Writing the Book of the World

Theodore Sider

Draft of August 7, 2011

The central theme of this book is: realism about structure. The world has a distinguished structure, a privileged description. For a representation to be fully successful, truth is not enough; the representation must also use the right concepts, so that its conceptual structure matches reality's structure. There is an objectively correct way to "write the book of the world".

Realism about *predicate* structure is fairly widely accepted. Many—especially those influenced by David Lewis—think that some predicates (like 'green') do a better job than others (like 'grue') at marking objective similarities, carving nature at the joints. But this realism should be extended, beyond predicates, to expressions of other grammatical categories, including logical expressions. Let "there schmexists an *F*" mean that the property of being an *F* is expressed by some predicate in some sentence of this book. 'Schmexists' does not carve at the joints; it is to the quantifier 'there exists' as 'grue' is to 'green'. Likewise, the question of joint-carving can be raised for predicate modifiers, sentential connectives, and expressions of other grammatical categories. (Structure is a generalization and extension of Lewisian naturalness.)

I connect structure to fundamentality. The joint-carving notions are the fundamental notions; a fact is fundamental when it is stated in joint-carving terms. A central task of metaphysics has always been to discern the ultimate or fundamental reality underlying the appearances. I think of this task as the investigation of reality's structure.

Questions about which expressions carve at the joints are questions about how much structure reality contains. Whether reality contains causal, or ontological, or modal structure is a matter of whether causal predicates, quantifiers (or names), and modal operators carve at the joints. Such questions lie at the center of metametaphysics. Those who say that questions of ontology are "merely verbal", for example, are best regarded as holding that reality lacks ontological structure. Such deflationary metametaphysical stances are thus themselves metaphysical stances. There is no ametaphysical Archimedean point from which to advance deflationary metametaphysics, since any such metametaphysics is committed to at least this much substantive metaphysics: reality *lacks* a certain sort of structure.

A subsidiary theme is: ideology matters. There is an unfortunate tendency, perhaps encouraged by bad terminology, to psychologize Quine's notion of ideology: to regard a theory's choice of primitive notions—its ideology—as a merely psychological or linguistic or conventional matter (in contrast to the entities it postulates—its ontology—which is part of its objective content). Philosophers reject their opponents' ideology in psychological/semantic terms:

"I don't understand what you mean by that." And when introducing their own ideology, the hurdle to be passed is again psychological/semantic: primitive notions must be "intelligible". But there is a squarely metaphysical issue concerning any proposed piece of ideology (including logical and quasi-logical ideology such as modal operators or second-order quantifiers): does reality contain the requisite structure? If it does, then "intelligibility" in previously "understood" terms is not required for successful reference to and theorizing about that structure, no more in metaphysics than in physics.

A shift of focus from psychological/semantic to metaphysical constraints on ideology is at times liberating for metaphysics, but it also keeps our feet on the ground, by restraining the tendency to evade ontological commitments by adding to ideology. A fundamental theory's ideology is as much a part of its representational content as its ontology, for it represents the world as having structure corresponding to its primitive expressions. And the world according to an ideologically bloated theory has a vastly more complex structure than the world according to an ideologically leaner theory; such complexity is not to be posited lightly.

Fixating on ontology while ignoring ideology is both too narrow and incautious.[1] It is too narrow because the goal of metaphysics is to give a fundamental description of the world, and doing so requires more than merely saying what there is. It is incautious because it uncritically assumes that quantificational structure is fundamental. If quantificational structure is indeed fundamental (as I think it is), ontology deserves its place in fundamental metaphysics. But if quantificational structure is not fundamental, then ontological inquiry deserves little more attention within fundamental metaphysics than inquiry into the nature of catcher's mitts.

A final theme is a "pure" conception of metaphysics, free of certain encumbrances. One encumbrance is doing metaphysics primarily in modal terms. Against this, there is a growing consensus that modal notions are too coarse for metaphysics, and that notions in the vicinity of "fundamentality", "in virtue of", and the like, should not be understood in modal terms. A second encumbrance is linguistic entanglements. Here too, there is a growing consensus: that it is not so important for metaphysical and linguistic theory to neatly mesh. The fundamental metaphysics underlying a discourse might have a structure quite unlike that suggested by the discourse. Whereas a good linguistic theory must

[1] Dorr (2004, section 1) and Schaffer (2009*a*) make related complaints.

fit the suggested structure, good metaphysics must fit the underlying structure.[2]

This book presented an organizational challenge. Theory-then-applications would have been neatest, but the concept of structure is unfamiliar enough that readability demanded early applications. My compromise was to intermingle. Chapter 1 introduces the concept of structure and describes in a preliminary way how it will be applied. Chapter 2 begins to present the theory, arguing that structure is primitive and objective, and defending an epistemology of structure. Chapters 3–5 turn to applications, showing how structure illuminates explanation and laws, reference, epistemology, physical geometry, substantivity, and metametaphysics. Chapters 6–8 return to theory, arguing that expressions of any grammatical category (not just predicates) can be evaluated for structure, addressing various abstract questions about how structure behaves, and criticizing certain rival concepts (such as truthmaking and ground). Chapters 9–12 return to applications, showing how the metaphysics of four domains—ontology, logic, time, and modality—looks when conceptualized in terms of structure. Chapter 13 concludes with a sketch of a "worldview": a comprehensive metaphysics cast in terms of structure. As a guide to those who wish to read selectively:

The metaphysics of structure: chapters 1, 2, 6–8;

Applications: chapters 3–5, 9–12;

Metametaphysics: chapters 4–5, 9, and (to a lesser extent) 10–12;

Mix of first-order and meta- metaphysics: chapters 9–13.

I am grateful to many people for helpful discussions and feedback: Frank Arntzenius, Elizabeth Barnes, Paul Boghossian, Craig Callender, Ross Cameron, David Chalmers, David Copp, Troy Cross, Louis deRosset, Janelle Derstine, Cian Dorr, Ant Eagle, Andy Egan, Matti Eklund, Adam Elga, Matt Evans, Delia Graff Fara, Mike Fara, Hartry Field, Hilary Greaves, Liz Harman, Allan Hazlett, Eli Hirsch, Thomas Hofweber, Paul Horwich, Alex Jackson, Carrie Jenkins, Boris Kment, Tora Koyama, Uriah Kriegel, Heather Logue, Ofra Magidor, Ishani Maitra, Colin Marshall, Farid Masrour, Andy McGonigal, Ian McKay, Joseph Melia, Ulrich Meyer, Alan Musgrave, Daniel Nolan, Jill North, Tim O'Connor, Laurie Paul, Zach Perry, Agustín Rayo, Tony Roark,

[2]Kit Fine's (1994*a*; 2001) recent work has been especially influential in forging both consensuses.

Dan Rothschild, Stephen Schiffer, Michael Schweiger, Adam Sennet, Alan Sidelle, David Sosa, Ernie Sosa, Joshua Spencer, Jason Stanley, Irem Kurtsal Steen, Steve Steward, Sharon Street, Zoltán Gendler Szabó, Amie Thomasson, Jason Turner, Ryan Wasserman, Brian Weatherson, Ralph Wedgwood, Bruno Whittle, Tim Williamson, Tobias Wilsch, Chris Wüthrich, Stephen Yablo, and Dean Zimmerman. I'm especially grateful to Karen Bennett, Gideon Rosen, Jonathan Schaffer, and Robbie Williams for extensive and challenging comments (which, I fear, I have not fully addressed). Thanks also to Oxford University Press and to Blackwell Publishing for permission to include bits of Sider (2003), Sider (2009), and Sider (2007*a*).

I'd also like to thank Kit Fine, John Hawthorne, and Phillip Bricker. I've learned much from talking to Kit about fundamentality in the past few years, and from thinking through his writings on the subject. John read large portions of the manuscript and gave me many insightful comments, as well as pushing me, years ago, to go beyond the predicate. Phil directed my dissertation, which was on Lewisian naturalness. He taught me the power of this idea, how to apply it to the philosophy of space and time, and much, much more. My intellectual debt to Phil is massive.

Finally, it should be obvious how much this book owes to David Lewis. His ideas on natural properties and relations have always seemed to me among his best: powerful, correct, revolutionary yet deeply intuitive.

Contents

Chapter 1

Structure

Metaphysics, at bottom, is about the fundamental structure of reality. Not about what's necessarily true. Not about what properties are essential. Not about conceptual analysis. Not about what there is. Structure.

Inquiry into necessity, essence, concepts, or ontology might help to illuminate reality's structure. But the ultimate goal is insight into this structure itself—insight into what the world is like, at the most fundamental level.

1.1 Structure: a first look

Discerning "structure" means discerning patterns. It means figuring out the right categories for describing the world. It means "carving reality at its joints", to paraphrase Plato. It means inquiring into how the world fundamentally is, as opposed to how we ordinarily speak or think of it.

Consider three objects: two electrons in identical intrinsic states, and a cow. It is the most natural thing in the world to say that the electrons are perfectly similar to each other, and that neither is perfectly similar to the cow. The three objects should be divided into two groups, one containing the electrons, the other containing the cow. The electrons *go together*, and neither goes with the cow.

Or imagine a universe that is entirely full of fluid. A plane divides the universe into two halves, one in which the fluid is uniformly red, the other in which the fluid is uniformly blue (figure 1.1). Now imagine a group of people who encounter this universe, but accord no special status to the dividing blue-red plane. Instead of thinking of the universe as divided into the red and

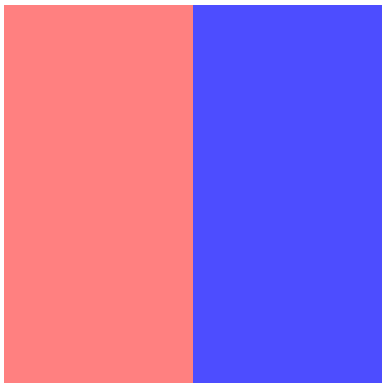

Figure 1.1: The red–blue world

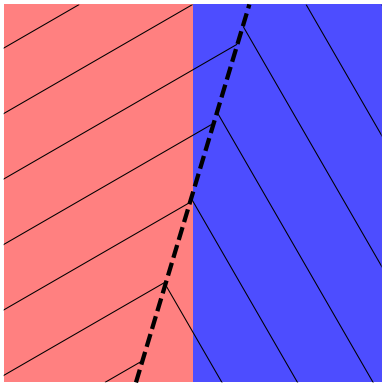

Figure 1.2: Bizarre carving of the red–blue world

blue halves, they think of it as being divided in half by a different plane, marked by the dashed line in figure 1.2. And they do not use predicates for red and blue. Instead, they have a pair of predicates that they apply uniformly within the two regions separated by their dividing plane. These predicates (whose extensions are indicated by diagonal hash lines in the diagram) cut across the predicates 'red' and 'blue'. The regions to the left of the dashed line they call "bred"; the regions to the right they call "rue".

It is almost irresistible to describe these people as *making a mistake*. But they're not making a mistake about where the red and blue regions are, since they make no claims about red or blue. And they make no mistakes when they apply their own concepts. The regions that they call "bred" are indeed bred, and the regions they call "rue" are indeed rue. The problem is that they've

got the wrong concepts. They're carving the world up incorrectly. By failing to think in terms of the red/blue dividing plane, they are *missing something*. Although their beliefs are true, those beliefs do not match the world's structure.

1.2 Philosophical skepticism about structure

All is well until we encounter a philosopher, who, as usual, asks some uncomfortable questions. Why do the two electrons "go together", the philosopher wants to know? Yes, they share many features in common: each has 1.602×10^{-19} C charge, 9.109×10^{-31} kg mass, and so on. But there are plenty of features that the electrons do not share. They are in different locations, travel at different velocities, and are parts of different wholes. And why doesn't the cow go together with the electrons? If all three are located in North America, then all three share the feature *being located in North America*. And all three share the feature: *being an electron or a cow*.

The philosopher continues: what is wrong with carving the red–blue world along the diagonal plane? What is wrong with grouping the bred things together and the rue things together? All bred things really are bred; they all share the feature of *being on the left side of the diagonal plane*. One might protest that not all bred things are alike, since some are red and some are blue; but the philosopher will reply that carving the world along the vertical plane is no better on this score. Not all red things are alike, since some are bred and some are rue.

In fact, once we get the hang of the philosopher's way of thinking about "features", we can see that *any* two objects share infinitely many features, and also differ with respect to infinitely many features. For consider any objects x and y. Where F_x and F_y are any features of x and y, respectively, x and y share the feature: *being either* F_x *or* F_y. And they share the feature *being either* F_x *or* F_y *or 1 kg mass*. And they share the feature *being either* F_x *or* F_y *or 2 kg mass*. And so on. So they share infinitely many features. As for the infinitely many features with respect to which they differ, consider:

being F_x*, and located at L*

being F_x*-or-1-kg-mass, and located at L*

being F_x*-or-2-kg-mass, and located at L*

etc.

where L is some location occupied by x but not y. Object x has each of these features; object y lacks each.

The crux is obviously the philosopher's willingness to allow such "features" as *being either an electron or a cow*, and to treat them on a par with features like *being an electron* and *being a cow*. If we had nothing but the philosopher's features to go by, then indeed, we wouldn't be able to make any sense of a "correct" way to group our three objects, or of the electrons being more similar to each other than to the cow. If, on the other hand, we could make a distinction between *genuine* features—features that are fundamental, that carve nature at the joints, whose sharing makes for similarity—and the rest, then we could say what we want. Can we make this distinction?

Concepts and distinctions that resist definition in terms of the popular philosophical ideology of the day tend to be viewed with suspicion. Thus it was that throughout much of the twentieth century, philosophers tended not to speak of genuine features. Quine's extensionalism, for example, which dominated the 1950s and 1960s, allowed only a meager set of concepts to be used in drawing distinctions (roughly, those of first-order logic plus an array of scientific predicates). Noticing the presence of disjunction in the definitions of many philosopher's features, an extensionalist might begin an attempt to characterize genuineness by disqualifying features defined in this way. But what language do we use to evaluate whether a feature is "defined using disjunction"? Speakers of English must use 'or' to define the feature: *being an electron or cow*, but speakers of a language with a primitive predicate for this feature—'blurg', call it—can define the same extension without using 'or'. Indeed, if the language is strange enough, its speakers would need to use 'or' and other logical connectives to say things that in English may be said using simple predicates like 'cow' and 'electron', just as we must use logically complex predicates of English to say what they say using 'blurg'. The extensionalist attempt fails to characterize an appropriately language-independent notion of genuineness.[1]

In the 1970s, modality became kosher ideology, and there were renewed attempts to define concepts in the vicinity of structure. For instance, Roderick

[1]The paradigm of first-order logic had perhaps the following additional influence. The standard model theory of first-order logic treats the semantic values of (n-place) predicates as subsets of the (n-place Cartesian product of the) domain. Viewed from a purely set-theoretic perspective, the semantic values of the predicates 'is an electron' and 'is an electron or cow' are on a par: each is a subset of the domain.

Chisholm (1976, p. 127) and Jaegwon Kim (1982, pp. 59-60) tried to give a modal definition of the notion of an *intrinsic property*—a property that an object has just by virtue of what it's like in itself, independently of how it is related to other objects. They proposed, roughly, that a property is intrinsic if and only if it is possibly instantiated by an object that is alone in the world. But this definition was shown to be unacceptable. The property of *being alone in the world*, and the property of *either (being alone in the world and being green) or (not being alone in the world and being blue)*, satisfy the definition but are extrinsic (Lewis, 1983*a*).

(The 70s' fixation on modality was doubly unfortunate. Not only are modal tools too crude;[2] they're also distant from the subject matter of most of metaphysics. It is needlessly indirect to approach the question of what the world is like by asking what it must be like and what it might have been like.[3])

Since the 1980s many philosophers have become comfortable with a richer ideology, one that includes notions in the vicinity of "genuine feature", "intrinsic property", and the like. The zeitgeist has been that these notions are legitimate even if they cannot be defined in other terms. Two Davids have led the way.[4] David Armstrong (1978*a*,*b*) used the traditional doctrine of universals to draw the distinction between genuine and nongenuine features. Some predicates, like 'is an electron', perhaps, stand for universals, Armstrong said; but others do not: there simply is no universal of "being either a cow or an electron". Through sheer force of will as much anything, he put realism about genuine features on the map. But as our second David, David Lewis (1983*b*) showed, Armstrong embedded this insight in a quite independent dialectic: the traditional debate over the existence of universals and their role in a general analysis of predication. According to Lewis, we can incorporate Armstrong's insight by admitting a notion of "natural properties and relations" (those properties and relations that carve nature at the joints) without thinking of these as universals in the traditional sense, and without taking on the (misguided, according to Lewis) project of giving a general analysis of predication. The notion of a genuine feature was thus freed from unwanted entanglements.

Of course, *everyone* can agree that there is *some* difference between *being an electron* and *being either an electron or a cow*. If nothing else, ordinary English

[2]On which see, for instance, Fine (1994*a*); Restall (1996).

[3]I also suspect that the right account of how the world might have been and must be defers to how the world *is* (chapter 12).

[4]Earlier relevant work includes Quinton (1958); Quine (1969); Putnam (1975*d*); Bealer (1982, chapter 8).

has a single word for the former attribute. What distinguishes Armstrong and Lewis is that they regard the distinction as objective. Structure, too, is to be understood as objective. There are hard questions about what objectivity amounts to (some of which will be discussed in chapter 4), but the intuitive idea is clear: whether a property, word, or concept[5] carves at the joints has nothing to do with the place of the concept in human languages, conceptual schemes, biology, or anything like that. Thus "fundamental" (which I use more or less interchangeably with "joint-carving" and "part of reality's structure") signifies a metaphysical, rather than conceptual, sort of fundamentality. Humans may need to acquire other concepts first before they grasp joint-carving ones; and conversely, those concepts we acquire first, or most easily, may fail to carve at the joints.

1.3 Structure in metaphysics: a preview

The goal of this book is to push forward the front of realism about structure. I want to expand our conception of structure's importance, generalize the concept of structure, investigate its nature, use it as the foundation of "metametaphysics", and reconceptualize metaphysics in terms of it.

The connection to similarity is only the beginning of the importance of the notion of structure. As we will see, structure pops up throughout philosophy, in our thinking about reference, epistemology, spacetime, objectivity, and other matters.

Structure is particularly central to metaphysics. The heart of metaphysics is the question: what is the world ultimately, or fundamentally, like? And fundamentality is a matter of structure: the fundamental facts are those cast in terms that carve at the joints.

The truly central question of metaphysics is that of what is *most* fundamental. So in my terms, we must ask which notions carve *perfectly* at the joints. By using

[5] Subtleties will come later, but to forestall misunderstanding: 1. Structure is a worldly, not conceptual or linguistic, matter (my informal talk of "notion/word/concept X carves at the joints" notwithstanding). 2. 'Structure' is not a noun; structure is not an entity or stuff (this very sentence, and phrases like "how much structure the world contains", notwithstanding). 3. 'Structure' and its variants are not predicates—not of properties, nor of any other sorts of entities ("charge carves at the joints" notwithstanding). 4. My most basic notion of structure is absolute, although I allow a derivative notion that comes in degrees. 5. Structure includes distinguished monadic features (such as charge), not just relational ones (despite what may be suggested by the term 'structure').

'red' and 'blue', we carve more closely to reality's joints than do the speakers of the 'bred'/'rue' language. But we do not thereby carve perfectly at the joints; colors are presumably not perfectly fundamental. To carve perfectly, one must use the most fundamental concepts, expressing the facets of reality that underly the colors.

Which concepts are the perfectly fundamental ones? In my view, certain concepts of physics, logic, and mathematics.[6] But this thesis about structure is not built into the idea of structure, and defending it is not one of the main goals of this book. The great metaphysical disputes concern which theses of this sort are true; my goal is to explain what is at stake in such disputes, not to settle them. Is mentality part of reality's fundamental structure? (Modal theses in the philosophy of mind, such as psychophysical supervenience, are crude ways of getting at what clearly was the issue all along: whether reality is fundamentally mental.) Do mathematical entities exist, in the fundamental sense of 'there exist', and if so, what are the fundamental features of those entities? Do causal or nomic notions have any place in a fundamental description of the world? These are questions about structure.

*Meta*metaphysics—inquiry into the status of metaphysics—will be central in this book. Is the pope (or Robinson Crusoe, or a twelve-year-old boy) a bachelor? Intuitively, the question is merely verbal or conceptual. To answer it, all we need to do is investigate our concept of a bachelor; intuitively, all that is at stake is how we use the word 'bachelor'. In contrast, the question of whether there is any lithium in a certain region on Mars has nothing to do with word use or concepts; it is *substantive*. This rough and ready notion of substantivity needs to be clarified; after all, the statement that Robinson Crusoe is a bachelor is no more *about* our concept of a bachelor than the statement that there is lithium in the region is about our concept of lithium. Nevertheless, there is a strong intuitive contrast between the two questions.

The opponents of metaphysics (and even some renegade practitioners) tend to regard many metaphysical questions as being—to some extent, anyway—like the question of whether the pope is a bachelor. True believers, on the other hand, tend to think of their favorite metaphysical questions as being substantive, like the question about lithium. In my view, whether a question is substantive—in one important sense of 'substantive'—depends largely on the extent to which its terms carve at the joints; to the extent, that is, that the question concerns the world's fundamental structure. The central metametaphysical questions

[6]Plus the concept of structure itself! See section 7.13.

are about *how much structure the world contains*.

Consider two properties:

> Being an unmarried male
>
> Being an unmarried adult male eligible for marriage.

It may well be that exactly one of these properties is (determinately) what we mean by 'bachelor'. So it may be that the question of the pope's bachelorhood has an answer. But neither of these two properties carves nature at the joints better than the other. The unmarried males don't go together any more than do the unmarried males eligible for marriage. A linguistic community that used the word 'bachelor' for the first property would not be getting at the world's structure any better than a community that used the word for the second property. And since the pope is an unmarried male who is ineligible for marriage, speakers of the first community speak truly when they say 'The pope is a bachelor', whereas speakers of the second community speak truly when they say 'The pope is not a bachelor'. So, intuitively, the only question facing us is: which sort of linguistic community do *we* inhabit? Which of two equally good ways to talk is *our* way to talk?

The question of whether there is lithium in the region near Mars has a very different status. Suppose that the region does indeed contain lithium. We can imagine a linguistic community that uses the word 'lithium' exactly as we do, but with one exception: their word does not apply to the lithium (in our sense of the word) in the region. So 'There is lithium in the region' counts as true in our language, and false in theirs. But here the parallel with the previous paragraph ends. The lithium in the region is just like the lithium elsewhere, so the imagined linguistic community fails badly to carve nature at the joints. They fail to group together things that, objectively, go together. The question of whether there is lithium in the region is *not* just a question of which of two equally good ways to talk is *our* way to talk.

Few would deny that the question of the pope's bachelorhood is insubstantial in a way that the question about lithium is not. But in metaphysics, things are far less clear. Consider questions of ontology, for example. There has been much discussion recently of whether tables and chairs and other composite material objects exist. It is generally common ground in these discussions that there exist subatomic particles that are "arranged tablewise" and "arranged chairwise"; the controversy is over whether there exist in addition tables and chairs that are composed of the particles. Is this really a substantive debate

about the world? Most of the ontologists engaged in the debate think so—or really, presuppose so. But Eli Hirsch, Hilary Putnam, and other "ontological deflationists" have argued that the debate is in some sense merely verbal or conceptual. The "metaontological" question here ultimately boils down, I think, to a question of structure: whether quantificational notions like 'there exists' carve at the joints. What the ontological deflationists have in effect been saying is that reality would need fundamental quantificational structure in order for the question of whether there exist tables and chairs to be worth asking, and that this structure is, in fact, missing. I oppose ontological deflationists in chapter 9, but they deserve credit for raising an important and difficult question—a question that is in a way more foundational than the first-order question of what there is.

There are similar foundational questions throughout metaphysics. Do modal concepts carve at the joints (chapter 12)? (Here my answer is *no*; modality is not the core of metaphysics that some take it to be.) Do tensed concepts (chaper 11)? (Again, *no*; but seeing the issue as concerning reality's temporal joints helps to illuminate what are otherwise extremely perplexing questions.) Do logical concepts (chapter 10)? (Here I say *yes*. Certain debates over the "correct" logic are genuine, and are not linguistic or conceptual; they are as substantive as ontological debates.)

More generally, metaphysicians regularly speak of what is "really" or "genuinely" the case. (Often they feel guilty about it, but don't know how to stop.) As Kit Fine (2001, 2009) has emphasized, such talk is central to metaphysics, but in dire need of explication. When a nominalist says that there do not *really* exist abstract entities like properties, while granting that frogs share more properties in common with crocodiles than they share with humans, the '*really*' is essential; otherwise she contradicts herself. Those who think that "time is like space" say that there is no "genuine" or "objective" distinction between past, present, and future, but they do not deny that there once were dinosaurs. Again, if 'genuine' and 'objective' are dropped then the position becomes incoherent. These claims are not merely about what is *true*; they are about what is true at the *fundamental* level.

If the concept of structure is to play this role in metametaphysics, it must be generalized beyond Armstrong's notion of a universal and Lewis's notion of natural properties and relations. For many metaphysical questions are not about universals, properties, and relations. The crucial expressions in ontology, logic, and modality do not stand for universals, properties, or relations; these expressions are quantifiers and operators, not predicates. Our conception of

structure, therefore, must allow us to ask, of expressions of any grammatical category, whether they carve at the joints.

Call a language "fundamental" if all of its expressions carve at the joints. Realism about structure leads to realism about fundamental languages. On the generalized conception of structure, in order to be fundamental, it is not enough that a language have the right predicates. It must also have the right logical apparatus. Will a fundamental language contain quantifiers? The sentential connectives of propositional logic? Modal or tense operators? The realist about structure thinks that these questions have objective answers. There is a privileged way to "write the book of the world".

Chapter 2

Primitivism

I CANNOT DEFINE 'structure'. As we will see, a rich characterization can be given: connections to other concepts, theses about its behavior, and an official regimentation for talking about it. But none of this will add up to a definition. Indeed, I will argue in section 7.13 that structure is perfectly fundamental.

2.1 Understanding

I know from bitter experience that philosophers are wary of this primitivism. Many times I have been asked (to murmuring general approval): "What on earth do you *mean* by 'structure'??".

Let's be realistic about the extent and value of definitions. Philosophical concepts of interest are rarely reductively defined. Still more rarely does our *understanding* of such concepts rest on definitions.

On what does our understanding of philosophical concepts rest? Sometimes there is a perceptual basis: we directly experience space and time, perhaps. But the perceptual model is mostly unhelpful (think of modality, logic, laws of nature, identity over time, morality, justice, knowledge, justification, …). We generally "understand" philosophical concepts to the extent that we know what role they play in our thinking. (Understand "role" here very broadly, so as to include particular "cases"—we judge Gettier's (1963) Smith not to know that Jones owns a Ford or Brown is in Barcelona—as well as general inferential patterns—we think of identity as a transitive relation; we think of inability to refrain from action as an excuse.)

Philosophers sometimes slip into a magical-grasp picture of understanding. An opponent wields a crucial term. She will not be bullied into equating it with some combination of preferred terms. An inward search for a mystical mental state of UNDERSTANDING comes up empty. The opponent is pronounced confused or obscure.

Philosophical terms *can* be unclear: when they have been given no clear theoretical role to play. But 'structure' has a relatively clear role—given in this book and elsewhere. What more is wanted? The perceived magical grasp of more familiar concepts like modality, in-virtue-of, or law of nature, is due solely to the fact that we've become accustomed to talking about them. The theoretical roles backing those concepts are no richer or better specified than the role backing structure. Philosophy is not just the building of theories on previously existing concepts. We also build new concepts, by building theories that use them.

This is not to say that all there is to meaning and reference is inferential role. Meaning and reference may well be determined by external factors that transcend inferential role (see section 3.2). So even if structure's inferential role is richly specified, the concept may nevertheless fail to refer to anything. But that's true of any philosophical concept: the world may simply fail to contain anything—or any unique thing—fitting the inferential role associated with the concept. My hope is that this unhappy possibility is not realized.

2.2 Primitivism supported

And my *argument* that the unhappy possibility is not realized is simply the overall argument of this book: recognizing structure improves our understanding of the world—our understanding of:

(Objective) similarity (chapter 1);

Intrinsic properties (Lewis, 1986*b*, pp. 61–2);[1]

Laws of nature and explanation (section 3.1);

Reference (section 3.2);

[1]Lewis defined "duplicates"—intuitively, perfectly intrinsically similar objects—as pairs of objects whose parts can be mapped one-to-one preserving parthood and perfectly natural properties and relations; and he then defined intrinsic properties as those that can never differ between a pair of duplicate objects, whether in the same or in different possible worlds.

Induction and confirmation (section 3.3);

The intrinsic structure of space and time (section 3.4);

Substantivity (chapter 4);

Epistemic value (section 4.5);

Metametaphysics (chapter 5);

Disputes about time, modality, ontology, logic (chapters 9–12)

(In the first four items I follow David Lewis.) Structure is a *posit*, a posit that is justified by its ability to improve our theories of these matters.

Posits are most justified when they're unifying. When a single posit can be viewed as underlying multiple phenomena, this counts in favor of the posit. But here we must distinguish two kinds of unification. Compare two unifying features of Newtonian mechanics. Newtonian mechanics contains two fundamental laws governing mass: the second law of motion, which specifies the acceleration of a body as a function of its mass and the net force acting on it, and the law of gravitation, which specifies the gravitational force between a pair of bodies as a function of their masses and spatial separation. The first unifying feature is the identification of inertial and gravitational mass: the very same notion of mass is asserted to be involved in both laws. This first unification is one of ideology, of the set of our undefined words/concepts/notions. Instead of containing two notions of mass, one for the second law ("resistance to acceleration") and one for the law of gravitation ("tendency to produce gravitational force"), Newtonian ideology contains just one notion: mass. The second unification is the derivation of the planets' elliptical orbits. Here there was unification of fundamental principles. The orbiting of the planets was shown to require no new fundamental laws, since elliptical orbits were shown to follow from the the second law and the law of gravitation. While both sorts of unification seem to count in favor of a posit, too much of the former sort without any of the latter seems rarely to be pursued. We like to keep our posits few in number, but we also want them to obey a small number of fundamental laws, from which much else can be derived.

To be sure, metaphysics isn't rocket science. In metaphysics we lack precisely formulated, sufficiently specific fundamental laws from which myriad important and precise consequences may be rigorously derived. Still, we can be guided by physics here. The posit of structure will be *un*justified if its alleged

"applications"—its connections to similarity, intrinsicality, and the rest—are in the end unified by nothing beyond a bare assertion that a single notion of structure plays the needed role in each case. That would be unification of the first sort without unification of the second sort. But positing structure is not like this sort of bare assertion, since the applications are all intertwined. For example, many of the applications can be seen as flowing from a single principle connecting structure to epistemic value (section 4.5). Also, the applications just seem intuitively appropriate. Similarity, intrinsic structure of spacetime, substantivity, and the rest really do seem to be connected to fundamentality in the ways to be proposed.[2] This claim is difficult to justify, but I hope the reader will agree once the applications have been laid out.

2.3 Epistemology

A typical follow-up to "What do you *mean* by 'structure'?" is "How are we supposed to *know*—or even, reasonably believe—anything about structure?" Unless structure is defined in more familiar terms, it's thought, facts about structure become epistemically inaccessible.

The epistemic worry leads to a further worry about understanding. Suppose we reject the magical-grasp picture of understanding, as urged above. Perhaps we replace it with a more inferentialist picture. But then, if facts about structure are unknowable, structure-talk becomes inferentially isolated, in which case our understanding of such talk is again threatened.

But why think that primitivism about structure has such drastic epistemic consequences? The dialectic here is a familiar one. A "realist" resists downsizing the facts in some domain—reducing them or regarding them as subjective—and so is accused of making the facts in that domain unknowable. Her response is fallibilist epistemology: the subjective or reductive facts that her opponents offer as replacements are instead fallible guides to the upsized facts.

Most philosophers are comfortable with taking the realist side of this dialectic in the most familiar case: they reject Berkeleyan idealism, phenomenalism, and other downsized conceptions of the external world, and instead regard

[2] This vague claim raises difficult issues. The intuitive conception of structure as involving *fundamentality* doesn't seem inert; if the entire theory of this book were replaced with its ramsey sentence, omitting all mention of fundamentality, something would seem to be lost. But in what does this intuitive conception consist? Is it simply further principles connecting fundamentality to other notions? Or does it somehow resist being captured in principles?

ordinary evidence as a fallible guide to upsized facts. But their ranks dwindle as the subject matter becomes more metaphysical. The reason for this is simple: many of our models of the nature of reasonable—albeit fallible—belief about the external world do not apply straightforwardly to beliefs about more metaphysical matters. For example, we do not seem to be in causal contact with the facts debated by metaphysicians in the same way that we are in causal contact with more familiar facts about the external world.

But the models that immediately disallow reasonable belief in metaphysics are too simplistic, and as a result are in trouble anyway. Our causal contact with the facts of logic, mathematics, and particle physics, for example, is quite unlike our causal contact with the facts of the everyday external world. The ray of hope for the metaphysician is this: when the models become more sophisticated, allowing for reasonable belief in logic, mathematics, and particle physics, perhaps they will also allow for reasonable belief in metaphysics as well.

The epistemology of metaphysics is far from clear; this any metaphysician should concede. For what it's worth, as a general epistemology of metaphysics I prefer the vague, vaguely Quinean, thought that metaphysics is continuous with science. We employ many of the same criteria—whatever those are—for theory choice within metaphysics that we employ outside of metaphysics. Admittedly, those criteria give less clear guidance in metaphysics than elsewhere; but there's no harm in following this argument where it leads: metaphysical inquiry is by its nature comparatively speculative and uncertain.

This Quinean thought suggests an epistemology for structure in particular. Quine's advice for forming *ontological* beliefs is familiar: believe the ontology of your best theory. Theories are good insofar as they are simple, explanatorily powerful, integrate with other good theories, and so on. We should believe generally what good theories say; so if a good theory makes an ontological claim, we should believe it. The ontological claim took part in a theoretical success, and therefore inherits a borrowed luster; it merits our belief. This all is familiar; but a believer in structure can say more. A good theory isn't merely likely to be *true*. Its ideology is also likely to carve at the joints. For the conceptual decisions made in adopting that theory—and not just the theory's ontology—were vindicated; those conceptual decisions also took part in a theoretical success, and also inherit a borrowed luster. So we can add to the Quinean advice: regard the ideology of your best theory as carving at the joints. We have defeasible reason to believe that the conceptual decisions of successful theories correspond to something real: reality's structure.

The term 'ideology', in its present sense, comes from Quine (1951*a*, 1953*b*). It is a bad word for a great concept. It misleadingly suggests that ideology is about ideas—about *us*. This in turn obscures the fact that the confirmation of a theory confirms its ideological choices and hence supports beliefs about structure. A theory's ideology is as much a part of its worldly content as its ontology.

The familiar Quinean thought is that we search for the best—simplest, etc.—theory that explains our evidence. My addition to this thought—though it may have been implicit all along—is that this search is ideological as well as doctrinal; we search simultaneously for a set of concepts and a theory stated in terms of those concepts. We solve for the best and most explanatory pair $\langle I, T_I \rangle$ of ideology I and theory T_I in terms of that ideology. We do not hold fixed our initial ideological choices ('fire', 'air', 'water'…) since there may be limits to how good a theory can be formulated in those terms. Many of the most dramatic advances in science are ideological; a new ideology (such as that of Minkowskian spacetime) can dissolve intractable problems and enable new, more powerful theories.

(Sometimes our evidence does not support a unique pair $\langle I, T_I \rangle$. This is not in itself worrisome; we do not know everything, after all. But in some cases, it is hard to see what evidence *could* be mustered in favor of one pair rather than another. For example, should our fundamental theory of part and whole take 'part' or 'overlap' as primitive? Should our fundamental logical theory take conjunction and negation, or instead, disjunction and negation, as primitive? In such cases it's hard to see how to choose, and indeed, hard to believe that there could be a single correct choice. We will return to this issue in section 10.2.)

The Quinean thought rationalizes commonly held beliefs about what is fundamental. Nearly everyone agrees that physical notions like mass and spatiotemporality are fundamental. These beliefs are reasonable because those notions are drawn from highly successful theories.

The Quinean thought also rationalizes changes in beliefs about what is fundamental. The special theory of relativity led to (at least) two such changes. First, we came to regard electromagnetism as a single fundamental force, rather than regarding electricity and magnetism as separate fundamental forces.[3] And second, we came to regard spacetime as lacking absolute spatial and temporal separation. These changes weren't ontic: changes in which entities are accepted.

[3] See Maudlin (1996, pp. 131–3).

Nor were they merely doctrinal: changes in view, but phrased in the old terms. The changes were rather ideological: we revised our fundamental ideology for describing the world. The changes are rationalized by the Quinean thought because the fundamental ideology of the special theory of relativity differs from the fundamental ideology of Newtonian physics: in place of electrical, magnetic, spatial, and temporal ideology, the special theory has unified ideology for electromagnetism and unified ideology for spatiotemporal metrical structure.

The Quinean thought about ontology is sometimes put in terms of indispensability: believe in the entities that are indispensable in your best theory. The analogous thought about ideology may be similarly put: regard as joint-carving the ideology that is indispensable in your best theory. This is fine provided "indispensable" is properly understood, as meaning: "cannot be jettisoned without sacrificing theoretical virtue". The indispensability argument for mathematical entities is not refuted by just any nominalistic alternative to platonist mathematical physics. The nominalistic alternative must be attractive as a scientific theory; it must compete with the platonist theory for being simple, explanatory, and so on. Similarly, consider rewriting a given theory of mass and charge in terms of *schmass* and charge, where the schmass of an object is its mass if it has unit negative charge and twice its mass otherwise. The rewritten theory has the same consequences about charge and mass as the original, so 'charge' and 'mass' are in a sense dispensable in physics. But the resulting theory is far worse as a theory. What were syntactically simple generalizations in the old theory are no longer simple in the new.

We have been exploring the positive side of the Quinean approach to ideology: we can *support* claims about joint-carving by showing that the ideology in question is part of a good theory. The approach has a negative side too. Good theories must be as simple as possible, and part of simplicity is having a minimal ideology. So we can *oppose* claims about joint-carving by exhibiting good theories that do not contain the ideology in question.

The demand for minimal ideology recalls a familiar trade-off between ontology and ideology. We often face a choice between reducing our ontology at the cost of ideological complexity, or minimizing ideology at the cost of positing new entities.[4] If ideology is psychologized, the trade-off is one of apples for oranges: whether to posit a more complex world or a more complex mode of expression.[5] But on the present approach, both sides of the trade-off

[4]See, for example, Quine (1976*a*).

[5]See Oliver (1996, section 2).

concern worldly complexity. A theory with a more complex ideology posits a fuller, more complex, world, a world with more structure. Thus ideological posits are no free lunch.

"Believe the ontology and ideology of your best theory" is schematic in various ways. One in particular is worth mentioning: should the special sciences be counted as part of our "best" theory? Saying yes leads to an expansive conception of the fundamental; saying no—my preferred answer—leads to a more austere conception. The defender of the latter answer must concede that our understanding of the world would be severely impoverished without the special sciences, but will insist that since facts about the special sciences hold "in virtue of" more fundamental facts in some sense (chapter 7), they needn't be cited in our "best" theory. This is a difficult issue, which I will not attempt to resolve here.

We have, then, an epistemology for structure. Claims about structure can be supported by evidence, and so are not inferentially isolated, and so are not in danger of unintelligibility. This epistemology is admittedly superficial and birds-eye. Then again, so are the epistemological models that are claimed to preclude reasonable belief in metaphysics. What is needed for progress in these issues is a more sophisticated and detailed understanding of the epistemology of our more theoretical endeavors, such as mathematics and theoretical physics, including their foundations.

2.4 Against reduction

Primitivism about structure would be unnecessary if structure could be reduced. This section will argue briefly against a few reductive approaches. But the matter cannot really be decided by a few quick arguments, since it turns on the question of which global metaphysics is most attractive. My primary aim in this book is to exhibit the attractions of my own approach, rather than to refute others. (Also, much of my approach could be adopted by a reductionist, provided the reduced notion of structure is sufficiently objective and capable of broad application.)

Consider first reductions to putatively fundamental concepts. One can object to such reductions in two ways: extensionally and systematically. One can argue that a reduction fails to generate a reasonable extension for 'structure'. Or one can argue, on global, systematic grounds, that structure itself, rather than the proposed reducing concept, should be taken to be fundamental.

Consider, for example, the proposal that a structural property or relation is one that figures in some law of nature, where the notion of a law of nature is taken to be fundamental. In my view, logical and mathematical notions, as well as the notion of structure itself, carve at the joints; but it is unclear whether these notions figure in the laws of nature. A notion of structure that is too closely tied to lawhood will not be general enough to do all the work it needs to do. This is an extensional objection. But there is also a systematic consideration: it is better to posit fundamental structure than fundamental lawhood. (One could posit both, but that would be overkill.) Later on I will say more against all forms of fundamental modality, including fundamental lawhood; but in brief: modal notions are generally of dubious explanatory value. Adding the notion of law to physical theory, for example, doesn't seem to enhance its explanatory power. Also, the claims of physicists do not bifurcate naturally into laws and mere facts (which is "spacetime is a four-dimensional Lorentzian manifold"?), so introducing a fundamental notion of law imposes a distinction on their inquiry that isn't there to begin with.[6]

For similar extensional and systematic reasons, we should not define structure in terms of a fundamental notion of causation. Causation is a particularly unsavory fundamental posit—at least if the posit is intended to closely match our ordinary concept of causation. It takes only a glance at the recent literature on causation to appreciate how arbitrary and baroque our ordinary concept of causation is.[7] One might posit a fundamental sort of causation that is only distantly related to our ordinary concept, perhaps on the grounds that the

[6]See section 3.1 on laws and chapter 12 on modality.

[7]This literature, it seems to me, needs to make up its mind whether it is about fundamental metaphysics or conceptual analysis. (Hall (2006) and Lewis (2004) are notable for being clear which they are pursuing—the latter.) Above all, it is important not to shift uncritically between the two conceptions of the project, since they call for different methodologies. If the project is conceptual analysis, then heavy reliance on thought experiments is appropriate; but it is far less clear that one can insist on premises like: absences cannot be causes; abstract entities like facts cannot be causes; there must be uniform causal relata; causation cannot be contextual or normative or extrinsic. One could claim that these premises are part of our ordinary concept; but conceptual analysts generally regard intuitive judgments about particular cases as being far more diagnostic than intuitive judgments about such general principles. At the very least, one should evaluate these premises as alleged components of our concept, and not as if they are supported by considerations of simplicity or fit with a favored fundamental metaphysics. If, on the other hand, the project is to investigate the fundamental metaphysics of the causal relation, then it must be argued that there is such a thing—that there is a fundamental relation of causation; and heavy reliance on thought experiments must be abandoned, or else somehow justified. See also Paul (2009, section 1; 2010, sections 1–3).

sciences demand recognition of some sort of fundamental scientific causation. The case for fundamental causation is, I suspect, strictly weaker than the case for fundamental laws of nature; but at any rate, the objections to defining structure in terms of fundamental scientific causation are the same as those in the previous two paragraphs.

Consider, next, a proposal to define structure in terms of supervenience, where supervenience is then defined in terms of a putatively fundamental notion of metaphysical necessity.[8] Call a set of properties and relations *complete* iff all properties and relations supervene (globally, across all metaphysically possible worlds) on it; and call a set *minimally* complete iff it is complete and contains no proper subset that is also complete. The proposal I want to attack is this: the set of structural properties and relations can be defined as the minimally complete set of properties and relations.

Here too there is the systematic objection that structure is a better fundamental posit than any modal notion. But there are also extensional objections. In Sider (1996*b*, section 3.3) I gave two.[9] First, it is unclear whether the structural properties and relations are complete: nonqualitative properties and relations (such as the property of being identical to Ted) may not supervene on them, and it's hard to exclude nonqualitative properties and relations in the analysis without helping oneself to some notion in the vicinity of structure. Second, and more importantly, even if the set of structural properties and relations *is* complete, it may not be minimally complete. Perhaps both earlier-than and later-than are structural relations, for example. Third, and still more importantly, even if the set of structural properties and relations is minimally complete, it won't be the only such set: certain other sets containing "grueified versions" of the structural properties and relations will also be minimally complete. Begin with a set containing charge and mass, say, and replace mass with schmass (see above). Since an object's mass supervenes on its charge and schmass, any set that supervenes on the former set will supervene on the latter set; so the latter set is complete. And minimally so: since it doesn't contain mass properties, removing the schmass properties would disrupt its

[8]To say that some properties supervene on others is to say that in some sense, the former properties cannot vary independently of the latter properties. Supervenience is normally defined in terms of possible worlds and individuals, which are then defined in terms of necessity (and ancillary notions). For an overview of these issues, see McLaughlin and Bennett (2005) on supervenience and Divers (2002) on possible worlds.

[9]The proposal was there taken as a definition of Lewisian perfect naturalness. See also Lewis (1983*a*); Marshall (2009).

completeness.

There is a response to my third argument that I did not consider in my (1996*b*) paper. Perhaps we can single out the set of structural properties and relations, from amongst all the minimally complete sets, as that set that "enables the best Lewisian laws", in the following sense. Lewis used his notion of naturalness to give a reductionist, "Humean" theory of laws of nature, a souped-up version of the old regularity theory.[10] Neglecting complications involving chance, he defined a law as any generalization in the "best system"; that is, in the deductive system, stated in a language whose predicates stand for natural properties and relations, that best balances the virtues of strength and simplicity. A system is strong depending on how much (in some suitable measure) information it entails; a system is simple depending on how simply (in some suitable measure) it can be axiomatized. Now, Lewis pursued this project under the assumption that naturalness was primitive (or near enough). But the suggestion I want now to consider is that we reductively define structural properties and relations in terms of goodness of deductive systems and supervenience, as follows. Say that a deductive system is *based on* a set of properties and relations iff each primitive predicate in that system expresses a property or relation in the set. And say that one set, S_1, *enables better laws than* another set, S_2, iff some system based on S_1 is better—i.e., better balances strength and simplicity—than any system based on S_2. (Perhaps the laws of the systems should also be required to nontrivially utilize all the members of the sets.) With any luck, there exists some minimally complete set that enables better laws than every other minimally complete set. If so, structural properties and relations may be defined as the members of this set.

This proposal is worth thinking about more, but the other objections given above remain. The systemic objection remains: better to posit basic structure than basic modality. And the first two of my (1996*b*) objections remain. The first might be addressed by simply taking 'qualitative' as primitive, and the second might be addressed by claiming that neither *earlier-than* nor *later-than* is structural (those who deny that time has an intrinsic direction will already want to say this). But other objections in the vicinity of the second objection remain. Since supervenience is defined modally, mathematical properties and relations will supervene trivially on any set whatsoever, given the common dogma that mathematical facts are necessary. Thus, these properties and relations will not be present in any *minimally* complete set. But our conception of structure would

[10] See Lewis (1973*b*, pp. 73–4; 1983*b*, pp. 366–8; 1986*c*, pp. 121–4; 1994).

be impoverished by their exclusion. The distribution of structural properties and relations is supposed to give the fundamental facts about the world, and we might well want to say that the fundamental facts include mathematical facts.[11]

We have considered reductions to putatively fundamental law, cause, and necessity. What of reduction to one of these notions construed nonfundamentally? The systematic objections, anyway, would be avoided. But the reduction would be circular if the nonfundamental notion were in turn reduced to structure. And I suspect it would need to be. (Consider, for example, Lewis's account of law and cause, and my account of necessity in chapter 12.)

2.5 Against subjectivity

Primitivism would also be unnecessary if structure were tied to *us* in some way—to human language, biology, history, or psychology. According to this view, what distinguishes the class of the electrons, as opposed to the class of the electron-or-cows, is just that humans have a simple predicate for the former class, find it psychologically more natural to think in terms of 'electron', and so on; it doesn't go any deeper than that. This kind of subjectivism about structure could be taken as a reduction or, alternatively, as expressivism or some other form of noncognitivism.

Speaking just for myself, this is incredible. It is really, really hard to believe that the fact that electrons go together, in a way that electrons-or-cows do not, is merely a reflection of something about us.

But this is autobiography, not argument. The *argument* here, such as it is, is that any subjectivity in the notion of structure would infect all the domains in which structure is applied. If structure is just a reflection of our language (or whatever) then so are the facts about similarity, intrinsicality, laws of nature, the intrinsic structure of space and time… And this is incredible.

At its last step the argument again reverts to autobiography. Certain philosophers will rightly remain unconvinced, for example "antirealists" of various stripes—pragmatists, Kantians, logical positivists, and so on.

A certain "knee-jerk realism" is an unargued presupposition of this book. Knee-jerk realism is a vague picture rather than a precise thesis. According to the picture, the point of human inquiry—or a very large chunk of it anyway, a chunk that includes physics—is to *conform* itself to the world, rather than to *make*

[11] See also section 10.2.

the world. The world is "out there", and our job is to wrap our minds around it. This picture is perhaps my deepest philosophical conviction. I've never questioned it; giving it up would require a reboot too extreme to contemplate; and I have no idea how I'd try to convince someone who didn't share it.

2.6 The privilege of the physical

A final consideration in favor of primitivism about structure may be advanced. It is based on knee-jerk realism.

Let Γ be the set of true sentences in the language of completed physics, and consider two sets of propositions. The first set, P, is the set of propositions expressed by the members of Γ, under their intended interpretation. Thus P is the set of physical truths. The second set, S, consists of "scrambled" propositions. To arrive at S, reinterpret all nonlogical symbols of the language of physics under some arbitrary permutation μ of the totality of objects (see section 3.2 for a full description of this sort of reinterpretation), and let S be the propositions expressed by the members of Γ thus reinterpreted. The members of S are all true, since the members of Γ are true under their intended interpretation, and reinterpretation under a permutation preserves truth (again see section 3.2).

The consideration is then this. Even though both P and S consist of *true* propositions, knee-jerk realism requires recognizing that there is something *better* about P. Believing the members of P would be better than believing the members of S; P constitutes a better description of reality than S. To deny these things would be to admit that there's nothing mandatory about physics, that other perspectives on the world are "just different", not worse. Knee-jerk realism is incompatible with the thought that it's just *optional* to think in physical terms, that it would be just as good to pick wholly arbitrary carvings of the world (meanings under μ) and think in those terms.

Knee-jerk realism further requires that the betterness be *objective*. It isn't merely that the propositions in P have a simpler description in our language, or are more useful if one is trying to stay alive, let alone build rocket ships (though this usefulness is good *evidence* for the betterness.) The betterness, it is natural to conclude, is that only the propositions in P are cast in joint-carving terms.

Are there less metaphysical terms in which one might characterize the inferiority of S? It's hard to see what they might be. For example, P is not

more "complete" than *S* in any interesting sense. Neither set contains all the true propositions. Granted, *S* is silent on questions of physics, but *P* is silent on questions of "shmysics" (the subject matter of *S*); what we need is some objective sense in which the first omission is worse.

At this point an opponent might try out some of the moves of section 2.4. She might say that *P* is better because it contains laws, or because it is associated with a complete set of properties and relations ("associated" in the sense that *P*'s members are expressed by simple sentences in a language with predicates for the members of the complete set; "complete" in the modal sense introduced in section 2.4). Systematic reasons against appealing to modality—whether nomic or metaphysical—were given; but also, the appeals are intuitively off-target. The deficiency of *S*, surely, inheres in its categorical, real-world, nature, and should be explicable without bringing in the ghosts of what might have been.

Knee-jerk realism, then, requires that the physical description of reality be objectively privileged. And a natural account of the privilege is that physical notions carve at the joints.

Chapter 3

Connections

> … as I bear [the distinction between natural and unnatural properties] in mind considering various topics in philosophy, I notice time and again that it offers solutions to my problems.—David Lewis (1983*b*, p. 343)

Structure is central in and around metaphysics. We saw in chapter 1 how it connects to similarity. The next three chapters explore further connections.

Each connection is a proposal, stated in terms of structure, for how to think about a certain topic. Although none of these proposals is mandatory for the believer in structure, each is natural and attractive.

This chapter discusses explanation and laws, metasemantics, induction, and physical geometry. These connections (especially induction and physical geometry) will play only a peripheral role in the rest of the book. The following two chapters discuss more central connections. Chapter 4 connects structure to questions about the status (substantivity, conventionality, objectivity) of disputes in general, and chapter 5 connects structure to questions about the status of metaphysical disputes in particular—to metametaphysics.

3.1 Explanation and laws

Many connect laws of nature with fundamentality, in one way or another. Primitivists about lawhood sometimes define fundamental properties as those that are involved in the laws, or else claim that "fundamental properties are

those that are involved in the laws" is a substantive principle connecting two primitive notions. Conversely, a primitivist about fundamentality—in my case, about structure—might define laws in terms of structure.

The Lewisian approach to laws of nature is an example of the final option. As we saw in section 2.4, Lewis defined laws as generalizations in the best system—the deductive system, cast in a language whose predicates express natural properties and relations, that best balances the virtues of simplicity and strength. The restriction on the language of the best system is essential; otherwise, as Lewis (1983*b*, p. 367) points out, a simple and maximally strong theory could be given with a single, simple axiom, $\forall x F x$, where F is a predicate true of all and only things in the actual world. All true generalizations would be counted as laws.

The metaphysical core of Lewis's theory is its Humean, negative part: the rejection of metaphysically fundamental laws of nature. Fundamentally speaking, the world is anomic. The best reason to accept this Humean core is parsimony.[1] We ought, other things being equal, to keep fundamental theories simple; and metaphysically fundamental laws would add complexity without adding to understanding. The claims of physics aren't explanatorily enhanced by adding that those claims are laws.[2]

The positive part of Lewis's theory is its definition of lawhood. This part is attractive, even more so if it is augmented as follows. Adding more content to a system generally makes it stronger but also more complex. Since the best system must balance strength and simplicity, additions are justified only if the benefit of the added strength outweighs the demerit of the added complexity. Question: how much complexity can be tolerated to gain a given amount of strength? That is, how much does complexity "cost"? Different answers correspond to different notions of law. The more expensive complexity is made, the simpler the best system will need to be, thus making fewer generalizations count as laws. Special science generalizations, for example, are more complex than those of physics when they involve notions that carve at the joints less well than do physical notions, and when they include ceteris paribus conditions (whatever that means exactly, it surely detracts from simplicity). So let a "middling" assignment of cost to complexity be one that counts the generalizations of physics as laws, but only barely. This corresponds to a sense of 'law' in which there are laws

[1] Lewis's reason is different: metaphysically fundamental laws would require "necessary connections" (Lewis, 1983*b*, p. 366); but see section 12.5.

[2] See also section 12.1.

of physics, but in which certain special-science generalizations do not count as laws.[3] If complexity is instead made cheaper, then those special-science generalizations will count as laws. (And if complexity is made expensive, so that very little complexity is tolerated, then even physical generalizations no longer count as laws; all that remain are laws of metaphysics and logic. See sections 10.3 and 12.5.)

Once augmented, Lewis's theory of law is good as far as it goes. But I doubt that the notion of law is quite as central as philosophers (or metaphysicians, anyway!) think. Consider these claims: spacetime is a four-dimensional Lorentzian manifold; the universe began with an initial singularity; the universe began in a state of very low entropy. It's a stretch to call these laws, yet they're perfectly central to physics. If the point of the notion of law is its role in a rational reconstruction of science, then we should broaden our focus, to include these non-laws as well.[4] And the centrality of laws arguably diminishes once we move beyond physics to the special sciences. It has been argued, for example, that there are no laws of biology.[5]

The core insight of Lewis's account of laws can be generalized beyond the case of laws: good scientific theories, whether or not they cite laws, must be cast in joint-carving terms. We may put this in terms of explanation: "theories" based on bizarre, non-joint-carving classifications are unexplanatory even when true.[6] Theories whose basic notions fail badly to carve at the joints fail badly as theories, even if they are exemplary from an "internal" point of view, for their inner workings fail to mirror the inner workings of the world. We know on cardinality grounds that there are functions from the motions of the planets to past stock market performance under which the motions correspond to the fluctuations of the Dow Jones industrial average to date. But if someone were

[3] I have in mind special-science generalizations that are physically contingent—perhaps because they depend on certain physically contingent "initial conditions".

[4] It might be held instead that the point of the notion of law is to play some role in systematic metaphysics. For example, laws are commonly taken to play a constitutive role in the analysis of counterfactuals: we "hold constant" the laws in determining what would have happened under a counterfactual supposition. Even so, we should still broaden our focus beyond laws, since surely we hold constant these non-laws when evaluating counterfactuals.

[5] See Hamilton (2007, section 2) for a survey.

[6] Hirsch (1993, chapter 3, section 7a) argues that explanations cast in joint-carving terms can be recast, without explanatory loss, in a priori necessarily equivalent non-joint-carving terms. But the recast explanation will have "syntactic" demerits, such as being highly disjunctive. So let us refine our claim about explanation: good explanations must be cast in joint-carving terms when stated in syntactically ideal form.

actually to produce such a function, no one would regard it as being explanatory (and no one would expect the correlation to continue). In contrast, explanations that cite facts about the geometry of spacetime or the initial singularity *are* genuine (whether or not the cited facts are laws), in part because the cited facts are stated in joint-carving terms.

This dimension of theoretical excellence is best exemplified by theories cast in perfectly fundamental terms—theories of fundamental physics, for example. But it is exemplified, to varying degrees, by special-science theories as well, since the notions of the special sciences carve at the joints reasonably well.

3.2 Reference magnetism

One of the "problems" Lewis used his notion of naturalness to solve was the problem of radical semantic skepticism (1983*b*; 1984). The problem is one in metasemantics. How do words (or thoughts—but let's stick to words) get their meanings? What "semantic glue" attaches them to the world? There are different views about the nature of the semantic glue, but on nearly all of them, the glue doesn't seem to be sticky enough; it apparently cannot secure meaning with sufficient determinacy. Most roughly put: what I mean by 'pig' is surely determined by such facts as that I've always said 'pig' when in the presence of pigs; but why do such facts determine that by 'pig' I mean pigs, rather than pigs-I've-encountered-in-the-past, or pigs-in-my-immediate-vicinity, or pigs-before-2011-A.D.-or-cows-afterwards or …?[7]

To bring out the problem more fully, we must consider concrete proposals for the nature of the glue. Following Lewis (1984), I'll illustrate how the problem goes for a toy descriptivist theory of the nature of semantic glue. According to this descriptivism, there is a set of sentences, *S*, such that our words mean whatever they must in order for the sentences in *S* to come out true. (This theory is schematic; different versions specify different sets *S*. *S* might, for example, be taken to include "definitional sentences", whatever that means exactly.)

The sentences in *S* do not, on their own, provide enough semantic glue. Let (F) be an intuitively false sentence of our language that is logically consistent with *S*—"Some pigs have wings", say. We had better be able to say that (F) is false (otherwise, as Fodor would say, it's the end of the world). Now, (F)

[7]The problem derives ultimately from Wittgenstein (1958), and in the form presented here from Putnam (1978, part IV; 1980; 1981, chapter 2) and Kripke (1982).

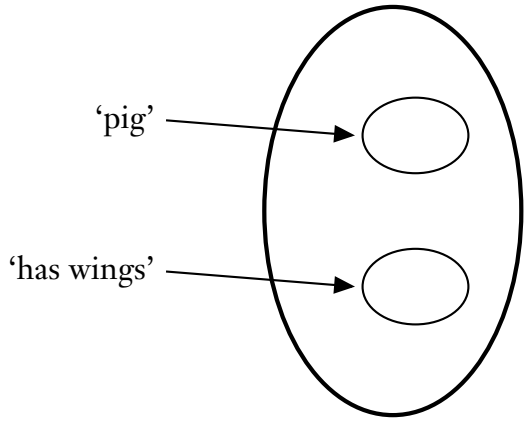

Figure 3.1: Interpretation 1

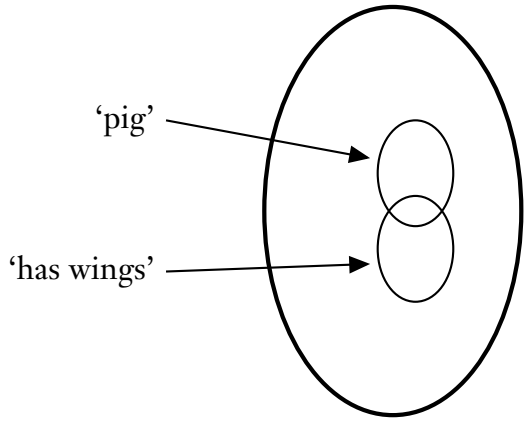

Figure 3.2: Interpretation 2

turns out false *in an interpretation* iff nothing is both in the extension of 'pig' and also in the extension of 'has wings' in that interpretation (figure 3.1). And (F) turns out true in an interpretation iff these extensions overlap in that interpretation (figure 3.2). So if (F) is to turn out false, the *correct* interpretation of our language must be of the former sort; interpretations like interpretation 2 are *incorrect*—such interpretations do not reflect what we really mean.[8] But if—as the descriptivist says—all that is required of a correct interpretation is that the sentences in *S* come out true under that interpretation, then we are

[8]Let us ignore the complicating factor that there will exist multiple correct interpretations, because of benign sorts of semantic undetermination like vagueness.

pretty much *guaranteed* to be able to construct a correct interpretation like interpretation 2. All we need to do is assign extensions to predicates so that every sentence in *S*, plus (F) as well, turns out true. We might, for instance, begin by assigning the set of hard-boiled eggs to 'pig' and assigning the set of edible things to 'has wings'. This makes (F) true since some hard-boiled eggs are edible. Now, suppose (S) contains the sentence 'Every pig is an animal'. The interpretation we are constructing must count this sentence true as well. But this is easy to accomplish: simply assign the set of eggs to 'is an animal'. Since every hard-boiled egg is an egg, 'Every pig is an animal' comes out true: the bizarre assignment to 'pig' is "cancelled out" by the equally bizarre but compensating assignment to 'is an animal'. We then continue in this way, assigning compensating bizarre extensions so that every other member of *S* comes out true as well.

The extensions assigned by this bizarre interpretation to 'pig', 'has wings', 'animal', and other predicates are clearly not what those predicates really mean, by any ordinary standard of "really mean". But our descriptivist says that to be correct (i.e., to reflect what we really mean), an interpretation need only make every member of *S* come out true. So the bizarre interpretation 2 would count as a correct one, and so we could not say that 'Some pigs have wings' is false! Something has plainly gone wrong, and it is pretty clear what that is. Descriptivism is false as stated; there must be more to the correctness of an interpretation than merely making certain specified sentences come out true.

The argument assumes that it is possible to continue selecting meanings for the nonlogical expressions in our language so that *every* member of *S* (and (F) as well) turns out true. But this will always be possible, except in special cases. Let *C* be the class of entities that we are using to interpret our words, and consider any abstract model—in the logician's sense—in which (F) as well as all members of *S* are true, and in which the domain is no larger than *C*. (The only cases in which there would not exist such a model are: i) if (F) is inconsistent with *S*; or ii) if *S* and (F) are consistent but only have models that are larger than *C*. And we lose no important generality by stipulating that the sentences *S* are consistent with (F) and do not, in concert with (F), logically require the universe to be larger than *C*.[9]) Since the domain of this abstract model is no larger than *C*, it can be mapped by some one-to-one function μ

[9]If the sentences in *S* and (F) are all first-order then, by the downward Löwenheim-Skolem theorem, it would be enough to stipulate that *C* is infinite. And even if *S* contains second-order sentences, we can just stipulate that it does not contain the logically complex sentences necessary to force domains larger than *C*.

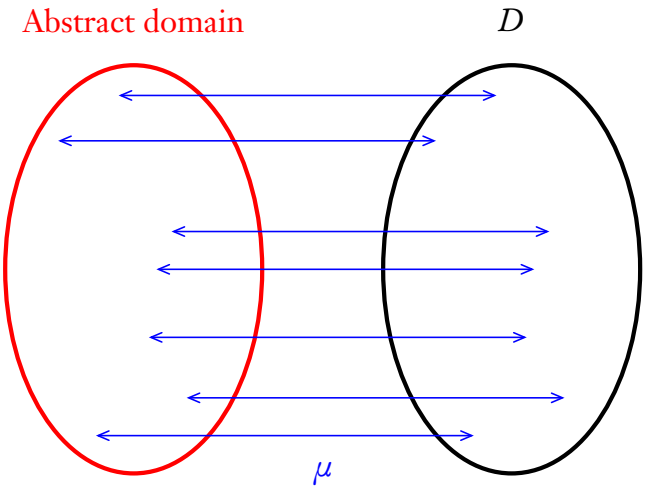

Figure 3.3: Abstract domain mapped one-to-one onto D

onto some subset, D, of C, as in figure 3.3. But then we can use μ to construct our desired interpretation 2 in D. Each nonlogical symbol in our subject's language has a meaning, m, in the abstract model; let the meaning of each such symbol in interpretation 2 be the object or set of 'tuples $\mu[m]$ in D to which m corresponds under the mapping. For example, 'pig' has an extension, E, in the abstract model, so the meaning of 'pig' in interpretation 2 will be the set $\mu[E]$ of members of D to which members of E are mapped by μ (figure 3.4). Now, it is an elementary fact from model theory that if a sentence is true in a model, it is true in any model constructed via a one-to-one mapping as above. So, since all the members of S, plus (F), are true in the abstract model, they are true in interpretation 2 as well.[10]

The argument shows that a correct interpretation must do more than make specified sentences come out true. What more? Lewis's answer is that correct interpretations must, as much as possible, assign natural properties and relations (or their extensions) to predicates. The meanings under interpretation 1 are (or correspond to) at least somewhat natural properties: being a pig, being edible, and so on. But there is no guarantee that the meanings assigned by interpretation 2 are natural to any significant degree, since they were constructed from the arbitrarily chosen function μ.

[10]The argument just given assumed an extensional conception of meaning. Williams (2005, chapter 5) shows how to rework it under richer conceptions of meaning.

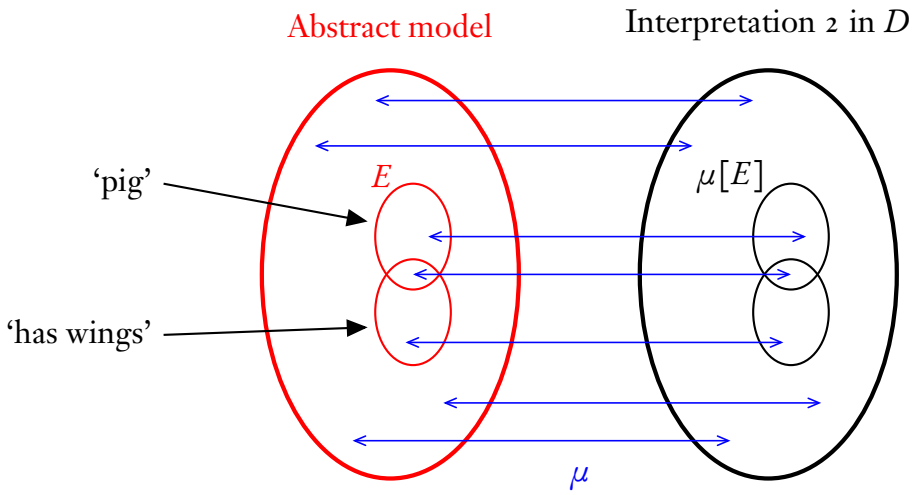

Figure 3.4: Interpretation 2 induced by μ

Lewis's proposal must be distinguished from the proposal that the sentence 'Predicates stand for natural properties and relations' is to be included in the set, S, of meaning-determining sentences. This proposal would not solve the problem. For so long as this sentence is consistent with (F) and the rest of the sentences in S, we will again be able to construct our interpretation 2. The resulting interpretation will misinterpret 'predicate', 'stand for', 'natural', 'property', and 'relation' so that 'Predicates stand for natural properties and relations', like the other sentences in S, come out true. Lewis's idea is not to stick with the original descriptivist constraint on correct interpretations ("correct interpretations are those that make every sentence in S true") and add more sentences to S; it is rather to modify descriptivism by adding an additional constraint. That constraint is not that 'Predicates stand for natural properties and relations' must come out true on a correct interpretation; it is rather, and more simply, that *predicates must stand for natural properties and relations in a correct interpretation*.

Lewis's constraint on reference is "externalist"; reference is not determined merely by us. We, of course, do part of the work. Our patterns of use of language determine which sentences go into S. But the degree to which an interpretation counts sentences in S as being true is just one factor in determining whether that interpretation is correct. The other factor is the naturalness of that interpretation's semantic values. And this second factor has nothing to do

with us. Natural properties and relations are "intrinsically eligible meanings"; they are "reference magnets".

The doctrine of reference magnetism has been received warily. Some regard it as an ad hoc response to a problem, with no independent backing. Others regard it as occult metaphysics, as the postulation of an irreducible "semantic force". (Perhaps the term 'reference magnetism', never used by Lewis but increasingly the doctrine's customary name, encourages this thought.[11]) Neither charge is founded. Following J. Robert G. Williams (2007, section 2), we can derive the doctrine of reference magnetism from a well-motivated and more general doctrine about theoretical virtue. This doctrine is the one defended in section 3.1: explanatory theories must be cast in joint-carving terms.

As I will develop it, the crucial assumption of the derivation is that *reference is an explanatory relation*—one can explain certain facts by citing what words refer to. But if reference were given by a bizarre interpretation, then reference-involving "explanations" would not in fact be explanatory, since they would be cast in badly non-joint-carving terms. Hence reference is not given by a bizarre interpretation.

Thus the exclusion of bizarre interpretations is of a piece with something we already do. We already regard "theories" based on bizarre classifications as being explanatorily useless; reference magnetism is just the extension of this point to metasemantics. Attempting to explain a community's linguistic behavior by citing a relation based on the permuted interpretation 2 would be like attempting to explain the behavior of the stock market in terms of an arbitrary correlation with the motions of the planets.

Let's look at the derivation of reference magnetism in more detail. Reference is a theoretical concept of a certain special science: semantics. A special science, quite generally, attempts to explain a certain target domain of facts by means of certain theoretical concepts. In the case of semantics the target domain is certain aspects of human thought, behavior, and communication, and the theoretical concepts include reference and truth. To use an example of Vann McGee's (2005, section 4), suppose a high-school teacher writes on the chalkboard the sentence 'Maiasaurs were highly social animals that traveled in herds of as many as 10,000.' Why did precisely that pattern of marks appear on the chalkboard? It clearly has *something* to do with the connection between those marks and maiasaurs. In particular, it surely involves the fact that the marks constitute

[11] The name was introduced by Hodes (1984, 135) for a related view (which he rejected).

words that *refer* to maiasaurs and their properties.

Some philosophers reject reference-based explanation. This is a big issue, not to be resolved here. But two small points. First, some claim that conventional meaning encodes far fewer referential properties than mainstream semantics thinks, but even they usually grant *some* explanatory role to reference—or at least to reference-by-a-speaker.[12] Second, granting the existence of deflationary concepts of truth and reference that cannot take part in explanations of the sort described above is compatible with also accepting nondeflationary concepts of truth and reference that can take part in those explanations.[13] At any rate, I will assume that some reference-based explanations are indeed explanatory.

If reference-based explanations are to be explanatory, then the reference relation must be a joint-carving one. Not a perfectly joint-carving relation, presumably; but at the very least, not a wildly non-joint-carving relation. And this excludes the bizarre interpretations. For only a wildly non-joint-carving relation would relate a linguistic population to the semantic values of a bizarre interpretation.

Caveat: suppose there were a *perfectly* joint-carving reference relation. Suppose, that is, that part of the absolutely fundamental story of the world included a specification of a relation of reference between certain sounds or inscriptions made by human beings and parts of the world. Such a fundamental relation could perhaps, as a brute fact, relate words to bizarre semantic values.[14]

But reference is surely not metaphysically fundamental.[15] As a result, it's hard to see how a reasonably joint-carving reference relation could relate us to the bizarre semantic values. The point may be seen initially by making two strong, crude assumptions about "reasonably joint-carving". Assume first that a notion is reasonably joint-carving iff it has a reasonably simple and

[12] See, for example, Chomsky (2000); Pietroski (2003); Wilson and Sperber (2004).

[13] See McGee (2005).

[14] A defender of this metaphysics might argue that it is *unlikely* that fundamental reference relates us to the bizarre semantic values. The simplest hypothesis, it might be claimed, is that we bear the fundamental reference relation to reasonably joint-carving semantic values; the reasoning here, it might be claimed, is analogous to any choice of a simplest hypothesis about the behavior of fundamental properties and relations. (Kripke (1982, pp. 38–9) complained that a simplicity response to his Wittgensteinian skeptic ignored the fact that the skeptical problem is not primarily epistemic, but rather is one of what constitutes semantic facts. But on the view here envisioned, semantic facts are fundamental, and the only remaining problem *is* epistemic.)

[15] See, however, Hawthorne (2006*a*, section 17).

nondisjunctive definition in terms of the *perfectly* joint-carving notions, and second that the perfectly joint-carving notions are those of physics. Then surely no reasonably joint-carving relation that is to play the role of reference could relate a human population to bizarre semantic values. For the bizarre semantic values themselves have no simple basis in the physical, nor do they stand in any physically simple relations to human populations. Given any relation that does relate us to bizarre semantic values, there is surely some other relation with a simpler basis in the physical that relates us to nonbizarre semantic values.

The two assumptions of the previous paragraph are undoubtedly too crude, but the point is independent of them. Whether a notion is reasonably joint-carving—enough to take part in special-science explanations—has *something* to do with how it is based in the fundamental. So reference must have the right sort of basis in the fundamental if it's to be explanatory. It's highly unclear what exactly the "right sort" of basis is (this is in essence the question of the relation between special-science concepts and physics), but it's quite clear that a relation connecting us to bizarre semantic values would have the wrong sort of basis—for the same reason that arbitrary correlations between the motions of the planets and the stock market have the wrong sort of basis.

Conceiving of reference magnetism in Williams's way risks restricting its scope. In addition to replying to the Putnamian semantic skeptic we discussed earlier, Lewis also used reference magnetism to reply to Kripke's (1982) Wittgensteinian skeptic about the meaning of mathematical language, who asks why by 'plus' we mean *plus* rather than *quus*, where quus is a function like plus except that it assigns to all pairs of sufficiently large numbers the value 5. Lewis's answer is that plus is a more natural function than quus; this becomes, in Williams's hands, the claim that a semantic theory assigning quus as a semantic value is less explanatory. But now, some will claim that abstract entities cannot take part in genuine explanations, perhaps because abstracta are causally inert. For myself, I reject this conception of explanation. Explanations come in many sorts, not all of them of the "pushing-and-pulling" variety. But at any rate, the constraint on reference magnetism's scope here seems like welcome discipline, not unwanted restriction.

The problem of radical semantic skepticism was presented above as it confronts descriptivist metasemantics. But other metasemantic theories face the problem as well, and reference magnetism—understood in Williams's way—can defend these other metasemantic theories as well, so long as they conceive of reference as an explanatory relation.

For example, Ruth Millikan (1989), who thinks of reference as a biological

phenomenon, also faces the problem. Why does the true reference relation, rather than numerous gerrymandered alternatives we could construct, count as the relation that has been selected by evolution for the purpose of storing and communicating information? Answer: *any* hypothesis which says that a gerrymandered relation plays a role in an evolutionary explanation is to be rejected simply because the relation is gerrymandered. The correct answer on behalf of Millikan to the skeptical worry is thus akin to any biologist's answer to the skeptical worry that it is facts about *gruegenes*, rather than genes, that explain inheritance. (The array of gruegenetic properties a thing has is determined by the physical makeup of the spatiotemporally nearest pair of bluejeans; on cardinality grounds, there are sufficiently many different gruegenetic properties to enable us to concoct "laws" of gruegenetics which really do correlate gruegenetic properties with inheritance as fine-grainedly as the real laws of genetics do with real genetic properties.) Similar remarks apply to causal theorists like Jerry Fodor (1987) (unless causation itself is taken to solve the problem[16]).[17]

When understood in the Williamsian way, we might think of reference magnetism as a thesis of *meta*-metasemantics, rather than a metasemantic theory in its own right: reference magnetism may be combined with any metasemantics you like. A metasemantics is a metaphysical proposal about the nature of the reference relation; schematically: "reference is a relation of such-and-such a type". Millikan holds that reference is a relation that was chosen by natural selection to achieve a certain goal; the descriptivist says that reference is a relation that assigns values under which sentences in *S*—"definitional" sentences, let us say—come out true; and so on. Each proposal faces the problem of semantic skepticism: there are many relations of the

[16]Causal theories say that reference is a certain sort of causal relation, and it's arguably built into the nature of causation that only reasonably joint-carving relations are causal. So causal metasemantics may not need reference magnetism. Even so, joint-carving remains crucial to metasemantics, via its connection to causation. See the discussion of causal theories below.

[17]These remarks defend only against the extreme undetermination threatened by the Putnamian argument, *not* against the kind of underdetermination about which Fodor (or for that matter, Quine in *Word and Object* (1960*c*)) is worried. The simple causal theories that Fodor rejects because they don't allow for misrepresentation generally employ notions that are just as joint-carving as those he himself employs in his own theory. Whether 'rabbit' determinately means *rabbit* rather than *undetached rabbit part* is similarly not settled by reference magnetism. (Not that reference magnetism is irrelevant; it may play a role in the story of why 'identity' means identity, which in turn is relevant to 'rabbit' meaning rabbit.) Note that subsentential undetermination of Quine's sort wouldn't result in indeterminacy in, e.g., whether (F) is true.

proposed type. How to cut down on this multiplicity? The answer of reference magnetism is to reinterpret a metasemantic proposal that reference is a relation of type T as the proposal that reference is a reasonably joint-carving relation of type T.

One way a relation could fit one of the modified proposals is for it to be defined in terms of the concept of joint-carving. (Assuming, that is, that the concept of joint-carving itself is reasonably joint-carving—see sections 7.11.1 and 7.13.) For example, the modified descriptivist proposal says that reference is a reasonably joint-carving relation that assigns values under which definitional sentences come out true; here is a relation that fits that proposal: $R(w, x) =_{\text{df}}$ w is a word and x is the semantic value assigned to w by the interpretation I that maximizes the combination of i) joint-carvingness of assigned semantic values; and ii) truth assigned to the sentences that are definitional in the linguistic population that uses w. But this is not the only way that a relation could fit one of the modified proposals: the relation could be defined in terms of reasonably joint-carving concepts other than the concept of joint-carving itself.

If reference magnetism is merely a thesis of meta-metasemantics, it can survive the demise of individual metasemantic theories based on it. Consider again descriptivism. To be more concrete, consider what we might call "simple charity-based descriptivism", a close cousin of the reference-magnetism-enhanced descriptivism discussed in the previous paragraph. According to simple charity-based descriptivism, the correct interpretation of a language is that interpretation that maximizes the combination of two factors, each of which comes in degrees: "eligibility"—determined by the degree of naturalness of the semantic values it assigns; and "charity to use"—the number (on some suitable measure) of sentences believed (or reasonably believed, or …) by the speakers of the linguistic community that come out true under that interpretation.[18] Since it's the combination of eligibility and charity that must be maximized, this theory implies that an imperfectly charitable interpretation can nevertheless be the correct interpretation, if it assigns sufficiently joint-carving properties. A highly eligible interpretation can "trump" the superior charity of rival interpretations.

[18]This view is often associated with Lewis; see especially:

> … overall eligibility of referents is a matter of degree, making total theory come true is a matter of degree, the two desiderata trade off. The correct, 'intended' interpretations are the ones that strike the best balance. (Lewis, 1984, pp. 227–8)

> … we need some give and take between the eligibility of referents and the other factors that make for 'intendedness', notably satisfaction of appropriate bits of theory. (Lewis, 1983*b*, p. 372)

However, Lewis's full view, as developed in (1969) and (1975), is more complex.

Now, this prediction is in some cases correct, especially for "theoretical" terms—terms that are, intuitively, *intended* to stand for joint-carving meanings. When a term like 'mass' is introduced in physics, it's intended to stand for a fundamental physical magnitude, and so if there's a joint-carving property in the vicinity then that property is meant by 'mass', even if it doesn't quite fit the physicists' theory of 'mass'. But for nontheoretical terms, the prediction of trumping sometimes seems wrong.

For example, imagine a linguistic community that uses the word 'amulet' in the same way we do, except that it just so happens that—with a single exception—the objects they call 'amulets' are all and only the instances of some highly joint-carving property—gold, say.[19] The single exception is that they call one silver ornament an amulet as well. Now, the property gold carves very well at the joints—much more so than the property *gold or silver*, or the property *ornament or small piece of jewelry thought to give protection against evil, danger, or disease*, to quote the dictionary. Moreover, nearly everything the community says about "amulets" comes out true if 'amulet' is assigned the property *gold*. So such an assignment seems to best combine the virtues of eligibility and charity; its superior eligibility trumps the superior charity of interpretations that also count the silver ornament as an "amulet". The simplistic account thus seems to imply that 'amulet' means *gold* in this linguistic community. But this is absurd. The silver ornament obviously counts as an "amulet"; the fact that it lacks a joint-carving property shared by all other objects called 'amulets' is irrelevant.[20]

Now, this particular example is easily addressed with minor refinements. For example, charity to use might be strengthened to require counterfactual robustness: charitable interpretations must make sentences that would be believed in certain counterfactual circumstances come out true in those circumstances. Since the community in question is disposed to call appropriate ornaments 'amulets' whether or not they're made of gold, the assignment of *gold* to 'amulet' no longer counts as charitable. But the underlying point of the example cannot be so easily dismissed: the fact that gold carves much better at the joints than *being an ornament or small piece of jewelry thought to give protection against evil, danger, or disease* just seems *irrelevant* to the question of which is meant by 'amulet'. No matter how "charity to use" is understood, examples of this sort will surely emerge. Consider, for instance, this example from John Hawthorne:

[19] *Being a maximal continuous portion of gold*, to be more exact.

[20] I learned of such examples from Matti Eklund, John Hawthorne, and Robbie Williams.

if there just happens to be a physically special line through the Ural mountains, that line isn't thereby the determinate boundary of Europe. The candidacy of a meaning for 'Europe' that takes this line into account is not enhanced by its superior eligibility.

It would seem, then, that for nontheoretical terms, the joint-carvingness of candidate semantic values plays a complex role. Highly joint-carving candidates don't automatically trump; simple charity-based descriptivism is too simplistic. But on the other hand, joint-carving isn't completely irrelevant either: the gruesome candidates of the semantic skeptic must still be excluded. Fortunately, even if we reject simple charity-based descriptivism, we can still appeal to reference magnetism to exclude the gruesome candidates. We can exclude them simply because they're gruesome—simply because they carve so badly at the joints. No reference relation connecting us to the gruesome candidates could take part in an explanatory theory. And we can say this even if we're uncertain what the true metasemantic theory is. In fact, we can say this even if we doubt that the truth about metasemantics can be captured in any simple formula—not an overly pessimistic doubt, given how few simple reductive theories of complex macro-phenomena have ever been given.

I have argued that reference magnetism solves the problem of radical semantic undetermination, is a consequence of a more general claim that explanatory theories must be stated in joint-carving terms, and does not presuppose descriptivism, but rather may be combined with any metasemantic theory that regards reference as explanatory. These conclusions collectively provide a reason to recognize structure: we need structure to answer the semantic skeptic. But there is a challenge to this reasoning: a causal metasemantics also lets us answer the skeptic. According to causal metasemantics, 'pig' means pigs rather hard-boiled eggs, say, because it is pigs rather than hard-boiled eggs that cause our uses of 'pig'.[21] See? No need for reference magnetism!

But a pure causal theory is likely to be insufficiently general. We need to rule out incorrect interpretations of mathematical and logical language, for example—as put forward by Skolemite skeptics about set-theoretic language, say, or Kripke's Wittgenstein—just as we need to rule out incorrect interpretations of 'pig' and 'has wings'; and it is hard to see how a pure causal account could do this. Likewise for terms of theoretical physics for hypothesized properties that are instantiated only under conditions present only at the

[21] In its most simplistic form, anyway. See Devitt and Sterelny (1999, chapters 4–5) for a more sophisticated discussion.

time of the big bang, or even conditions that have never been (and will never be) present at all.[22] Likewise for predicates for spacetime structure. Likewise for many of the concepts of philosophy, to be discussed in subsequent chapters (though how much structure there is in these areas is debatable). Reference magnetism, on the other hand, can be combined with a broader, not purely causal, metasemantics to rule out bizarre interpretations in these cases.[23] This is not to say that causation has no role to play. It's perfectly compatible with reference magnetism, construed as a metametasemantics, that causation is one ingredient of the semantic glue.

Further, the notion of causation required by causal metasemantics may presuppose the notion of structure.[24] To bring this out, consider an intuitively incorrect interpretation of our language that matches the correct interpretation up until some specified time, 3000 A.D., say, but then goes haywire afterward. For the sake of definiteness, here is one such interpretation. Indulge in the harmless simplifying assumption of "super-substantivalism", according to which the physical world consists purely of spacetime, and let ν be a one-to-one mapping over points of spacetime that maps each point at or before (the first instant of) 3000 A.D. to itself, and maps each point after 3000 A.D. to the simultaneous point 10^{10} km in some one chosen direction. Thus, ν leaves everything before 3000 A.D. alone, and applies a Leibnizian shift to everything afterward. Now, the intuitively correct interpretation of our physical vocabulary assigns semantic values involving points of spacetime to physical vocabulary. So ν induces a "shifted" interpretation: simply replace each point p in each semantic value with $\nu(p)$. What makes the intuitively correct interpretation rather than the shifted one correct? The shifted interpretation is derived from the correct one by a one-to-one mapping, and therefore renders the same sentences true. Furthermore, the shifted semantic values are instantiated by exactly the same spacetime points before the year 3000 A.D. as are the correct semantic values. The interpretations diverge only in the far future. So on the face of it, the shifted values seem to be just as causally responsible for past usage as the correct ones, in which case the pure causal metasemantics fails.

Veterans of the literature on natural kinds will be quick to notice that

[22] Note that the Ramsey-Lewis method for defining theoretical terms (Lewis, 1970*b*) has no hope of working unless the property quantifiers are restricted to properties that carve at the joints.

[23] See Lewis (1983*b*) on mathematical semantic skepticism, chapter 10 on logical joint-carving, and section 3.4 on semantic determinacy for spacetime language.

[24] Compare Hirsch (1993, pp. 63–5).

according to a "robust" conception of causation, the shifted semantic values will not be causally responsible for past usage—indeed, not causally responsible for anything at all. On this conception, even if a property *P* is involved in relations of counterfactual dependence, relations of nomic necessity or sufficiency, or what have you, it still may not be causally efficacious. If I hold a *grue* emerald in front of you, you will have green sensations; my holding the grue emerald plus background conditions necessitates those sensations; if I hadn't held up the grue emerald you wouldn't have had those sensations; and so on. (An object is "grue" iff it is green and first observed before 3000 A.D. or blue and not first observed before 3000 A.D. (Goodman, 1955, chapter III).) Nevertheless, so the story goes, my holding up the grue emerald did not cause you to have the sensations. Only my holding up of the green emerald caused you to do this.[25]

What we have learned is that causation must be robustly conceived, if it's to play the needed role in metasemantics. But how is robust causation to be understood? A robust cause might be defined as a joint-carving property involved in an event that causes something. Or instead as any property involved in an event that causes something, where our theory of causation invokes structure.[26] Or instead as any property that figures in a law of nature, where our theory of lawhood invokes structure.[27] Each of these approaches invokes the notion of structure. Only a primitivist, structure-free approach to causation (or laws) yields a structure-free approach to metasemantics; and against such approaches, recall section 2.4.

3.3 Induction and confirmation

We need structure to make sense of learning from experience.

The simplest model of learning from experience is that we remember past experiences, we expect the future to be like the past, and so we form appropriate expectations about the future. This model requires the notion of structure to be plausible. For the philosophical skeptic of section 1.2 will be quick to

[25] Davidsonians about causation must rephrase: my holding of the grue emerald did not cause the sensations in virtue of its being a holding up of a grue thing. See Davidson (1970); Kim (1989).

[26] Whether in its theory of events, or its theory of the causal relation, or both. See Lewis (1986*a*, 1973*a*, 1979).

[27] Even the primitivist accounts of Armstrong (1983), Dretske (1977), and Tooley (1987) presuppose sparse universals.

point out that any possible future is "like" the past along *some* dimension of similarity. The model had better say that the future is like the past in some *genuine* dimension, some dimension that respects nature's joints.

"The future is like the past" is too crude a concept on which to base a theory of learning from experience. We do a little better with the following concept: "observation *o* confirms sentence *S*". Analogs of the above worries about similarity then confront confirmation-based theories of learning from experience.

Which observations confirm a generalization 'All *F*s are *G*s'? A natural answer is the "Nicod principle": observations of *F*s that are *G*s confirm 'All *F*s are *G*s'. But suppose that an observation confirms any logical equivalent of any sentence that it confirms. Then, as Hempel (1945) pointed out, the observation of red roses confirms 'All ravens are black' (given the Nicod principle it confirms 'All nonblack things are nonravens', which is logically equivalent to 'All ravens are black'.) And as Goodman (1955, chapter III) pointed out, Nicod's principle implies that observations of green emeralds before 3000 A.D. confirm 'All emeralds are grue' (since green emeralds observed before 3000 A.D. are grue.) But anyone who believed that all emeralds are grue would expect emeralds observed after 3000 A.D. to be blue.

These conclusions can be avoided by restricting Nicod's principle in some way—most crudely, to predicates that carve at the joints.[28] Since 'is nonblack', 'is a nonraven', and 'grue' fail to carve at the joints, the restricted principle does not apply to generalizations containing them. In Goodman's terminology, only terms that carve at the joints are "projectible".

'Observation *o* confirms sentence *S*' is less crude than 'the future is like the past', but it is still too crude a concept for a realistic theory of learning from experience. A realistic theory will need to consider relative confirmation, and so surely must consider quantitative measures of confirmation, and so must surely employ the concepts of probability theory.[29] It is natural to do this within a Bayesian framework, according to which i) a rational subject's beliefs at a time consist in a probability distribution over all propositions, whose values measure the subject's "degrees of belief" ("subjective probabilities", "credences") in the propositions; and ii) the rational subject updates her degrees

[28]Compare Quine (1969), though Quine held a deflationary account of joint-carving (under the rubric of "natural kinds").

[29]Indeed, it has been suggested that if one attends to the comparative notion of confirmation, Hempel's puzzle, anyway, evaporates. See Fitelson (2006) for a survey.

of belief by conditionalizing on propositions describing her experiences.[30] In this framework one can introduce various quantitative measures of confirmation, and can characterize relative confirmation.[31]

Bayesianism proper tells the rational subject what method she should use for updating whatever degrees of belief she began with: conditionalization. But it says very little about what those initial degrees of belief ought to be like (beyond the minimal demand that they must satisfy the axioms of the probability calculus, and perhaps be "nondogmatic" in never assigning the values 0 and 1 except to necessary falsehoods and truths). Now, *subjective* Bayesians say that Bayesianism proper is all there is to rationality. That is, provided a subject has updated her beliefs by conditionalization (and at each moment obeyed the minimal synchronic demands), she has done all that rationality requires. *Objective* Bayesians, on the other hand, add further synchronic requirements on credences beyond the minimal ones. One's credences must, perhaps, assign higher probabilities to "simple" hypotheses, other things being equal, or obey some sort of principle of indifference, dividing credence equally over symmetric possibilities.

This difference between objective and subjective Bayesianism matters greatly, since the effect that evidence has on a Bayesian conditionalizer depends heavily on her prior probability distribution. Bizarre prior probability distributions will result in bizarre responses to evidence. Consider, for example, making a series of pre-3000 A.D. observations of green emeralds. Intuitively, this should result in increasing confidence that emeralds observed after 3000 will likewise be green. This increasing confidence is indeed forthcoming for a Bayesian if she begins with an appropriate prior probability function Pr—one that assigns high probability to emeralds observed after 3000 being green conditional on earlier observed emeralds being green. But if she begins instead with a prior probability function $\Pr'$ that assigns high probabilities to emeralds observed after 3000 being *blue* conditional on earlier observed emeralds being green, then observing the green emeralds will result in increasing confidence that emeralds observed after 3000 will be blue. Garbage in, garbage out.

The pure subjective Bayesian will say that rationality says nothing about the choice between Pr and $\Pr'$. If one person begins with Pr, and another begins

[30] That is, if the subject begins with probability function Pr, and has an experience described by proposition e, then she will subsequently have a probability function that assigns to any proposition, p, the probability that Pr assigned to p conditional on e, i.e., $\Pr(p/e)$, i.e., $\frac{\Pr(p \wedge e)}{\Pr(e)}$.

[31] See Fitelson (2006); Sober (1994).

with Pr′, and each then receives exactly the same evidence, which evidence includes observations before 3000 of many green emeralds (and no blue ones), and each updates her beliefs by conditionalization, then at the end of this process, neither has behaved more rationally than the other. This despite the fact that the second will become increasingly certain that emeralds observed after 3000 will be blue!

Subjective Bayesians embrace this conclusion. Fascinatingly, this descendent of Hume's notorious attitude toward induction is not uncommon in contemporary formal epistemology. Is this because the field draws the formally inclined, and the problem of constraining priors has proved formally intractable? At any rate, any Bayesian who wants to say that the second person has responded irrationally to the evidence must find fault with Pr′. (Since the subject used the correct rule of updating, the fault must lie with her prior degrees of belief.)

How might Pr′ be criticized? One (vague, in need of refinement) strategy would invoke "simplicity": Pr′ is worse than Pr because its description is less simple. The probabilities of Pr that concern emerald color can be given by simple rules that are uniform across all emeralds, such as: "for any emerald, the probability of that emerald being green, conditional on many other emeralds being green, is high". But to specify Pr′ we need more complex rules giving different conditional probabilities for an emerald's being green, depending on the times at which earlier green emeralds were observed.

But suppose we speak the language of 'grue' rather than the language of 'green'.[32] In this language, Pr′ can be specified with simpler rules, such as "for any emerald, the probability of that emerald being grue, conditional on many other emeralds being grue, is high", whereas Pr will require more complex rules, specifying different conditional probabilities for an emerald's being grue, depending on the times at which earlier grue emeralds were observed. Whether a probability function is simple depends on the language in which it is described. So a meaningful simplicity-based criticism of Pr′ requires some sort of restriction on the language in which we evaluate simplicity.

Similar remarks apply to attempts to constrain priors using some version of the principle of indifference. Any version of this principle says to distribute credence equally over "symmetric" possibilities. But which possibilities count

[32] This language replaces 'green' and 'blue' by 'grue' and '*bleen*', where an object is bleen iff it is blue and first observed before 3000, or green and not first observed before 3000. In a sense this language equals the 'green'/'blue' language in descriptive power; one simply says 'grue' instead of 'green' and 'bleen' instead of 'blue' for objects first observed before 3000, and 'bleen' instead of 'green' and 'grue' instead of 'blue' for other objects.

as symmetric will depend on what language we use to describe the possibilities. If we use the 'grue' language, an otherwise reasonable principle of indifference might recommend Pr′ over Pr. Now, even with a suitable language picked out, there are serious obstacles. To take a familiar example, suppose a factory is making cubes of varying sizes. If I have no information specific to a given cube produced by the factory, how should I assign prior probabilities to propositions of the form *the side length of the cube is between* l_1 *and* l_2? In direct proportion to the difference $l_1 - l_2$, one wants to say. And how should I assign prior probabilities to propositions of the form *the face area of the cube is between* a_1 *and* a_2? In direct proportion to the difference $a_1 - a_2$, one wants to say. But these two answers are incompatible (van Fraassen, 1989, 303–4). So selecting the right language doesn't, on its own, solve the problem of formulating principles of indifference. But without an appropriate language, we cannot even get started on a solution.

To constrain prior probability distributions, then, we need some way to pick out appropriate languages for evaluating simplicity, symmetry, and related notions. And—to finally get to the point—it seems reasonable to pick them out by using the notion of structure. Now, even given the notion of structure, there are nontrivial questions about how exactly to pick out the appropriate languages. For example, *how* well do the predicates in the appropriate languages need to carve at the joints? Again, the appeal to structure is the beginning of a solution, not the end of one.

The argument of this section has been that structure fills a need in epistemology. A reply would be that epistemology does not demand structure objectively construed; a conception of "structure" tied to human history, biology, psychology, or interests would do. My reply to such challenges elsewhere is that subjectivity in structure would infect all notions to which structure is connected (similarity, intrinsicality, duplication, laws of nature, and so on). But in this case, the infected notion would be epistemic value. And perhaps we should embrace the idea that values in general are not objective.

But even if epistemic value is subjective along *some* dimensions, we shouldn't embrace the idea that it's subjective along all dimensions. Let Pr be a rational credence function we ought to adopt and Pr′ be one that we ought not to adopt. Intuitively speaking, we might embrace subjectivity in the "oughtiness" of the obligation while rejecting subjectivity in the distinction between Pr and Pr′. The objective facts might not mandate that we have our notion, or any notion, of epistemic obligation, but might nevertheless mandate that we choose Pr over Pr′ if we do have our notion (or anything like our notion). We will return

to this issue in section 4.2.

3.4 Intrinsic structure in physical spaces

We need structure to understand talk in physics of the "intrinsic" structure of space, time, spacetime, and other spaces.[33]

NonEuclidean geometries were discovered in the early nineteenth century, proved consistent relative to Euclidean geometry later that century, and applied in physics by Einstein in the early twentieth century, in his general theory of relativity. Taken at face value, Einstein's claim that physical spacetime is curved is a substantive claim in direct conflict with the assumption of flat spacetime implicit in classical Newtonian physics (and in the special theory of relativity as well). But taken at face value, this claim raises various philosophical questions: epistemic, semantic, and metaphysical.

Let us approach the questions through the simpler case of spatial, rather than spacetime, curvature. Imagine a series of experiments, carried out with rigid measuring rods and the like, that apparently show that space in a certain region is curved. The epistemic questions arise because we do not observe space directly; what we observe is things *in* space, such as measuring rods. Effects attributable to curved space could instead result from systematic distortions to the rods. While this may at first appear to be mere Cartesian demonry, compare two alternative hypotheses. According to the first hypothesis, the measurements result from curvature. According to the second, space is flat but there are "universal forces" that affect all matter, cannot be blocked, and systematically shrink and expand the measuring rods so that their lengths are exactly as the first hypothesis predicts.[34] Unlike Cartesian skeptical hypotheses, the second hypothesis is not scientifically absurd. The epistemological question, then, is: What reason could we have for attributing distortions in our measuring rods to spatial curvature rather than to universal forces?

The semantic questions concern how the meaning of spatial language could be fixed in such a way that it would remain an open question whether space is curved. To simplify, pretend that all spatial facts may be expressed using Tarski's predicates:[35]

[33] See also Bricker (1993); Sider (1993*a*, chapter 9).

[34] See Reichenbach (1958, chapter 1).

[35] These are the predicates from Tarski's axiomatization of Euclidean geometry; see Tarski (1959); Tarski and Givant (1999). Really, though, the fundamental metrical facts should

point x is *between* points y and z

points x and y are *congruent* to points z and w

One sort of semantic question arises only given an extreme empiricist philosophy of language. If every meaningful predicate must be associated with verification conditions for its application, then given the previous paragraph, the predicates 'between' and 'congruent' would seem not to be meaningful.

Other semantic worries will have force even for nonverificationists. If Einstein is right and Newton is wrong about curvature, then the referents of 'between' and 'congruent' must satisfy nonEuclidean rather than Euclidean axioms. But in addition to having an interpretation under which they satisfy nonEuclidean axioms, 'between' and 'congruent' also have an interpretation in which they satisfy Euclidean axioms. (The Euclidean axioms are true in the abstract model whose domain is $\mathbb{R}^3$ and in which the predicates are interpreted in the obvious way. Since the set of physical points of space has the same cardinality as $\mathbb{R}^3$, the model in $\mathbb{R}^3$ induces a model in physical space.) So if Einstein is to be right and Newton wrong, it must be that only one of these assignments is the *correct* assignment, the *intended* interpretation of 'between' and 'congruent'. But what determines that one of these assignments is the *correct* interpretation? We cannot specify the intended interpretation of 'congruent' by saying that "it is to apply to x, y, z, and w when the *distance* between x and y is the same as the distance between z and w", for 'distance' is in the same boat as 'congruent'; how is its intended interpretation determined?

The metaphysical questions concern the same issue as the semantic ones, only more directly. In what would the fact that spacetime is curved consist? In the fact that the congruence and betweenness relations satisfy nonEucledian axioms, is the obvious reply. But since there are continuum-many spacetime points, there exist "Euclidean-congruence" and "Euclidean-betweenness" relations that satisfy Euclidean axioms. So in what sense is spacetime *really*

---

probably not be taken to be direct point-to-point distance comparisons as in Tarski's system, but should rather be local metrical facts, from which distances along paths may be recovered. On the other hand, the standard development of a local metric should probably not be taken at face value, since it grounds metric structure in the metric tensor, a mathematical object involving real numbers. Surely the fundamental distance facts are purely about points (as they are in Tarski's account). Thus what we really want is a synthetic geometry from which one can prove representation theorems about the metric tensor (see Field (1980); Mundy (1987) for two approaches to representation theorems). I do not know whether such an account exists.

Euclidean? What makes the "real" betweenness and congruence relations, as opposed to their Euclidean counterparts, "physically significant"?

We face the same questions when reading physics textbooks on special or general relativity that speak of the "intrinsic structure" of physical space, time, and spacetime. In classical physics, we are told, spacetime is flat, and there is a "well-defined" relation of simultaneity, whereas in Minkowski spacetime there is no such relation of simultaneity—there is no "distinguished" notion of simultaneity. But of course, there are relations—sets of ordered pairs, anyway—between space-like separated points of Minkowski spacetime that foliate the spacetime; many such relations. What does it mean to say that none of these relations is "distinguished"?[36]

Geometrical conventionalists like Henri Poincaré (1952, Part II) and Hans Reichenbach (1958, Chapter 1) give a deflationary answer to the semantic questions that answers the epistemological questions, and which implicitly assumes a deflationary answer to the metaphysical questions. Return to the question of how to specify the intended interpretation of 'congruent'. According to Reichenbach, a theoretical predicate like 'congruent' requires a "coordinative" definition, a definition that correlates the predicate with something that is (relatively) observable. An example of a coordinative definition would be the definition of straight lines through spacetime as the paths of light rays in vacuum. We might think to give a coordinative definition of congruence in terms of measuring rods: points of space are congruent when they can be the endpoints of a single measuring rod. But measuring rods can be distorted by forces. Might we then define congruence in terms of measuring rods unaffected by forces? The problem is that 'force' is itself a term in need of a coordinative definition, since one can no more directly measure forces than congruence. What Reichenbach says, in essence, is that one must simultaneously give coordinative definitions of 'force' and 'congruent' in terms of measuring rods: congruent points are those picked out by the endpoints of a measuring rod that is not subject to forces. Since this coordinative definition constrains two terms, there is a certain amount of freedom in assigning meanings to those terms. One can understand 'force' and 'congruent' so that space is Euclidean but there are universal forces acting on all objects that produce the measurements that we make, or one can understand 'force' and 'congruent' so that space is nonEuclidean and there are no universal forces.[37] In fact, physicists

[36] See North (2009) on structure in physics generally.

[37] Better: we can understand 'force' and 'congruent' so that the sentence 'space is Euclidean'

have preferred the latter course, and so we speak of space as being curved. But this is in part a matter of arbitrary definition; physicists could have chosen the former course. They chose the latter only because the resulting physics was simpler. So according to conventionalists like Reichenbach, it is at best misleading to say that space itself is curved, that space is *intrinsically* curved. 'Space is curved' is true to say given the linguistic choices that physicists have in fact made. But those choices were arbitrary, and moreover are inextricably tied to the conventional choice of whether to speak of universal forces.

Given this view about the semantics of spatial language, the epistemological questions are immediately answered. How do we know that space is curved, rather than being flat but accompanied by compensating universal forces? We know this simply by knowing which conventions our linguistic community has adopted. (Better: knowledge of linguistic conventions plus empirical observation tells us that 'space is curved' is the right description. If different linguistic conventions had been adopted, the same observations would have called for the description 'space is flat'.)

Reichenbach (unsurprisingly) does not address the metaphysical question. But it seems clear that he would regard talk of space as being "intrinsically" flat or curved as misguided, given the need for coordinative definitions. Insofar as conventionalists have a metaphysics of spatial structure at all, it is that space is not intrinsically structured (or perhaps that talk of intrinsic structure makes no sense).

What might an anti-conventionalist—realist—account of physical geometry look like?[38] Here is what the realist *wants* to say. About semantics, spatial predicates like 'between' and 'congruent' can be understood *purely spatially*; they need no coordinative definitions. And they are attached to particular relations over physical points, which satisfy nonEuclidean axioms (assuming that Einstein is right). As for metaphysics, space is intrinsically structured; the genuine betweenness and congruence relations are privileged in a way that Euclidean-betweenness and Euclidean-congruence are not. The naïve and natural picture of physical geometry one gets from physics is thereby vindicated. The epistemological problems then confront us head-on. But these problems should be solved the way we realists solve all such problems of theory being underdetermined by observation: by appealing to criteria—"simplicity" is a

comes out true, etc.

[38]Nerlich (1976), especially chapter 9, defends realism about spatiotemporal structure, and distinguishes this realism from realism about the existence of entities (chapter 5, §6). See also Bricker (1993).

common placeholder—that choose between observationally indistinguishable theories. (If such criteria do not deliver a verdict, we remain agnostic until some new test or consideration breaks the stalemate.)

A realist about structure has a clear path to this realism about physical geometry. Metaphysically, the distinction enjoyed by the genuine betweenness and congruence relations is that they are part of reality's distinguished structure: they carve perfectly at the joints, unlike any relations of Euclidean-betweenness and Euclidean-congruence. Semantically, given any reasonable metasemantics for theoretical terms, 'between' and 'congruent' attach to betweenness and congruence rather than to any Euclidean-betweenness and Euclidean-congruence relations precisely because only the former are part of the world's genuine structure. Reichenbach was led to his position by his insistence on the need for coordinative definitions, deriving ultimately from an internalist and highly empiricist approach to meaning. But a more reasonable metasemantics will allow a role for a nonobservational and externalist determinant of meaning: the world's structure.[39]

More generally, questions about metric, affine, topological, and other structure of space, spacetime, and other physical spaces are questions about the distinguished structure of those spaces. There is a substantive purely spatiotemporal fact of the matter as to whether spacetime is Galilean, neoNewtonian, Minkowskian, or a curved Lorentzian manifold.[40] The fact is given by the joint-carving features of points of spacetime. If, for example, the joint-carving features of points of physical spacetime are as described by Minkowski's theory, then physical spacetime is Minkowskian, and there is "no physically distinguished relation of simultaneity" in the sense that there is no joint-carving relation that foliates the spacetime (nor can any foliation be defined from the joint-carving relations over points).

Spacetime could have had a structure that would have vindicated a kind of geometrical conventionalism. Suppose spacetime had lacked distinguished metrical structure—suppose there had been joint-carving topological features

[39]Grünbaum (1973) also defended conventionalism about metric structure, but based it on metaphysical considerations rather than on an empiricist account of meaning. An intrinsic metric, Grünbaum argued, would have to be definable from facts intrinsic to the points, but if space is continuous then there are no facts intrinsic to points that suffice for the definition. The realist about structure, however, regards the distinguished structure of space as constituting facts intrinsic to points from which a metric may be defined.

[40]Or even whether there is a fundamental four-dimensional spacetime at all, as opposed to a massively dimensional configuration space; see Albert (1996); North (2009, 2010).

but no joint-carving metrical features. Then no metric would have been distinguished from any other, and spacetime would have been a kind of amorphous "point soup". Reality might at the same time have lacked sufficient structure to define forces. In such a Reichenbachian world, we would have been free to choose either of a pair of coordinative definitions, simultaneously defining force and metric predicates. Neither choice would have carved reality at its joints better than the other. Metric and force predicates would require coordinative definitions in such a world, not because of general semantic considerations, but rather because the world would lack the structure needed to supply semantic determinacy. What reason do we have to think that *our* world has any more structure? The fact that physical theories with primitive metrical predicates have been so successful (section 2.3).