challenge to the counterfactual theory:

(4) When φ is a simple, positive sentence involving an empty proper name, *and ψ is a sentence in which this name does not occur*, no sentence of the form $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$

is non-vacuously true.

If even this much were true, we would still have to give up the counterfactual theory. For however we might specify the sentence ψ that occurs as the antecedent of the counterfactual, we cannot hope to include every single empty name in the language![7] (4) is more plausible than (1). It is still rather tempting to think that sentences like

(5) If there had been a planet orbiting closer to the Sun than Mercury, whose presence caused the variations in the orbit of Mercury that worried nineteenth-century astronomers, and which was referred to as 'Vulcan' after it had been discovered, Vulcan would have existed.

are non-vacuously true. But it is easier to convince yourself that (5) is not true than it is to convince yourself that 'If Vulcan had existed, Vulcan would have existed' is not true (or vacuously true). The way to do it is first to convince yourself that it is *possible* for something to satisfy the antecedent of (5) without being Vulcan, and then ask yourself why you shouldn't think that this is the possibility that would have been actual if the antecedent of (5) had been true.

This new problem for the counterfactual theory give us another reason to attend to the new conditional, explained in terms of the 'according to the fiction that. . . ,. . . ' operator, which I suggested introducing on p. 83 above. For there do seem to be cases in which empty names can be used to state truths about what is the case according to fictions in which they do not occur. Here are some examples:

[7] If there are lots and lots of predicates that are semantically defective even though they occur in simple, positive sentences that can correctly be asserted, we cannot hope to include all of them, either. But (4) would not be so plausible if it were extended to cover semantically defective predicates as well as empty names. It seems plausible that 'If there were succubi, there would be demons' is true, even though 'succubus' and 'demon' are both (arguably) semantically defective.

(6a) According to the fiction that there is a planet which lies beyond Pluto, is inhabited by monsters, and is generally referred to as 'Planet Z', Planet Z is a heavenly body.

(6b) According to the fiction that ancient Greek religion had its source in an encounter with a group of mischievous aliens who came to be worshipped as gods, Apollo was an alien.

Certainly we can't *always* use the language spoken within a fiction to make true reports of what is the case according to the fiction:

(6c) According to the fiction that people speak a language just like English, except that 'shrdlu' is a sentence that means the same as 'snow is white', shrdlu.

is not a true sentence, but a piece of nonsense. But proper names which are actually empty, but have referents according to a fiction, can apparently be imported with impunity from the world of the fiction into the language we use for talking about it.[8]

3.2 A metalinguistic version of the counterfactual theory

All things considered, it seems to me that the problems posed by empty names and semantically defective predicates for the counterfactual theory of correctness and its

[8]This phenomenon isn't limited to proper names. The sentence

(6d) According to the fiction that there are certain beings called 'Gorgons', one of whom bears the name 'Medusa', Medusa is a Gorgon.

also seems to be true. It's harder to find convincing examples of this sort of thing happening with verbs and adjectives, and as far as I can see it doesn't happen at all with conjunctions, quantifiers and the like.

close variants are not insuperable. Nevertheless, the foregoing discussion does vividly illustrate the sort of murkiness we have to deal with if we want to make essential use, in our theory of correctness, of the difficult reaches of language in which empty names and semantically defective predicates can make a non-trivial difference to truth-conditions.

By changing the form of the schematic theory of correctness we have been assuming up to now, we can greatly reduce our reliance on assumptions about the semantics of empty names and semantically defective predicates. Suppose that, instead of giving a theory of the form 'φ is correctly asserted iff $\ulcorner\bigcirc\varphi\urcorner$ is strictly and literally true', we instead give one of the form 'φ is correctly asserted iff $\bigcirc$ (φ is strictly and literally true).' The operator $\bigcirc$ could be analysed, as before, as $\ulcorner$If it were the case that ψ, it would be the case that...$\urcorner$ (or $\ulcorner\psi \rightarrowtail \ldots\urcorner$), where ψ is some sentence that logically entails folk mereology. This won't get rid of *all* our worries, since we are still using the semantically defective predicate 'is part of' in the antecedent. But we no longer need to use any empty names, or any semantically defective predicates other than those that occur explicitly in ψ. The only counterfactuals we ever need to evaluate concern the semantic properties sentences would have had if had been the case that ψ. These aren't the easiest of counterfactuals: whereas on the old approach, the assertability of the principles of folk mereology was a more or less trivial matter, on the new theory we have to translate between the language we would have spoken if it had been the case that ψ and the language we actually speak before we can tell what it is actually correct for us to assert. Still, if the right counterfactuals are true, what does it matter if they are a bit harder to evaluate than the counterfactuals used by the old theory?[9]

[9]Can this strategy be used to circumvent the problem of semantically defective

A more serious problem for this approach concerns the operator 'strictly and literally speaking'. On the natural understanding of this operator, the sentence ⌜Strictly and literally speaking, φ⌝ can correctly be asserted iff φ is strictly and literally true. 'Strictly and literally speaking' can govern parts of sentences as well as whole sentences. For example, if I am right that strictly and literally speaking, everything there is is microscopic, I can correctly say 'The only parts of the Eiffel Tower that strictly and literally exist are microscopic'. This is not at all hard to explain for the old kind of theory: the interaction of 'strictly and literally speaking' with the 'according to the fiction' operator seems to be just like the much-studied interaction between 'actually' and 'necessarily'. But on the new theory, it is much harder to see what could be going on. If there had been lots of composite objects, so that 'everything is microscopic' was strictly and literally false, surely 'strictly speaking, everything is microscopic'

---

predicates as well as the problem of empty names? One idea would be to replace the antecedent ψ,

> Whenever x is part of y and y is part of z, x is part of y, and ...

with its Ramsey sentence:

> There is a binary relation P such that (whenever x and y instantiate P and y and z instantiate P, x and z instantiate P, and ...)

Provided that we can explain why if there had been such a relation it would have been expressed by the English predicate 'is part of', this does the trick. But this quantification over relations makes me uncomfortable. I would prefer a statement of the principles of the fiction that could be embraced wholeheartedly by a Nominalist who denied the existence of properties and relations. To be honest, this is mostly because I am interested in developing a fictionalist account of property-talk along the same lines as the present account of talk about composite material objects (see chapter 6). If the predicate 'instantiates' itself is semantically defective—something we might well want to hold, if we hold that there aren't really any properties—we will have no choice but to use it in stating the principles of our fiction of properties. The Ramsifying strategy will be no help at all.

would also have been strictly and literally false. In our actual language, ⌜strictly and literally speaking, φ⌝ is strictly and literally true just in case φ is; surely this wouldn't have been any different in the language we would have spoken had there been lots of composite objects. It might be suggested that 'strictly and literally speaking' has the purely pragmatic function of forcing a shift in the context to one in which it is strict and literal truth in the *actual* world that matters for correctness. But it's not clear how this is compatible with the capacity of 'strictly and literally' to modify parts of sentences. The sentence 'The only parts of the Eiffel Tower that strictly and literally exist are microscopic' isn't strictly and literally true at the actual world (since the name 'The Eiffel Tower' is actually empty); moreover, it wouldn't have been strictly and literally true if there had been lots of composite objects, including a referent for the name 'The Eiffel Tower'. It's not easy to see how any theory of a metalinguistic sort could explain what makes it correct to assert this sentence.[10]

3.3 The principles of the mereological fiction

I have been very lazy up to now about actually specifying the "principles of the mereological fiction" that occur in the counterfactual theory of correctness and its variants. Counterfactual conditionals are flexible enough, and ordinary practice is

[10]One possible solution would be to add, as an extra conjunct to the antecedent of the counterfactual, that 'strictly and literally speaking' means something different from what it actually means, something closer to the actual meaning of 'If there hadn't been any composite objects'? (This solution bears a family resemblance to Kendall Walton's idea (1990, chapter 11) that when we speak of fictional characters existing and not existing, we speak within a fiction according to which 'exists' expresses a property that only some things have.) Perhaps this can be made to work, but it strikes me as rather awkward. For one thing, it seems to be correct to say ' "strictly and literally speaking" is quite different in meaning from "if there hadn't been any composite objects" ': we will somehow have to explain this away.

fluid enough, that there is no terribly urgent need for us to be precise or explicit in stating these principles. For instance, we could just analyse $\ulcorner\bigcirc\varphi\urcorner$ as $\ulcorner$If there were lots of composite objects, it would be the case that $\varphi\urcorner$. This analysis doesn't obviously give the wrong results. One could argue that that if there had been lots of composite objects, objects arranged chairwise would have composed chairs, piled-up plates would have composed chairs... and so on for all the other principles of folk mereology. In other words, one could argue that worlds where folk mereology is true are closer to the actual world than other kinds of worlds where there are lots of composite objects, e.g. worlds where only things that are disconnected and far-apart compose anything. Nevertheless, it is worth seeing whether we can find some a more precise and explicit way to articulate the principles of a mereological fiction, one which relies as little as possible on our ability to evaluate delicate questions about the truth-values of counterfactuals. Of course, in doing this we will be characterising a highly specific and idealised version of our messy ordinary practice. But this will put us in a better position to understand the range of different standards of correctness that are available, the considerations for and against adopting them, and the extent to which they play a role in governing our actual practice.

The idealised fiction has eight principles. The first three are just the axioms of Mereology:

(M1) Whenever there are some things, there is at least one thing that they compose.

(M2) Whenever there are some things, there is at most one thing that they compose.

(M3) Whenever x is part of y and y is part of z, x is part of z.

The other five principles make it explicit how the things that exist strictly and literally speaking are to be embedded in the world of the fiction. (Up to now I've been deliberately using this stilted phrase 'strictly and literally' so as to make it clear that it's a term of art, but it would be cumbersome to keep on doing so now. Let me instead start using 'really' to express the same meaning. 'x is real' abbreviates 'x really exists'; 'x is fictional' abbreviates 'x doesn't really exist'.[11])

(M4) Every real thing exists.

(M5) Every real thing is intrinsically just as it really is; any real things are intrinsically, and in their relations to one another, just as they really are.

(M6) All fictional things are characterless and exactly alike, except insofar as they have different parts and are parts of different things.

(M7) Everything overlaps some real things.

(M8) Every real thing is simple.

These principles are just about as explicit as they could possibly be. That doesn't mean, of course, that the task of saying what would be the case if they were true is a mechanical or algorithmic one, even given a detailed and exhaustive particle-by-particle description of the actual world. Supposing there are actually some particles arranged catwise, then if M1–M8 had been true, there would have been a cat; but to know this, one must know the meaning of the word 'cat'. However, the sort of judgment we are called upon to make when we ask whether there would have been

[11]For the moment, we can suppose 'really' to be synonymous with the familiar 'actually' operator of modal logic. I'm avoiding the word 'actually' because of a departure from the counterfactual analysis which I'll be proposing in section 3.4

cats if M1–M8 had been true is quite different in character from, and less prone to subjective variation than, the sort of judgment we have to make when we ask whether there would have been cats if there had been lots of composite objects. It is natural to think of the latter judgment as involving the question whether, among worlds where there are lots of composite objects, those that contain cats are closer to the actual world than those that don't contain cats. By contrast, the work of singling out a world is already done for us by M1–M8.[12] When we ask whether there would have been cats if they had been true, we are asking whether there are cats at *that* world. Differences in our understanding of 'cat' are relevant; but differences in our understanding of closeness are not.[13] Because of this, and because of the the principles' simplicity, the

[12]Let w and w' be two worlds at which M1–M8 are all true. We can define a function f from the things that exist at w to the things that exist at w', as follows. If x really exists, let $f(x) = x$; otherwise, let $f(x) =$ the unique thing at w' which has all the same real things as parts that x has. (x has some real parts by M7 and M8; they all exist at w' by M4, and uniquely compose something there by M1 and M2.) f is one-to-one, by M2, and onto by M7 and M8. Since f maps real things to themselves, it preserves the intrinsic and relational character of real things, by M5. It preserves the parthood relation, by M1, M2 and M3. By M6, this means it preserves all there is to the qualitative nature of fictional things.

Thus, the only thing that is left not to be preserved by f is the *identity* of the fictional things, i.e. the things that don't exist at the actual world. For my part, I prefer to think of worlds in such a way that worlds can't differ as regards which nonactual things there are: there aren't any nonactual things, so all a world can "say" is that there are so many nonactual things, which do such-and-such. So f is the identity function. However, even non-modal-realists can make sense of a way of talking about worlds on which they sometimes do differ as regards the "haecceities" of non-actual things (see Sider MS). If you prefer this way of thinking about worlds, you should say that M1–M8 don't uniquely single out a world, but only an equivalence class of worlds differing only as regards the haecceities of nonactual things. But for our purposes this is just as good as singling out a world, since the truth values of propositions expressible by us can't differ between such worlds.

[13]Nolan (1997) suggests that, to accommodate the fact that we can reason non-trivially about counterfactuals with impossible and even logically inconsistent antecedents, the standard possible worlds approach to counterfactuals should be ex-

fiction whose principles are M1–M8 is very neat from a formal point of view.[14]

Now that we have the principles of this fiction before us, we are well-equipped to understand the many different ways in which other acceptable versions of the mereological fiction might differ from this one. I will consider possible modifications to M1–M8 in order.

M1 has some surprising consequences, which many have found counterintuitive. It entails, for example, that if I put a seven-pound computer and a one-pound book on the desk, then there will also be spatially disconnected object that weighs eight pounds on the desk. Certainly, there are many contexts in which it would be inappropriate to mention such a thing: for example, the right answer to the question 'What is the heaviest thing on your desk?' is 'The computer'. We could account for this by saying that, in these contexts at least, the operative fiction is one in which M1 is replaced by some statement of a more restrictive answer to the Special Composition Question.

tended so as to include *impossible* worlds, including worlds at which logical inconsistencies are true. On this way of thinking about worlds, the worlds where M1–M8 hold include some worlds where there are cats, and some worlds where there aren't cats; it's just that the latter worlds are inconsistent, whereas the former are not (worries about the semantic defectiveness of 'is part of' aside). Nolan proposes that in general, consistent worlds are closer to the actual world than inconsistent ones. Provided we agree on this principle, no *further* divergences in our understanding of the standards of closeness will lead us to disagree about what would have been the case if M1–M8 had been true.

[14]Truth according to this fiction can be characterised in a different way: it is the property a sentence gets to have if a strict and literal truth results when it is run through the singulars-to-plurals algorithm described on p. 29 above. So for example, 'Every cat is on some mat' is fictionally true iff 'Whenever there are some things arranged catwise, they are on some things arranged matwise' is strictly and literally true. But what do we mean by 'arranged catwise'? The fictionalist semantics has a ready answer, which respects the intuition that we understand 'arranged catwise' by understanding 'is a cat': some things are arranged catwise just in case, according to the fiction, they compose something that is a cat.

This is not the only possible explanation, however. There is also the possibility that a *tacit restriction on quantification* is in play: we can say that there is no eight-pound object on the table for the same sort of reason that we can say 'there are exactly two things on the table, one that weighs seven pounds and one that weighs one pound', ignoring all the proper parts of the computer and the book. How are we to decide between the two explanations? I'm not sure. Perhaps we should see whether people generally react to the objection 'What about the sum of the computer and the book?' by saying something like 'Oh, I wasn't counting *that* sort of thing, or by saying something more like 'But the computer and the book don't have a sum!' But even if the former reaction was the dominant one, it wouldn't be decisive evidence for the explanation in terms of quantifier restriction: maybe when I use the expression 'the sum of the computer and the book', I change the context to one governed by a fiction which includes M1.[15]

M2, too, has consequences which have been found counterintuitive. If some things never compose more than one thing, it can never happen that some things compose a statue that is valuable, beautiful, and easy to destroy, while also composing a piece of metal that is neither valuable, beautiful, nor easy to destroy. At least, this follows un-

[15]There may be no genuine difference between the two explanations. One possible account of the divergence between strict and literal truth and assertability which is involved in quantifier-domain restriction treats this phenomenon in much the same way that I have been treating talk of composite objects. Thus, the assertability of 'there is no beer' in a context in which it is used to express that there is no beer in the house is explained by the strict and literal truth of some sentence of the form '○ there is no beer'—like 'In the house, there is no beer', or 'Restricting our attention to things in the house, there is no beer'. Perhaps this 'in the house' operator can be analysed along the lines of 'If there were no things besides the things in the house...' or 'According to the fiction that things outside the house don't exist...'. (In their survey of approaches to quantifier-domain restriction, Stanley and Szabo (2000) argue that no analysis of this sort can be correct; but I'm not convinced.)

less there is a shift in meaning between the first and second occurrences of 'valuable', 'beautiful' and 'easy to destroy' in that sentence—it isn't obviously contradictory to suppose that one and the same thing might be valuable/beautiful/easy to destroy *as a statue* but not *as a piece of metal*. This, in a nutshell, is David Lewis's strategy for defending the principle that composition is unique in the face of our intuitions about the differences between statues and pieces of metal, and other such problem cases (1971; 1986b, section 4.5). If all the things we say which appear to entail the falsehood of M2 can be explained away in this fashion, there is no need to suppose that we ever speak according to a fiction in which M2 fails. However, this strategy can clearly only take us so far. If in any context it is correct to say outright that some statue is composed by the same things as some distinct piece of metal, we must of course admit that M2 fails in the fiction governing that context. And, as Kit Fine (MS) has shown, there are some rather natural ways of talking about statues and pieces of metal on which it is completely clear that we *treat* them as distinct, even if we never explicitly say that they are distinct. Contexts in which the facts about what we can correctly assert are as Fine intuits them to be must also, I think, be explained using fictions in which M2 fails. I am inclined to hold the boring view that there are some contexts that do work like this, other contexts in which we can correctly say that composition is unique, and many contexts in which it's just not determinate what we should say.

M3 looks obvious when you first see it; but some of its consequences have an odd sound to them. If Jasper is part of a human pyramid, and Jasper's brain is part of Jasper, does that mean that Jasper's brain is part of a human pyramid? It's not clear what we should say. But even if it isn't correct in some context to say 'Jasper's brain is part of a human pyramid', I doubt that that means that context is governed

by a different sort of mereological fiction. It would be simpler, I think, to say that 'part' is ambiguous: there is the strict, transitive notion of parthood which occurs in the principles of the fiction, and then there is the looser notion expressed by 'part of' in this context. It is natural to define the loose notion in terms of the strict one, as 'salient part' or 'relatively well-demarcated part' or some such thing.[16]

M4 should be uncontroversial. We *could* adopt a practice of 'pretending away' certain things that really exist: someone who really hated electrons (or things arranged electron-wise) might find it soothing to speak according to the fiction that there aren't any. But it's obvious that our ordinary practice of talking about composite objects doesn't work like this. The same goes for **M5**. There are certainly some predicates which can apply to a real thing in the fiction but not in reality: 'is part of something else', 'is such that there are composite objects', and the like. But predicates which concern only what the thing is like *in itself*, and in relation to other real things, should be unaffected.

The point of **M6** is to rule out gratuitous differences among fictional objects. When some things xx are exactly like some things yy, we want to ensure that it is true in the fiction that the thing composed by xx is exactly like the thing composed by yy. M6 is obviously not supposed to rule out that according to the fiction, some composite things are chairs/statues/red/conscious/..., and others are not. The assumption,

[16]If we adopted the opposite approach, defining the "strict" notion of parthood as the transitive closure of the "loose" one, we could no longer take it for granted that sentences like 'x is a loose part of y iff x is a salient strict part of y' were true in the fiction, as we could if they were just analytic truths. So for them to be true in the fiction, they would have to be added as further principles; any vagueness or contextual variability as regards what it takes for something to be a "loose" part of something else would be treated as vagueness or contextual variability in the principles of the fiction. This seems pointless, though it is not unworkable.

rather, is that composite things get to be chairs, etc., according to the fiction, *in virtue of* what the fiction says about what their actual parts are and how they are arranged. The plausibility of this assumption depends on whether the fiction includes M2 (the uniqueness of composition). Given M2, the assumption is quite reasonable. It is easy to see how the facts about the arrangement of certain real things could be sufficient to determine whether, in a fiction in which they compose one and only one thing, that thing is a chair/a statue/red/conscious.[17] By contrast, in a fiction in which M2 is false, it seems that M6 must be given up as well. Even though a statue and a piece of metal are exactly alike as far as parthood is concerned, they are not exactly alike in all respects: for example, only the former is a statue.[18] We will need to put other principles in place of M6: we want it to follow from the principles of the fiction, for instance, that when some things are arranged statue-wise, they compose at least one statue and at least one thing that isn't a statue. It would be unsatisfactory

[17]Remember that there is no presumption of physicalism here: a property-dualist might want to claim that the plural property (or "multigrade relation") *being arranged conscious-wise* is fundamental and unanalysable.

[18]This isn't absolutely inevitable. If you were happy to count it as *indeterminate* which of them was the statue and which the piece of metal, you could agree that the statue and the piece of metal are exactly alike in all respects despite the fact that one of them is a statue and the other one is not. Given the right "penumbral connections" (Fine 1975) between this indeterminacy in 'statue' and the indeterminacy of other words, it could still be determinately true that the statue, but not the piece of metal, is valuable, beautiful, fragile.... We could talk this way; but it seems much more natural to suppose that according to the fiction there is some relevant qualitative difference between the two in virtue of which one of them is determinately a statue.

Another way to hang on to M6 would be to deny that the statue and the piece of metal are "exactly alike as far as parthood is concerned". It might be proposed for example that the statue has the piece of metal as a part, while the piece of metal doesn't have the statue as a part. Thus, the property *having a proper part which has all the same real parts as one* might distinguish the statue from the piece of metal. However, it is not exactly clear why this should suffice to make the statue but not the piece of metal a *statue*.

if we had to take this, and indefinitely many more such claims, as principles of the fiction in their own right. Surely if this thing is a statue, while this other thing with all the same parts is not a statue, this is so in virtue of some more basic difference between the two. What we want is a new basic predicate of the fiction in terms of which to characterise some sort of structure which goes beyond what can be captured using the predicate 'is part of'.

As an example of what this new structure might look like, we can consider Kit Fine's metaphysic of "qua-objects" (Fine 1982, 2000). According to Fine, whenever an object has a property, there is a new object which has the first object as its "basis", and the property as its "gloss": the statue, perhaps, has the piece of metal as its basis and the property of being statue-shaped as its gloss. No object has more than one basis or gloss. This gives us not only a multiplicity of objects composed by the same things, but a way to characterise an appropriately fundamental sort of difference among the objects from which one could plausibly take the facts about which of them are statues, etc., to follow. We could adopt the predicates 'is the basis of' and 'is the gloss of' alongside 'is part of' as basic predicates of the fiction, even if we preferred to replace Fine's principle that an appropriate qua-object exists whenever a thing has a property with something much more restrictive. We could then replace M6 with the principle that all fictional things are alike except insofar as they have different parts, are parts of different things, have different bases, are the bases of different things, and have different glosses.

The purpose of **M7** is to rule out worlds in which extra things are gratuitously added to the real world. If there are really only so many simple particles, it should be true in the fiction that there are only that many particles. A fiction in which M7 was not true—a fiction, that is, according to which there were some objects wholly

disjoint from all real things—would surely not count as a *mereological* fiction. It may be, of course, that the sort of fictionalism I have been advocating with regard to composite things is also true with regard to some other domain of entities: in part II, for example, I will argue that this is the right way to think about complex properties, relations and propositions. If so, the present discussion must be regarded as merely provisional; but however we expand the fiction, we will surely want to retain some appropriately modified equivalent of M7.

M8 makes the fiction formally very neat; but it is not the only possible principle that could be adopted to govern the locations of real things in the world of the fiction. What reason is there to adopt M8 instead of its competitors? Given my conclusions in 2 about what the real world is likely to be like, it's not hard to justify the view that it's likely to be *true* in the fiction that all real things are simple. If the only things that really exist are spacetime points and the like, it's presumably correct to say that none of them has any proper parts. However, to justify taking M8 to be a *principle* of the fiction, we must decide what we would want to say in the unlikely event that the actual world contains some things which, for whatever reason, one might be tempted to take as composite objects. *Prima facie*, there seem to be some ways for the actual world to turn out which would make it fictional of some real things that they are composite. Suppose that—thanks to the failure of microphysics to be nomologically complete—there are some extra, non-microphysical, causally efficacious things, which we can and do refer to; moreover, we are disposed to apply the predicate 'complex' to them. If that's how we talked, surely we would be *correct* in doing so. This suggests that M8 needs to be replaced with a more complicated principle, from which it would follow that if the actual world were like that, some real things would be complex.

However, I think we will do better to resist this suggestion. The best thing

to say, I think, is that if the real world were like that, 'part', 'simple', and other mereological vocabulary would be ambiguous. In one sense, 'part' is an exclusively fictional predicate: taken in this sense, M8 is true in the fiction. In another sense, 'part' expresses a relation that really does obtain among some really existing things. The facts about the extension of 'part' in this sense are the same in the real world as in the world of the fiction. So when 'a' is a name for one of the non-microphysical, causally efficacious things, 'a is complex' will be fictionally true when taken in the latter sense, but not when taken in the former sense.[19] Perhaps there is a third sense of 'part', on which it expresses the disjunction of the first two senses—or better, the transitive closure of the disjunction. In this sense too, 'a is complex' will be fictionally true.

Here is a positive argument that M8 gives the right answer about a world at which there are a lot of epiphenomenal things, even if they are structured in a way that would make it natural to think of them as in some sense complex. Suppose that according to the fiction these real epiphenomena were composites (in the "fictional" sense) of microphysical things. Then according to the fiction they would be things

[19]An analogy might help to make the situation clear. Suppose that a very absent-minded author of science-fiction writes a story called *Quark!*, set in the near future, in which some aliens conquer the world using powerful guns which fire deadly particles of a hitherto unknown sort, called 'quarks'. The author chose the name 'quark' because of its euphonious sound, forgetting all about of the established use of this word among physicists. What is the case according to *Quark!*? It's a "realistic" story, intended for a knowledgeable audience of science-fiction readers, so presumably it's true in the story that quarks are very common particles, just as it is in reality, and just as it would have been in the story if the author had chosen a different name for the fictional particles. On the other hand, it's also true in the story that quarks are dangerous and found only inside the aliens' guns. But the story isn't contradictory! Rather, when it used to say what is the case according to the story, the word 'quark' becomes ambiguous: it has a "real world" sense, on which quarks are common, and a "fictional" sense on which they are rare and deadly.

with causal powers. For surely, according to the fiction, it is sufficient for something to be causally efficacious for it to be composed (in the "fictional" sense) of parts which are causally efficacious (see section 2.5 above). So these things are in reality epiphenomena; but if we spoke according to a fiction in which they were treated as composite things (in the "fictional" sense) the correct thing to say would be that they are not epiphenomena. That seems like a bad result. Epiphenomena are *weird*: if we were to talk according to a fiction in which things that are really epiphenomenal have causal powers, we would be hiding our heads from the true strangeness of the world. The practice of talking according to such a fiction ought not be embraced as a useful manner of speaking by someone who knew what the world was really like; rather, it ought to be rejected as liable to perpetuate a grave mistake.

This argument can be extended into an argument that no matter what the real world is like—even if microphysics isn't nomologically complete, so that entities which are in some sense "composed" of microphysical things don't have to be epiphenomena—it should be true in the fiction that no real thing is composed (in the "fictional" sense) of any other real things. Should a non-microphysical thing with radically emergent properties be counted in the fiction as a composite of microphysical things? No: if it were, it would turn out according to the fiction to be involved in various causal transactions *just* in virtue of the involvement of its microphysical parts. In addition to the amazing causal powers the thing really has, the fiction would ascribe it a lot of other powers inherited from its parts. So the fiction wouldn't be merely adding some extra stuff while leaving the facts about real things alone: it would be falsifying the the causal relations among real things. That's not the sort of fiction we would have any good reason to adopt as a guide to our practice.[20]

[20]This argument only tells against fictions in which real things have other real

That concludes my survey of different possible ways of specifying the details of the the mereological fiction. The survey has been a partial one in some respects. In particular, I have in effect been assuming that 'is part of' is a *time-independent* notion. Many philosophers (e.g. Thomson 1983) reject the idea that something could be part of something else *simpliciter*, as opposed to at this or that time: for them, a metaphysically perspicuous notation for mereology would have to treat 'is part of' as a three-place predicate, with an extra argument place for times. If we wanted to settle on a set of principles for a fiction in which parthood was time-dependent, we would face all the same choices that we face in developing the time-independent fictions I have been considering, and many more besides: we will need to find some new principles to tell us under what circumstances one and the same thing is composed by xx at t and also by yy at t', when $t \neq t'$. All the famous philosophical conundrums about identity through time will arise for us as questions about the structure of the mereological fiction. Not that there's anything wrong with that! When we are first introduced to the Ship of Theseus and the rest of the puzzles, we feel in our bones that these aren't really questions about what the world is like, but questions about how

---

things as parts (in the "fictional" sense): it poses no problem for fictions in which real things have *fictional* proper parts. Perhaps we sometimes talk according to this sort of fiction. If you don't believe that there are really any objects that only last for an instant, you might want to understand philosophers' talk of "time-slices" as governed by a fiction according to which real things have fictional instantaneous proper parts. This sort of fiction might even be useful for understanding the talk of ordinary people who never explicitly say anything about time-slices. Suppose that the only things that really exist are enduring particles, all of which last for a very long time. In that case, it is also true according to the fiction given by principles M1–M8 that all composite objects last for a very long time. So none of the things that exist according to that fiction is a good candidate to be a soap-bubble or a butterfly or a person. One way to modify those principles so that they generate an abundant supply of short-lived objects would be to allow for fictional time-slices of real particles; soap-bubbles and so forth could then be taken to be composed of these.

to talk, which we could if we wished resolve by stipulation. Of course there are many other ways to make philosophical sense of this feeling; but saying that the puzzle cases are cases in which we are torn between different versions of the mereological fiction seems to be as good a way as any.

3.4 Necessity according to the fiction

Even if we abstract away entirely from the problems arising from empty names and semantically defective predicates which I discussed in section 3.1, there is an independent reason to think that it is sometimes correct to assert sentences which, strictly speaking, couldn't possibly be true. The aim of our fictionalist theory of correctness is to explain how it can be correct for us to treat the axioms of folk mereology as completely unproblematic, analytic truths, as we do when we treat 'the plates compose a pile' as if it were synonymous with 'the plates are piled up'. But our treating these axioms as if they were completely unproblematic analytic truths is not just a matter of our taking them for granted as premises in this way. We also explicitly comment from time to time on the special status enjoyed by the principles. If I were to say 'There couldn't possibly be some plates that were piled up but failed to compose a pile' in an ordinary context—not the present context, in which the strict and literal truth of folk mereology is in question, and our standard practice is suspended!—I suppose you would treat what I said as correct. More generally, whenever P is a logical consequence of folk mereology, ⌜It is metaphysically necessary that P⌝ can correctly be asserted in ordinary contexts, despite the fact that (assuming that the logic of 'it is metaphysically necessary that' is S5) it is metaphysically impossible for

this to be the case when P is false.[21]

It is natural to explain the fact that the principles of folk mereology can correctly be said to be metaphysically necessary by reference to the logic of the operator $\bigcirc$ which determines correctness. $\bigcirc$, however it is to be understood, *commutes* with the operator 'it is metaphysically necessary that' ($\Box$): that is, $\ulcorner\Box\bigcirc\varphi\urcorner$ is logically equivalent to $\ulcorner\bigcirc\Box\varphi\urcorner$. Since $\ulcorner\bigcirc P\urcorner$ is analytic when P is a logical consequence of folk mereology, $\ulcorner\Box\bigcirc P\urcorner$ is true, and hence $\ulcorner\bigcirc\Box P\urcorner$ is true as well.

This need not force us to give up the analysis of $\bigcirc$ as 'If it had been the case that $\psi\ldots$'. We could accommodate the point by including the right claims about metaphysical necessity in the antecedent ψ, so that ψ logically entails that folk mereology is not just true but metaphysically necessary. Assuming that metaphysical necessity obeys the S5 law, this means we will be dealing with a counterfactual whose antecedent is not just false but metaphysically impossible—perhaps even analytically inconsistent. It is not always easy to make determinate sense of such counterfactuals; but when the impossibility is simply due to the fact that the antecedent falsely says that certain things are necessary, there is no great difficulty.[22] How would things have

[21]If this sounds like an inappropriate thing to say, given that folk mereology isn't *really* necessary, and in fact is probably false, this fact can be explained by saying that the use of this technical philosophical expression 'metaphysically necessary' has a tendency to shift the context to a more philosophical one in which strict and literal truth is required for correctness.

[22]If we want to give a Lewis-Stalnaker style semantics for counterfactuals with impossible antecedents, we will need to introduce some entities to play the role of "impossible worlds" (see Nolan 1997). When the only impossibility we are trafficking in is the impossibility involved in supposing truths which are in fact contingent to be necessary, it is easy to find entities to do the job, given an account of *possible* worlds. We can take a world to be an ordered pair, the first member of which is a possible world, and the second member of which is a set of possible worlds; $\ulcorner$it is metaphysically necessary that $\varphi\urcorner$ is true at $\langle w, S\rangle$ iff φ is true at $\langle w', S\rangle$ whenever $w' \in S$. (Something more complicated than this would be required if we wanted

been if it had been metaphysically necessary that kangaroos lacked tails? Much as they would have been if it had been merely contingently true that kangaroos lacked tails, I suppose; except that, for instance, it would also have been nomologically necessarily true that kangaroos lacked tails; the sentence 'it is metaphysically necessary that kangaroos lack tails' would have been true; and so forth.

A less conservative way to respond to the need to make $\bigcirc$ and $\Box$ commute would be to replace the counterfactual analysis with an analysis in terms of a different conditional $\rightarrowtail$, for which

(7) $\quad \Box(\varphi \rightarrowtail \psi) \equiv (\varphi \rightarrowtail \Box\psi)$

was a logical truth. If $\ulcorner\bigcirc\varphi\urcorner$ is defined as $\ulcorner\psi \rightarrowtail \varphi\urcorner$, it follows directly from (7) that $\ulcorner\Box\bigcirc\varphi\urcorner$ is logically equivalent to $\bigcirc\Box\varphi$, without our needing to assume anything at all about ψ. I can't see that any conditional we have in English clearly obeys (7); but it doesn't seem so hard to introduce one. We can simply stipulate that $\ulcorner\psi \rightarrowtail \varphi\urcorner$ is to be synonymous with $\ulcorner$If it had been the case that ψ, it would have been the case that $\varphi\urcorner$, when φ is nonmodal; when the consequent includes modal vocabulary, the meaning of $\rightarrowtail$ departs from the meaning of the counterfactual conditional *just enough* to make (7) come out true. This isn't entirely precise, but I think it is good enough to be going on with.[23]

worlds at which S5 was false; but these aren't needed for our present purposes.)

[23]Suppose that $\rightarrowtail$ also obeys the following laws:

$$(\varphi \rightarrowtail (\alpha \wedge \beta)) \equiv (\varphi \rightarrowtail \alpha) \wedge (\varphi \rightarrowtail \beta)$$

$$\sim(\varphi \rightarrowtail \psi) \equiv (\varphi \rightarrowtail \sim\psi)$$

Using (7) together with these two principles, we can reduce every sentence of the form $\varphi \rightarrowtail \psi$ to a sentence in which only atomic sentences occur on the right-hand-side of $\rightarrowtail$. But we know that such sentences are equivalent to the corresponding counterfactuals.

This sort of departure from the strict counterfactual theory of correctness might seem unwarranted if this were the only problem we had to deal with. But there are a couple of other factors which might give us independent motivation for replacing the counterfactual conditional with some new conditional better suited to our fictionalist purposes. One is the problem of empty names and semantically defective predicates which I have already discussed (see p. 83 above). The other concerns sentences involving the 'actually' operator, like 'Actually, there are tables' and 'Possibly, there is a table ten times bigger than some actual table'. Surely these sentences can correctly be asserted in many contexts. But as 'actually' is normally understood by philosophers, the sentences ⌜If it had been the case that ψ, it would have been the case that actually, there are tables⌝ and ⌜If it had been the case that ψ, it would have been the case that possibly, there is a table ten times bigger than some actual table⌝ both logically entail 'there are tables', and hence can't be strictly and literally true if there aren't strictly and literally speaking any tables. So these sentences couldn't correctly be asserted if the counterfactual theory were true.[24]

This doesn't tell us what to do when ψ contains modal operators occurring inside the scope of quantifiers. And indeed, our instructions about how to understand $\rightarrowtail$ don't give us a formulaic way to answer interesting questions about essences when they occur on the right hand side of $\rightarrowtail$. We know that 'snow is black $\rightarrowtail$ everything is necessarily such that snow is black' is true, but the rules don't tell us whether 'snow is black $\rightarrowtail$ something is necessarily human' is true. But the instruction to answer questions like this by departing from the facts about what the essences of things would have been if snow had been black "just enough" to make (7) true seems clear enough, even if it vague.

Things are less neat if the law of 'conditional excluded middle', $\sim(\varphi \rightarrowtail \psi) \equiv (\varphi \rightarrowtail \sim\psi)$, fails. But I think we can harmlessly assume that it holds: see footnote 25 below.

[24]Note that sentences involving 'actually' pose no problem if we adopt the metalinguistic variant of the counterfactual theory of correctness I suggested in section 3.2: this is the flip side of the problems that approach runs into with the 'strictly and literally' operator. If there had been composite entities, 'actually there are composite entities' would have been true, even though there would not actually have been com-

To fix this problem, we need the operator $\bigcirc$ to be "impermeable" to 'actually'. In other words, we want $\ulcorner\bigcirc$ actually $\varphi\urcorner$ to be synonymous, not with $\ulcorner$actually $\varphi\urcorner$, but with $\ulcorner$actually $\bigcirc\varphi\urcorner$. Thus 'actually' is inert in sentences whose only modal operator is $\bigcirc$: since $\ulcorner$actually $\varphi\urcorner$ is logically equivalent to φ, $\ulcorner\bigcirc$ actually $\varphi\urcorner$ is logically equivalent to $\ulcorner\bigcirc\varphi\urcorner$.

I think this is clearly the sort of thing that can be achieved just by *stipulating* that $\bigcirc$ is to be understood in such a way as to commute with 'actually'. This seems not so much a substantial change in the meaning of $\bigcirc$ as a change in the conventions for manipulating the symbol 'actually'. 'Actually' clearly isn't the sort of operator that has a rich meaning all of its own, like 'necessarily': it is merely a bookkeeping device, which plays the same sort of indispensable but humble role as variables and parentheses. We can tinker however we please with the rules governing 'actually' without worrying that we will somehow overstep the bounds of comprehensibility.[25]

posite entities. If I'm wrong to think that the proponent of the original counterfactual theory can deal with the problem of 'actually' by engaging in some trivial conceptual innovation, this is another good reason to adopt the metalinguistic view.

Does the metalinguistic view also let us do away with the need for counterfactuals with impossible antecedents? I don't think so. If we make the antecedent of the counterfactual say not that such-and-such is metaphysically necessarily the case, but only that such-and-such sentences involving 'metaphysically necessary' are true, the worlds that matter to the truth of the counterfactual will be ones where the word 'metaphysically necessary' means something different in our mouths from what it actually means. For one thing, it's not clear that this gives the right results about other languages: even if 'metaphysically necessary' had had some more restricted meaning, wouldn't sentences like 'The French sentence "Nécessairement, il y a des objets composites" ' still have been false?

[25]Perhaps it will help make it clear how all this fits together if I sketch a model-theory for a propositional language with the four operators '$\Box$', '$\bigcirc$', 'actually' and 'really'. Take a model to be a set of worlds W, together with an accessibility relation $<$ on W, to be used in the interpretation of '$\Box$', and a function $f : W \to W$, to be used in the interpretation of '$\bigcirc$'. In the standard "two-dimensional" model theory for 'actually' (see, e.g. Davies and Humberstone 1980), truth-values are assigned to

So far I have been concentrating on the interaction of $\bigcirc$ with the operator 'it is metaphysically necessary that'. But very similar claims are true for other operators. In particular, it seems to be correct in ordinary contexts to say things like 'No empirical evidence is required for us to be justified in being certain that piled-up objects compose piles' and 'it doesn't make sense to suppose that piled-up objects might

---

sentences relative to ordered pairs of worlds. I will extend this approach by assigning truth-values to sentences relative to triples $\langle w, w', n\rangle$, where w and w' are worlds, and n is a natural number. ψ is a logical consequence of φ iff whenever φ is true at some $\langle w, w, 0\rangle$ in some model, ψ is true there too. Suppose we are given an assignment of truth-values to atomic sentences at worlds. We extend this into an assignment of truth values to sentence letters at triples according to the following rules:

(i) When P is atomic, P is true at $\langle w, w', n\rangle$ iff P is true at $f^n(w')$—i.e. at w' if $n = 0$, at $f(w')$ if $n = 1$, at $f(f(w'))$ if $n = 2$, etc.;

(ii) $\ulcorner\varphi \wedge \psi\urcorner$ is true at $\langle w, w', n\rangle$ iff both φ and ψ are.

(iii) $\ulcorner\sim\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is not.

(iv) $\ulcorner$Necessarily $\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w'', n\rangle$ whenever $w' < w''$;

(v) $\ulcorner$Actually $\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w, n\rangle$;

(vi) $\ulcorner\bigcirc\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w', n + 1\rangle$;

(vii) $\ulcorner$Really $\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w', 0\rangle$.

This model theory gives us all the relations of commutation that we want. $\bigcirc$ and $\Box$ commute; so do $\bigcirc$ and 'actually', $\Box$ and 'really', and 'actually' and 'really'.

Since we are modelling $\bigcirc$ using a function from worlds to worlds, $\ulcorner\sim\bigcirc\varphi\urcorner$ comes out equivalent to $\ulcorner\bigcirc\sim\varphi\urcorner$. Whether this would be valid if $\bigcirc$ were analysed using counterfactuals along the lines I have been suggesting depends on whether we take $\ulcorner\sim\psi \mathrel{\Box\!\!\rightarrow} \varphi\urcorner$ to be equivalent to $\ulcorner\psi \mathrel{\Box\!\!\rightarrow} \sim\varphi\urcorner$, as Stalnaker (1968) does, or reject this inference rule, with Lewis (1973b). I think both ways of understanding counterfactuals are intelligible, and that given the choice Stalnaker's is the more convenient for our purposes. Whenever we would say $\ulcorner$Neither $\psi \mathrel{\Box\!\!\rightarrow} \varphi$ nor $\psi \mathrel{\Box\!\!\rightarrow} \sim\varphi\urcorner$ if we were using Lewis-style counterfactuals, let us instead say $\ulcorner$It is indeterminate whether $\psi \mathrel{\Box\!\!\rightarrow} \varphi\urcorner$ (see Stalnaker 1981).

fail to compose a pile'. This suggests that $\bigcirc$ interacts in the same way with the operators 'It is knowable a priori that...' and 'It is analytic that...' as with 'It is necessarily the case that...'.[26] Again, we have the options of dealing with this by keeping the counterfactual analysis but with a more complicated antecedent, or replacing the counterfactual conditional with a new conditional which gives us the relevant inferences for free.

The fact that certain things which are not really analytically true are analytically true according to the fiction is useful: it lets us explain how composite objects can be causally efficacious according to the fiction. In chapter 2, I gave a purely conceptual argument from the premise that microphysics is nomologically complete to the conclusion that all non-microphysical objects are epiphenomena. Surely if it's strictly and literally true that microphysics is nomologically complete, it will also be true according to the mereological fiction that microphysics is nomologically complete. Why, then, does it not follow that according to the fiction, all non-microphysical objects are epiphenomena—so that it is not even *correct* to say things like 'the window broke because it was struck by a brick'?

Here is a sketch of an answer. Since ⌜according to the fiction, it is analytically true that φ⌝ can be true even when φ is not, strictly and literally speaking, an analytic

[26] $\bigcirc$ might also commute with the operator 'x believes that...'—or better, 'xx collectively believe that...'. Certainly if my particles collectively believe that according to the fiction, φ, it should be true according to the fiction that they—and hence also the thing they compose—believe that φ. The reverse inference is not so clearly valid: even though it is clearly true according to the fiction that ordinary people believe all sorts of things, we might well balk at attributing beliefs involving this unheard-of operator to ordinary people (or rather, to particles arranged ordinary-person-wise). I am inclined to say that while there may be a notion of *explicit* or *occurrent* belief for which the inference fails, there is also a more permissive notion of belief on which it is just fine to attribute beliefs about what is the case according to the fiction to particles arranged ordinary-person-wise.

truth, it can happen certain pairs of predicates are analytically equivalent according to the fiction without being strictly and literally analytically equivalent—the pair 'are arranged molecule-wise' and 'are such that the thing they compose is a molecule', for example. It is a general fact about the concept of causation that in a very wide range of cases, when it is analytically true that φ iff ψ, the inference from ⌜such-and-such occurred because φ⌝ to ⌜such-and-such occurred because ψ⌝ is valid.[27] So if 'the atoms stayed in the same small region because they were arranged molecule-wise' is true in the fiction, so is 'the atoms stayed in the same small region because the thing they composed was a molecule'. But here is another general fact about causation: from the premise ⌜such and such occurred because φ(the F)⌝, one can, in a very wide range of cases, infer ⌜the F is causally relevant to the fact that such-and-such occurred⌝. Using this inference rule, we can conclude that it is fictionally true that the molecule is causally relevant to the atoms' staying where they are. And this holds in general: when a fact about some things is causally relevant to some state of affairs, and sufficient according to the fiction for them to compose something, the thing they compose will also turn out to be causally relevant to the same state of affairs, according to the fiction.

[27]Is this inference *always* valid? The most problematic cases are those like the inference from 'the gun fired because the trigger was pulled' to 'the gun fired because ((the trigger was pulled or snow is white) and (the trigger was pulled or snow is not white))'; it would certainly be odd to assert the conclusion, although the oddness may not be due falsehood. Even if we do think the conclusion is false, that doesn't mean the inference rule is invalid. David Lewis suggested to me that the effect of dragging in 'snow is white' might be to cause some contextual parameter relevant to the interpretation of 'because', such as the "implicit contrast class", to shift: if this were right, we could still hold that the inference rule is valid within any one context.

Part II

Complex Attributes

Chapter 4

The existence of complex attributes

4.1 The question

Sometimes—or so some philosophers think—it happens that there are three properties p, q_1 and q_2, such that p is the *conjunction* of q_1 and q_2. What is it for a property to be the conjunction of two other properties? It is not at all easy to say. One thing we do know is this:

(1) If p is the conjunction of q_1 and q_2, then the things that instantiate p are exactly the things that instantiate both q_1 and q_2.

And surely a property which merely *happened* to fulfil this condition would not be the conjunction of q_1 and q_2; so (1) can be strengthened to

(2) If p is the conjunction of q_1 and q_2, then it is nomologically necessary that the things that instantiate p are exactly the things that instantiate both q_1 and q_2.

Perhaps we can strengthen this condition still further in the same direction:

(3) If p is the conjunction of q_1 and q_2, then it is *metaphysically* necessary that the things that instantiate p are exactly the things that instantiate both q_1 and q_2.[1]

It is an interesting question whether the conditional in (3)—or some variant of (3) involving some even stronger notion of necessity—could be replaced with a biconditional. I won't presuppose anything about the answer to this question.

When a property is the conjunction of two other properties, it is a *complex* property; metaphorically speaking, it is "built up" from its conjuncts; they are in some sense its "constituents". Here are some other ways in which a property might be complex:

> It might be the *disjunction* of two other properties. If there are such properties as the property of being red, the property of being purple and the property of being red or purple, the latter is the disjunction of the other two.

[1]There is a strong reading of (3) on which it entails that

> If p is the conjunction of q_1 and q_2, then it is metaphysically necessary that p, q_1 and q_2 all exist

—i.e., that only metaphysically necessarily existent beings can have or be conjuncts. Philosophers who conceive of metaphysical necessity in such a way as to rule out metaphysically necessarily existent beings might find this too strong. However, the strong reading of (3) is not obligatory, or even especially natural. On a weaker reading, all that follows is this:

> If p is the conjunction of q_1 and q_2, then it is metaphysically necessary that if p exists and is instantiated, then q_1 and q_2 exist, and if q_1 and q_2 exist and are co-instantiated, then p exists.

This still gives us "necessary connections between distinct existences"—but the connections are not different in kind from those often held to obtain between people, sperms and eggs.

It might be the *negation* of some other property. If there are such properties as the property of being red and the property of not being red, the latter is the negation of the former.

It might be an *existential quantification* of some binary relation. If there are such things as the property of loving something, the property of being loved by something and the relation of loving, each of the former is an existential quantification of the latter.

It might be a *universal quantification* of some binary relation. If there are such things as the property of loving everything, the property of being loved by everything and the relation of loving, each of the former is a universal quantification of the latter.

It might be an *application* of some binary relation to some object. If there are such things as the property of loving Rome, Rome, and the relation of loving, the first of them is an application of the third of them to the second of them.

It might be the *reflexivisation* of a binary relation. If there are such things as the property of loving oneself and the relation of loving, the former is the reflexivisation of the latter.

It might be an *expansion* of some proposition. If there were such things as the property of being such that snow is white and the proposition that snow is white, the former is an expansion of the latter.

It is not just properties that could be complex in this sense. A relation could be complex in the very same sense, e.g. by being the conjunction, disjunction or nega-

tion of some other relation or relations, an existential quantification of some relation of higher degree, or an expansion of a property. There are also sorts of complexity that apply only to relations: for example, one binary relation could be the *converse* of another binary relation, standing to it as the relation *loving* stands to the relation *being loved by*, if there are any such things. Propositions, too—at least on one way of thinking about them—seem to admit of many of the same sorts of complexity. A proposition could be the conjunction, disjunction or negation of some other propositions; it could be the existential quantification or the universal quantification of a property, or the application of a property to an object. I will use the word 'attribute' as a synonym for 'property or relation or proposition'. We can extend the notion of degree to attributes in general by counting properties as attributes of degree 1, and propositions as attributes of degree 0.

Now that we have all these notions before us, a number of questions present themselves. Under what circumstances do two attributes have a conjunction? a disjunction? Under what circumstances does an attribute have a negation? an existential quantification? Under what circumstances is there an attribute which is an application of an attribute to an object? Under what circumstances does a relation have a reflexivisation? a converse?...

One possible answer to these questions is what I will call *Universalism about complex attributes*, in allusion to the corresponding doctrine about the composition of particulars (see p. 2 above). According to Universalism, any two attributes have a conjunction and a disjunction, provided they are of the same degree; every attribute has a negation; all properties and relations have existential quantifications and universal quantifications; and so on. Many philosophers who have written about properties have taken Universalism for granted.

Another possible theory of these matters is *Nihilism about complex properties*. According to the Nihilist, it never happens that there are distinct (i.e. non-identical) attributes p, q_1 and q_2 such that p is the conjunction or the disjunction of q_1 and q_2; it never happens that there are distinct attributes p and q such that p is the negation of q; it never happens that there is a property p and a binary relation r such that p is an existential quantification of r; and so on. Nihilism is defended by Frank Ramsey (1925) and D. H. Mellor (1991, 1992).

Universalism and Nihilism are not incompatible. If there are no attributes—if Nominalism is true—then both Universalism and Nihilism are true. Conversely, if both Universalism and Nihilism are both true, it follows that there are no attributes.[2]

Universalism and Nihilism obviously aren't the only possible views on the topic of complex attributes. Many intermediate positions are conceivable, although few have ever been defended. A prominent example of an intermediate view is that of Armstrong 1978b. Armstrong holds that there are no negations or disjunctions. There are some conjunctions, but on the other hand not just any two properties have a conjunction: two properties have a conjunction just in case they are co-instantiated.[3]

[2]Suppose there is a property, p. If Universalism is true, then p has a negation. But the negation of p must be distinct from p, since it is instantiated by all and only those things which do not instantiate p. Hence there are two properties one of which is the negation of the other; so Nihilism is false. Similar arguments work for relations and propositions.

[3]This seems to be as good a place as any for me to note that there is an influential view—defended by Lewis (1983a), and in recent work by Armstrong (1997)—according to which the word 'property', and related words like 'relation' and 'instantiate', are systematically ambiguous. If so, then Universalism, Nihilism, and other such doctrines are ambiguous as well: in fact, Lewis and Armstrong hold that in one sense Universalism is true, while in another sense, something closer to Nihilism is true. The former is sometimes called the "abundant" sense; the latter, the "sparse" sense—but of course these labels are only appropriate if Lewis and Armstrong are right to assert that attributes in the former sense are abundant, and attributes in the

4.2 Sharpening the question

As stated, Universalism and Nihilism are both quite vague: it is not very clear how 'and so on' is to be understood. This is not much of a problem as far as Universalism is concerned. There are various ways to develop precise theories which entail the fragment of Universalism I have already set forth, and which don't cry out for further strengthening in the same way the fragment does. The main task we face in trying to come up with such a theory is that of inventing a richer and more expressive vocabulary for talking about the different ways in which attributes could be complex, which applies in perfect generality to relations of arbitrary degree. The details of this project are a bit technical, and remote from our present purposes, so I will postpone further discussion of this to the appendices.

The vagueness of our official formulation is more of a problem as far as Nihilism is concerned. No matter how we extend the given statement of Nihilism so as to mention other ways of building complex attributes besides those listed on pp. 113–114, the letter of our formulation will still fall short of the spirit of Nihilism. It will always be possible for some philosopher to come along with a new theory of attributes which is compatible with the letter of the formulation, but which features some newly-invented operation on attributes which clearly belongs in the same category as the operations of conjunction, disjunction, etc. For example, a philosopher might deny the existence of existential quantifications and conjunctions, and yet hold that it sometimes happens that two binary relations have a "product", a relation which is

latter sense are sparse.

Since I don't agree with this view, I will treat the relevant words as unambiguous. Those who hold the ambiguity thesis are free to adopt whichever uniform disambiguation they prefer—after all, I intend everything I say to be true on *all* disambiguations.

guaranteed (as a matter of metaphysical necessity) to hold between any x and y iff there is some z such that the first relation holds between x and z, and the second holds between z and y. If there was such a thing as the ternary relation *being an x, y and z such that r_1 holds between x and z, and r_2 holds between z and y*, then the product of r_1 and r_2 would presumably count as an existential quantification of this relation. But provided that the believer in products does not believe in any such ternary relation, there is no conflict between belief in products and disbelief in existential quantifications.

To respond to the challenge to articulate the "spirit" of Nihilism, the Nihilist must identify some objectionable factor that is common to conjunctions, disjunctions, etc., apart from their appearing on a certain list. One possible way of doing this is to point out that, thanks to analytic truths like (3), conjunctions, disjunctions, etc. are *metaphysically redundant*. Roughly speaking, an attribute is metaphysically redundant if there are some metaphysically necessary truths which allow one to "read off" the facts about which things instantiate the attribute from the facts about which things instantiate other attributes. To be more precise, let's say that an n-ary attribute a is metaphysically redundant iff there is some true instance of the schema

(4) There are $y_1 \ldots y_m$, all distinct from a, such that it is metaphysically necessary that for all $x_1 \ldots x_n$, ($x_1 \ldots x_m$ instantiate a iff $\varphi(x_1 \ldots x_n, y_1 \ldots y_m)$).

where '$\varphi(x_1 \ldots x_n, y_1 \ldots y_m)$' is to be replaced by some open sentence whose only free variables are among 'x_1'...'x_n', 'y_1' ...'y_m'; whose only piece of nonlogical vocabulary is the predicate 'instantiate'; and all of whose quantifiers are restricted to things other than a.[4] It follows from (3) that every conjunctive property is metaphysically

[4]I will treat 'instantiate' as a multigrade predicate capable of taking any number

redundant—let $n = 1$, $m = 2$, and take $\varphi(x_1, y_1, y_2)$ to be 'x_1 instantiates y_1 and x instantiates y_2'. Likewise for disjunctions, negations, etc.[5]

Let *Strong Nihilism* be the thesis that there are no metaphysically redundant properties or relations. Strong Nihilism can be expressed perspicuously as a schematic theory, whose axioms are all and only sentences of the form

---

of arguments greater than one: 'x instantiates y' is well-formed, but so are 'x_1 and x_2 instantiate y' (which is synonymous with 'x_1 bears y to x_2'), 'x_1, x_2 and x_3 instantiate y', etc.

[5]Perhaps not absolutely every *application* is metaphysically redundant. If there could be a property p which was an application of a binary relation r *to* p *itself*—so that p was the property of bearing r to p—p would not be metaphysically redundant. At least, it would not follow from the counterpart of (3) for applications,

> If p is the application of r to x, then it is metaphysically necessary that p is instantiated by all and only those things which bear r to x.

that p was metaphysically redundant.

For similar reasons, we do not have a guarantee that absolutely all *existential quantifications* are metaphysically redundant. Let us grant that if p is the existential quantification of r (in the second argument place), it follows that it is metaphysically necessary that for all x, x instantiates p iff x bears r to something. This is not enough to establish that p is metaphysically redundant, since the quantifier on the right-hand-side is not restricted to things other than p. Whether p is metaphysically redundant, in our sense, depends on whether the condition 'x bears r to something' is metaphysically necessarily equivalent to a condition in which the quantifier is restricted to things other than p, such as 'x bears r to something other than p'. If it was metaphysically possible for there to be an x that bore r *only* to p, p might not be metaphysically redundant.

It is plausible that most relations couldn't possibly have their own existential quantifications among their relata; the existential quantification of any such relation is metaphysically redundant. But it is not uncontroversial that all relations are like this. George Bealer (MS) has recently proposed a novel hybrid of Dualism and Functionalism in the philosophy of mind, according to which each mental property p is the property of having some property or other which φs—i.e., the existential quantification of the relation *being an* x *and* y *such that* x *instantiates* y*, and* y φs—and moreover, for many of the things that instantiate p, p is the only property which φs that they instantiate.

(5) There do not exist a, $y_1 \ldots y_m$, all distinct, such that a is an n-ary attribute, and it is metaphysically necessary that for all $x_1 \ldots x_n$, ($x_1 \ldots x_n$ instantiate a iff $\varphi(x_1 \ldots x_n, y_1 \ldots y_m)$).

where '$\varphi(x_1 \ldots x_n, y_1 \ldots y_n)$' is to be filled in as before. Assuming that (3) and similar claims are analytic, Strong Nihilism entails that there are no conjunctions, disjunctions, etc.[6]

It is worth mentioning some theories even stronger than Strong Nihilism, which a Nihilist might also want to endorse. Let *Very Strong Nihilism* be a theory just like Strong Nihilism, except that the operator 'it is metaphysically necessary that' is replaced throughout with 'it is nomologically necessary that'. What Very Strong Nihilism says is that there are no *nomologically redundant* attributes, where a nomologically redundant n-ary attribute is one for which there is some true instance of the schema

[6]There is another reason, apart from the one discussed in footnote 5, to worry that Strong Nihilism fails to capture the full spirit of Nihilism. A Nihilist might want to deny the existence not only of metaphysically redundant properties as I have defined them, but also of properties which would have counted as metaphysically redundant if only we spoke an infinitary language powerful enough to articulate the necessary and sufficient condition for instantiating the property. One way to rule out even such infinitarily metaphysically redundant properties is to formulate a strengthened notion of redundancy in terms of supervenience. According the revised definition, an attribute is metaphysically redundant iff the facts about which things instantiate it supervene (with metaphysical necessity) on the facts about the distribution of other attributes. We could explain the relevant notion of supervenience using possible-worlds talk: a property p is metaphysically redundant iff for any two possible worlds w and w', if x has p at w and does not have p at w', then there is some property or relation a distinct from p, and some $y_1, \ldots, y_n$ such that $y_1, \ldots, y_n$ instantiate a at w but not at w'. But it seems best not to get involved in all the difficulties associated with possible-worlds talk just for the sake of ruling out the bizarre hypothesis that there are infinitarily complex properties but no ordinary finite ones. I hope that the arguments for Strong Nihilism which I will be considering will carry over without much change to the case of infinitarily complex properties.

(6) there are $y_1 \ldots y_m$, all distinct from a, such that it is nomologically necessary that for all $x_1 \ldots x_n$, ($x_1 \ldots x_n$ instantiate a iff $\varphi(x_1 \ldots x_n, y_1 \ldots y_m)$),

where φ is subject to the same restrictions as before.

Let *Extremely Strong Nihilism* be a theory just like Strong Nihilism, except that the modal operator is omitted altogether. What Extremely Strong Nihilism says is that there are no *extensionally redundant* attributes, where an extensionally redundant n-ary attribute is one for which there is some true instance of the schema

(7) there are $y_1 \ldots y_m$, all distinct from a, such that for all $x_1 \ldots x_n$, ($x_1 \ldots x_n$ instantiate a iff $\varphi(x_1 \ldots x_n, y_1 \ldots y_m)$),

where φ is again subject to the same restrictions as before. Since ⌜it is metaphysically necessary that φ⌝ entails ⌜it is nomologically necessary that φ⌝, and ⌜it is nomologically necessarily that φ⌝ entails φ, Very Strong Nihilism entails Strong Nihilism, and Extremely Strong Nihilism entails Very Strong Nihilism.

4.3 The temptation to think that Nihilism is inconsistent

Thinking of the question what complex attributes there are as a question with many consistent answers, including Nihilism, does not come easily. When one first learns to talk about properties, the central thing one learns is that ⌜has the property of φing⌝ is to be treated as synonymous with ⌜φs⌝, and ⌜lacks the property of φing⌝ is to be treated as synonymous with ⌜does not φ⌝.[7] For example, we hear the sentence

[7]Two caveats. First: we may want to make an exception for semantically defective predicates like 'is a unicorn', 'is a snark', 'has the same hat size as Sherlock Holmes', 'orbits Vulcan'. I will tacitly assume in the rest of this section that we are not dealing with predicates of this distinctive sort. Second: we do not take ⌜has the property of

'Syracuse has the property of being cold' as if it were just a cumbersome stylistic variant of 'Syracuse is cold'. The derivation of the former for the latter doesn't strike us as the sort of inference on which worries about justification could even begin to get a grip.[8] We learn similar things when we learn to talk about relations and propositions. ⌜...bears the relation of φing to...⌝ is to be treated as synonymous with ⌜...φs ...⌝; ⌜The proposition that φ is true⌝ is to be treated as synonymous with φ. I will refer to expressions like ⌜the property of φing⌝, ⌜the relation of φing⌝ and ⌜the proposition that φ⌝ as *abstract terms*.

This practice commits us to treating something close to Universalism as not just true, but analytically true. By substituting expressions we are committed to treating as synonymous, we can derive

(8) Either something has the property of φing or everything lacks the property of φing

from the uncontroversially analytic truth

(9) Either something φs or everything doesn't φ.

But (8) logically entails

⌜φing⌝ to have all the same scope possibilities as ⌜φs⌝. Whereas 'The king of France doesn't live in Paris' seems to be true on one reading and untrue on another; 'The king of France has the property of not living in Paris' seems to be unambiguously untrue.

[8]When the substitution of ⌜has the property of φing⌝ for ⌜φs⌝ is made inside the scope of a so-called "propositional attitude" expression like 'believes that' or 'says that', we don't invariably treat it as truth-preserving. When a professed Nominalist says 'the table does not have the property of being rectangular', we do not report her as having said that the table was not rectangular. But this doesn't show that we are not treating the expressions as synonyms, since notoriously, even the substitution of 'lawyers' for 'attorneys' can seem not to be truth-preserving in certain contexts (see Mates 1950).

(10) The property of φing exists.

So our practice commits us to treating every instance of (10) as an analytic truth. By similar reasoning, we must treat the instances of

(11) The relation of being an $x_1 \ldots x_n$ such that $\varphi(x_1 \ldots x_n)$ing exists; and

(12) The proposition that φ exists

as analytic truths.

If even a few instances of these schemata are true, Nihilism must be false. For example, from 'the property of being a soldier exists', 'the property of being a statesman exists' and 'the property of being a soldier and a statesman exists', one can surely validly infer that the property of being a solider and a statesman is the conjunction of the property of being a soldier and the property of being a statesman. So there is a property which is the conjunction of two other properties; so Nihilism is false.[9] Someone could, in theory, believe all the instances of (10), (11) and (12) while denying that *any* two properties have a conjunction, that *every* property has an existential quantification, etc. Perhaps there are some properties which are not the referents of any abstract terms which lack conjunctions. But any two properties which *are* the referents of such terms do have a conjunction: for if the property of

[9]To derive the denial of Strong Nihilism from some instances of (10), we need to rely on the corresponding instances of the schema

> If the property of φing exists, then it is metaphysically necessary that it is had by all and only the things which φ.

Thus, the property of being a soldier and a statesman is a property which is metaphysically necessarily had by exactly the things that have the property of being a soldier and the property of being a statesman; so the property of being a solider and a statesman is metaphysically redundant.

φing, the property of ψing, and the property of φing and ψing all exist, the latter is the conjunction of the first two. The same goes for all the other sorts of complex attributes. Thus, given (10), (11) and (12), Universalism must at least be true when restricted to those attributes that are the referents of English abstract terms.

So where do we stand?

The first thing to be said is this: since Nihilism is *not* inconsistent, our ordinary practice of talking about properties involves treating as analytic many claims and inferences that are not really analytic. Why isn't Nihilism inconsistent? Well, for one thing it's entailed by the claim that nothing whatsoever exists, and this surely is not inconsistent—it is obviously false, yes, but not for the same sort of reason that it's obvious there aren't any quadrilateral triangles or married bachelors. Not that we have to rest our case on that: if it's consistent for there to be nothing at all, surely it is also consistent for there to be any number of things but no attributes. There may be some interesting analytic sentences of the form ⌜if there is anything at all, there is something that φs⌝—try taking ⌜φs⌝ as 'is not a dog'—but 'is an attribute' palpably doesn't have the very undemanding sort of meaning a predicate would have to have to make this sentence analytic.[10] Nihilism, understood at face value, just isn't the

[10]Someone could, I suppose, claim that the denial of Nihilism follows analytically from 'there are attributes'. It might even be held analytic that if there are attributes at all, then Universalism must be true—in which case Universalism itself would be analytic, since it is true by default if there are no attributes. Of course we could decide to use the word 'attribute' in this way, thereby forcing those Nihilists who previously would have characterised themselves as believers in attributes to introduce some other word to talk about the distinctive attribute-like entities in their ontology. Likewise, we could use the word 'attribute' in such a way as to make Nihilism or even Nominalism analytically true, thereby forcing linguistic innovation on those who previously would have characterised themselves as deniers of Nihilism and Nominalism. But this seems like an unhelpful way to use the terminology at our disposal: the underlying metaphysical questions are still there, and we might as well allow ourselves

sort of claim that could conceal a hidden contradiction. Thus, while it seems to be conceivable (though to my mind unlikely) that there is some English predicate φ for which ⌜φs⌝ and ⌜has the property of φing⌝, or ⌜does not φ⌝ and ⌜lacks the property of φing⌝, really are analytically equivalent, there is no predicate for which both of these equivalences hold.

The second thing to be said is this: the fact that it is an integral part of our practice to treat these expressions as synonymous when they are not strictly and literally speaking synonymous suggests that the standard that governs our practice—the minimal standard a sentence has to meet in order to be correctly asserted—is not that of strict and literal truth, but something less demanding. To help confirm this conclusion, we can imagine another encounter with the god I introduced in section 1.2. The god says 'There are no complex attributes'; you have no choice but to believe him. Do you then abandon the ordinary way of talking about properties, relations and propositions? This wouldn't be so hard, if you were still allowed to talk in the ordinary way about attribute-like entities such as features, conditions, circumstances, concepts, notions, kinds, types, sorts, theories, hypotheses, claims, statements, beliefs, views, principles.... But that's cheating: suppose you know that the god's claim is to be interpreted in such a way as to apply to all these entities as well. In that case, I think it's clear that the answer is no. One has only to glance at some works of philosophy written by professed Nominalists and Nihilists—for example, this dissertation—to realise that it would be *much too hard* to do without talk of all these entities. Thus, after your meeting with the god, it is not *incorrect* for you to talk about properties in the ordinary way. What about those who haven't had the benefit of divine revelation? Have they been incorrect whenever they said or wrote anything

to use ordinary language when we are discussing them.

which took the ordinary way of talking about properties for granted? I think you have no choice but to judge them by the same standards you apply to yourself, now that you know there aren't any complex attributes. To accuse them of making incorrect assertions would be to hold them to a pragmatic standard too exacting for anyone to meet.

If strict and literal truth is not required for an assertion about attributes to be correct, what *is* required? If we want to answer this question in the same way that I answered the corresponding question about composite objects in chapter 1, we will want to look for some sentence ψ, such that φ is correctly asserted whenever ⌜If it had been the case that ψ, it would have been the case that φ⌝ is strictly and literally true. One natural idea would be to take ψ to be the conjunction of all the instances of the following three schemata

> **Property Schema:** For any x, x instantiates the property of φing iff x φs;
>
> **Relation Schema:** For any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate the relation of being a $y_1 \ldots y_n$ such that $\varphi(y_1 \ldots y_n)$ iff $\varphi(x_1 \ldots x_n)$;
>
> **Proposition Schema:** The proposition that φ is true iff φ.

But we have to be careful. The Property Schema has contradictory instances, such as the notorious one where ⌜φing⌝ is 'not instantiating itself': before we go any further, we will have to decide what we are going to say about these. I will return to the question what the principles of a fiction of complex attributes might be in chapter 6.

4.4 Making up one's mind

Conceptual considerations alone aren't going to resolve the debate between Nihilism and its rivals. How else are we to decide what to believe? I have already said how I see this sort of epistemological landscape in section 1.4, so I will be brief. As before, the point of this discussion is not so much to settle anything as to set things up for the argument of chapter 5.

(i) It might be held that we have *direct empirical evidence* for the existence of certain complex attributes. (I assume no-one could hold that we have direct empirical evidence *against* the existence of any complex attributes!) There are two possibilities. First, we might have *perceptual* evidence for believing in complex attributes: one can *just see* that there are complex attributes, just as many philosophers believe that one can just see that there are tables and chairs, and I believe that one can just see that there are things arranged tablewise and chairwise. For example, it might be held that I can see that the table before me has the property of being brown and rectangular, the property of being brown, and the property of being rectangular; from this I can validly infer that there is at least one conjunctive property. Second—and perhaps more plausibly—it might be that we have *introspective* evidence for believing in complex attributes. It does seem natural to say, for example, that I can tell by introspection that I believe the proposition that there are no flying penguins, and hence infer that the proposition that there are no flying penguins exists. But for a proposition to be the proposition that there are no flying penguins is for it to be the negation of an existential quantification of the conjunction of the properties *flying* and *being a penguin*. So perhaps introspection gives me good reason to believe that the properties *flying* and *being a penguin* exist; that they have a conjunction; that it

has an existential quantification; and that its existential quantification has a negation.

To anyone remotely sympathetic towards Nihilism, these claims will seem to involve a flagrant attempt to pass off bits of ambitious philosophical theory as experiential data. Once one has been apprised of the difference between believing that the table has the property of being brown and square *strictly and literally speaking*, and believing that the table has the property of being brown and square *according to the fiction of complex properties*, shouldn't it be obvious that it is only the second, more cautious belief that is justified by experience? There certainly is something dogmatic about insisting on the stronger characterisation when the weaker one is available. But then again, this sort of dogmatism may well be the right response to a sceptic who insists that our experience doesn't justify us in believing that there are external objects, but only in believing that *as far as our experiences are concerned, it is as if* there are external objects. One can't win debates by making this sort of move, but one might still be right.

(ii) I won't say much here about the idea that we might have indirect empirical evidence for the existence of complex attributes, of the same sort as our evidence for the theoretical entities of science. Scientists do indeed say things that appear to presuppose the existence of complex attributes, as do we all. But this kind of talk is normally confined to expository remarks: it doesn't normally feature in precisely-stated theories in the same way that talk of composite physical objects often does, as I discussed in section 1.4.2 above. Thus, there isn't the same *prima facie* case for believing in complex attributes on scientific grounds as there is in the case of composite physical objects. There are, indeed, a few sciences—biological taxonomy, for example—in which attribute-talk plays a more central role. But it is crazy to think that we should believe in complex attributes just because taxonomists find

it useful to say things which appear to be committed to such entities as species and genera. Typical taxonomists have given no more thought to the metaphysics of properties than any ones else. The idea that their testimony on such matters carries special weight just because they are scientists carries the generally sensible policy of deference to science to an absurd extreme.[11]

(iii) Supposing that all the empirical evidence is neutral as regards the existence of complex attributes, what then should we believe? I say that the right, if uninformative, answer to this question is that we should have those beliefs that are default reasonable beliefs; we should apportion our credence among the different theories in

[11]Whereas it should be easy to do good science without mentioning complex attributes per se, it is famously hard to do science without mentioning mathematical entities. Because of this, it is widely held that we have scientific reason to believe (some version of) set theory. If we do, then arguably we have equally good reason to believe in complex attributes: for it is arguable that set theory *entails* Universalism, on at least one legitimate interpretation of 'property' and related vocabulary like 'instantiate'. Given some initial interpretation of this vocabulary on which Universalism comes out false, there are many ways to extend the interpretation in such a way as to count certain set-theoretic constructions out of the original properties as properties in the new sense, in such a way that under the new interpretation Universalism is a consequence of set theory. (Or we might instead set things up so that some theory intermediate between Nihilism and Universalism comes out true, as in the theory of "linguistic ersatz structural universals" discussed in Lewis 1986a.)

I *think* I know enough about the intended meaning of the predicate 'instantiate' to be able to see that it doesn't express some complex and arbitrary set-theoretic condition; so I doubt that the interpretation of Universalism on which it is entailed by set theory is really a legitimate interpretation. But to give a fully satisfactory response to this argument, I would have to say more about the epistemology of set theory than there is space to say here. (Although, given what I have said about the epistemology of belief in composite objects and complex attributes, my views about set theory shouldn't be too hard to extrapolate.)

In any case, even if this argument for the existence of complex attributes is sound, there is always the further question which theory of complex attributes is true *when we restrict our attention to those attributes which are not sets*. Everything I say in the sequel will be just as relevant to this question as to the original one.

proportion to their intrinsic plausibility (see section 1.4.3). And what will end end up believing, if we follow this advice correctly? I say that we will end up believing that Nihilism—Very Strong Nihilism, to be precise—is true, or at least close to being true. Other things being equal, one ought to believe that the world is a simple place rather than a complicated one. But the existence of redundant properties—whether metaphysically redundant, or merely nomologically redundant—is a sort of complexity. So, other things being equal, one ought to believe that there are no redundant properties, or at least very few of them. And other things *are* equal: including redundant properties in one's theory of the world does nothing to make the theory more worthy of belief in any other respect. So, one ought to believe, to a pretty high degree, that there are no redundant properties, or at least very few of them.

I don't want to rest much weight on the power of this argument to convince those who hold some other view of the intrinsic plausibility of the various theories of complex attributes. I fervently hope, and privately believe, that my intuitions about default reasonableness are roughly in line with those of many others. But it is all too easy, once this sort of argument has been made, to respond by tailoring one's epistemology to one's metaphysics. One can declare that one doesn't see why the distinctive sort of simplicity of Nihilism should be any consideration in favour of it; or that this consideration in its favour is outweighed by some other consideration, such as the consideration that it fails to entail the existence of complex attributes. I hope that the argument of chapter 5 will help us advance beyond this sort of stalemate.

(iv) There is another way of thinking about the task of deciding what to believe about questions for which we has no relevant empirical evidence, according to which what you should believe depends on some other sort of fact about you which can vary even when your relevant empirical evidence doesn't: the beliefs you start out with,

or the beliefs you find natural, or your intuitions.

At first sight it looks as if the proponent of this sort of approach to epistemology won't have much trouble arguing that *some* people at least ought to believe in complex attributes. Nevertheless, there are difficulties: the problems which I raised in section 1.4.4 for the attempt to justify belief in folk mereology by these means also arise here, with greater force. First: to use these methods to decide what to believe about complex attributes, one needs to know whether one believed in complex attributes before one considered the question philosophically, or whether one finds it natural to believe in complex attributes, or whether one intuits that there are complex attributes. This may not be such an easy task: certainly it is much less clear that we are really committed to the existence of complex attributes than it is that we are really committed to composite objects. Second: even if we grant that there is prima facie justification for belief in complex attributes, there is an argument that this justification should be defeated when we realise that this belief rests on a false lemma, namely our assumption that the expressions that our practice of talking about attributes licenses us to treat as synonymous really are synonymous, so that something like Universalism is an analytic truth.

Appendices to chapter 4

4A Stating Universalism

A Universalist claims that any two attributes of the same degree have a conjunction; that every attribute has a negation; that every binary relation has an existential quantification, *and so on*. In this appendix I will consider the question how this 'and so on' is to be interpreted. This topic is interesting both for its own sake, and because—as I will suggest in chapter 6—we might want to borrow some or all of the principles of Universalism to use as principles of a fiction of complex attributes.

To set forth Universalism as a precise theory, we need a richer, more expressive vocabulary for talking about the different ways in which attributes can be complex. The vocabulary introduced in section 4.1 is lacking in two main respects. First: there seem to be two different ways in which a property could be an existential quantification of a binary relation. The property could stand to the relation as the property of *loving something* would stand to the relation *loving*, if there were any such things; or it could stand to it as the property of *being loved by something* would stand to the relation of *loving*, if there were any such things. We say that in the former case the property is the existential quantification of the relation *in its first argument place*, while in the second case the property is the existential quantification of the relation *in its second argument place*. We need to decide how to understand this difference. What we want is a canonical notation for expressing the n different ways for an $n-1$-ary attribute could be the existential or universal quantification of an n-ary attribute, or the application of an n-ary attribute to some object.

Second: the notions of *converse* and *reflexivisation* only have clear significance

when applied to binary relations, but we obviously need to recognise operations of a similar sort which apply to relations of higher degree. We also need to find a way to generalise the notion of *expansion* to attributes of arbitrary degree.

The most straightforward way of remedying these deficiencies is to make liberal use of mathematical entities. As far as the first problem is concerned, we can take the official predicate expressing existential quantification to be the three-place predicate 'u is the existential quantification of v with respect to the nth argument-place': for short, '$\mathrm{EQ}(u, v, n)$'. Likewise, we will write '$\mathrm{UQ}(u, v, n)$', for 'u is the universal quantification of v with respect to argument place n', and '$\mathrm{APP}(u, v, x, n)$', for 'u is the application of v to x in argument place n'. The axioms governing these predicates in our Universalist theory fall naturally into two groups. First, we have some axioms which can plausibly be regarded as analytic truths:

(U1) If $\mathrm{EQ}(u, v, n)$ or $\mathrm{UQ}(u, v, n)$ or $\mathrm{APP}(u, v, x, n)$, then n is a number, and v is an attribute whose degree is at least n, and u is an attribute whose degree is one less than the degree of v.

(U2) If $\mathrm{EQ}(u, v, n)$, and the degree of v is $m + 1$, then necessarily, for all $x_1 \ldots x_m$, ($x_1 \ldots x_m$ instantiate u iff for some y, $x_1 \ldots x_{n-1}, y, x_n \ldots x_m$ instantiate v).

(U3) If $\mathrm{UQ}(u, v, n)$, and the degree of v is $m + 1$, then necessarily, for all $x_1 \ldots x_m$, ($x_1 \ldots x_m$ instantiate u iff for all y, $x_1 \ldots x_{n-1}, y, x_n \ldots x_m$ instantiate v).

(U4) If $\mathrm{APP}(u, v, y, n)$, and the degree of v is $m + 1$, then necessarily, for all $x_1 \ldots x_m$, ($x_1 \ldots x_m$ instantiate u iff $x_1 \ldots x_{n-1}, y, x_n \ldots x_m$ instantiate

v).[12]

Second, we have an axiom which expresses the distinctive Universalist commitment to the existence of *arbitrary* existential quantifications, universal quantifications, and applications:

(U5) Whenever n is a number, and v is an attribute whose degree is greater than or equal to n, and x is any object, there is a unique u_1 such that $\mathrm{EQ}(u, v, n)$, and a unique u_2 such that $\mathrm{UQ}(u_2, v, n)$, and a unique u_3 such that $\mathrm{APP}(u_3, v, x, n)$.

We can use some slightly more elaborate mathematical machinery to remedy the second deficiency, by introducing a general notion of *transformation* of which the notions of converse, reflexivisation and expansion are all special cases. Just as the predicates expressing existential quantification, universal quantification and application have argument places for numbers, so the predicate expressing transformation has an argument place for *sequences* of numbers—i.e., functions whose domain and range are initial subsets of the positive integers. We will speak of an attribute u being a transformation of an attribute v with respect to a sequence f; for short, '$\mathrm{TR}(u, v, f)$'. On the intended interpretation of this predicate, it is governed by the following analytic truths:

[12]If we attempt to write down the instance of U2, U3 or U4 where $m = 0$, we find ourselves with something ungrammatical: the clause 'instantiate u' won't have a subject. To remedy this, let's agree to understand 'instantiate u' as a strange way of writing 'u is true'. So, for example, the instance of U2 where $m = 0$ is to be taken as equivalent to this:

> If $\mathrm{EQ}(u, v, 1)$, and the degree of v is 1, then necessarily, u is true iff there is a y such that y instantiates v.

(U6) Whenever $\mathrm{TR}(u, v, f)$, u and v are attributes, and f is a function from the positive integers $\leq$ the degree of v to the positive integers $\leq$ the degree of u.

(U7) Whenever $\mathrm{TR}(u, v, \langle k_1 \ldots k_n \rangle)$ and $\mathrm{Deg}(u) = m$, necessarily, for any $x_1 \ldots x_m$, $x_1 \ldots x_m$ instantiate u iff $x_{k_1} \ldots x_{k_n}$ instantiate v.[13]

The notions of converse, reflexivisation and expansion can be defined in terms of transformation as follows:

r_1 is the converse of $r_2 =_{df} \mathrm{TR}(r_1, r_2, \langle 2, 1 \rangle)$ and $\mathrm{Degree}(r_1) = 2$.

p is the reflexivisation of $r =_{df} \mathrm{TR}(p, r, \langle 1, 1 \rangle)$ and $\mathrm{Degree}(p) = 1$.

Property p is the expansion of proposition $q =_{df} \mathrm{TR}(p, q, \langle\rangle)$ and $\mathrm{Degree}(p) = 1$.[14]

Binary relation r is an expansion of property $p =_{df} \mathrm{Degree}(r) = 2$ and either $\mathrm{TR}(r, p, \langle 1 \rangle)$ or $\mathrm{TR}(r, p, \langle 2 \rangle)$.

Finally, we can state a single axiom which expresses the Universalist commitment to the existence of *arbitrary* converses, reflexivisations, expansions and much else besides:

(U8) Whenever v is an attribute, m is a positive integer, and f is a function from the positive integers $\leq$ the degree of v to the positive integers $\leq m$, there is a unique u such that the degree of u is m, and $\mathrm{TR}(u, v, f)$.

[13] This is an axiom schema, whose instances are got by replacing m with some numeral, and replacing each of $k_1 \ldots k_n$ with one of the numerals $1, 2, 3, \ldots, m$.

[14] $\langle\rangle$ is the empty sequence, i.e. the function whose domain is the empty set—the empty set itself, on the standard reduction of functions

Our first attempt at a precise statement of Universalism, then, comprises U1–U4, U5, U6–U7, U8, and the following familiar principles:

(U9) When u is a conjunction of v_1 and v_2, u, v_1 and v_2 are attributes of the same degree n, and necessarily, for any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate u iff $x_1 \ldots x_n$ instantiate v_1 and $x_1 \ldots x_n$ instantiate v_2.

(U10) When u is a disjunction of v_1 and v_2, u, v_1 and v_2 are attributes of the same degree n, and necessarily, for any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate u iff $x_1 \ldots x_n$ instantiate v_1 or $x_1 \ldots x_n$ instantiate v_2.

(U11) Whenever two attributes are of the same degree, they have a unique conjunction and a unique disjunction.

(U12) When u is a negation of v, u and v are attributes of the same degree n, and necessarily, for any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate u iff $x_1 \ldots x_n$ do not instantiate v.

(U13) Every attribute has a unique negation.

The theory whose axioms are U1–U13 is a natural resting point for the Universalist. There are stronger theories, but this theory doesn't *cry out* for further strengthening in the way that weaker theories do. This is explained by the fact that the theory provides for a structure in the realm of attributes that is isomorphic to the structure of open and closed formulae described in first-order logic. To be more precise: suppose that we have some predicates $F_1 \ldots F_n$ which we know correspond to attributes, in the sense that the terms $\ulcorner$the property of being an x such that $F_i(x)\urcorner$—or when appropriate, $\ulcorner$the relation of being an $x_1 \ldots x_m$ such that $F_i(x_1 \ldots x_m)\urcorner$—all

have referents. In that case, it follows from U1–U13 (together with some mathematics) that all terms of the form ⌜the property of being an x such that $\varphi(x)$⌝, ⌜the relation of being an $x_1 \ldots x_n$ such that $\varphi(x_1 \ldots x_n)$⌝ and ⌜the proposition that φ⌝ have referents, whenever φ is an expression built up out of some or all of $F_1 \ldots F_n$ using truth-functional connectives, first-order variables and quantifiers, and names.

Proof: Let's adopt Bealer's convention that when φ is a first-order formula, '$[\varphi]$' stands for 'the proposition that φ', '$[\varphi]_x$' stands for 'the property of being an x such that φ', and '$[\varphi]_{x_1 \ldots x_n}$' stands for 'the relation of being an $x_1 \ldots x_n$ such that φ' (Bealer 1982). Suppose we start out with a first order language with predicates $\mathbf{F}^1_1, \mathbf{F}^1_2, \ldots, \mathbf{F}^2_1, \ldots$, variables $\mathbf{x_1}, \mathbf{x_2} \ldots$ and names $\mathbf{a_1}, \mathbf{a_2} \ldots$. Suppose, furthermore, that every expression of the form ⌜$[\mathbf{F}^n_i \mathbf{x_1} \ldots \mathbf{x_n}]_{\mathbf{x_1} \ldots \mathbf{x_n}}$⌝ is another name in the language. We'll show how to interpret arbitrary terms of the form ⌜$[\varphi]_{v_1 \ldots v_n}$⌝, where $v_1 \ldots v_n$ are any distinct variables, and φ is a formula of the language which may contain names and free variables apart from $v_1 \ldots v_n$. These terms will be analysed as definite descriptions involving the special predicates EQ, UQ, APP, TR, CONJ, DISJ and NEG, as well as mathematical vocabulary. The analysis proceeds recursively:

> ⌜$[\varphi \wedge \psi]_{v_1 \ldots v_n}$⌝ abbreviates ⌜the x such that CONJ(x, $[\varphi]_{v_1 \ldots v_n}$, $[\psi]_{v_1 \ldots v_n}$)⌝, where x is some new variable. Likewise, ⌜$[\varphi \vee \psi]_{v_1 \ldots v_n}$⌝ abbreviates ⌜the x such that DISJ(x, $[\varphi]_{v_1 \ldots v_n}$, $[\psi]_{v_1 \ldots v_n}$)⌝.
>
> ⌜$[\sim\varphi]_{v_1 \ldots v_n}$⌝ abbreviates ⌜the x such that NEG(x, $[\varphi]_{v_1 \ldots v_n}$)⌝.
>
> When y is a variable that is not one of $v_1 \ldots v_n$, ⌜$[\forall y \varphi]_{v_1 \ldots v_n}$⌝ abbreviates ⌜the x such that UQ$(x, [\varphi]_{y v_1 \ldots v_n}, 1)$⌝, and ⌜$[\exists y \varphi]_{v_1 \ldots v_n}$⌝ abbreviates ⌜the x such that EQ$(x, [\varphi]_{y v_1 \ldots v_n}, 1)$⌝. When y is one of $v_1 \ldots v_n$, we must first

replace it uniformly with some new variable that is not one of $v_1 \ldots v_n$, and then apply this rule.

When y is the first variable other than $v_1 \ldots v_n$ that occurs free in φ, $\ulcorner[\varphi]_{v_1 \ldots v_n}\urcorner$ abbreviates $\ulcorner$the x such that $\mathrm{APP}(x, [\varphi]_{yv_1 \ldots v_n}, y, 1)\urcorner$.

When a is the first name that occurs in φ, and φ contains no free variables, $\ulcorner[\varphi]_{v_1 \ldots v_n}\urcorner$ abbreviates $\ulcorner$the x such that $\mathrm{APP}(x, [\varphi^y_a]_{yv_1 \ldots v_n}, a, 1)\urcorner$, where y is some new variable, and 'φ^y_a' stands for the result of replacing every occurrence of a in φ with an occurrence of y.

When $k_1 \ldots k_m$ is a sequence of positive integers each of which is $\leq n$, $\ulcorner[\mathbf{F}^m_i v_{k_1} \ldots v_{k_m}]_{v_1 \ldots v_n}\urcorner$ abbreviates $\ulcorner$the x such that $\mathrm{Deg}(x) = n$ and $\mathrm{TR}(x, [\mathbf{F}^m_i \mathbf{x_1} \ldots \mathbf{x_m}]_{\mathbf{x_1} \ldots \mathbf{x_m}}, \langle k_1 \ldots k_m \rangle)\urcorner$.

These instructions suffice to eliminate all terms of the form $\ulcorner[\varphi]_{v_1 \ldots v_n}\urcorner$ except for the basic terms $\ulcorner[\mathbf{F}^n_i \mathbf{x_1} \ldots \mathbf{x_n}]_{\mathbf{x_1} \ldots \mathbf{x_n}}\urcorner$. It follows straightforwardly from axioms U1, U5, U6, U8, U9, U10, U11, U12 and U13, and mathematics, that all the definite descriptions introduced by this method have unique referents. Moreover, if we add the new multigrade predicate 'instantiate', together with all instances of the axiom schema

$$\forall y_1 \ldots y_n (y_1 \ldots y_n \text{ instantiate } [\mathbf{F}^n_i \mathbf{x_1} \ldots \mathbf{x_n}]_{\mathbf{x_1} \ldots \mathbf{x_n}} \equiv \mathbf{F}^n_i (y_1 \ldots y_n))$$

it follows by U2, U3, U4, U7, U9, U10, U12 and mathematics that, more generally,

$$\forall y_1 \ldots y_n (y_1 \ldots y_n \text{ instantiate } [\varphi]_{v_1 \ldots v_n} \equiv \varphi \begin{smallmatrix} y_1 & \ldots & y_n \\ v_1 & \ldots & v_n \end{smallmatrix}).$$

4B Notational Variants

Now that we have this theory in hand, we can consider some possible variations in how we express it.

(i) Many of the "axioms" of the theory are really axiom-schemata: this is unavoidable if we want to use the multigrade predicate 'instantiate' in stating the axioms. However, we can do without axiom-schemata if we instead use a two-place predicate applying to attributes and *sequences* of things. To distinguish this predicate from 'instantiates', let us say that attributes are *exemplified* by sequences when they are *instantiated* by the elements of the sequences. The intended connection between the two notions is as follows:

$\langle x_1 \dots x_n \rangle$ exemplifies u iff $x_1 \dots x_n$ instantiate u.

So that we don't need to make a special case for propositions, let's add that

$\langle\rangle$ exemplifies u iff u is true.

Using this notion, we can replace the axiom-schemata U2, U3, U4, U7, U9, U10 and U12 with single axioms. Here are two representative examples:

U2′ If $\mathrm{EQ}(u, v, n)$, and the degree of v is $m + 1$, then necessarily, whenever s is a sequence of length m, s exemplifies u iff there is a sequence s' which exemplifies v, such that s is derived from s' by deleting the nth element.

U7′ Whenever $\mathrm{TR}(u, v, f)$, necessarily, a sequence s is in the extension of u iff its length is the same as the degree of u, and $s \circ f$ is in the extension of v.

(ii) In the opposite direction, we might want to avoid using mathematical entities at all in our statement of Universalism. The theory of appendix 4A has its intended

effect of entailing that if there are any attributes at all there are lots of them only in conjunction with the mathematical theory of numbers and sequences; if there are no numbers or sequences of numbers, then the crucial axioms U5 and U8 are trivially true. But intuitively the question what sorts of complex attributes there are seems like it should be independent of the question whether there are numbers and sequences. So we might want to look for a theory which entails the existence of "lots" of attributes in the appropriate sense while remaining neutral on mathematical questions.

It is easy to avoid dragging in mathematical entities if we don't mind having to use infinitely many undefined predicates in stating our theory. We need only replace the three-place predicate predicate '$\mathrm{EQ}(u, v, n)$' with an infinite family '$\mathrm{EQ}_1(u, v)$', '$\mathrm{EQ}_2(u, v)$', '$\mathrm{EQ}_3(u, v)$... of two-place predicates. Similarly, we replace '$\mathrm{UQ}(u, v, n)$' with '$\mathrm{UQ}_n(u, v)$'; '$\mathrm{APP}(u, v, x, n)$' with '$\mathrm{APP}_n(u, v, x)$', and '$\mathrm{TR}(u, v, \langle k_1 \ldots k_m \rangle)$' with '$\mathrm{TR}_{k_1 \ldots k_m}(u, v)$'. Finally we must replace the two-place predicate 'n is the degree of u' with an infinite family of one-place predicates, Deg_1, $\mathrm{Deg}_2 \ldots$. It is an entirely mechanical exercise to rewrite U1–U13 using this new notation.

(iii) However, it's not very nice to have to state one's theory using infinitely many undefined predicates. Especially if one is claiming the theory to be strictly and literally true, one will have to face awkward questions about how we finite beings could even understand such a theory. In fact, however, it turns out that all those infinitely many predicates are unnecessary. With a bit of technical wizardry, one can state a theory with a reasonable claim to be a version of Universalism using only finitely many undefined predicates. The basic idea is that of the predicate-functor logic described in Quine 1960a; Bealer (1982) puts it to work in stating a theory of complex attributes. My version of the story is a slight variant of Bealer's.

Here's how it works. First, note that we really don't need all the different pred-

icates EQ_1, $\mathrm{EQ}_2 \ldots$. We can make do with EQ_1: instead of saying that a property is the existential quantification in the second argument place of a binary relation, we can say that it is the existential quantification in the first argument place of the converse of that relation, which plausibly amounts to the same thing. In general, we can existentially quantify any argument place we like in a relation by first permuting the relation so that the argument place we want to quantify is the first one, leaving the order of the other argument-places unchanged, and then quantifying in the first argument place. For the same reason, we really only need UQ_1 and APP_1.

To reduce the number of transformation predicates we need to take as primitive, we can rely on the following plausible principle about the identity of attributes derived by successive transformations with attributes derived by single ones:

If $\mathrm{TR}(u_1, u_2, f)$ and $\mathrm{TR}(u_2, u_3, g)$, then $\mathrm{TR}(u_1, u_3, f \circ g)$.

or, to put things in terms of the primitive predicates which we are trying to analyse away,

If $\mathrm{TR}_{j_1 \ldots j_m}(u_1, u_2)$ and $\mathrm{TR}_{k_1 \ldots k_n}(u_2, u_3)$, then $\mathrm{TR}_{j_{k_1} \ldots j_{k_n}}(u_1, u_3)$.

This means that if we can find some class of sequences of postive integers such that all sequences of positive integers can be derived from the members of that class by composition, we will be able to define all of the many two-place transformation predicates in terms of the predicates corresponding to the sequences in that class. One such class comprises:

(i) All sequences of the form $\langle 2, 3, 4, \ldots, n, 1 \rangle$;

(ii) All sequences of the form $\langle 2, 1, 3, 4, \ldots, n \rangle$; and

(iii) All sequences of the form $\langle 1, 1, 2, 3, \ldots, n\rangle$.[15]

Thus, we can reduce our initial supply of undefined transformation predicates to the predicates corresponding to this much smaller, though still infinite, class of sequences.

To reduce this infinite class of predicates to a finite number, we can make use of the fact that $\mathrm{TR}_{k_1 \ldots k_m}(u, v)$ is only ever true when the degree of v is m. Thus, we can let the degree of v do the work of selecting a sequence of the appropriate length; so we really only need three predicates, one for each of the basic classes of sequences we just discussed. Call these predicates PERM, CONV and REFL, respectively. We define the transformation predicates that remain undefined in terms of these and the predicates Deg_1, Deg_2,. . . :

$$\mathrm{TR}_{2,3,4,\ldots,n,1}(u, v) =_{df} \mathrm{PERM}(u, v) \text{ and } \mathrm{Deg}_n\, v$$

$$\mathrm{TR}_{2,1,3,4,\ldots,n}(u, v) =_{df} \mathrm{CONV}(u, v) \text{ and } \mathrm{Deg}_n\, v$$

$$\mathrm{TR}_{1,1,2,3,\ldots,n}(u, v) =_{df} \mathrm{REFL}(u, v) \text{ and } \mathrm{Deg}_{n+1}\, v.$$

Given these definitions, U8—or rather, all the instances of the axiom schema that replaces U8 when we adopt the subscript notation—follows from the following simple axiom:

U8″ Whenever v is an attribute, there is a unique u_1 such that $\mathrm{PERM}(u_1, v)$,

[15] I don't have a proof of this, but it's easy to convince yourself that it's true. First, all the permutations of the first n positive integers can be derived by successive compositions from $\langle 2, 3, 4, \ldots, n, 1\rangle$ and $\langle 2, 1, 3, 4, \ldots, n\rangle$. Second, all the non-repeating sequences, including gappy ones, can be derived by composition from the permutations of initial subsets, since every non-repeating sequence is an initial segment of some permutation. Finally, it's pretty clear that the sequences $\langle 1, 1, 2, 3, \ldots n\rangle$ are all we have to add to the non-repeating sequences to get all the repeating sequences when we close under composition: we can permute the elements we need to double up into first place, apply $\langle 1, 1, 2, 3 \ldots n\rangle$ as many times as we need, and then permute the results back into the positions where we want them to end up. See Quine 1960a.

and a unique u_2 such that $\mathrm{CONV}(u_2, v)$, and a unique u_3 such that $\mathrm{REFL}(u_3, v)$.

Finally, we can define away the predicates Deg_n, in terms of a single two-place predicate, 'u is of the same degree as v'. To do this, we first define 'the degree of u_1 exceeds the degree of u_2 by one' as 'there is a v which is of the same degree as u_2, such that $\mathrm{EQ}_1(v, u_1)$'. Next, we define 'the degree of u is zero' (i.e. 'u is a proposition') as 'there is no v such that the degree of u exceeds the degree of v by one'. Finally we inductively define '$\mathrm{Deg}_n(u)$' as 'there is a v such that the degree of v is $n-1$, and the degree of u exceeds the degree of v by one'.

4C Strengthenings

In this section I will describe two stronger Universalist theories which generate what we might describe as *infinitely complex* attributes. The first strengthening can be added to any of the theories we have encountered so far. The second, however, only works for the thoroughly mathematical theory described in item (i) of 4B.

First, we could replace the three-place predicates 'u is the conjunction of v_1 and v_2', 'u is the disjunction of v_1 and v_2', with predicates taking one singular and one plural argument: 'u is the conjunction of vv', 'u is the disjunction of vv'. This is justified from the linguistic point of view. The expression 'the conjunction of p_1 and p_2' seems to contain as a constituent the plural term 'p_1 and p_2'. The position occupied by that plural term can be meaningfully occupied by other plural terms: 'the conjunction of p_1, p_2, and p_3', 'the conjunction of my favourite properties', 'the conjunction of the properties that I instantiate' are all perfectly well-formed.

If we understand the plural notions of conjunction and disjunction at all, we can

see that they are governed by the following analytic truths:

U9* If u is the conjunction of w, then u is an attribute of the same degree as all the members of w, and necessarily, for any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate u iff $x_1 \ldots x_n$ instantiate all of the members of w.

U10* If u is the disjunction of w, then u is an attribute of the same degree as all the members of w, and necessarily, for any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate u iff $x_1 \ldots x_n$ instantiate any of the members of w.

These notions can be used to state stronger versions of Universalism. One could go all the way, and make it an axiom of one's theory that

U11* Whenever there are some attributes w, all of the same degree, they have a conjunction and a disjunction.

Or one could require that w satisfy some further condition in order to have a conjunction and/or a disjunction. Plausibly, it already follows from the Universalist theory of appendix 4A that some attributes have a conjunction and a disjunction when they are all of the same degree *and finite in number*.[16] So it is by entailing the existence of infinite conjunctions and disjunctions that a theory employing the plural notions can get to be stronger than the theory of appendix 4A.

Infinite conjunctions and disjunctions are not just idle curiosities. They play an important part in our ordinary way of talking about attributes. For one thing, our language seems to contain the expressive resources to produce complex formulae which, if they correspond to attributes at all, must correspond to infinite conjunctions

[16]This follows from a plausible generalised principle of associativity, according to which the conjunction of (the conjunction of w and the conjunction of $u u$) is the conjunction of w and $u u$.

and disjunctions. For example, we have the versatile ellipsis dots '. . . ', which enable us to form predicates like 'is a hero, or the son of a hero, or the son of the son of a hero, or. . . '. If there is such a thing as the property of being a hero, or the son of a hero, or . . . , what could it be if not the disjunction of the properties *being a hero*, *being the son of a hero*, *being the son of the son of a hero*. . . ?

Moreover, even though the use of abstract terms is central to our ordinary practice of talking about attributes, there is a bit more to the practice than that, and some of the other things we say about attributes appear to require infinite conjunctions and disjunctions. For instance, we have the idea that for each thing, there is the *way* that thing is. The way a thing is would seem to be the conjunction of all its properties—or perhaps, in certain contexts, the conjunction of all its properties of a certain sort (e.g. all its intrinsic properties). We also speak of "ways for things to be" that are not the way anything actually is. These might be identified with uninstantiated conjunctions of properties w, where w are such that whenever a property (of the right sort) is not one of w, its negation is. And an appealing account of *possible worlds* would take them to be certain infinite conjunctions of propositions, with the actual world being the conjunction of all true propositions.

The second sort of strengthening only makes sense in the context of the thoroughly mathematical version of Universalism described in part (i) of appendix 4B. In that theory, each attribute is assigned a natural number as its degree; attributes of degree n are exemplified by sequences of length n. In the strengthened theory, attributes can have *transfinite ordinals*, as well as finite ones, as their degrees. A relation of transfinite degree α is exemplified by sequences of length α—which we might identify with functions from the ordinals less than α to objects, or with orderings of order-type

α, or in some other way.

By making a few small modifications to the axioms of the theory, we can generalise the various building operations so that they apply to relations of transfinite degree. First, we must replace the references to natural numbers in the axiom governing transformation with references to arbitrary ordinals:

U6* Whenever $\mathrm{TR}(u, v, f)$, u and v are attributes, and f is a function from the ordinals $<$ the degree of v to the ordinals $<$ the degree of u.

U8* Whenever v is an attribute, m is an ordinal, and f is a function from the ordinals $<$ the degree of v to the ordinals $< m$, there is a unique u such that the degree of u is m, and $\mathrm{TR}(u, v, f)$.

U7′ does not need to be modified.

U8* guarantees the existence of vast numbers of relations of transfinite degree. Just as we can "expand" a property into an $n + 1$-ary relation by adding n extra argument places at the beginning or at the end, likewise we can expand a property into an ω-ary relation by adding an ω-sequence of extra argument places at the end, or expand it into an $\omega + 1$-ary relation by adding an ω-sequence of extra argument places at the beginning. Relations built up in this way are pretty dull, in that all but finitely many of their argument-places are "inert"; but they are available to serve in turn as conjuncts or disjuncts of relations that aren't dull in this way.

We need to tinker a bit with the notions of existential quantification, universal quantification and application in order to generalise them in the right way to relations of transfinite degree. For example, we want to be able take a relation of degree ω, and existentially quantify all of its argument places, making a proposition which is true iff there is some ω-sequence of things that stand in that relation. Our current

notation for existential quantification, which allows us to quantify only one argument place at a time, won't let us do that: if you existentially quantify just one argument place of a relation of degree ω, the result is still a relation of degree ω. What we need is a notion of existential quantification which lets us quantify with respect to many argument places at once. We can do this by allowing '$\mathrm{EQ}(u, v, x)$' to be true when x is any set of ordinals. We replace the old principle U2′ with:

U2* If $\mathrm{EQ}(u, v, x)$, then necessarily, for any sequence s, s exemplifies u iff there is some sequence s' which exemplifies v, such that s is derived from s' by deleting all members whose position in the sequence is a member of x.

Similarly for universal quantification:

U3* If $\mathrm{UQ}(u, v, x)$, then necessarily, for any sequence s, s exemplifies u iff whenever s is a sequence derived from s' by deleting all members whose position in the sequence is a member of x, s exemplifies v.

A corresponding change is needed in the notion of application: we want to be able to take a relation of degree ω and apply it to an ω-sequence of objects so as to yield a propostion—or a property, or an n-ary relation, or an ω-ary relation, depending which argument places we choose to apply the objects in the sequence to. To do this, we can allow '$\mathrm{APP}(u, v, s, x)$' to be true when s is a sequence of objects, and x is a set of ordinals:

U4* If $\mathrm{APP}(u, v, s, x)$, then necessarily, a sequence t is in the extension of u iff there is a sequence t' in the extension of v, such that when you delete all members of t' whose position in the sequence is a member of x, the result is t; and when you delete all members of t' whose position in the sequence is *not* a member of x, the result is s.

Using these revised notions, we can state an appropriately generalised axiom which guarantees the existence of arbitrary existential and universal quantifications and applications:

U5* Whenever x is a set of ordinals and v is an attribute, there is a unique u such that $\mathrm{EQ}(u, v, x)$ and a unique u' such that $\mathrm{UQ}(u', v, x)$. Moreover, when s is a sequence whose length is the order type of the ordinals less than the degree of v which are not members of x, there is a unique u'' such that $\mathrm{APP}(u'', v, s, x)$.

If the only effect of this strengthening was to generate new relations of transfinite degree, it would be a mere curiosity. But, thanks to the revised operations of existential quantification, universal quantification, and application, which let us build attributes of finite degree out of attributes of transfinite degree, the new theory generates a much richer, more fine-grained supply of attributes of finite degree than the old theory did. As the old theory stood to first-order logic, so the new theory stands to a very powerful infinitary logic with arbitrary infinite conjunctions and disjunctions and quantifier-blocks of arbitrary length.[17] This means, for example, that if we start with just two attributes, a property *φing* and the binary relation *identity*, we

[17] Mathematical logic studies the infinitary languages $\mathcal{L}_{\kappa,\lambda}$, which allow conjunctions of disjunctions of up to κ many formulae, and blocks of quantifiers of up to length λ (see Dickmann 1975). The "language" of attributes posited by our new theory is more powerful than any of these languages, since there is no limit on the number of attributes we can conjoin and disjoin, and no limit to the number of argument-places we can quantify at once. It is even more powerful than $\mathcal{L}_{\infty,\infty}$, the union of all the languages $\mathcal{L}_{\kappa,\lambda}$, since in $\mathcal{L}_{\infty,\infty}$ we are still limited to taking the conjunctions and disjunctions of *sets* of formulae, whereas, if we adopt the completely unrestricted principle U11* governing the existence of conjunctions and disjunctions, we will have conjunctions and disjunctions even of attributes which are so numerous that they are not the members of any set.

can built up the proposition *that there are exactly κ things that φ*, for any infinite cardinal κ.[18] Another illustration of expressive power: suppose we have some ordinary attributes of finite degree, one of which is *identity*. Then, given any model for a first-order language which takes these attributes as predicate-letters—subject to the constraint that an n-ary attribute is to be treated as a n-ary predicate-letter, and *identity* is the identity predicate—and arbitrary things as names, we can construct a proposition which is necessarily true iff this model is an *accurate description of the world*, under the obvious interpretation on which names stand for themselves and predicate-letters express themselves.[19]

[18]The proposition we want will be the existential quantification of all the argument places of a transfinite relation r of degree κ, which is, necessarily, exemplified by a sequence s iff (i) every member of s instantiates *φing*, and (ii) s is non-repeating, and (iii) everything that instantiates *φing* occurs somewhere in s. We can define r as the conjunction of three relations corresponding to these three conditions:

(i) r_1, the conjunction of all κ-ary relations q for which there is some $\alpha < \kappa$ such that $\mathrm{TR}(q, \varphi ing, \langle\alpha\rangle)$.

(ii) r_2, the conjunction of all κ-ary relations q for which there are some $\alpha, \beta < \kappa$ such that $\mathrm{TR}(q, nonidentity, \langle\alpha, \beta\rangle)$, where *nonidentity* is the negation of *identity*.

(iii) r_3, the universal quantification in the $\kappa + 1$st argument place of the disjunction of all $\kappa + 1$-ary relations q for which there is some $\alpha < \kappa$ such that $\mathrm{TR}(q, t, \langle\kappa, \alpha\rangle)$, where t is the binary relation *being an x and y such that if x φs, $x = y$*, i.e. the disjunction of *identity* with the negation of the binary expansion of *φing*.

[19]If the cardinality of the model is κ, the proposition that corresponds to it will again be derived by existentially quantifying all the argument places of a κ-ary relation r, where r is a big conjunction of κ-ary transformations of the initial set of attributes and their negations, together with the relation *being a κ-sequence of things in which everything there is occurs exactly once*, defined as in footnote 18.

Of course, if we *really believe* in all these sets and attributes, we'll have to say that none of these propositions are true, since there are more things than there are

Why might we be interested in such a profusion of attributes? First of all, those of us who know a little set-theory seem to have the resources to formulate sentences and expressions, like 'there are beth-one spacetime points', which arguably must correspond to attributes whose construction involves some sort of powerful infinitary building operation if they correspond to attributes at all.[20] Secondly, transfinite relations might prove useful even as regards our account of the attributes corresponding to perfectly ordinary expressions which don't explicitly involve any reference to infinity. Suppose that we want to hold, as good Universalists, that all expressions of the form ⌜the property of φing⌝ refer, and yet also want to hold that there is some small set of *basic* attributes out of which all other attributes are constructed. The more powerful the building operations we have at our disposal, the more plausible it will be that attributes fit to be the referents of all those property abstracts can somehow be constructed out of the basic ones. I will take up this question in more detail from a fictionalist perspective in section 6.2 below. Thirdly, transfinite relations might be relevant to those aspects of our ordinary practice of talking about attributes that go beyond our use of abstract terms. *Ways for things to be*, and possible worlds, are

members of any set. There are even more *nonsets* than there are members of any set, since for each cardinal number κ there is the proposition that there are κ things that φ, and no two of these propositions are identical. If we want a true proposition derived from a model, we will need the proposition to say 'this model is an accurate description of such-and-such fragment of the world'.

In chapter 6 I will propose a way of being a fictionalist about complex attributes on which this problem doesn't arise.

[20]Of course, you might think that 'there are beth-one spacetime points' is to be analysed using an ordinary first-order sentence in the language of set theory. This is certainly the way the expression is first introduced to us. But even if you really believe in sets, it seems plausible to think that the expression has by now acquired enough of a life of its own to make 'there are no sets but there are beth-one spacetime points' consistent.

supposed to be *maximally specific* attributes, which specify how things are *in complete detail*. When we think about what this means, it is natural to draw an analogy with the way in which a *model* specifies in complete detail how things are as regards the vocabulary it interprets. We feel that pairs of models which are not isomorphic generally correspond to different ways for things to be, even when they are elementarily equivalent (are models of all the same first-order sentences). Thus our ordinary thought about *possibilities* seems to presuppose that the realm of attributes has the sort of fine-grained structure captured by the theory of relations of transfinite degree.

Chapter 5

Against complex attributes

5.1 Epiphenomenal properties

In this chapter, I will present an argument which I hope has some prospect of advancing the debate about complex properties beyond the impasses we encountered in section 4.4. This argument will be similar in its general structure to the argument I gave in section 2.2 for the claim that if there are composite objects, they are epiphenomenal. But I will proceed at a more methodical pace, since questions concerning the causal roles of attributes aren't as easy to have clear intuitions about as questions concerning the causal roles of particulars.

The core of the argument is this:

P1 If there are more than a few nomologically redundant attributes, then almost all of them are epiphenomenal.

P2 There are no epiphenomenal attributes.

C Therefore, if there are nomologically redundant attributes, there are few of them.

If this argument is sound, then Very Strong Nihilism is at least *close* to being true. I haven't explained 'few' yet, so this is all terribly vague. But just to give you a rough idea what I have in mind, I should say that on its intended interpretation, P1, together with a plausible additional premise, will entail that for a singular term of the form ⌜the property of φing⌝ to refer to an attribute that is not epiphenomenal, φ must be a short expression in the language of physics.

What does it mean to say that an attribute is epiphenomenal? I will try not to presuppose any particular answer to this question. I think that insofar as the notion of an epiphenomenon is clear at all, we know what it would be for an arbitrary entity to be one. If pressed to explain the notion, I would say something like this: for any entity x, x is epiphenomenal iff x isn't causally efficacious; x is causally efficacious iff things happen *because of* x; things happen because of x iff there is some way of filling in the schema ⌜φ because $\psi(x)$⌝, where 'because' has its strictly causal sense, so as to yield a truth. When x is an attribute, the most natural sentences to play the role of $\psi(x)$ are sentences which say which things instantiate x. So it is a sufficient condition for a property not to be epiphenomenal that things are as they are in some respect because something instantiates the property, or because something doesn't instantiate the property, or because such-and-such things instantiate the property and such-and-such things don't instantiate the property....

This seems like a reasonable result. It is sometimes said that properties are not located in space and time, and that they are therefore not causally efficacious. But I think that those believers in properties who say this sort of thing generally would grant that many properties "matter" causally, in the sense that it makes a difference which things instantiate them. Certainly I know of no good argument that *all* properties have to be epiphenomenal, in my sense of the word. Thus, P1 does contradict a view

that most believers in complex attributes have held, even if it sounds innocuous.

I have nothing original to say in support of P2. I take it that P2 is a default reasonable belief, but I don't have much of a clue how to give a convincing argument for this claim. However, the relevance of P1 to the question what to believe about complex attributes is not limited to its role as a premise in this argument: in addition, and more importantly, it undermines many of the defences of the rationality of belief in complex attributes that we considered in section 4.4.

In particular, P1 undermines the claim that we have empirical justification for believing in complex attributes. On the intended interpretation, P1 (together with the plausible premise about physics) entails that if there is such a thing as the property of being brown, it is epiphenomenal. And if nothing about my visual system is as it is *because* the table has the property of being brown—if properties of this sort never play any role in causing anything to happen in anyone's visual system—surely I can't know just by looking that the table has the property of being brown. Even if my visual experience gives me *prima facie* justification for believing in the property of being brown, this justification is defeated if I have good reason to think that if the property of being brown exists, it is epiphenomenal. The appeal to introspection is undermined in a similar way. If nothing ever happens in me *because* I bear the special belief relation to the proposition that there are no flying penguins, how could I find out that I bear that relation to that proposition by introspection? Introspection is indeed mysterious and poorly-understood, but it is not magic.

P1 also undermines the other ways of defending the rationality of belief in complex attributes. The belief that if there are any complex attributes, more than a few of them are epiphenomenal seems to be a default reasonable belief; it seems to be something we all start out believing; it seems to be a belief that comes naturally;

and it seems quite intuitive. No matter which of these authorities we appeal to in support of the belief that there are complex attributes, that authority is at odds with itself. So some further justification is required for retaining the belief that there are complex attributes in the face of P1, and it is not clear what this justification could be.

The remainder of this chapter will be devoted to arguing for P1. The argument is quite complicated, and depends on intuitions about possible cases which not everyone will share. I wish it could be otherwise; but I suspect that these are features which any good argument for an interesting general claim about causation would have to have. Causal concepts are peculiarly difficult in this respect: almost all our competence with them comes in the form of intuitions about possible cases.

5.2 Epiphenomenal properties at a simple world

In an effort to keep things manageable, I will begin by investigating the connection between nomological redundancy and epiphenomenality on the assumption that there aren't very many attributes.

Say that a (conceptually possible) world is *economical* iff none of the attributes that exist at that world are nomologically redundant. It will be helpful if we have a specific example in mind of a world that is both economical and governed by simple laws.[1] So let's focus on Simple World, a world of enduring point-particles inhabiting

[1]I am going to be very blatant in helping myself to talk of *worlds* and *laws*, even though I don't think there really are any such things. I think it would be possible, but very laborious, to conduct the whole discussion without having recourse to such figures of speech, using counterfactuals and the operator 'it is nomologically necessary that', as well as some devices of substitutional quantification. A less laborious way to make everything perfectly strict and literal would involve judicious application

a four-dimensional spacetime manifold. There is a binary relation *occupation*, which each particle bears to some spacetime points. (In fact, we can just use the word 'particle' to mean 'thing that bears the occupation relation to something', and the word 'point' to mean 'particular that is not a particle'.) The rest of the relations at Simple World are the spatiotemporal relations among points which generate the structure of a four-dimensional manifold. Three relations are enough for this task: the binary relation *being earlier than*, the ternary relation *betweenness* and the quaternary relation *congruence*. ('x, y, z and w' instantiate *congruence*' can be thought of as meaning 'the distance between x and y is the same as the distance between z and w'.) The laws governing these relations are those of four-dimensional Euclidean geometry, as set forth for instance by Hilbert (1899) (or better, Neo-Newtonian geometry, as in Field 1980). These four relations are the only attributes at Simple World. There are no properties: all the particles are exactly alike, except that they occupy different points. The rest of the laws of Simple World describe the shapes of the paths traced out by the particles: we can suppose these laws to be the laws of Newtonian gravitational theory, with all the particles assigned the same mass.[2] It is clear, I think, that none of the attributes of Simple World is epiphenomenal.[3]

---

of the 'according to the fiction of complex attributes' operator which I introduced in section 4.3. At least we incorporate the axioms of a powerful, infinitary version of Universalism, of the sort described in appendix 4C, into the fiction, some of the propositions that exist according to the fiction will be well-suited to playing the role of worlds and laws.

[2]Field (1980) shows how laws of this sort can be elegantly stated without bringing in mathematical entities.

[3]I should note that there are some theories about laws of nature according to which the description of Simple World is inconsistent. According to these theories, for it to be nomologically necessary that φ, for any φ or at least any interesting φ, some special attribute such as the "necessitation" relation of Armstrong 1983 has to exist. Since Simple World only has the four relations I mentioned, Simple World

Next consider a world—call it World X—just like Simple World except for the addition of a new property p, subject to the following law:

(1) For any x, x instantiates p iff x is a spacetime point which lies midway between two occupied spacetime points.

(x is midway between y and z iff x, y and z instantiate *betweenness*, and x, y, x and z instantiate *congruence*.) Imagine God looking down on World X, seeing the occupied spacetime points coloured in red, while the spacetime points instantiating p are coloured in yellow. When he focuses on just the red points, he sees that familiar Newton's-laws pattern. When he focuses on the yellow points, he sees a funny web-like pattern that doesn't look like much of anything, until he notices that the yellow points are exactly those that are midway between two red points. It is obvious—I hope!—that this property p is an epiphenomenon.

Since p is epiphenomenal, there must be something wrong with the following argument: 'Suppose x is a point that instantiates p. Since it is a law of nature that a point instantiates p just in case it is midway between two occupied points, if x had failed to instantiate p, it would not have been midway between two occupied points. Thus, some point that is in fact occupied would have to have not been occupied had x not instantiated p. But since the occupation relation is a straightforwardly causally

has no laws of nature, or at least no laws of nature that are not also metaphysically necessary truths.

I am inclined to think that the core of my argument for P1 could be preserved even if one of these theories were true. For even if there must be certain nomologically redundant propositions for there to be any laws at all, these propositions are going to be quite simple, given that the laws themselves are quite simple. As will emerge below (p. 173), this means that on the intended interpretation of P1, the existence of such propositions is compatible with the claim that there are "few" nomologically redundant attributes.

efficacious one, many other things would have gone differently in that case. So, many other things would have gone differently if x had not instantiated p. So things are as they are partly *because* x instantiates p; so p is not an epiphenomenon.' Even if we aren't sure where this argument goes wrong, we know it goes wrong somewhere. One possibility is that this is one of those cases where the the inference from ⌜If it hadn't been the case that φ, it wouldn't have been the case that ψ⌝ to ⌜ψ because φ⌝ fails. Another possibility is that the counterfactual we are considering is false: if the particle had failed to instantiate p, it would not have been the case that the instances of p are all and only the things that that are midway between two occupied points, despite the fact that this is actually a law of nature.[4]

A better counterfactual test for p's being epiphenomenal is to ask how things would have been different if some point which in fact instantiates p had failed to instantiate p *while still being midway between two occupied points.* The only sensible answer to this question, or so it seems to me, is that things would have been just as they actually are in every other way. So this test confirms the intuitive judgment that p is an epiphenomenon.

5.3 Metaphysical and nomological redundancy

It was part of the description of World X that (1) was nomologically necessary; I didn't specify, however, whether (1) was *metaphysically* necessary in addition. But the description I gave was enough to conceptually entail that p was an epiphenomenon. So the question whether (1) is metaphysically necessary is irrelevant to the question whether p is an epiphenomenon.

[4]I address the question which of these responses is correct in footnote 8 of chapter 2 (p. 53).

Why would anyone think otherwise? They might be influenced by an argument parallel to the argument from counterfactuals that I dismissed in the penultimate paragraph of section 5.2. The inference from ⌜It is metaphysically necessary that φ⌝ to ⌜If it had been the case that ψ, it would have been the case that φ⌝ is widely held to be valid, unlike the corresponding inference from ⌜It is nomologically necessary that φ⌝. If this is right, then if it is metaphysically necessary that a point instantiates p just in case it is midway between two occupied points, it follows that if x had failed to instantiate p, it would not have been midway between two occupied points. The rest of the argument goes on as before.

If considerations of metaphysical necessity were relevant to causal questions in the way assumed by this argument were any good, the objection that one's theory entails that entities of a certain sort are epiphenomenal would be all too easy to sidestep just by adopting some unusual essentialist thesis. For example, there would be an easy way for a substance dualist to avoid being an epiphenomenalist about minds, without having to deny that particles in peoples' brains are governed by the same simple and exceptionless laws of physics as particles anywhere else. All the dualist need do is adopt the thesis that it is essential to any mind that its mental state corresponds with the physical state of a given body; and essential to any body that its physical state corresponds to the mental state of its associated mind. By an analogue of the above argument, it would would follow that minds were not epiphenomena, but every bit as causally efficacious as their associated bodies. But this is all wrong: the objection that one's theory posits epiphenomenal entities isn't that easy to avoid!

So there must be something wrong with the argument. I've already argued that we should reject the inference from ⌜It is metaphysically necessary that φ⌝ to ⌜If it had been the case that ψ, it would have been the case that φ⌝, so I won't repeat this

argument here (see p. 79 above). But even if the inference rule is valid—or at least valid on one way of understanding the counterfactual—the argument might be bad for some other reason. Perhaps the inference from 'if x had lacked p, many other things would have gone differently' to 'p is causally efficacious' is bad. Or perhaps the description I gave of World X was enough to analytically entail that (1) is not metaphysically necessary.

5.4 Other simple worlds

World X is typical of worlds which extend Simple World by adding a single nomologically redundant property. For the vast majority of ways of filling in the right hand side of (1), the intuition that the extra property is epiphenomenal is just as strong. The more complicated and arbitrary we make the law, the more intuitively obvious it is that the property is an epiphenomenon.

One might be tempted to conclude from this survey of cases that at every world that is derived from some economical world by adding a single nomologically redundant property, the property in question is epiphenomenal. This is close to being true, but it is not entirely true. Suppose the economical world we start with is just like Simple World, except that there is a special extra law which says that if a particle ever comes to be midway between two other particles, it will immediately blink out of existence. To be more precise:

(2) If a particle x occupies a point y that is midway between two other occupied points, x does not occupy any point that is later than y.

Let World Y be derived from this variant of Simple World by adding a nomologically redundant property p, subject to just the same law (1) as at World X. It is not at all

clear that p is an epiphenomenal property at World Y. Given the new law, it is natural to think of the p-instantiating points at World Y as special "danger zones", with the power to annihilate the particles which comes to occupy them. When a particle ceases to exist, it does so in the first instance because it occupies a p-instantiating point; the point instantiates p, in turn, because it is midway between two occupied points. But if (*per impossibile*) the point had failed to instantiate p despite being midway between two occupied points, it would not have been deadly, and the particle would not have ceased to exist. Thus p passes our counterfactual test for being causally efficacious at World Y.[5]

What is the relevant difference between World X and World Y, in virtue of which p is epiphenomenal at World X but not (or not determinately) an epiphenomenon at World Y? At both worlds, p obeys (1) and so is nomologically redundant. At both worlds, there are also many other laws involving p, which follow from (1) together with laws that make no mention of p. At World Y, there is a very salient example of such a law:

(3) If a particle ever occupies a point which instantiates p, then it does not occupy any later point.

But at World X too there are laws other than (1) that involve p. What they are depends on the details of the laws of Simple World; a typical example might look something like this:

[5] I don't say that this way of thinking of World Y is *inevitable*. I find I can flip back and forth between thinking of p as a causally efficacious property which makes the points that instantiate it into "danger zones", and thinking of it as an epiphenomenal property which happens to be a useful way of tracking a purely extrinsic source of danger. This suggests that the best thing to say might be that it is indeterminate whether p is epiphenomenal at World Y. But the question is a delicate one, and I would not want to rest anything important on this claim.

(4) If there are exactly three particles, and one of them occupies a point that instantiates p, then its acceleration at that point is zero.

So wherein lies the difference? It seems to me that the following answer fits both the intuitive data and the psychological process which generates the intuitions: If one wanted to give a succinct axiomatisation of the laws of World Y, one might well choose (3) as one of the axioms. By contrast, if one wanted to give a succinct axiomatisation of the laws of World X, one would certainly not want to mention (4). In fact the only law involving p that would be any use as an axiom at World X is (1) (or something logically equivalent to it). Thus, p is not at all *integrated* into the structure of the laws of World X; it is, as it were, an afterthought—a "nomological dangler", in J.J.C. Smart's apt phrase (Smart 1959). And that's what makes it epiphenomenal.

Let's call laws that are axioms in simple axiomatisations of the set of all nomologically necessary truths—stated in a language which contains names for all attributes, and whose only nonlogical predicate is 'instantiate'—"basic laws".[6] And let's call any law of the form

(5) For all $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate a iff $\varphi(x_1 \ldots x_n)$,

where a is not mentioned in the expression $\varphi(x_1 \ldots x_n)$, and $\varphi(x_1 \ldots x_n)$ is nomologically equivalent to an expression in which all quantifiers are restricted to things other

[6]This "best system" analysis of the distinction between basic and derived laws was suggested to me by Jonathan Schaffer; in unpublished work he develops a theory of causation in which the notion of a basic law plays a central role. The analysis doesn't presuppose any particular view of what it is for something to be a law of nature in the first place. However, it would be very natural to combine the view with the "best system" analysis of the distinction between laws of nature and other truths, as presented for instance in Lewis 1994. According to this analysis, a sentence expresses a law of nature iff it is a consequence of the system of axioms which best balances simplicity and informativeness—thus, there is an analytic guarantee that the system of laws of nature *can* be axiomatised reasonably succinctly.

than a, a *redundancy law for* a. Our reflections on World X and World Y suggest that whether an attribute is epiphenomenal depends on the role it plays in the basic laws. While we are not yet in a position to give a full theory of how this works, we can at least state a sufficient condition for an attribute to be epiphenomenal:

S An attribute a is epiphenomenal if the only basic laws in which a is mentioned are redundancy laws for a.

This condition is met by property p at World X, but not at World Y.[7]

It follows from this theory that if we start with an economical world w, and extend it into a world w^+ by adding a single new attribute, governed by a redundancy law

(6) For all $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate a iff $\ldots$,

then *almost all* of the ways of filling in the dots will make the new attribute be an epiphenomenon. For it not to be an epiphenomenon, the expression on the right hand side must play a very special role vis-a-vis the basic laws of w. It's not easy to give an informative account of exactly what playing this special role amounts to.[8] But at

[7] If, as I suggested in footnote 5, it is indeterminate whether p is epiphenomenal at World Y, we might want to refine our understanding of S in such a way as to make it indeterminate whether the condition is satisfied by p at World Y. We could do this by treating 'the basic laws' as indeterminate in denotation between each of the sufficiently simple sets of axioms, rather than denoting their union.

[8] There is a temptation to think that the expression has to actually appear in some basic law of w; but this is a mistake. We could expand Simple World by adding a new binary relation r, governed by the redundancy law

R For any x_1, x_2, x_1 bears r to x_2 iff there is a y such that x_2 is midway between y and o, and x_1 occupies y.

(o is some arbitrarily-selected point.) S correctly fails to imply that r is an epiphenomenon. For at this world, one can state the laws just as simply by mentioning r throughout instead of *occupation*; only the values of the constants have to be changed. (And since we are working with Field-style intrinsic formulations of the laws, this isn't

least this much is pretty obvious: for a to be causally efficacious at w^+, the expression on the right-hand side of (6) must be of *roughly the same order of simplicity* as the expressions that occur in the basic laws of w.

5.5 Epiphenomenal properties at more complicated worlds

When we consider worlds derived from economical worlds by adding more than one nomologically redundant attribute, things get a bit more involved. Let's look at World Z, which extends Simple World by adding two new properties p and q, subject to the laws

(1) For any x, x instantiates p iff x is a spacetime point which lies midway between two occupied spacetime points.

(7) For any x, x instantiates q iff x is a spacetime point which lies midway between two spacetime points, each of which in turn lies midway between two occupied spacetime points.

If we had added either of these properties on its own, it would have been a completely clear case of an epiphenomenon. But with both of them in the picture, it is not so clear what to say. Given (1), (7) is nomologically equivalent to the simpler law

(8) For any x, x instantiates q iff x is a spacetime point which lies midway between two spacetime points which instantiate p.

---

really any change at all.) Nevertheless, the expression on the left-hand-side of R does not appear in anything that is even a remote candidate to be a law of Simple World: at Simple World, point o is just like any other point, and "midway between" has no special significance.

It is natural to regard (1) and (8) as the basic laws, and (7) as a derived law: if this is correct, S does not imply that p is epiphenomenal. And intuitively, p does seem to have a certain limited sort of causal efficacy. It is natural to think that the explanation of why any given point instantiates q goes in two steps: it instantiates q because it is midway between two p-instantiating points; they instantiate p, in turn, because they are midway between two occupied points. Thus, the facts about which things have p are causally relevant to the facts about which things have q. If (*per impossibile*) p had been instantiated by points other than those midway between occupied points, different things would have instantiated q—(8) would still have been true, while (7) would not.

I think we can grant that this is the right thing to say about World Z without denying that *in the epistemologically relevant sense we are interested in*, both p and q should count as epiphenomenal. Here is an analogy which shows how an entity can be an epiphenomenon in the relevant sense even if it has a certain very limited causal efficacy.[9] Suppose we start with a world where epiphenomenalist substance dualism is true—there are *souls* in addition to physical things, and the state of each soul is nomologically determined by the state of its corresponding body—and add a new category of *spirits*. Every spirit corresponds to a soul, and the state of a spirit is nomologically determined by the state of its corresponding soul in the same way that the state of a soul is determined by the state of its corresponding body. We must agree that in this new world, souls are not *entirely* epiphenomenal. Each soul is a necessary intermediary in the causal chain whereby the state of a certain body determines the state of a certain spirit. If, *per impossibile*, the state of some soul had failed to correspond to the state of its corresponding body, the state of the spirit that

[9]I have already presented this analogy in footnote 4 of chapter 2.

corresponded to that soul would have been different as well. Nevertheless, the souls are epiphenomenal in the sense we care about. Nothing about the physical world, or about any other soul, or about any spirit other than the one associated with it, is at it is because of a given soul. Likewise, p is epiphenomenal in the sense we are interested in, because the only facts that depend on p are the facts about which things have the entirely epiphenomenal property q.

How can we extend S so that it gives us a more useful sufficient condition for an attribute to be epiphenomenal in the sense we are interested in, one which for instance classifies property p at World Z as epiphenomenal? One idea would be to deem an attribute epiphenomenal whenever the only basic laws in which it is mentioned are redundancy laws, irrespective of whether they are redundancy laws *for it*. But this is too strong. We can imagine a world that stands to World Z much as World Y stands to World X: at this world, whenever a particle occupies a point that instantiates q, it immediately blinks out of existence. At such a world, q is not (or at least not determinately) epiphenomenal; nor is p, since p plays a role in determining the extension of q. But in the simplest axiomatisation of the laws of this world, p is mentioned only in the redundancy biconditionals (1) and (8).

I think it is pretty clear what we should say about this. An attribute can get to be causally efficacious either directly, by being mentioned in some basic law that is not a redundancy law; or indirectly, by "inheriting" the causal efficacy of the attributes in whose redundancy laws it is mentioned. If it doesn't get to be causally efficacious in either of these ways, the attribute is epiphenomenal. Thus, we can state a sufficient condition for a property to be epiphenomenal as an infinite conjunction:

S′ An attribute a is epiphenomenal if

(i) All the basic laws in which a occurs are redundancy laws, and

(ii) Whenever a occurs in a basic law which is a redundancy law for an attribute b, b satisfies (i), and

(iii) Whenever a occurs in a basic law which is a redundancy law for an attribute b, b satisfies (ii), and

(iv) ...

This condition correctly counts both p and q as epiphenomenal at World Z, but not at the world where q-instantiating points are deadly to particles.

S′ is a stringent condition: it makes it very hard for nomologically redundant attributes to avoid being epiphenomena. If w is an economical world, and w^+ extends w by adding some new attributes $a_1 \ldots a_m$, each governed by a law of the form

(6) For all $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate a_i iff ...

in which the only attributes mentioned on the right-hand side are those of w, then it will almost always be the case that almost all of the new attributes are epiphenomena. For one of the a_is not to satisfy S′, the expression on the right-hand side of the law that introduces it must play a very special role vis-à-vis the laws of w—and to play this role it must *at the very least* be roughly of the same order of simplicity as the expressions that occur in the basic laws of w. In fact we can omit the condition that w should be an economical world: even if w already contains some nomologically redundant attributes, the new attributes added at w^+ will still be epiphenomenal if the redundancy laws that govern them are considerably more complicated than any of the basic laws of w.[10]

[10]I would like to be able to give a proof of this claim, but I can't see where to begin. What I do have, however, is a response to an argument which purports to show that

This conclusion seems quite intuitive. When you extend Simple World by adding a *single* property governed by a redundancy law that is much more complex than the basic laws of Simple World, it is very clear that the new property will be epiphenomenal. And there is no intuitive obstacle to our combining many of these epiphenomenal properties in the same world; how could just adding more of them make them not be

---

it is false. What the argument tries to show, specifically, is this: whenever w_1 extends w_0 by adding a nomologically redundant property p, one can find a w_2 which extends w_1 by adding a second nomologically redundant property q, such that at w_2, p does not satisfy S′.

Here's the argument. Suppose that the nomologically redundant property at w_1 is governed by the law

A For all x, x instantiates p iff $\varphi(x)$;

At w_2, we introduce a second property q, governed by the law

B For all x, x instantiates q iff (L iff there is no y such that $\varphi(y)$)

where L is the conjunction of some set of basic laws for w_0. All the laws of w_2 follow from A, B and

C Either something has p and nothing has q, or nothing has p and everything has q.

This set of axioms is quite simple; and in it, p is mentioned in a law, C, which is not a redundancy biconditional. Therefore S′ does not imply that p is an epiphenomenon at w_2.

The right response to this argument is just to deny that the system whose axioms are A, B and C is a sufficiently simple axiomatisation for these to count as (candidates to be the) basic laws of w_2. There is a simpler system: it comprises A,

B′ For all x, x instantiates q iff there is no y such that $\varphi(y)$

and L, the conjunction of the basic laws of w_0. Even if simplicity is measured just by counting symbols, this system is a good bit simpler. And to the extent that symbol-counting is an inadequate measure of simplicity, one effect of the departure should be to exact an additional penalty on systems like {A, B, C} which are derived by gratuitous logical trickery from systems with a more straightforward structure.

epiphenomena?[11] It is true that as we add more nomologically redundant attributes to Simple World, it gets harder to have an immediate intuition about what's going on as far as causation is concerned: it's just too confusing to try to hold all those attributes in one's mind at once. That's why I took us on this detour through all these funny simple worlds, rather than just announcing that P1 struck me as intuitive and leaving it at that. But the fact that S′ fits our intuitions so well at worlds simple enough to have clear intuitions about, together with the fact that the difference between worlds with lots of nomologically redundant attributes and worlds with few of them is intuitively not relevant to whether the attributes in question are

[11]It might be suggested that, to test whether a given nomologically redundant property at a world that extends an economical world by adding several nomologically redundant properties is epiphenomenal, we should see whether it would be epiphenomenal at a world in which all the other nomologically redundant properties were deleted. This test arguably gives the right result in the cases we have considered so far. But it doesn't always give the right result. Consider a variant of Simple World where instead of the single relation *occupation*, there are two relations $occupation_1$ and $occupation_2$. Each particle either occupies$_1$ some points, or occupies$_2$ some points; no particle does both. Suppose we add a nomologically redundant relation r_1 to this variant world, governed by the law

> For all x, y, x bears r_1 to y iff x occupies$_1$ a point midway between p and y

(where p is some arbitrarily selected point). At this world, r_1 is epiphenomenal; there is no possible excuse for mentioning it in any interesting basic law. But now suppose that we add a second relation r_2, governed by the law

> For all x, y, x bears r_2 to y iff x occupies$_2$ a point midway between p and y.

Now the situation is very different. Just as in the corresponding case discussed in footnote 8, the laws of this world can be axiomatised as easily using r_1 and r_2 as using $occupation_1$ and $occupation_2$. In fact the very same laws will do, since the paths traced out by particles under r_1 and r_2 are derived by a linear transformation from the paths they trace out under $occupation_1$ and $occupation_2$.

epiphenomena, constitutes a good case for accepting S′ even at complicated worlds.

5.6 Higher-order attributes

A long time ago, I set out to argue for the following claim:

P1 If there are more than a few nomologically redundant attributes, then almost all of them are epiphenomenal.

The conclusion of the previous section can be regarded as a precisification of this claim. Let me restate that conclusion in such a way as to make this clearer. Say that a world w *extends* a world w^- iff (i) the only entities that exist at w but not w^- are attributes, and (ii) attributes that exist at both worlds have the same extension at both worlds, and (iii) for each attribute a that exists at w but not at w^-, there is a law of nature of the form

For any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate a iff ...

where the only attributes mentioned on the right-hand side are those that exist at w^-. Using this notion, we can state the main conclusion of section 5.5 as follows:

P1′ Whenever an attribute a is nomologically redundant at a world w, a is epiphenomenal, *unless*: whenever w^- is a world that w extends, either a exists at w^-, or there is some law of the form

For any $x_1 \ldots x_n$, $x_1 \ldots x_n$ instantiate a iff ...

where the right-hand side mentions only the attributes that exist at w^-, and is not substantially more complex than the expressions which feature in the basic laws of w^-.

P1′ describes an interesting tendency for nomologically redundant attributes to be epiphenomenal, which holds across a wide range of possible worlds. However, it is not yet strong enough for our purposes, because is completely silent about what happens at worlds which contain nomologically redundant properties but are not extensions of any worlds other than themselves. I can see only one way for it to happen that a world contained nomologically redundant attributes without being an extension of any world other than itself, and that is for the world to contain some *higher-order* properties and relations, instantiated by the nomologically redundant attributes.

It is conceivable that a nomologically redundant attribute which would otherwise have been completely epiphenomenal could get to be causally efficacious by instantiating a higher-order relation. For example, we can imagine a world very like World X, except that in addition to the four relations from Simple World and the nomologically redundant property p, there is also a new binary relation, "*loving*". Certain particles love p; other particles don't. And even though it doesn't matter at all which things *instantiate* p, it does matter which things *love* p: when two particles both love p, there is an extra repulsive force between them (they are rival lovers). It seems wrong to classify p as an epiphenomenon at this world. It is epiphenomenal *as far as the facts about which things instantiate it are concerned*, but it gets to be causally efficacious by being the object of the particles' love.

The causal efficacy of p at this world depends on the fact that there is no law of nature of the form

> For any x, x loves p iff ...,

where neither p nor *loving* is mentioned on the right-hand side. If we just had a property of "loving p" instead of the relation of loving, the property wouldn't be

nomologically redundant. When we consider worlds where there are laws of this sort, we will (almost always) find that we no longer want to say that p is causally efficacious. Consider, for example, a world which is just like World X, except for the new relation of *loving*, governed by the law that a particle loves p iff there are seven particles that are all the same distance away from it. We are no more tempted to think of p as causally efficacious at this world than we would be tempted to think of a property that was, as a matter of law, *instantiated* by all and only those particles for which there are seven particles that are all the same distance away from it as causally efficacious. In fact, this seems to be a general test for whether a property gets to be causally efficacious by standing in higher-order relations: see whether you would count the property as causally efficacious if the predicate expressing the higher-order relation was replaced in the laws by 'instantiate'. Thus, a full discussion of the causal efficacy attributes can get to have by standing in higher-order relations would just recapitulate our earlier discussion of sort of causal efficacy an attribute gets to have when it matters which things instantiate it.[12]

I won't go through all the steps of this discussion, but will merely assert a strengthened version of P1′, which I claim we will end up with if we give an account of the sort of causal efficacy conferred by higher-order relations which parallels the account the ordinary sort of causal efficacy which we have already given. To do this, I will need

[12]Even if I am wrong about this, the conclusion that all but a few nomologically redundant attributes are epiphenomenal *as far as the facts about which things instantiate them are concerned* is still pretty bad news for the opponents of Nihilism. It seems pretty intuitive, for example, that if there is such a thing as the property of being a dog, certain creatures behave as they do because they instantiate it. As far as I can see, of the strategies for justifying belief in complex attributes which I discussed in section 4.4, the only one which wouldn't be undermined by the weaker conclusion is the appeal to introspection.

to articulate a sense in which a world can be an "extension" of another world even if it contains new attributes which instantiate higher-order relations. Let's say that w^- is a *nomological core* of w iff (i) the only entities that exist at w but not w^- are attributes, and (ii) attributes that exist at both worlds have the same extension at both worlds, and (iii) under the laws of nature of w, the facts about the distribution of the attributes that exist at w^- nomologically determine the facts about the distribution of all attributes (including higher-order relations). The appropriate successor of P1′ can be stated in terms of this notion:

P1″ Whenever an attribute a is nomologically redundant at a world w, a is epiphenomenal, *unless*: whenever w^- is a nomological core of w, either a exists at w^-, or there is some law of the form

> For any $x_1 \ldots x_n$, $\varphi(x_1 \ldots x_n, a)$ iff $\psi(x_1 \ldots x_n)$
>
> where $\varphi(x_1 \ldots x_n, a)$ is an atomic sentence in which the predicate is 'instantiate', and the variables 'x_1'…'x_n and the name 'a' occur, and the names of any other attributes that exist at w may also occur; and $\psi(x_1 \ldots x_n)$ is an expression in which only the attributes that exist at w^- are mentioned, which is not substantially more complex than the expressions which feature in the basic laws of w^-.

This is the closest I will come to a precise statement of the "intended" interpretation of P1. P1″ sets a very demanding condition which a nomologically redundant attribute has to fulfil in order to escape being an epiphenomenon. If nomologically redundant attributes are even close to being as numerous as they are according to the Universalist about complex attributes, the number of nomologically redundant attributes which fulfil the condition must be very small.

In section 5.1 I said that on its intended interpretation, P1, together with a "plausible" additional premise about the actual world, would entail that if any term of the form ⌜the property of φing⌝ refers to a non-epiphenomenal property, φ must be a short expression in the language of physics. Now that we have specified the intended interpretation more precisely, we can see what the additional premise needs to be:

(9) The actual world has a nomological core w^- such that, whenever there is a law of the form

> For all x, $\varphi(x)$ iff $\psi(x)$
>
> —where φ is an expression in English, and ψ is an expression in a language whose only predicate is 'instantiates' and whose only names are the names of the attributes that exist at w^-, which is not substantially more complicated than the basic laws of w^-—φ is a short expression in the language of physics.

The plausibility of this claim doesn't exactly leap out at one; but properly understood, it is fairly uncontroversial. Almost everyone who believes in attributes at all believes that there are some attributes, which are either expressed by the predicates of fundamental physics or not expressed by any predicates in our current language, such that the facts about the distribution of these attributes determine, with nomological necessity, the facts about the distribution of all the attributes there are. (The question whether this determination is *metaphysically* necessary as well as nomologically necessary is much debated—it is the question whether physicalism is true, on one definition of 'physicalism'—but the weaker claim is generally embraced even by the most vehement opponents of physicalism.) Moreover, almost everyone agrees

that the laws governing the distribution of these attributes are extremely simple and natural—at least as simple and natural as the laws of our current best theories in fundamental physics. (The exceptions to this are sceptics about the unobservable like van Fraassen (1980).) When one looks at these theories, it is very plausible that the degree of complexity involved in the statement of the laws is slight, compared to the degree of complexity that would be involved in the statement of any law that gave nomologically necessary and sufficient conditions, in terms of the chosen attributes, for an object to be yellow, or a dog, or in pain, or even an oxygen atom. . .—any predicate, in fact, other than short expressions in the language of physics. And that's all that (9) says: just take w^- to be the world one gets when one deletes all but the chosen attributes from the actual world.

For all I have argued in this chapter, it may be that canaries look yellow because they are yellow; that Fido barks because he is a dog; that I yawn because I am tired; that blood cells are red because they contain oxygen atoms. But it is not the case that canaries look yellow because they have the property of being yellow; that Fido barks because he has the property of being a dog; that I yawn because I have the property of being tired; or that blood cells are red because they contain things that have the property of being an oxygen atom.[13]

[13]Of course, if the main argument of chapter 2 is sound, there aren't any such things as canaries, dogs, people or blood cells. Moreover, there is a good case, completely independent of the question what theory of complex attributes is true, for the nonexistence of the relation *parthood* and the properties *being yellow*, *being a dog*, *being tired* and *being an oxygen atom*—on the grounds that the predicates 'is part of', 'is yellow', 'is a dog', 'is tired', 'is an oxygen atom' are all semantically defective. (See footnote 2 of chapter 3.)

Chapter 6

Fictions of complex attributes

6.1 The advantages of unreflectiveness

I suspect that there are fewer than $10^{10^{100}}$ things, all told. If there are in fact fewer than $10^{10^{100}}$ things, this suspicion of mine is true. Anyone who believes what I suspect to be the case believes something true, namely the proposition that there are fewer than $10^{10^{100}}$ things.

This innocent-seeming paragraph is in fact quite mysterious. However we work out the details of the fiction of complex attributes, there will surely turn out to be many more than $10^{10^{100}}$ things according to the fiction. In fact there will be infinitely many things in the world of the fiction: for example, the propositions *that there is at least one thing*, *that there are at least two things*, *that there are at least three things*.... So in a context in which correctness is determined by truth in the fiction of complex attributes, 'there are fewer than $10^{10^{100}}$ things' should be unassertable, since it is not fictionally true.

No problem there, as yet. Anyone who takes a fictionalist approach to some part

of our ordinary conversational practice can and should recognise that we can slip in and out of fiction-governed talk with the greatest of ease—even within the same sentence. But what are we to make of the sentence 'The proposition that there are fewer than $10^{10^{100}}$ things is true'? We know that it is not strictly and literally true, since the expression 'the proposition that there are fewer than $10^{10^{100}}$ things' doesn't refer to anything. But given that the fiction is one according to which there are more than $10^{10^{100}}$ things, there also seems to be a problem with the claim that the sentence is *fictionally* true. If the Proposition Schema:

The proposition that φ is true iff φ

holds in the fiction, then according to the fiction, the proposition that there are fewer than $10^{10^{100}}$ things is *not* true.

To explain what is going on in this case, we must suppose that the Proposition Schema does *not* always hold in the fiction according to which I was talking in the last sentence of the first paragraph of this chapter. The inference $\ulcorner\varphi$; therefore the proposition that φ is true$\urcorner$ and its converse do not preserve truth in this fiction. Rather, the relevant inference rules are as follows. From a sentence φ, uttered in a context which is *not* governed by the fiction, one can infer $\ulcorner$the proposition that φ is true$\urcorner$, where this is now uttered in a context which *is* governed by the fiction. Likewise in reverse: the inference from $\ulcorner$the proposition that φ is true$\urcorner$ to φ is only valid when it is accompanied by a change from a fiction-governed context to one governed by the standard of strict and literal truth. In other words, the following schema holds:

(1) φ iff according to the fiction, the proposition that φ is true.

Hence the Proposition Schema fails, according to the fiction, whenever really φ but not according to the fiction φ, or really not φ but according to the fiction φ.

This point about propositions applies equally to the case of properties and relations. To report my suspicion that I coexist with fewer than $10^{10^{100}}$ things, I might ostentatiously say 'A property I suspect myself to have is that of coexisting with fewer than $10^{10^{100}}$ things'. We need an account of the fiction governing this utterance on which one can consistently suppose that I have that property, in the fiction. The obvious explanation is that the predicate 'coexists with fewer than $10^{10^{100}}$ things' is analytically equivalent to the predicate 'is something which according to the fiction has the property of coexisting with fewer than $10^{10^{100}}$ things'. Thus, the schemata that govern the instantiation of properties and relations in the fiction are

(2) For any x, x φs iff according to the fiction, x has the property of φing.

(3) For any $x_1 \ldots x_n$, $\varphi(x_1 \ldots x_n)$ iff according to the fiction, $x_1 \ldots x_n$ instantiate the relation of being a $y_1 \ldots y_n$ such that $\varphi(y_1 \ldots y_n)$.

So it can happen that according to the fiction something has the property of φing but does not φ: this will be so whenever something really φs but does not φ according to the fiction.

You end up saying some rather odd-sounding things when you speak according to a fiction for which (1)–(3) hold. For example, if there aren't actually any propositions, it will be correct to assert the sentence

(4) The proposition that there are no propositions is true.

This certainly feels paradoxical enough that one would feel uncomfortable asserting it. But the more I think about this question, the more it strikes me that if there really weren't any propositions, (4) would be exactly the right thing to say, at least in many contexts. It is interesting that the following sentences seem much more palatable than (4):

(5a) The theory that there are no propositions is true.

(5b) The doctrine that there are no theories is true.

(5c) The proposition that there are no doctrines is true.

If one of these sentences occurred in the context of a serious paper on the metaphysics of propositions (or theories, or doctrines) we wouldn't bat an eyelid. We would react just as if the author had said 'there are no propositions/theories/doctrines': if we agreed that there were no propositions/theories/doctrines, strictly and literally speaking, we would take the assertion to be correct. This suggests that our uneasiness about (4) has a superficial source in our desire not to say things that sound overtly paradoxical. Once we get clear about the distinction between correctness and strict and literal truth, the oddness of (4) is liable to evaporate.

While it isn't correct in all contexts, it seems to me that the way of talking about attributes captured by (1)–(3) is quite common. We fall into it naturally in those circumstances when we aren't being philosophically self-conscious about our apparent reference to attributes, but are merely using attribute-talk as the indispensable aid to clear communication that it is. And even when we *are* being self-conscious about the metaphysics of properties, our need to use attribute-talk to get our point across is often so strong that we often find ourselves having to talk about 'theories', 'claims', 'statements', 'concepts', 'notions', 'features'... in the sort of way that only makes sense in a fiction governed by analogues of (1), (2) and (3). To mark the fact that they play this important role, I will refer to fictions about complex attributes for which (1), (2) and (3) hold as *unreflective* fictions.

Even if we never made the sorts of claims about attributes that only make sense within unreflective fictions, it would still be worth paying attention to these fictions,

and seeing how much of our ordinary attribute-talk could be reconstructed in terms of them, for two reasons. First, we might find that thinking about this simple sort of fiction puts us in a better position to see how to construct a more elaborate fiction, adequate to those occasions when one can correctly say things that don't make sense within an unreflective fiction. In section 6.3 I will explore a way of understanding this more elaborate sort of fiction on which unreflective fictions still turn out to have foundational importance.

Second, many unreflective fictions have a important feature which sets them apart from any fiction in which all the instances of the original Property Schema, Relation Schema and Proposition Schema are true: they are consistent.

In a fiction in which all the instances of the Property Schema

> For any x, x instantiates the property of φing iff x φs

were true, it would be true in particular that

(6) According to the fiction, for any x, x instantiates *non-self-instantiation* iff x doesn't instantiate itself.

The 'according to the fiction' operators I have considered so far have all been ones which preserve classical logical entailments. Suppose that the 'according to the fiction' operator in (7) works like this. Then, since *non-self-instantiation* exists according to the fiction, it follows by universal instantiation that

(7) According to the fiction, *non-self-instantiation* instantiates itself iff it doesn't instantiate itself.

or equivalently

(8) According to the fiction, *non-self-instantiation* both instantiates itself and doesn't instantiate itself.

From this in turn it follows that

(9) According to the fiction, φ

is true for all φ. Obviously an operator which invariably produced truths no matter what sentence it was attached to would be of no use at all in making sense of the standard of correctness that governs ordinary talk of properties. True, thanks to the efforts of relevance logicians and other logical heretics, we know of ways to avoid this "explosion" while still being able to hold that *most* classically valid inferences are preserved under 'according to the fiction'. But the explosion isn't the only problem with fictions in which (8) is true. There is also the fact that some of the things which are true according to these fictions just don't seem to be correct to assert: for instance, '*non-self-instantiation* both does and doesn't instantiate itself'; 'the proposition that *non-self-instantiation* both does and doesn't instantiate itself is true'; 'some contradictory propositions are true'. We could learn to regard these as correct assertions; but at present, we don't.

Paraconsistent fictions in which all the instances of the Property Schema are true are certainly worth exploring.[1] So are fictions which block the inference from (7) to

[1]If we wanted to make sense of some important aspect of ordinary talk using a fiction within which contradictions are sometimes true, we would have a nice way to split the difference between logical orthodoxy and the "dialethist" view that some contradictions are true, held for example by Graham Priest (Priest 1998). We can deny—we should be outraged at the very suggestion—that any contradictions are *strictly and literally* true. But we might nevertheless agree that some contradictions are *fictionally* true, and hence can correctly be asserted. The question whether contradictions are ever *true* then comes down to the dispute between semantic and pragmatic interpretations of fictionalism, which as I suggested in section 1.3 may not be all that significant.

(8). But it is also interesting to see how far one can get in understanding our practice in terms of the more familiar, classically consistent sort of fiction. And the answer, I think, is that one can get quite far: (2), unlike the Property Schema, can hold even in a classically consistent fiction. The instance of (2) which corresponds to (6) is

(6′) For any x, x x doesn't instantiate itself iff according to the fiction, x instantiates *non-self-instantiation*.

This leaves us free to say whatever we like about whether, according to the fiction, *non-self-instantiation* instantiates itself, provided that we deny that *non-self-instantiation* exists, strictly and literally speaking. And surely we should deny this. Even apart from the argument for Nihilism about complex properties which I presented in chapter 5, there is a rather strong argument against the existence of *non-self-instantiation* in particular. The premise of this argument is the following analytic-seeming claim:

(10) For any x, if x is *non-self-instantiation*, then x is instantiated by all and only those things that do not instantiate themselves.

We should admit that *non-self-instantiation* is a mere creature of fiction; hence (2) is completely silent as regards the question which properties it instantiates, according to the fiction.[2]

[2]Even if we believed that *non-self-instantiation* really existed, (2) still wouldn't lead us into contradiction. All that would follow is that the fiction misrepresents *non-self-instantiation*'s relation to itself: if *non-self-instantiation* really doesn't instantiate itself, then it does instantiate itself according to the fiction, and if it really does instantiate itself, then it doesn't instantiate itself according to the fiction.

6.2 The principles of the fiction of complex attributes

To explain what it takes for a sentence about attributes to be correctly asserted in an "unreflective" context, we need an "according to the fiction" operator for which all the instances of (1), (2) and (3) are analytic truths. One way to define such an operator is to use the counterfactual conditional (or one of the related conditionals described in sections 3.1 and 3.4). If we adopt this strategy, our task is to come up with a sentence ψ, "the principles of the fiction", for which the operator 'If it had been the case that $\psi\ldots$' behaves the way we want 'according to the fiction' to behave.

One obvious way to do this is to take ψ to be the conjunction of all the instances of the following three schemata:

> **Unreflective Proposition Schema:** For all real $y_1 \ldots y_n$, (the proposition that φ is true iff really, φ).
>
> **Unreflective Property Schema:** For all real $y_1 \ldots y_n, x$, (x instantiates the property of φ-ing iff x really φs).
>
> **Unreflective Relation Schema:** For all real $y_1 \ldots y_n, x_1 \ldots x_m$, ($x_1 \ldots x_m$ instantiate the relation of being m things such that φ iff $x_1 \ldots x_m$ are really m things such that φ).

Since these schemata have infinitely many instances, we will have to take ψ to be an infinite conjunction, or what amounts to the same thing, a universal substitutional quantification with all the instances of the schemata as substituends. This doesn't bother me, since I think I understand substitutional quantification perfectly well.[3]

[3]A caveat: sentences derived from these schemata by replacing φ with a *semanti-*

The point of the initial quantifier 'for all real $y_1 \ldots y_n$' is to make sure that claims like these are among the principles of the fiction:

> For all really existing y, the proposition that y is happy is true iff really, y is happy.
>
> For all really existing y and x, x instantiates the property of loving y iff x really loves y.

If we didn't have these principles, the schemata (1)–(3) wouldn't hold in full generality: for instance, we wouldn't be able to replace φ with 'according to the fiction, the proposition that φ is true' when φ contained free variables. But clearly our ordinary way of talking about attributes does allow us to quantify into abstract terms.

Unlike the principles (M1–M8) of the mereological fiction I described in section 3.3, the unreflective attribute schemata leave some questions about the world of the fiction unresolved, even when we are given all the facts about the real world. In particular, they say nothing about the instantiation of properties and relations by entities that don't really exist—which, if the arguments of chapter 5 is sound, include all or almost all complex attributes. One could consistently extend the unreflective attribute schemata by adding any of the following additional principles: no fictional thing instantiates any property; every fictional thing instantiates every property; every fictional thing instantiates exactly those properties which are instantiated by some pig. If we don't settle such questions by adding new principles, claims about the instantiation of properties and relations by fictional things will all be indeterminate in

cally defective expression like 'is a unicorn' or 'orbits between Mercury and Vulcan' or 'Vulcan is a planet' should not be counted as genuine instances (see chapter 4, footnote 7). For it strikes me as correct to say that if there is no such thing as Vulcan, there is no such thing as the property of orbiting between Mercury and Vulcan.

(fictional) truth-value.

Since our aim is to say what it takes for an sentence about attributes to be correctly asserted in an "unreflective" context, it is just fine to leave this indeterminacy unresolved. Any context in which interesting things could correctly be said about the instantiation of properties by entities, such as complex attributes, which don't really exist, would not be an unreflective one. For surely in any such context, it would be correct to assert the sentence

(11) The property of being a red planet has the property of being instantiated by Mars

even if there was really no such thing as the property of being a red planet. (For the moment let's put aside all worries about the existence of composite objects like Mars.) But in any ordinary context in which (11) could correctly be asserted, it would be equally correct to assert

(12) Mars has the property of instantiating the property of being a red planet.

If correctness is governed by an unreflective fiction, (12) can be correctly asserted only if

(13) Mars has the property of being a red planet

is strictly and literally true; and (13) can only be strictly and literally true if there is really such a thing as the property of being a red planet.[4]

[4]Even given that it's not correct to make specific claims like (11) about the properties of fictional things, one might suppose that certain general claims about their properties could still be made. For example, it might be correct to say, e.g., that *everything* instantiates the property of self-identity, or that nothing instantiates both a property and its negation. If we wanted claims like these to be fictionally true,

The unreflective fiction whose principles are all the instances of the unreflective attribute schemata makes perfectly good sense. Nevertheless—by contrast, say, with the principles of the mereological fiction which I presented in section 3.3—the principles of this fiction give us very little insight into the *structure* of the world they describe. They gives us no way to answer to the question *in virtue of what* the property of being φ is the property of being φ. What is it about this entity that makes 'the property of being red' or 'the property of being intelligent' or 'the property of being a dog' a good name for it? One could get used to talking according to a fiction in which questions like these never have answers.[5] Likewise, one could get used to talking according to a fiction in which it was true that some fictional things were statues, while others were pieces of metal, not in virtue of any deeper structural difference between them, but simply because this was explicitly stated in the principles of the fiction (see p. 98 above). But this is a very unnatural way of thinking. Yes,

---

we could easily add them as further principles. That would leave it indeterminate, for instance, whether the property of being a red planet had the property of being instantiated by Mars, but determinate that if it instantiates it it doesn't instantiate the property of not being instantiated by Mars.

However, I think a better account of what makes it correct to say things like 'Everything instantiates self-identity' is that the quantifier is implicitly restricted to things that *really* exist. To put it differently, the 'according to the fiction' operator governs not the whole sentence, but only the predicate 'instantiates self-identity'. Since we regularly switch back and forth between fiction-governed and non-fiction-governed discourse in any case, it wouldn't be at all surprising if we sometimes say things that are best made sense of by putting an "according to the fiction" operator in front of only part of a sentence.

[5]As have, for example, Paul Horwich (1990; 1998) and Stephen Schiffer (1996). Schiffer claims to be a realist about attributes, but his bald insistence that they are nevertheless "language-created" puts him in the category of philosophers whom I am inclined to interpret as not really disagreeing with what *I* mean when I say 'there are no attributes' (see footnote 9 to chapter 1).

some claims about fictional entities must be basic truths of the fiction, which don't hold in virtue of anything else—we shouldn't expect, for example, to have an answer to the question 'in virtue of what do some things have this particle as a part while others don't?' But 'is part of' seems like a natural stopping point in a way that 'is a statue', 'is a piece of metal', 'is the property of being a red planet', 'is the proposition that snow is white', 'is the property of being a dog' and so forth just don't.[6]

[6]Moreover, if one doesn't give any account of the structure in virtue of which things get to be statues, one will be hampered in one's ability to do any sort of satisfying *semantics* within the fiction. Why does 'dealbh' in Irish apply to statues and not to pieces of metal? If statues are statues in virtue of some distinctive sort of ontological structure, we can see how this question might be answered by giving an interesting general theory about how the use of a word determines what sorts of things it will apply to. But if the distinction between statues and pieces of metal is fictionally basic, the most we can hope for in foundational semantics is an unsatisfactory theory of the form 'If a word is used like *this:* [long description of the pattern of use characteristic of 'statue' and 'dealbh' follows] then it applies to all and only statues'.

This problem becomes much more important and general when the fiction we are concerned with is the fiction of complex attributes. If we take all claims of the form $\ulcorner x$ is the property of φing$\urcorner$ as basic, we will have to give up the whole project of foundational semantics. Whenever any expression, in any language, *expresses* a certain property, relation or proposition, one can ask 'What is it about the way this expression is used that makes it express that property/relation/proposition rather than some other one?' If one has an account of the deep structure of the world of attributes, one will be in a position to look for an illuminating general account of the way in which patterns in the use of an expression determine what it expresses. Without any such account, on the other hand, one will have to rest content with a completely fragmentary account, a big conjunction of sentences of the form $\ulcorner$If a predicate is used like *this:* [long description of a pattern of use follows] then it expresses the property of φing$\urcorner$. Some philosophers (e.g. Horwich 1998) embrace the view that this is as far as we can ever get in foundational semantics. This is one of the most central claims of *deflationism* (Field 1994a). But I find this view deeply, and unwarrantedly, pessimistic. In any case, it is worth seeing whether we can set up the principles of the fiction in such a way as not to prejudge the issue.

These remarks about foundational semantics also apply to the fixing of mental content. Only if we have some account of the structure of the realm of attributes can we hope to give an interesting general theory about the ways in which goings-on in a

The first step towards coming up with a more satisfactory list of principles for an unreflective fiction of complex attributes is to include the axioms of some Universalist theory of complex attributes as principles of the fiction. Any of the versions of Universalism I discussed in the appendices to chapter 4 would do for this job. But since we are only talking about a fiction, we need have no qualms about helping ourselves to mathematical entities. We can simply include the axioms of some appropriately powerful set theory as further principles of the fiction, so that the existence of numbers, sequences, etc. in the world of the fiction doesn't depend on the question whether there are really any mathematical entities.[7] And, given that we can help ourselves to all the mathematical entities we need, we might as well adopt the strongest of the available Universalist theories, the one from appendix 4C that allows for arbitrary infinite conjunctions and disjunctions and relations of transfinite degree. We will find this strength helpful in our quest for set of principles which characterises the structure of the realm of complex attributes in an illuminating way while still entailing all the instances of the unreflective attribute schemata.

A few small but crucial changes must be made to the axioms of the Universalist

person's head and environment make it true that the person believes (or desires-true) this proposition rather than that one. In the absence of such a theory, all we can hope for is a big conjunction of fragmentary, proposition-by-proposition theories.

[7]Even if there are some mathematical entities, we can ensure that there are *more* of them in the fiction than in reality if we include the principle that there is a set whose members are all and only those things which *really* exist as one of the principles of the fiction. For many purposes it is indispensable to be able to assume the existence, in the fiction, of such a set. For instance, I noted in section 4C that when we admit relations of arbitrary transfinite degree and arbitrary infinitary conjunctions and disjunctions, the propositions we can build starting with some initial set of attributes include some which correspond up to isomorphism with *models* of the language in which those attributes are the predicates. To have a guarantee that one of these propositions is *true*, we need to be able to assume that there is such a thing, according to the fiction, as the set of all real things.

theory before they will be fit serve as principles for an *unreflective* fiction of complex attributes. First, according to the old principle for universal quantification, the universal quantification of a property p is true iff *everything* instantiates p. But according to the Unreflective Proposition Schema, the proposition that everything is transient is true iff it is *really* the case that everything is transient, i.e. iff every real thing is really transient. Therefore, the proposition that everything is transient—i.e. the universal quantification of the property of being transient—is true iff every *real* thing instantiates the property of being transient. So the rule for universal quantification needs to be changed to:

(14) When a proposition p is the universal quantification of a property q, p is true iff everything *that really exists* instantiates q.[8]

In the same way, we will want to change the rule for existential quantifications so that the existential quantification of p is true iff some *real* thing instantiates p.

Second, according to the old principle for *application*, when p is a property and x is *anything whatsoever*, there is such a proposition as the application of p to x, and it is true iff x instantiates p. If we adopted this as a principle of the unreflective fiction, we would have to countenance a whole host of propositions derived by application to fictional things. This would be odd, and out of keeping with the spirit of an unreflective fiction. Much better, I think, to do away with all such propositions by

[8]Of course this holds not just for the universal quantifications of properties, but for universal quantifications in general. The principle U3* from appendix 4C must be modified so as to read

U3** If $\mathrm{UQ}(u, v, x)$, then for any sequence of real things s, s exemplifies u iff whenever s is a sequence of real things derived from s' by deleting all members whose position in the sequence is a member of x, s exemplifies v.

restricting the scope of the operation of application, so that x must be something that *really exists* for there to be such a thing as the application of p to x. This also lets us hold onto the very natural principles that the universal quantification of a property p is true iff all the applications of p are true, and the existential quantification of p is true if any of the applications of p are true.

With these two changes, the axioms of the Universalist theory of appendix 4C will do just fine as principles of the fiction.[9] Of course they can't be the *only* principles of the fiction. They are compatible with the claim that there are no attributes at all, whereas we want it to be an analytic truth that according to the fiction there are attributes. To fill in the gap, we will still have to take *some* instances of the unreflective attribute schemata as principles of the fiction. But we won't need all the instances of the schemata, or even most of them. Now that we have the axioms of Universalism, we only have to take a few instances of the schemata as principles to get the rest for free as fictional truths. Which instances can we throw away, and which do we need to keep?

Given the strength of the Universalist theory we are adapting, it is clear that we can throw away all instances of the schemata in which the expression φ has the sort of logical complexity that can be captured in first-order logic, or infinitary extensions

[9]If we wanted the principles of the fiction to be strictly silent about all questions having to do with the instantiation of properties and relations by fictional entities, we should also weaken the standard principle governing negation, so that it reads 'For any *real* thing x, x instantiates the negation of p iff x doesn't instantiate p'. If we don't make this revision, it will still be indeterminate which properties any fictional thing instantiates, but it will be determinate that it instantiates exactly one of any two properties one of which is the negation of the other. As I have already emphasised (on page 185 above) it simply doesn't matter what we say about the instantiation of properties by fictional things, given that we're dealing with an unreflective fiction. So we might as well say whatever is most convenient.

of first-order logic. For example, if it is true in the fiction that there are such things as the property of being a soldier and the property of being a statesman, and that any two properties have a conjunction, it will thereby also be true in the fiction that there is such a thing as the property of being a soldier and a statesman. For, given the way we use the term 'the property of being a soldier and a statesman', we know that if there is such a thing as the conjunction of the property of being a soldier and the property of being a statesman, that thing is the referent of 'the property of being a soldier and a statesman'. Moreover, if it is true in the fiction that

> For any x that really exists, x instantiates the property of being a soldier iff x is really a soldier, and
>
> For any x that really exists, x instantiates the property of being a statesman iff x is really a statesman,

then it will also be true that

> For any x that really exists, x instantiates the property of being a soldier and a statesman iff x is really a soldier and a statesman,

thanks to the principle that a conjunction is instantiated by the things that instantiate both its conjuncts. The same goes for all abstract terms involving the sort of logical complexity visible to (infinitary extensions of) first-order logic.[10]

But these are not the only instances of the unreflective attribute schemata that are made redundant when we incorporate the Universalist principles. For example, it is plausible that the sentence

[10]See the proof on pp. 137–138 above.

(15) For any x that really exists, x instantiates the property of being a hydrogen atom iff x is really a hydrogen atom

will automatically be true in the fiction, provided that

(16) For any x that really exists, x instantiates the property of being composed entirely of an atomic nucleus that contains exactly one proton together with all the electrons that orbit it iff x is really composed entirely of an atomic nucleus that contains exactly one proton together with all the electrons that orbit it

is. And, very plausibly, (16) in turn is automatically true in the fiction if

(17) For any x that really exists, x instantiates the property of being an x such that [there is a y such that (y is part of x, and y is an atomic nucleus, and there is exactly one z such that z is part of y, and z is a proton, and for all w, if w is part of x, then either w overlaps y, or there is a v such that v is an electron that orbits y, and w overlaps v)] iff x is really such that [there is a y such that (...)].

is. But the predicate in (17) has the sort of logical complexity visible to first order logic. So, provided that the instances of the unreflective attribute schemata corresponding to the predicates 'is an atomic nucleus', 'is part of', 'is a proton', 'overlaps', 'is an electron' and 'is identical to' are true in the fiction, (17) is true in the fiction as well; and hence, so are (16) and (15).[11]

[11]I'm still bracketing my scepticism about composite objects. If I wasn't, I would classify 'part' as a semantically defective predicate, and hence deny that there is such a thing as the relation of parthood even according to the fiction of complex attributes; I would probably to take the same line as regards 'is a hydrogen atom'. Our ordinary

I think the predicate 'is a hydrogen atom' is typical in this respect. Speaking according to the fiction, there is a small set of *basic* attributes from which all the attributes there are—including the referents of all abstract terms—can be built up. Thus, we only need to count a few instances of the unreflective attribute schemata as principles of the fiction, alongside the principles of some appropriately powerful Universalist theory of complex attributes, to guarantee the fictional truth of all the instances of the schemata.

This claim is much more plausible when the Universalist theory which we build into the fiction provides for at least one, and preferably both, of the forms of infinite complexity described in appendix 4C. Using devices of natural language which go beyond the expressive power of first-order logic—and I assume that there are quite a few of these—we can easily form abstract terms for attributes which would have to be taken as basic if we didn't have the infinitary building operations. Thus, to take a simple example, I hereby introduce the predicate 'friend*' by the following definition: x is a friend* of y iff x is a friend of y, or x is a friend of a friend of y, or x is a friend of a friend of a friend of y, or.... Suppose we know that it is true in the fiction that there is such a thing as the relation *being a friend of*. If we had only taken the axioms of a finitary version of Universalism as principles of the fiction, it would be left open whether there was such a thing as the relation *being a friend* of*. But if we provided in addition for infinite disjunctions, the existence of that relation would follow automatically. Now, if we can introduce predicates like 'friend*', which correspond to infinitely complex attributes if they correspond to complex attributes

talk about the property of being a hydrogen atom is governed by a hybrid fiction which combines a mereological fiction with a fiction of complex attributes: see section 6.3 below.

at all, by explicit definition, it is plausible that many of the predicates that we already have have the same status. When one considers how complicated and nuanced our use of the predicates 'is a volcano', 'is a tree' and 'is a human being' is, it seems very likely that it is only by means of powerful infinitary operations that properties worthy of the names '*being a volcano*', '*being a tree*' and '*being a human being*' can be constructed out of any small set of basic attributes.

These examples raise a difficulty that does not (or at least does not obviously) arise for 'is a hydrogen atom': they are vague. When φ is a vague predicate of ordinary language like 'bald' or 'vehicle' or 'table' or 'dog', it often doesn't seem to be possible to find an interesting logically complex expression ψ—even an infinitary one—in which neither φ nor any of its trivial synonyms occurs, such that 'The property of φing is the property of ψing' can correctly be asserted. Does this mean that we are going to have to include a great many instances of the unreflective attribute schemata involving vague words among the principles of the fiction?

No. Even if there is no interesting logically complex expression ψ such that ⌜The property of being a dog is the property of ψing⌝ is *determinately* true in the fiction, there may be many such expressions $\psi_1 \ldots \psi_n$ such that ⌜The property of being a dog is either the property of $\psi_1$ing, or the property of $\psi_2$ing, or ... or the property of $\psi_n$ing⌝ is determinately true, even though ⌜The property of being a dog is the property of $\psi_i$ing⌝ is in each case neither determinately true nor determinately false. In other words, the term 'the property of being a dog' could be indeterminate in reference between various "precise" properties, each of which is built up in the ordinary way from the basic attributes. Provided that this referential indeterminacy is correlative to the indeterminacy of 'is a dog', 'For any x that actually exists, x instantiates the property of being a dog iff x is really a dog' will still be determinately true in the

fiction.[12]

This seems to me the best way to treat abstract terms involving vague words. The main alternative is to treat 'the property of being a dog' as referring determinately to one property, attributing the indeterminacy of 'instantiates the property of being a dog' to the predicate 'instantiates' instead. One could talk this way if one wanted to, I suppose: but one would then face the question *in virtue of what* this property was such as to be neither determinately instantiated nor determinately not instantiated by this thing. Unless we can say something in precise language about what the property is like and how it stands towards the thing, the fact that 'instantiates' neither determinately applies nor determinately fails to apply to them seems like a pointless mystery. But I have no idea what this more fundamental level of description

[12]A similar strategy lets us endorse even those instances of the unreflective attribute schema which involve predicates for which *expressivism* (emotivism, non-cognitivism, non-factualism...) is true. It is often taken for granted that an expressivist about 'good' must deny that there is such a thing as the property of being good. Perhaps people are assuming some argument along the following lines: (i) expressivism is false for 'instantiates'; so (ii) for each property, the question which things instantiate it is a factual one; so (iii) if there is such a thing as the property of being good, it is a factual question which things instantiate it; so (iv) if there is such a thing as the property of being good, it is a factual question which things are good. But the step from (ii) to (iii) seems to be invalid, for the same reason that the corresponding inference is invalid when 'factual' is replaced with 'non-vague'. Even if it was a factual question, concerning each person in the room, whether he or she is more than six feet tall, the question whether the *best* person in the room is more than six feet tall could still be an non-factual one. On the only well-developed semantics for expressivism, that of Allan Gibbard (1990), it seems that 'the property of being good' will be similar to 'the best person in the room' in this respect. On Gibbard's semantics, the expressive character of words like 'good' consists in their expressing different ordinary, descriptive properties relative to different *complete systems of norms* (this characterisation of the view is due to Field 1994b). But surely it's a factual matter, hence true relative to all systems of norms, that 'the property of being good' refers to the property that 'good' expresses. So 'the property of being good' refers to an ordinary, descriptive property, but it's a norm-relative matter which property this is.

could be.[13]

What is a minimal set of predicates such that, if the instances of the unreflective attribute schemata that correspond to those predicates are taken as principles of the fiction alongside the axioms of Universalism, all the other instances of the schemata will follow automatically? An interesting question, and a hard one. In fact, this is one guise of what I take to be the most basic question of metaphysics: the question *which predicates are primitive*—i.e. *objectively* primitive, as opposed to being treated by us as primitive in this or that context. This question is prior to all the more familiar metaphysical questions of the form 'are there any Fs?'—the disagreement we have when we agree as regards which predicates are primitive, but disagree about how the world should be described in terms of them is less fundamental, and more tractable, than the disagreement we would have if we disagreed about the question of primitive predicates. One rather misleading and partial way to ask the question is to ask which *ontological categories* there are: for if a predicate "corresponds to an ontological category", surely it is either itself primitive, or at least has some very simple definition in terms of primitive predicates. As Quine put it,

> The quest of a simplest, clearest overall pattern of canonical notation is

[13]One suggestion is that the fundamental description would use the predicate 'instantiates to degree x': the vagueness of 'instantiates' is due to our failure to fix a precise threshold. But this doesn't work, since it doesn't give us the "penumbral connections" we need (see Fine 1975). Suppose for example that we have a creature, Fang, who is determinately either a wolf or a dog, but not determinately a wolf or determinately a dog. Hence, according to the unreflective fiction, Fang doesn't determinately instantiate the property of being a wolf, and doesn't determinately instantiate the property of being a dog, but nevertheless Fang determinately either instantiates the property of being a wolf or instantiates the property of being a dog. This can't be explained just by appeal to degrees of instantiation and vague thresholds.

> not to be distinguished from a quest of ultimate categories, a limning of the most general traits of reality.[14]

However, only monadic predicates can "correspond to ontological categories", whereas primitive predicates can be, and surely are, polyadic. Even if we agree that the only ontological category there is is the category of particulars, or substances—or better, if we deny that there are any ontological categories—there are many ways in which we might still disagree about the question which predicates are primitive.[15]

[14]Quine 1960b, p. 161. Here Quine is assuming, controversially, that the right method of *answering* the question which predicates are objectively primitive has something to do with the pursuit of simplicity and clarity. He may also intend, not so much to suggest that the "quest of a simplest, clearest overall pattern of canonical notation" should be approached in the serious metaphysical fashion normally associated with the "quest of ultimate categories", as to suggest that the latter quest should be approached with the same pragmatic, subjectivistic, anything-goes attitude normally associated with the former. If that is the suggestion, I want no part in it.

[15]Suppose that there is some notion of necessity on which the claim ⌜Necessarily, something φs iff it ψs⌝ is equivalent to the claim ⌜According to the fiction, the property of φing = the property of ψing⌝. Suppose, further, that the principles of the fiction provide for both of the varieties of infinite complexity discussed in appendix 4C. Then the question we are considering is equivalent to the question which predicates constitute a *minimal supervenience base* for everything, in the sense of 'supervenience' corresponding to that sense of 'necessarily'.

Certainly there are some perfectly acceptable ways of talking about attributes, and some senses of necessity, on which the claim that the property of φing is the property of ψing is understood to follow from the claim that necessarily, something φs iff it ψs. But if our aim is to explicate the question which predicates are objectively primitive, we should be especially interested in versions of the fiction in which this isn't the case. For there does seem to be the potential for serious factual disagreement between two philosophers who have different lists of primitive predicates, each of whom recognises that the others' predicates constitute a supervenience basis for everything: e.g. Philosopher A's list might include 'blue' and 'green', while Philosopher B's list included 'grue' and 'bleen'. Not that *all* differences between competing lists of primitive predicates correspond to genuine metaphysical disagreements: surely, for example, the question which of 'instantiates' or 'is instantiated by' is primitive is an entirely spurious one. But I am inclined to think that this is a very special, and rare, case.

When I talk in this vein about the importance of the question which predicates are objectively primitive, I get encouraging (and sometimes impatient) nods of recognition from some philosophers; but from many others all I get are blank stares and furrowed brows. I sympathise with those in the latter category: there is much work to be done before it is really clear what the question means, and in what sense it can be a basic question of metaphysics as opposed to a shallow question about our language. Recasting the question as the question which instances of the modified attribute schemata need to be taken as basic principles in an unreflective fiction of complex attributes might help a bit, but many deep questions still remain unanswered. I am not going to undertake to answer them here. What I will do is briefly lay out what seem to me to be the most important sorts of answer to the question, just to convey a sense of the range of the debate. Throughout the discussion, I will treat the claim that the predicate 'φ' is primitive as equivalent to the claim that according to the fiction, the property/relation of φing is "basic".

One predicate which, relatively uncontroversially, belongs on the list of primitive predicates is the predicate 'is identical to'. At least, this is uncontroversial among those of us who hold the view of identity expressed by David Lewis in the following passage from *On the Plurality of Worlds*:

> Identity is utterly simple and unproblematic. Everything is identical to itself; nothing is every identical to anything else except itself. There is never any problem about what makes something identical to itself; nothing can ever fail to be. And there is never any problem about what makes two things identical; two things never can be identical. (Lewis 1986b, pp. 192–193)

However, even we are dimly aware that there is another tradition in thought about identity, on which it is supposed to be a fundamental metaphysical question what

"criterion of identity" of "principle of individuation" governs entities of this or that sort. *Perhaps* what lies behind such questions is the premise that 'is identical to' is *not* a primitive predicate, so that some interesting claim about how the relation *identity* is built up from more basic properties and relations is true in the fiction. For example, someone might hold that identity is the relation

> being an x and y such that either x is a material object and y is a material object and x occupies all the same spacetime points as y; or x is a spacetime point and y is a spacetime point and x is between any other spacetime points z and z' iff y is; or x is a set and y is a set and x has the same members as y.

Alas, this view strikes me as perfectly intelligible, so I can no longer claim that I have no idea what people are talking about when they talk about criteria of identity.

Even the limited consensus about identity disappears when we ask which predicates besides 'is identical to' belong on the list of primitive predicates. Not only is there no agreement on the answer to this question; there isn't even any agreement on the general sort of methodology that is appropriate for finding out the answer. I will mention three important kinds of answer, considering some representative instances of each.

(i) A Realist answer:[16] according to the fiction, the basic relations are the binary relation *being an x and y such that x instantiates y*; the ternary relation *being an x, y, and z such that x and y instantiate z*; the quaternary relation *being an x, y, z, and w such that x, y and z instantiate w*; and so on. All other attributes can be built up out of these relations. For example, the relation *being exactly like* might be

[16] "Realist" in roughly the sense of Armstrong 1978a.

identified with the relation *instantiating all the same things as*. The relation *being an electron* might be identified with the relation *instantiating* e, i.e. the application of the binary relation *instantiation* to e, where e is a certain really existing property (or "universal").[17]

This is not the only sort of answer to the question which predicates are primitive which is Realist in spirit. For example, Armstrong (1997) should perhaps be interpreted as holding that the primitive predicates have something to do with the participation of universals and particulars in "states of affairs". Even among those Realists who want to take something like *instantiation* as primitive, there is room for dispute about exactly what the predicate should look like (see Dorr MS). And perhaps there are other, less complex primitive predicates besides whatever plays the

[17]Whereas a Nominalist who thinks that there aren't really any attributes will presumably want to say that according to the fiction, no real thing is an attribute, it is an interesting question for the Realist where real universals, like e, are to be located in the world of the fiction. If, according to the fiction, all attributes are built up from the *instantiation* relations, and e is an attribute according to the fiction, how is e built up from the instantiation relations?

One option for the Realist is to say that according to the fiction, e is the property *instantiating* e: it is the result of applying *instantiation* to itself. Another option is to claim that 'property', 'attribute', 'instantiate' and related terms are ambiguous between a "real-world" sense, and a sense in which they are purely fictional predicates. e is a property only in the real-world sense; in the fictional sense, no actual thing is a property. It would be natural to use the word 'universal' as a synonym for 'property or relation in the real-world sense'. (This is analogous to a suggestion I made on p. 99 above, about how one ought to think of the world of the mereological fiction if there are actually things which count in *some* sense as composite.)

I am inclined to favour the latter option, on the following grounds: I don't see how a Realist who thinks that 'instantiate' is primitive can justifiably rule out the possibility that there are *multigrade* universals: for instance, an entity a which is instantiated by some x, and also borne by some y to some z. But it would be arbitrary to identify a in the fiction with the property *instantiating* a, or with the relation *being an* x *and* y *such that* x *bears* a *to* y.

role of instantiation: 'is a universal', for example, or 'is an n-ary universal'.[18] What all these views have in common is their recognition of at least one primitive predicate F for which it is plausible that $\ulcorner F(x_1 \ldots x_n)\urcorner$ entails 'at least one of $x_1 \ldots x_n$ is a universal'.[19] They are thus naturally combined with the view that there actually *are* some universals: for it would be odd if any objectively primitive predicate failed to apply to anything.[20]

(ii) A Resemblance Nominalist answer (see Price 1962; Armstrong 1978a, ch. 5): the only basic attribute (apart from *identity*) is the quaternary relation *being an x, y, z and w such that x resembles y at least as much as z resembles w*—"comparative resemblance", for short. All other attributes are built out of this one. For example, the relation *being exactly like* might be identified with the relation *resembling at least as much as one resembles oneself*. The proponent of this sort of view has some trouble

[18]If we don't take 'is a universal' as primitive, we will have to somehow explain how to build up the property *being a universal* out of the basic instantiation relations or whatever takes their place. The obvious way to do this is to identify *being a universal* with *being instantiated by something*. But this commits us to holding that it is impossible for there to be uninstantiated universals; some have found this objectionable.

[19]Here I'm using 'universal' to mean 'thing that *really* is an attribute'.

[20]In the case of views, like the one I began with, on which $\ulcorner F(x_1 \ldots x_n)\urcorner$ entails 'at least one of $x_1 \ldots x_n$ is a universal' for *every* primitive predicate F, this can be strengthened into a powerful argument for the existence of universals. Suppose that (according to the fiction of complex attributes) all the basic attributes (except for *identity*) are uninstantiated. Then the facts about the instantiation of other attributes are very boring. Every property is either had by everything or had by nothing, except for properties of the form *being identical to a* ("haecceities") and other properties built out of them. To show that this isn't the case, all we need is an example of some predicate G such that $\ulcorner \exists x \exists y (G(x) \wedge \sim G(y))\urcorner$ is true, and such that it is obvious that (according to the fiction) the property *being an x such that $G(x)$* is not a haecceity or built up out of haecceities. There are obviously many such predicates: the world is not utterly homogeneous. (As a Nihilist about composition, I'm not going to accept any of the examples that spring immediately to mind, but I'll still accept 'is among some things which are arranged cat-wise', for instance.)

explaining how to construct properties, like *being an electron*, which a Realist would take to involve application to real universals. One possible strategy is to identify *being an electron* with *being exactly like a*—i.e. the application of *being exactly alike* to *a*—where '*a*' is a name for a certain "paradigm electron". (Presumably it is indeterminate which electron is the paradigm.) Generalising this, Price (1962) suggested in effect that many familiar properties might have the form *resembles each of* $a_1 \ldots a_n$ *at least as much as any of them resemble one another*, for some paradigms $a_1 \ldots a_n$. But, as Armstrong (1978a, section 5.4) and others have pointed out, this invocation of paradigms faces a serious problem. If there is an *a* such that *being an electron* is *being exactly like a*, then it is *necessary* that, if there are any electrons, *a* exists. But intuitively, even if any given actual electron, and indeed all of them, had failed to exist, there could still have been electrons.

The view that comparative resemblance is the only basic attribute (apart from identity) faces many other problems. The space of attributes built up from comparative resemblance and identity alone is simply too impoverished: there are too many ordinary abstract terms for which no plausible referents can be found anywhere in that space. However, there are other, more promising views of the same general sort. What these views have in common is the claim that the basic attributes are (i) not such that their instantiation entails the real existence of universals, and (ii) among those attributes which a Realist would take to be *structural* or *topic-neutral*, i.e. built up from the basic attributes without the use of application. The prospects for this sort of view look much better if it makes sense to claim that non-first-order predicates—for instance, predicates which take plural arguments rather than singular ones—are primitive. I think that this does make sense. The fact that it doesn't make sense on our current way of translating the question which predicates are primitive

into a question about the fiction of attributes just shows that we need to liberalise our conception of attributes—for example, we need to recognise "plural properties" (symmetric multigrade relations?) which can be instantiated by *some things* without being instantiated by any of them taken individually. I myself favour a view of this sort, in which the primitive plural predicate is 'are natural'—the predicate which a Realist might analyse as 'are all and only the instances of some universal' (see Lewis 1983a, especially footnote 9).[21] But defending the coherence and plausibility of this view is a very large task, beyond the scope of this survey. The view faces some deep difficulties. The worst of these, I think, is the "problem of paradigms", which is just as much a problem for this view as for the view that comparative resemblance is primitive. Unless I can somehow resolve or finesse this problem, I won't be able to give a credible account of the structure of non-topic-neutral properties like *being an electron*.

(iii) An *a posteriori* Nominalist answer: the list of primitive predicates is to be discovered empirically, by doing physics.[22] Thus, 'is an electron' might itself be a

[21]In fact, if we are to have any chance of being able to construct non-topic-neutral relations like *betweenness*—the sorts of relations that a Realist would characterise as built up by application to real *relations* (relation universals), the view needs to be more complicated than this. We need some way of capturing the sort of structure which a Realist captures by positing real relations. My preferred way of doing this is to take 'are natural' to be, not a plural, but a *per*plural predicate, taking an argument which stands to plural referring terms as plural referring terms stand to singular ones. (See footnote 26 of chapter 1).

[22]Someone could, of course, hold that the right method for discovering which predicates are primitive involves *a posteriori* inquiry, without holding that the methods of *physics* are the right methods to use. A dualist, for example, might think that some of the inquiry can be done simply just by introspection: we find out that predicates like 'blue' (or 'unique blue' or 'blue-seeming' or 'blue′') belong on the list of primitives. I am concentrating on the view that it is to be done by doing physics only for the sake of definiteness.

primitive predicate, although it might on the other hand turn out that (according to the fiction) the property of being an electron is built up from some still more basic properties and relations. If it's an a posteriori matter what the basic attributes are, it's also an a posteriori matter how any other attribute is constructed out of basic attributes. Thus *every* predicate works the way Kripke (1972) and Putnam (1975a) taught us 'water' and 'gold' work, if one inquires deeply enough into its meaning. Nevertheless, we can make general guesses from the armchair about the sort of structure any given attribute might turn out to have. For instance, it is plausible that on this sort of view, the relation *being exactly like* is something along the lines of *being an x and y such that x is an electron iff y is an electron, and x is a photon iff y is a photon, and...*.

I suspect that this kind of view is quite popular, although it is frustratingly hard to find explicit statements of it in the literature (Field 1992 comes pretty close). It may not be generally appreciated just how radical it is. The discovery that the property of being a hydrogen atom is the property *being composed entirely of an atomic nucleus that contains exactly one proton, and all the electrons that orbit it* forces us to deny that it is *possible* for there to be a hydrogen atom without there being any protons. In just the same way, if the *a posteriori* Nominalist is right about *being exactly alike*, something of this form is true:

> Necessarily, if there are no electrons, and no photons, and... and there are no x, y, and z such that x is between y and z, and ... then any two things are exactly alike.

And the predicate 'is exactly alike' is quite typical in this respect. For every predicate $\varphi(x_1 \ldots x_n)$, something like this is true:

> Necessarily, if there are no electrons, ... and no actual things exist, then either $\forall x_1 \ldots x_n \varphi(x_1 \ldots x_n)$, or $\forall x_1 \ldots x_n, \sim\varphi(x_1 \ldots x_n)$.[23]

Of course, this would be unproblematic if 'necessarily' meant 'nomologically necessarily'; and there is, indeed, a way (not a good way, in my opinion) of talking about property identity on which the claim that the property of φing = the property of ψing commits one only to its being nomologically necessary that the things that φ are the things that ψ (Putnam 1975b). But on the standard understanding of property-identity claims, the necessity in question is much stronger, at least as strong as metaphysical necessity. And surely it is very much harder to give up one's intuition that the world could still have been interesting and multifarious even if there were no electrons, etc., than it is to set aside any intuition one might have started out with that there could have been water without H_2O, or hydrogen atoms without protons!

This is far from being an exhaustive survey of the different possible answers to the question which predicates are objectively primitive. My aim, to repeat, has just been to illustrate the scope of the debate, and show how different views about this fundamental metaphysical question translate into different views about what the principles of the fiction need to be like.

The principles we have discussed so far—the axioms of Universalism, and the instances of the unreflective attribute schemata which contain objectively primitive predicates—give us the sort of the insight into the structure of the world of the fiction they describe that we would lack if we just took all the instances of the unreflective attribute schemata as principles. But they do not resolve all questions about what the

[23] '... and no actual things exist ...' is included so as to make sure that the claim holds even for predicates like 'is identical to Mars'.

structure actually looks like. Two lacunae are particularly noticeable. First, these principles say nothing at all about what attributes there *aren't*. Unless we put in some new principle to address this, no interesting claims of the form 'all attributes do such-and-such' or 'the only properties that do such-and-such are ones of so-and-so sort' are ever going to be determinately true in the fiction. For instance, no matter how exactly alike they might be in reality, it will never be determinately true in the fiction that any distinct x and y share all their non-haecceitistic properties. The natural way to fill this gap would be to add a principle of *completeness*, according to which the basic attributes, and the attributes built out of them by conjunction, disjunction, negation, transformation, quantification and application, are the only attributes there are.

Second, these principles leave open many questions about the identity of attributes. They don't tell us when, if ever, it happens that the same attribute can be built out of basic attributes in more than one way. We don't even have answers to the simplest of questions, like 'Is every attribute the negation of its own negation?' and 'Is the conjunction of the negations of some properties the same as the negation of their disjunction?'. For most purposes it doesn't matter what the answers to these questions are. Still, we would have a more comfortable and intuitive grasp of the structure of the realm of attributes if we added some new principles to the fiction in order to resolve them.

What might these new principles say? One natural idea is that they should say that it *never* happens that something is built up out of basic attributes in more than way. For example: if an attribute is the negation of some other attribute, it is not also an existential quantification, or transformation, or conjunction of any other attribute or attributes; if an attribute is the conjunction of some attributes, then it is not the

conjunction of any other attributes; if an attribute is one of the basic attributes, then it can't be built out of other attributes at all. Unfortunately, this attractive proposal turns out to be inconsistent with the Universalist principle U11*, which says that *any attributes whatsoever* have a conjunction and a disjunction, provided that they are of the same degree.[24] One might respond to this by proposing some restriction on U11*, though I think this would be a serious loss. Alternatively, one might look elsewhere for a principle capable of resolving open questions about identity. For instance, one could help oneself to an unexplained notion of necessity, and claim that attributes are identical iff they are necessarily coextensive. Or, more ambitiously, we could introduce a notion of entailment among attributes as a new fictional predicate, governed by new principles which say how it fits into the the structure we already know about; we could then add the further principle that mutually entailing attributes are identical.

[24]Consider the conjunction c of all properties that are non-self-conjuncts: that is, are not among any properties of which they are the conjunction. c exists according to U11*, and obviously must be a self-conjunct. So, in addition to being the conjunction of the non-self-conjuncts, there must also be some *other* properties, of which c is one, of which c is the conjunction; so the principle that if an attribute is the conjunction of some attributes it is not the conjunction of any other attributes fails. This problem might initially seem easy to resolve by making some slight emendation to the suggested principle about identity. In fact, however, it is intractable, and goes right to the heart of the view. Intuitively, the problem is this: U11* entails that there are more conjunctive properties than non-conjunctive properties, and more disjunctive properties than non-disjunctive properties. But the principle that every property has a negation and that no two properties share a negation entails that there are as many negative properties as there are properties altogether. So if we want to hold on to U11*, we will have to hold that almost always, when a property is the negation of some other property, it is also the disjunction of some other properties, and the conjunction of some still further properties. Similarly, most transformations, existential quantifications, universal quantifications, and applications are also conjunctions and disjunctions.

6.3 Fictions within fictions

An "according to the fiction" operator can be thought of as corresponding to a function from worlds to worlds.[25] What is true according to the fiction is what is true at the world which results when we feed the function the actual world as input. Now, whenever f and g are functions from worlds to worlds, $f \circ g$ is another function from worlds to worlds, equally suited to serving as the semantic value of an "according to the fiction" operator which might provide the correctness-conditions for some actual or possible practice. Thus, when we are seeking an account of some new practice—especially some practice which can be somehow be distinguished into different aspects—instead of immediately attempting to define an appropriate "according to the fiction" operator from scratch, we should see whether the job can be done with less effort by stringing together the operators we have already come up with in our accounts of simpler practices.

In ordinary contexts, we can correctly say things like 'the pile of plates in the sink has the property of being about to topple over', which presuppose the existence both of composite particulars and of complex attributes. To the extent that this property-talk is of an "unreflective" sort, we can understand this practice as governed by the combination of one of the mereological fictions discussed in section 3.3 with an unreflective fiction of complex attributes. To find out what the world of the fiction is like, you start with the actual world; move from there to the world of the mereological fiction by adding lots of composite objects; and then look at the world you get when you feed *that* world to the function which corresponds to the fiction of

[25]This is how such operators are represented in the model-theoretic semantics I gave footnote 25 of chapter 3 (p. 107).

complex attributes.[26]

I would like to be able to put this point by saying that the 'according to the fiction' operator appropriate to the combined practice can be defined as 'According to the mereological fiction, according to the unreflective fiction of complex attributes...', where these operators are in turn given conditional analyses along the lines discussed in sections 6.2 and 3.3. Unfortunately, because of the use of the operator 'really' in the principles in terms of which these operators were defined, this gives us the wrong results. The combined fiction will talk only about the instantiation of properties by things that *really* exist, whereas we wanted it to tell us also about the instantiation of properties by composites of real things, which don't really exist but do exist according to the mereological fiction. The problem is that 'really', as we have been understanding it, is an operator that undoes the effect of any number of 'according to the fiction' operators, taking us all the way back to the real world: $\ulcorner \bigcirc_1 \ldots \bigcirc_n$ really $\varphi \urcorner$ is logically equivalent to φ. What we want, on the other hand, is an operator which takes us back only one step, from the world of the combined fiction to the world of the mereological fiction. Fortunately, we can get around this difficulty by adapting an excellent piece of modal-logical technology invented by Harold Hodes (1984): it is called the "backspace" operator, written '$\downarrow$', and it works like a weaker version of the "actually" operator which can only undo the effect of one modal operator: whereas

[26]Why do it that way round, rather than applying the attribute fiction first and the mereological fiction second? Two reasons. First, if you did that, and the version of the mereological fiction you were working with included absolutely unrestricted composition (principle M1), you would end up with a world in which there were lots of mereological fusions of complex attributes with each other and with particulars: to say the least, it's not clear this corresponds to anything in our ordinary practice. Second, and more importantly, if you did it in that order, you'd get a world that could be described using the predicate 'is part of', but you wouldn't get the relation *parthood*, or any of the complex attributes built out of it.

$\ulcorner\Box\Box$actually $\varphi\urcorner$ is equivalent to φ, $\ulcorner\Box\Box\downarrow\varphi\urcorner$ is equivalent to $\ulcorner\Box\varphi\urcorner$. To make the operators defined in sections 3.3 and 6.2 behave in the right way when they are iterated, we can simply decide to understand 'really', when used in stating the principles of the fiction, as working like "backspace" rather than like "actually".[27]

One more wrinkle needs to be resolved before this is can be a satisfactory account of the combined practice. I have argued that 'is part of' should be regarded as a semantically defective predicate, at least in the sense in which it occurs in the principles of the mereological fiction. But I explicitly forbade semantically defective predicates from occurring in instances of the unreflective attribute schemata (footnote 3 above). So if the principles of the fiction of complex attributes are all instances of those schemata, it won't be true in the combined fiction that there is such a thing as the relation of parthood, or any of the attributes constructed out of it. The problem doesn't go away when we include the axioms of Universalism as principles of the fiction: plausibly, no property constructible out of attributes which correspond to predicates that aren't semantically defective could deserve the name 'the property of being a snark' or 'the property of orbiting between Mercury and Vulcan' or—if I'm right—'parthood'.[28] To resolve this problem, we need to take the version of the unreflective fiction that occurs within the scope of the 'according to the mereological

[27]To adapt the model-theoretic semantics from footnote 25 of chapter 3 to this new understanding of 'really', we need only modify the old clause in the truth-definition:

$\ulcorner$Really $\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w', 0\rangle$.

so as to read

$\ulcorner$Really $\varphi\urcorner$ is true at $\langle w, w', n\rangle$ iff φ is true at $\langle w, w', \max(0, n-1)\rangle$.

[28]Unless you wanted to say that these terms refer to *inconsistent* properties and relations. But that would be just as bad as not referring at all.

fiction' operator to be slightly expanded, so as to recognise the fact that 'is part of' and various other semantically defective predicates don't function as semantically defective when they are within the scope of 'according to the mereological fiction'. We must add 'For any real x and y, x and y instantiate *parthood* iff x is really part of y' as a new principle to the fiction. If the axioms of Universalism are also among the principles of the fiction, that should suffice to generate all the new attributes we need; if they aren't, we will need to add a new principle for each sentence or predicate that is really semantically defective, but not semantically defective within the mereological fiction.

If you can string *different* "according to the fiction" operators together to make new operators which generate new, more complicated fictions, what happens when you string together several instances of the same operator? That depends. If the operator in question obeys the S4 law, so that $\bigcirc\bigcirc\varphi$ is equivalent to $\bigcirc\varphi$, you don't get anything new. But the operators we have been considering don't obey the S4 law. Considered as functions on worlds, they take every world to a "bigger" world at which more things exist; as you apply the function again and again, the world you are considering gets bigger and bigger.

In the case of the mereological fiction, the bigger worlds we get when we iterate the fiction aren't terribly interesting. They contain several layers of mereological structure. If the principles of the fiction are M1–M8, this means that if there are actually n things, there will be 2^n things according to the single fiction, and 2^{2^n} things according to the double fiction. There are two senses of 'part' and other mereological vocabulary; Mereology is true on both interpretations. Real things are both simple_1 and simple_2; the objects introduced by the outer fiction are composed_1

of real things, but they are still simple$_2$.[29]

The fictions that result when one iterates an unreflective fiction of properties, by contrast, are quite interesting. They seem well-suited to providing an account of

[29]Here I'm assuming the view I advocated on page 99, that 'part' as used in stating the principles of the fiction should always be treated as a purely fictional predicate. Even if there are already things at the real world that in *some* sense have parts, we should distinguish this sense from the new, "fictional" sense introduced by the fiction. This means that when we have a fiction within a fiction, we will have *two* different fictional senses of 'part', one introduced by the outer fiction and one introduced by the inner fiction, in addition to any "real-world" sense of 'part' there might be.

If we didn't treat 'part' this way, the double mereological fiction with principles M1–M8 would be inconsistent. Principle M5 says that any two real things are related to one another just as they are in the fiction; but principle M8 entails that no real thing is part of any other real thing. If a real thing is ever really part of another real thing, these principles conflict—unless we are equivocating between two different senses of 'part'.

Perhaps I can make this clearer by continuing the analogy I began back in footnote 19 of chapter 3. Suppose that the story *Quark!* contains many realistic background details about what the world was like before the alien invasion, including the fact that the story *Quark!* was written. According to *Quark!*, *Quark!* itself was an eerily but entirely coincidentally prescient story, whose author happened to hit on the name "quarks" for his fictional deadly particles, little knowing that the same name would soon be used for some all-too-real deadly particles. So, according to *Quark!*, according to *Quark!*, there are quarks (since according to *Quark*, the words of the story *Quark!* are just what they actually are); and yet also, according to *Quark!*, according to *Quark!*, there aren't any quarks, for the same reason that there aren't any quarks according to the Sherlock Holmes stories or any other fictions written before the alien invasion. And yet, according to *Quark!*, *Quark!* is not an inconsistent story! Rather, the word "quarks" has, in addition to its "real-world" sense, two different senses when it is used to say how things are according to *Quark!* according to *Quark!*: it has a sense that goes with the inner fiction—this is the sense on which there are quarks according to *Quark!* according to *Quark!*—and another sense that goes with the outer fiction, on which there are no quarks according to *Quark!* according to *Quark!*. Likewise, when we wanted to say how things were according to *Quark!* according to *Quark!* according to *Quark!*, we would find ourselves needing to distinguish *three* fictional senses of the word "quarks" alongside its real-world sense; and so on. This shows, I think, that the multiplication of senses of words that we get when we talk about iterated fictions is a perfectly general and innocuous phenomenon.

correctness in those contexts which cannot comfortably be treated within the confines of the unreflective fiction. For example, suppose that Mars really exists and is a red planet, but there is really no such thing as the property of being a red planet. The sentence

(11) Mars has the property of having the property of being a red planet

won't be true in a single unreflective fiction, but it will be true in the double fiction. For by (2) (bearing in mind that 'really' is to be understood as 'backspace'), (11) is true in the double fiction iff

(18) Mars has the property of being a red planet

is true in the single fiction; and this in turn is true iff

(19) Mars is a red planet

is true, strictly and literally speaking. Likewise, the sentence

(20) Mars has the property of having the property of having the property of being a red planet

will be true in the triple fiction, but not in the double or single fiction.

In describing iterations of the mereological fiction, it was helpful to suppose that each "according to the fiction" operator introduces a new sense for the fictional predicate 'is part of'. If we make the same move for the distinctive predicates of the fiction of attributes, we can explain away certain oddities that arise when one iterates unreflective fictions of attributes. For example, suppose that strictly and literally speaking, everything has a mass. Then, in the single fiction, the proposition that everything has a mass is true, even though, in the fiction, there are many things that

don't have a mass, *viz.* properties, relations and propositions. This is just the oddness of unreflective fictions which we're already used to. But here's something odder: in the double fiction, the proposition that everything has a mass is *not* true, even though it *has the property of being true*. So far, the only examples we've had of cases where it's true in the fiction that x has the property of φing but does not φ have involved highly extrinsic predicates like 'is such that there are less than $10^{10^{100}}$ things'; now we find the same thing happening for 'is true', which is harder to swallow. Do we really have to swallow it? Not if we treat the term 'the proposition that everything has a mass' as ambiguous, having one sense when it occurs in the principles of the outer fiction and another sense when it occurs in the principles of the inner fiction. This can be attributed to a corresponding ambiguity in the predicate 'is the universal quantification of'. In the double fiction, the property of having a mass has a universal quantification$_1$ that is true, and a universal quantification$_2$ that is false. By using this strategy when we need to, we can set things up so that the double fiction departs from the single one only by adding some new attributes: the two fictions agree as regards the truth-values and instantiation of the attributes that exist according to both.[30]

If the only use of iterated unreflective fictions was in enabling us to say things of the form $\ulcorner x$ has the property of having the property of... of having the property of φing$\urcorner$, they wouldn't be very interesting. But in fact, these iterated fictions enable us to say a wide variety of things that can't be said in an unreflective fiction. For instance, we can talk about the properties of properties, correctly saying things like

[30]Thus there is no need for us to be particular about distinguishing different senses of 'true' and 'instantiates' corresponding to the different layers of fiction: the inner fiction's 'true' differs from that of the outer fiction only in its application to things that don't exist according to the outer fiction.

> The property of being a red planet has the property of being instantiated by Mars;
>
> The property of being a red planet has the property of having exactly one instance that orbits the Sun;
>
> The property of being a red planet stands in the relation *conjunction* to the properties *being red* and *being a planet*.

Moreover, we can have complex attributes which involve quantification over other attributes, themselves complex, as in

> The proposition that every property has a negation is true;
>
> Mars has the property of having some property whose instances tend to cause people to utter the sounds 'That's red';
>
> The property of being in pain is the property of having some property p_1 such that there are sixteen other properties $p_2 \ldots p_{17}$, such that p_1 tends to cause the things that have it to come to have p_2, and *being struck on the head with a mallet* tends to cause the things that have it to come to have p_1, and p_{14} is the conjunction of p_1 and p_{13}, and....

Any of these examples could be true in the double fiction; but there are more complicated sentences in the same vein that require us to ascend further than this. I conjecture that everything that can be correctly said about attributes in any normal context eventually comes out true, if we ascend far enough.

Bibliography

Note: As a convenience to the reader, this bibliography also functions as a partial substitute for an index: each entry is followed by a list of the pages on which that entry is cited.

Armstrong, David M. (1978a). *Nominalism and Realism*, volume 1 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press. (199, 201, 202)

— (1978b). *A Theory of Universals*, volume 2 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press. (8, 116)

— (1983). *What Is A Law Of Nature?* Cambridge: Cambridge University Press. (156)

— (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press. (116, 200)

Baxter, Don (1988). 'Many-One Identity.' *Philosophical Papers* 17: 193–216. (8)

Bealer, George (1982). *Quality and Concept*. Oxford: Clarendon. (137, 140)

— (1999). 'A Theory of the A Priori.' In James Tomberlin (ed.), *Philosophical Perspectives 13: Epistemology*, 29–55. Atascadero: Ridgeview. (38)

— (MS). 'Mental Causation.' Presented at the Metaphysical Mayhem conference, 2000. (119)

Boolos, George (1984). 'To Be Is To Be The Value Of A Variable (Or To Be Some Values Of Some Variables).' *Journal of Philosophy* 81: 430–449. (33)

Braun, David (1993). 'Empty Names.' *Noûs* 27: 449–469. (81)

Burgess, John P. and Gideon Rosen (1997). *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford: Clarendon. (12, 22, 29)

Davies, Martin and Lloyd Humberstone (1980). 'Two Notions of Necessity.' *Philosophical Studies* 38: 1–30. (107)

Dickmann, M. A. (1975). *Large Infinitary Languages: Model Theory*. Amsterdam: North-Holland Publishing Corp. (148)

Dorr, Cian (MS). 'Non-Symmetric Relations.' Forthcoming in *Oxford Studies in Metaphysics*, ed. Dean Zimmerman. (200)

Dorr, Cian and Gideon Rosen (2001). 'Composition as a Fiction.' Forthcoming in the *Blackwell Companion to Metaphysics*, ed. Richard Grandy. (22, 24, 76)

Ellis, Brian and Caroline Lierse (1994). 'Dispositional Essentialism.' *Australasian Journal of Philosophy* 72: 27–45. (61)

Evans, Gareth (1982). *The Varieties of Reference*. Ed. John McDowell. Oxford: Oxford University Press. (80)

Field, Hartry (1980). *Science Without Numbers*. Oxford: Blackwell. (34, 156)

— (1985). 'Can We Dispense With Spacetime?' In P. Asquith and P. Kitcher (eds.), *PSA 1984: Proceedings of the 1984 Biennial Meeting of the Philosophy of Science Association*, volume 2, 33–90. East Lansing: Michigan State University Press. Reprinted in Field 1989: 171–226. (32)

— (1988). 'Realism, Mathematics and Modality.' *Philosophical Topics* 19: 57–107. Reprinted in Field 1989: 227–281. (62)

— (1989). *Realism, Mathematics and Modality*. Oxford: Blackwell.

— (1992). 'Physicalism.' In John Earman (ed.), *Inference, Explanation, and Other Frustrations: Essays in the Philosophy of Science*, 271–291. Berkeley: University of California Press. (204)

— (1994a). 'Deflationary Theories of Meaning and Content.' *Mind* 103: 249–285. Reprinted in Field 2001: 104–156. (187)

— (1994b). 'Disquotational Truth and Factually Defective Discourse.' *The Philosophical Review* 103: 405–52. Reprinted in Field 2001: 222–258. (195)

— (2000). 'Apriority as an Evaluative Notion.' In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the A Priori*, 117–149. Oxford: Oxford University Press. Reprinted in Field 2001: 361–391. (35)

— (2001). *Truth and the Absence of Fact*. Oxford: Oxford University Press.

Fine, Kit (1975). 'Vagueness, Truth and Logic.' *Synthese* 30: 265–300. (97, 196)

— (1982). 'Acts, Events and Things.' In Werner Leinfellner, Eric Kraemer and Jeffrey Schank (eds.), *Language and Ontology: Proceedings of the Sixth International Wittgenstein Symposium*, 97–105. Vienna: Holder-Pichler-Tempsky. (98)

— (2000). 'Things and Their Parts.' In Peter French and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy XXIII: New Directions in Philosophy*, 61–74. Oxford: Basil Blackwell. (98)

— (MS). 'The Non-Identity of a Thing and its Matter.' Presented at the Metaphysical Mayhem conference, 2001. (95)

Fitch, G. (1993). 'Non Denoting.' In James Tomberlin (ed.), *Philosophical Perspectives 7: Language and Logic*, 461–86. Atascadero: Ridgeview. (81)

Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press. (195)

Grice, Paul (1957). 'Meaning.' *The Philosophical Review* 66. Reprinted in Grice 1989: 213–223. (19)

— (1989). *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Harman, Gilbert (1989). *Change In View: Principles of Reasoning*. Cambridge, Mass.: MIT Press. (38)

— (1996). 'Analyticity Regained?' *Noûs* 30: 392–400. (10)

Harper, William L., Robert Stalnaker and Glenn Pearce (eds.) (1981). *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: D. Reidel Publishing Company.

Hazen, Allen (1993). 'Against Pluralism.' *Australasian Journal of Philosophy* 71: 112–144. (33)

— (1997). 'Relations in Lewis's Framework Without Atoms.' *Analysis* 57: 243–248. (29)

Hilbert, David (1899). *The Foundations of Geometry*. Ed. Paul Bernays. Tenth edition. La Salle, Illinois: Open Court, 1971. Translated by Leo Unger. (156)

Hodes, Harold (1984). 'On Modal Logics Which Enrich First-order S5.' *Journal of Philosophical Logic* 13: 423–454. (209)

Horwich, Paul (1990). *Truth*. Cambridge, Mass.: MIT Press. (186)

— (1998). *Meaning*. Oxford: Oxford University Press. (10, 186, 187)

Hume, David (1739). *A Treatise of Human Nature*. Ed. L.A. Selby-Bigge and P.H. Nidditch. 2nd edition. Oxford: Clarendon, 1978. First edition by Selby-Bigge published in 1888. (v)

Kripke, Saul (1972). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1980. (68, 79, 204)

— (MS). 'Reference and Existence.' Unpublished manuscript; delivered as the Locke Lectures in 1973. (82)

Lewis, David (1971). 'Counterparts of Persons and their Bodies.' *Journal of Philosophy* 68: 203–211. Reprinted in Lewis 1983b: 26–46. (95)

— (1973a). 'Causation.' *Journal of Philosophy* 70. Reprinted in Lewis 1986c: 159–171. (50, 53)

— (1973b). *Counterfactuals*. Oxford: Blackwell. (108)

— (1978). 'Truth in Fiction.' *American Philosophical Quarterly* 15: 37–46. Reprinted in Lewis 1983b: 261–280. (16)

— (1979). 'Counterfactual Dependence and Time's Arrow.' *Noûs* 13: 455–476. Reprinted in Lewis 1986c: 32–66. (46)

— (1983a). 'New Work for a Theory of Universals.' *Australasian Journal of Philosophy* 61: 343–377. Reprinted in Lewis 1999: 8–55. (116, 203)

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

— (1984). 'Putnam's Paradox.' *Australasian Journal of Philosophy* 62: 221–236. Reprinted in Lewis 1999: 56–77. (10)

— (1986a). 'Against Structural Universals.' *Australasian Journal of Philosophy* 64: 25–46. Reprinted in Lewis 1999: 78–107. (129)

— (1986b). *On the Plurality of Worlds*. Oxford: Basil Blackwell. (62, 95, 198)

— (1986c). *Philosophical Papers*, volume 2. Oxford: Oxford University Press.

— (1991). *Parts of Classes*. Oxford: Basil Blackwell. (8, 68)

— (1993). 'Many, But Almost One.' In Keith Campbell, John Bacon and Lloyd Reinhardt (eds.), *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*, 23–42. Cambridge: Cambridge University Press. Reprinted in Lewis 1999: 164–182. (8)

— (1994). 'Humean Supervenience Debugged.' *Mind* 103: 473–490. Reprinted in Lewis 1999: 224–247. (162)

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

Markosian, Ned (1998). 'Brutal Composition.' *Philosophical Studies* 92: 211–249. (4)

Mates, Benson (1950). 'Synonymity.' *University of California Publications in Philosophy* 25: 201–226. (122)

Mellor, D.H. (1991). 'Properties and Predicates.' In *Matters of Metaphysics*, 170–182. Cambridge: Cambridge University Press. Reprinted in Mellor and Oliver 1997: 255–267. (116)

— (1992). 'There Are No Conjunctive Universals.' *Analysis* 52: 97–105. (116)

Mellor, D.H. and Alex Oliver (eds.) (1997). *Properties*. Oxford: Oxford University Press.

Morton, Adam (1975). 'Complex Individuals and Multigrade Relations.' *Noûs* 9: 309–18. (29)

Nolan, Daniel (1997). 'Impossible Worlds: A Modest Approach.' *Notre Dame Journal of Formal Logic* 38: 535–572. (50, 80, 92, 104)

Oliver, Alex and Timothy Smiley (2001). 'Strategies for a Logic of Plurals.' *Philosophical Quarterly* 51: 289–306. (33)

Price, H. H. (1962). *Thinking and Experience*. Cambridge, Mass.: Harvard University Press. (201, 202)

Priest, Graham (1998). 'What Is So Bad About Contradictions?' *Journal of Philosophy* 95: 410–426. (181)

Pryor, Jim (2000). 'The Skeptic and the Dogmatist.' *Noûs* 34: 517–549. (26)

Putnam, Hilary (1975a). 'The Meaning of Meaning.' In Keith Gunderson (ed.), *Language, Mind and Knowledge*, volume 7 of *Minnesota Studies in the Philosophy of Science*, 131–193. Minneapolis: University of Minnesota Press. (204)

— (1975b). 'On Properties.' In *Mathematics, Matter and Method*, 305–322. Cambridge: Cambridge University Press. (205)

— (1987). 'Truth and Convention: On Davidson's Refutation of Conceptual Relativism.' *Dialectica* 41: 41–67. Reprinted in Putnam 1990: 96–104. (17)

— (1990). *Realism with a Human Face*. Cambridge, Mass.: Harvard University Press.

Quine, Willard van Orman (1960a). 'Variables Explained Away.' *Proceedings of the American Philosophy Society* 104: 343–347. Reprinted in Quine 1966: 227–235. (140, 142)

— (1960b). *Word and Object*. Cambridge, Mass.: MIT Press. (197)

— (1966). *Selected Logic Papers*. New York: Random House.

Ramsey, F.P. (1925). 'Universals.' *Mind* 34: 401–417. Reprinted in Ramsey 1990: 8–33. (116)

— (1990). *Philosophical Papers*. Ed. D.H. Mellor. Cambridge: Cambridge University Press.

Russell, Bertrand (1919). *Introduction to Mathematical Philosophy*. London: Routledge. (xi)

— (1924). 'Logical Atomism.' In David F. Pears (ed.), *The Philosophy of Logical Atomism*, 157–181. La Salle, Illinois: Open Court, 1985. (v)

Salmon, Nathan (1998). 'Nonexistence.' *Noûs* 32: 277–319. (78, 81, 82)

Schiffer, Stephen (1996). 'Language-Created, Language-Independent Entities.' *Philosophical Topics* 24: 149–167. (186)

Shoemaker, Sydney (1980). 'Causality and Properties.' In Peter van Inwagen (ed.), *Time and Cause*, 228–254. Dordrecht: D. Reidel. Reprinted in Mellor and Oliver 1997: 228–254. (61, 73)

— (1998). 'Causal and Metaphysical Necessity.' *Pacific Philosophical Quarterly* 79: 59–77. (61, 73)

Sider, Theodore (1993). 'Van Inwagen and the Possibility of Gunk.' *Analysis* 53: 285–289. (62, 68)

— (2002). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press. (10, 32)

— (MS). 'The Ersatz Pluriverse.' Available online at `http://web.syr.edu/~trsider/papers/The_Ersatz_Pluriverse.pdf`. (92)

Smart, J.J.C. (1959). 'Sensations and Brain Processes.' *The Philosophical Review* 68: 141–156. (162)

Stalnaker, Robert (1968). 'A Theory of Conditionals.' In *Studies in Logical Theory: American Philosophical Quarterly Monograph Series, No. 2*. Oxford: Blackwell. Reprinted in Harper et al. 1981: 41–56. (108)

— (1981). 'A Defense of Conditional Excluded Middle.' In Harper et al. (1981), 87–104. (108)

Stanley, Jason and Zoltan Gendler Szabo (2000). 'Quantifier Domain Restriction.' *Mind and Language* 15: 219–261. (94)

Swoyer, Chris (1982). 'The Nature of Natural Laws.' *Australasian Journal of Philosophy* 60: 203–223. (61)

Thomson, Judith Jarvis (1983). 'Parthood and Identity across Time.' *Journal of Philosophy* 80: 201–220. (102)

van Fraassen, Bas (1980). *The Scientific Image*. Oxford: Clarendon Press. (175)

van Inwagen, Peter (1977). 'Creatures of Fiction.' *American Philosophical Quarterly* 14: 299–308. Reprinted in van Inwagen 2001: 37–56. (82)

— (1990). *Material Beings*. Ithaca: Cornell University Press. (2, 3, 5, 29, 74)

— (1994). 'Composition as Identity.' In James Tomberlin (ed.), *Philosophical Perspectives 8: Logic and Language*, 207–220. Atascadero: Ridgeview. (9)

— (2001). *Ontology, Identity and Modality*. Cambridge: Cambridge University Press.

Walton, Kendall (1990). *Mimesis as Make-Believe*. Cambridge, Mass.: Harvard University Press. (89)

Yablo, Stephen (2000). 'A Paradox of Existence.' In Anthony Everett and Thomas Hofweber (eds.), *Empty Names, Fiction and the Puzzles of Non-Existence*, 275–312. Stanford: CSLI Press. Available online at `http://www.mit.edu/~yablo/apex.html`. (12, 13, 29)

— (MS). 'Mathematics as Gamekeeping.' Available online at `http://www.mit.edu/~yablo/mgk.html`. (29)