Of Numbers and Electrons

Cian Dorr

Penultimate version: March 5, 2010
Forthcoming in *Proceedings of the Aristotelian Society*

1 The inductive case against nominalism

The sciences are full of theories which, in the course of making detailed claims about the physical world, say things which entail that there are mathematical entities like numbers and sets. According to an influential tradition stemming from Quine (1948) and Putnam (1972), good scientific reasoning—induction, broadly construed—requires us to believe some such theory, or some disjunction of such theories. And it is because of this that we ought to believe that there are mathematical entities. The belief that there are numbers is, according to this tradition, on a similar epistemological footing to the belief that there are electrons, viruses, quasars, etc.[1]

Some will regard this analogy as unhelpful because they think that we can know that there are mathematical entities in the same way—whatever it is—that we know that all dogs are dogs, or that all bachelors are unmarried.[2] Others may regard this analogy as unhelpful because they think that we can *directly perceive* that there are mathematical entities—such as sets of dogs standing in the street in front of us—just as we can directly perceive that there are red things; no inductive inference of any sort is required. Still others will regard the analogy as unhelpful because they think the epistemology of mathematics is *sui generis*, governed by its own distinctive, topic-specific norms—maybe just 'Believe in as many mathematical entities as you consistently can!'—very different from the topic-neutral norms that govern inductive reasoning in general.[3] On any of these views, the detour through the empirical sciences is at best unnecessary. Since my aim in this paper is to evaluate the distinctively inductive reasons for belief in mathematical entities, I will henceforth presuppose that all these views are wrong. I will take it for granted that we understand some notion of 'evidence', which serves as input to induction, and some notion of 'deductive entailment', which provides background constraints on the beliefs which are eligible for inductive support. And I will take it for granted that our evidence does not deductively entail that there are mathematical entities.[4]

Having internalised this assumption, one does not have to be an inductive sceptic to feel that the inference from evidence about the behaviour of physical objects to the conclusion that there are numbers or sets—entities as drastically unlike physical objects as anything could be—is eyebrow-

[1] I will not be discussing 'the indispensability argument for the existence of mathematical entities' as such, since I don't want to get hung up on the interpretation of the key term 'indispensable'. I hope that the ideas that lie behind this argument will nevertheless get a fair hearing.

[2] See, e.g., Wright 1983.

[3] See, e.g., Maddy 1997.

[4] Of course, Quine and Putnam are deeply sceptical about these epistemological categories. But in practice it is almost impossible to say anything substantive about the way in which empirical considerations bear on any theoretical question without assuming a background logic and a background conception of evidence. I hope that even those who regard our basic logical beliefs or basic perceptual methods as empirically revisable will be able to accept my assumptions about deductive entailment and evidence, if not as capturing deep epistemological facts, at least as a useful temporary device marking the distinction between the beliefs whose epistemic status we are currently trying to assess and the beliefs we are legitimately taking for granted.

raisingly audacious, as inductive inferences go. The hypothesis that there just aren't any such entities certainly doesn't *feel* very like the hypothesis that there are no electrons—there are no experiments that stand to numbers as, say, the Millikan oil-drop experiment stands to electrons. It is hard to get over the impression that someone who insisted on assigning substantial credence to the no-numbers hypothesis would be displaying prudent caution, rather than any kind of failure of rationality.

It is customary to use the framework of Inference to the Best Explanation (IBE) in setting up questions about what good inductive reasoning requires in cases like this. Expressed in these terms, the claim that our total evidence E provides us with good inductive reason to believe that there are mathematical entities becomes the claim that some theory that entails that there are mathematical entities is significantly better *qua* explanation of E than any theory that doesn't.[5] Even though this framework is rather creaky, and its relation to our best-developed formal theory of inductive reasoning (Bayesian conditionalisation) rather obscure, I will go along with this way of putting things: it seems reasonable to hope that there is *some* way of understanding 'better explanation' on which the reformulation is acceptable, though it is a further question what this might have to do with any meaning of 'explanation' we can independently understand. This framework has the dialectical advantage that it puts the ball in the court of those who *don't* think we have good inductive reasons to believe in numbers. For we already have a host of platonistic theories (theories that deductively entail that there are mathematical entities) which seem *pretty* good as explanations of large chunks of our evidence. Defenders of nominalism are committed to the claim that there is at least one nominalistic (non-platonistic) theory that is as good as (or at least not significantly worse than) all of these platonistic theories, *qua* explanation of our evidence. The challenge is to produce such a theory, or anyway to make it plausible that one could in principle be formulated.

The most famous response to this challenge is the programme initiated by Hartry Field in *Science Without Numbers* (Field 1980). There, Field formulated a version of Newtonian gravitational theory all of whose quantifiers can be understood as restricted to spacetime points, spacetime regions and particles. The consequences of Field's theory for the physical world are exactly those of the particular platonistic theory it purports to replace.[6] Field does not offer a detailed argument that his theory equals or exceeds that theory in respect of explanatory goodness. But at least it isn't *intuitively* the sort of theory whose truth would 'cry out for explanation'. It is hard to muster up any of the familiar feeling of explanatory satisfaction in contemplating a putative 'explanation' of the truth of Field's theory that consists in deriving it from the platonistic theory it aims to supplant.[7]

Field's project calls for a lot of hard work. While his methods evidently generalise far beyond the particular theory he chooses as his model, theories like general relativity and quantum mechanics pose distinctive challenges which have yet to be tackled convincingly.[8] It is too early to say whether we can find nominalistic theories in these domains which do as good a job as Field's Newtonian theory at providing the intuitive sense of explana-

[5] I am understanding IBE as the claim that we should believe *the* theory that best explains our evidence, assuming it is better than its competitors by a wide enough margin. This is somewhat different from the Quinean slogan that 'we should believe *our* best theories'. The slogan, perhaps problematically, makes the facts about what we should believe depend on which theories we have thought of. Couldn't it happen that although T is the best theory we have come up with, we *should* come up with a certain better theory T', and thus *should not* believe T? (Melia (1995) raises related concerns.) But this difference doesn't end up mattering much for my purposes. If we could find a workable general strategy for replacing platonistic theories with nominalist ones that are equally good, on either view we could conclude that *from now on* we should not believe in mathematical objects.

[6] This claim is true only for the version of Field's theory that uses something like second-order logic to express the idea that there is a region corresponding to every collection of points. It is not true for the first-order version of Field's theory (see Burgess and Rosen 1997: §§II.A.5.b, III.B.1.b). However, the claims about the physical world that follow from the platonistic theory without following from the first-order version of Field's theory are quite esoteric; certainly none of them are part of our evidence. So if we denied the legitimacy of anything beyond first-order logic, it would be hard to see how the additional consequences of the platonistic theory for the concrete realm would improve its claim to be well-supported by our evidence: if anything, the reverse seems more likely to be true.

[7] On the role of explanatory satisfaction in giving nontrivial content to IBE, see White 2005.

[8] In the case of quantum mechanics, the process of constructing anything as rigorous as a Field-style theory will inevitably incorporate some particular rigorous solution to the measurement problem. For this reason I am not convinced by the treatment in Balaguer 1996, which involves a highly problematic account of measurement in terms of primitive propensities.

tory satisfaction while avoiding anything which might strike someone as a cheap trick. And even if the programme goes as well as could be hoped in physics, it is a thorny question what additional burden defenders of nominalism need to meet with regard to the other sciences, and whether this burden can be met with equal success.

It is thus important to investigate what genuinely epistemic advantages, if any, all this honest toil has over theft. There are mechanical methods which take arbitrary platonistic theories as input and output empirically equivalent, nominalistic replacement theories. The outputs of these algorithms will certainly feel like cheats to those who are used to working within the constraints that make Field's programme difficult. But the stakes are high: we should do our best to subject these gut reactions to critical scrutiny before we accept them as an accurate guide to the standards of explanatory goodness.

One important constraint that Field imposes on his theories is that they make no use of modal operators. If we don't mind using them, a range of one-size-fits-all methods for generating nominalistic substitutes for standard scientific theories becomes available. I will mostly be discussing two very straightforward methods, although I hope that what I say about the explanatory goodness of the theories they generate will generalise to many more sophisticated approaches. The first method embeds the input theory T in the scope of a restricted possibility operator, as follows:

($T^\diamond$) Possibly, the concrete realm is just as it in fact is, and T.

$T^\diamond$ is just the claim that T is *concretely adequate*, or consistent with the totality of truths entirely about the concrete realm.[9] 'The concrete realm lives up to its side of the T-bargain' (Balaguer 1998). Or—to put it in a way that resonates with the tradition of 'fictionalism' stemming from Vaihinger (1924) and discussed by Putnam (1972: §8)—as far as the concrete realm is concerned, it is *as if* T were true.[10]

The second method requires us first to identify the *purely mathematical* component of T, call it M—the conjunction of the purely mathematical axioms employed by T. Typically with just a little artificiality we can take M to be something like ZFCU (= Zermelo-Fränkel set theory with choice + 'there is a set of all non-sets').[11] Having identified M, we use it in constructing a restricted *necessity* operator within which we embed our input theory T:

($T^\square$) Necessarily, if M and the concrete realm is just as it in fact is, then T.

According to contemporary fictionalists like Yablo (2001, 2005), something like $T^\square$ captures the 'real content' conveyed by ordinary utterances whose *literal* content is T: we can think of M as a story—'the story of standard mathematics'—and of $T^\square$ as the claim that T is true according to M, taking the facts entirely about the concrete realm to be 'imported' into the story in the same way that the facts of nineteenth-century geography are imported in settling what is the case according to the *Sherlock Holmes* stories. But even if we reject this hermeneutic claim, we might still conclude that $T^\square$ is no worse *qua* explanation of our evidence than T.[12]

---

[9] I take the notion of a proposition 'entirely about the concrete realm' to be clear enough for present purposes. Anti-Humean primitivist realists about laws of nature, counterfactuals, chances etc. should make sure to understand 'entirely about the concrete realm' in such a way that, for example, when the proposition that P and the proposition that Q count as 'entirely about the concrete realm', so do the proposition that it is a law that P, the proposition that if it were that P it would be that Q, the proposition that it is more likely that P than that Q, etc.

[10] Strictly speaking, the 'as if' claim looks to be a counterfactual conditional: the concrete realm is as [it would be] if T were true. Formally, this does not follow from $T^\diamond$, since 'possibly C and T' does not logically entail 'if it were that T, it would be that C'. But in fact, 'as if' claims in these kinds of debates seem to be universally treated as tantamount to claims like $T^\diamond$. Since counterfactuals are notoriously context-sensitive, this suggests that in the relevant contexts, $T^\diamond$ does suffice for the truth of the counterfactual. In possible-worlds terms, the contextually relevant closeness relation is such that if there are T-worlds concretely indiscernible from the actual world, they *ipso facto* count as closer to the actual world than any other T-worlds.

[11] In fact for normal scientific purposes we can get by with theories much weaker than full ZFCU. See Burgess and Rosen 1997: §I.B.1.b.

[12] A closely related approach, which arguably does a better job than $T^\square$ at capturing the intuitive content of the 'according to the fiction' claim, uses a counterfactual conditional instead of a strict one: If M were the case [and the concrete realm were just as it in fact is], then it would be the case that T (see Dorr 2007: §2; Dorr 2005: §§3–5). The standard approach to the logic of counterfactuals makes this formally weaker than $T^\square$: the closest

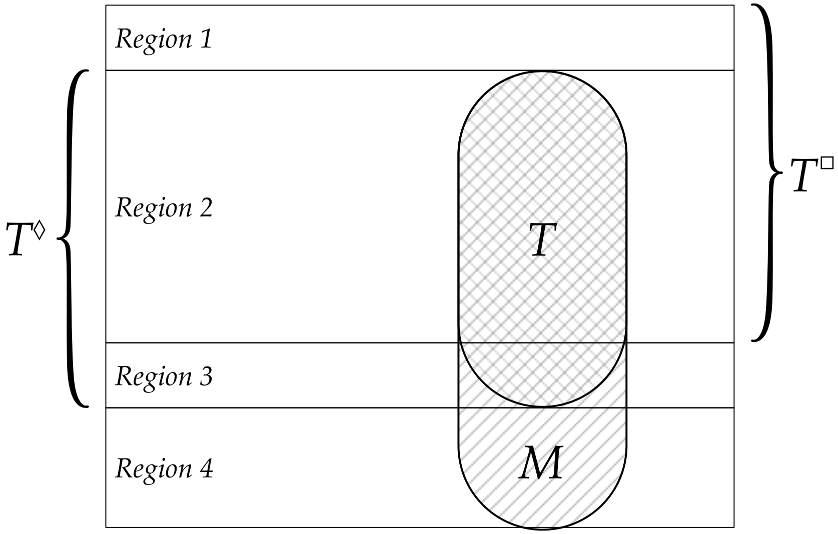

Figure 1. Possible-worlds representations of $T^\Diamond$ and $T^\Box$. Points on the same horizontal line represent concretely indiscernible worlds.

We can clarify the logical relation between $T^\Diamond$ and $T^\Box$ by thinking of them in terms of possible worlds, although of course those who don't believe in abstract entities will not want to take such glosses fully seriously. In these terms, $T^\Diamond$ is true at a possible world w iff some T-world is concretely indiscernible from w; $T^\Box$ is true at w iff every M-world that is concretely indiscernible from w is a T-world.[13] This is represented in figure 1. As the figure makes clear, there is no formal guarantee that $T^\Diamond$ and $T^\Box$ are true in the same worlds—for this to be the case, the sets of worlds labelled 'Region 1' and 'Region 3' would need to be empty. But as we will see in section 3, in many central cases of interest there is reason to think that these regions *are* empty, so that $T^\Box$ and $T^\Diamond$ are necessarily equivalent.

Ideally, those who want to deny that we have good inductive reason to believe in mathematical entities would like to have a general method that takes an arbitrary theory T as input and yields as output a theory that is

(i) deductively entailed by T;

(ii) *empirically equivalent* to T, at least in the minimal sense that it entails every proposition E that is part of our actual evidence and deductively entailed by T;[14]

M-and-C worlds could be T-worlds even if not all of them are. However, the counterfactual formulations introduce new complications, and in practice it is hard to think of any case where the alleged logical differences between counterfactuals and strict conditionals would matter.

[13] We should thus not understand 'in fact' as meaning the same as 'actually', in the standard philosopher's sense, since in that case $T^\Diamond$ and $T^\Box$ would, like 'actually ϕ', be necessary if true. Rather, 'in fact' works like the 'backspace' operator in Hodes 1984, which 'undoes' just one modal operator, so that $\Diamond \downarrow \phi$ is equivalent to ϕ, and $\Diamond\Diamond \downarrow \phi$ equivalent to $\Diamond\phi$.

How are we to analyse 'the concrete realm is just as it in fact is' in more fundamental terms? I don't think we need to answer this question at this stage in the dialectic: we obviously *do* understand this expression; if the only way of making nontrivial sense of this involves quantification over abstract objects of some sort, then we can establish the existence of such objects by deductive reasoning alone, and there is no point in arguing further about the inductive route. Nevertheless, note 43 below will consider some possible nominalistic analyses.

[14] A more adequate notion of empirical equivalence would take into account probabilistic relations between theory and evidence short of deductive entailment. For the sake of simplicity I will ignore this from the main text.

(iii) *nominalistic*—deductively consistent with there being no mathematical entities;

(iv) Explanatorily better than, or not significantly worse than, T.

If one had a method that satisfied all four of these desiderata, one could be assured that belief in mathematical entities could never be licensed by IBE—we could not justifiably believe some mathematical-entity-entailing T if there were guaranteed to be a nominalistic competitor of T that was almost as good *qua* explanation of our evidence. (Strictly, a theory-transforming method would not need to satisfy desideratum (i) to achieve this; but a method's having this feature will make it easier to argue that it satisfies desideratum (iv), since it is plausible that a worked-out version of IBE will give weaker theories, *ceteris paribus*, an advantage over stronger ones.) My central concern in this paper is with (iv)—whether the classes of modal theory-modifying methods whose most straightforward instances are $T^\Diamond$ and $T^\Box$ yield good theories as outputs, when given good theories as inputs. But first, I should say something about how $T^\Diamond$ and $T^\Box$ fare with respect to the first three desiderata. Sections 2 and 3 will sketchily address these questions for $T^\Diamond$ and $T^\Box$, respectively. These sections will also introduce some more sophisticated variants of $T^\Diamond$ and $T^\Box$ that one might turn to if one thought that $T^\Diamond$ and $T^\Box$ themselves failed in some serious way to satisfy the desiderata.

The epistemology proper will start in section 4, which will present and endorse a canonical argument for the explanatory badness of $T^\Diamond$. Sections 5–7 will consider how this form of argument bears on $T^\Box$. Finally, section 8 will glance at some other considerations which might suggest that $T^\Box$ is explanatorily bad, and section 9 will canvass some objections.

2 The adequacy of $T^\Diamond$

(i) That T deductively entails $T^\Diamond$ seems obvious: the inference from ϕ to $\ulcorner$Possibly $\phi\urcorner$ seems manifestly deductively valid, as does the inference from anything to 'the concrete realm is just as it in fact is'.[15]

(ii) Suppose T deductively entails some E which is part of our evidence. It would follow that $T^\Diamond$ deductively entails E if it were a priori—deductively entailed by everything—that E is entirely about the concrete realm (cannot differ in truth-value between concretely indiscernible worlds). For if it is a priori that E is entirely about the concrete, 'possibly, the concrete realm is just as it in fact is and E' must deductively entail E. Since T entails E, 'possibly the concrete realm is just as it in fact is and T' entails 'possibly, the concrete realm is just as it in fact is and E'.

How plausible is it that whenever E is part of our evidence, it is a priori that E is entirely about the concrete? This claim certainly does not follow from our assumption that our evidence does not deductively entail that there are mathematical entities. For a believer in mathematical entities might think that, as a matter of a posteriori necessity, some of the familiar properties of concrete objects that feature in our evidence (redness, squareness, etc.) consist in certain relations holding between concrete objects and non-concrete ones, which can vary even when all facts *entirely* about the concrete are held fixed. Perhaps it is part of our evidence that some objects are roughly equally long, and what it is for two objects to be roughly equally long is for a real number close to 1 to bear a certain primitive "being the ratio of the lengths of" relation to them. Perhaps we cannot even rule out a priori the even more radical hypothesis that some or all of the objects of our acquaintance are *themselves* mathematical objects, as in the 'hyper-Pythagorean' views entertained by Quine (1976) and Tegmark (2008).

Only on a fairly expansive conception of the scope of 'deductive entailment' could such hypotheses be ruled out a priori. Otherwise, $T^\Diamond$ is empirically equivalent to T only modulo an auxiliary hypothesis that entails, for each E that is part of our evidence, that E is entirely about the concrete realm. This is a limitation of the modal strategy. But I don't think it is a very serious limitation. The auxiliary hypothesis is plausible; and as far as I can see, none of the standard platonistic scientific theories that we have reason to take seriously derives any explanatory power from being consistent with its negation. The only exceptions are the hyper-

[15] One might deny this on the grounds that deductive reasoning cannot eliminate the hypothesis that modal concepts are radically defective. But even on the hypothesis of such defectiveness, it is more plausible to take 'possibly' as inert (so that $\ulcorner$Possibly $\phi\urcorner$ is equivalent to ϕ) than to reject all possibility claims.

Pythagorean theories, which achieve a somewhat attractive economy by entailing that there is nothing more to reality than the mathematical realm; they will eventually have to be compared on other grounds with the more familiar kinds of economical world-views available to nominalists.[16]

(iii) The most important reason to doubt that $T^\Diamond$ is nominalistic derives from what I will call the *necessity thesis*: that if there are no mathematical entities, it is metaphysically necessary that there are no mathematical entities. If the necessity thesis is a priori (deductively entailed by everything), and we interpret the modal operator in $T^\Diamond$ as expressing metaphysical possibility, then $T^\Diamond$ does after all deductively entail that there are mathematical entities, in virtue of entailing that it is metaphysically possible for there to be mathematical entities. It is thus useless to the nominalist.

Why might one believe the necessity thesis to be a priori? Most putative a priori arguments for the necessity thesis that I have come across use forms of reasoning which, if they worked, would support the stronger claim that it is metaphysically necessary that there are mathematical entities. Since we are assuming that the existence of mathematical entities cannot be established deductively, we can ignore these arguments. The most promising remaining argument that I know of is one that parallels Kripke's influential argument, in the appendix to *Naming and Necessity* (1972: pp. 156–58), for the claim that if there are no unicorns, it is metaphysically impossible for there to be unicorns. But Kripke's argument is not all that strong. It is not clear why a failed attempt to introduce a word as a natural kind term should, in effect, confer on it an empty intension rather than a more 'superficial' fallback intension, like that of 'horse-shaped animal with one horn on its forehead'. Similarly, one might think that if there are in fact no objects playing the characteristic 'structural role' of the numbers, 'number' would come to have the same (non-trivial) intension as 'entity playing such-and-such characteristic structural role'. The necessity thesis would be a weak reed upon which to rest a case against nominalism.

Even if we did accept the necessity thesis as a priori, we might hope to find a nominalistic reading of $T^\Diamond$ by reading 'possibly' as expressing something weaker than metaphysical possibility. For example, we might invoke a notion of conceptual possibility, cognate to the notion of 'deductive entailment' to which we have been helping ourselves. A worry about this approach concerns the status of 'theoretical definitions' of mixed mathematico-physical predicates (like 'is the ratio of the lengths of') in terms of purely physical predicates (like 'equal in length') and narrowly mathematical ones (like 'is a member of'). For example:

(*) r is the ratio of the length of a to that of b iff

for all functions f from physical line-segments to non-negative real numbers such that $f(x) = f(y)$ whenever x and y are equal in length, and $f(x) = f(y) + f(z)$ whenever x consists of two disjoint parts one of which is equal in length to y and the other of which is equal in length to z,

$f(a)/f(b) = r$.[17]

While it is quite plausible that (*) or something like it is metaphysically necessary, it is not so clear that there is any conceptual necessity in the

[16] What of the probabilistic notion of empirical equivalence mentioned in note 14? For this, the assumption that our evidence consists of propositions entirely about the concrete is not enough. We also need the claim that when A is any proposition entirely about the concrete and P is a conditional probability function representing rational prior credences, $P(A|T)$ is at least approximately equal to $P(A|T^\Diamond)$. It would be question-begging to assume that this is true for absolutely any T: if we take T to be 'there are mathematical entities', $T^\Diamond$ is trivial, so anyone who thinks that some evidence E entirely about the concrete would constitute good inductive evidence for the existence of mathematical entities will think that $P(E|T) >> P(E) = P(E|T^\Diamond)$. However, I can see no reason to doubt the assumption when T is a reasonably specific and detailed scientific theory. Suppose for example that T is deterministic, so that every proposition is equivalent modulo T to one about initial conditions. Even those who think that our evidence raises the probability that there are mathematical entities should still find it plausible that questions about the specific character of the T-compatible initial conditions are independent of questions about the existence of mathematical entities. There is a serious case to be made, which we are trying to assess, that our evidence about the concrete world supports the existence of mathematical entities by virtue of supporting some theory about the very general structural features of the world—dynamical laws and the like. By contrast, there isn't even a *prima facie* reason to think that the hypothesis of mathematical entities receives *further* support from the detailed aspects of our evidence that help us narrow down the location of the actual world within the space of possibilities left open by such a theory.

[17] This is similar to the definition of 'The distance from x to y is r' endorsed by Putnam (1972: p. 340).

vicinity; and it would be even less clear if we were dealing with some less familiar notions than that of length, like charge-density or the metric tensor. If these theoretical definitions are not conceptually necessary, and the original theory T makes no explicit use of 'purely physical' predicates like 'equal in length', then it will be much too easy for $T^\Diamond$ to be true taken as a claim about conceptual possibility: the facts entirely about the concrete realm place no conceptual constraint on the extensions of the mixed predicates, which are all that matters to the truth of T.

This problem does not arise when T is a rich theory which already contains some explicit theoretical definition like (*) for each of its mixed mathematico-physical predicates. One response to the worry, then, is simply to restrict the application of the method to such rich theories. I doubt that this limitation will be much help to the anti-nominalist. Even though standard scientific theories are not rich, this lack of specificity brings no evident explanatory benefits. Sometimes, it is true, a less specific theory is explanatorily better than all of its strengthenings—for example, a theory describing some deterministic dynamical laws may only be made worse by adding further specific information about initial conditions. But merely leaving it open how one's mixed predicates relate to purely physical and narrowly mathematical ones doesn't seem to achieve this kind of explanatorily beneficial generality. Thus it is hard to see how we would lose anything important if the menu of options to which we apply IBE were restricted to rich theories.

A different response to the worry is to look for some modality intermediate between conceptual and metaphysical, on which theoretical definitions like (*) are necessary even though the non-existence of mathematical entities is not, and use this in interpreting $T^\Diamond$ and $T^\Box$. This doesn't seem all that difficult. For intuitively, even if (*) and the non-existence of mathematical entities are both metaphysically necessary, the *source* or *explanation* of their necessity is different. If we can make sense of this thought, we can use it to pick out a deductively closed class of propositions that includes metaphysical necessities like (*) but does not include the proposition that there are no mathematical entities, even if it is metaphysically necessary. We could then use this class of propositions in specifying an interpretation of $T^\Diamond$ on which it is neither too weak nor too strong to be useful to the nominalist.[18]

Suppose, finally, that we both accept the necessity thesis as a priori and reject modalities more fine-grained than metaphysical necessity and possibility. In that case $T^\Diamond$ is no use as it stands. But we may still be able to modify it so as to achieve the desired effects.

One strategy for doing this is to substitute some nominalistically unproblematic predicates for the problematic ones (those whose meaning allows them to apply only to mathematical entities, in some of their argument places). The idea is to begin by replacing T with an isomorphic theory T^*, in which all the work that was done in T by problematic predicates is done by unproblematic ones, such as predicates applicable to ghosts, or angels, or inscriptions.[19] If T said that for all x and y there is a set that has just x and y as members, T^* can say that for all x and y there is a ghost that loves just x and y. If T said that for any two line segments there is a unique real number that is the ratio of their lengths, T^* can say that for any two line segments there is a unique ghost that haunts them in some distinctive way. Strictly speaking, this T^* is already a nominalistic theory, albeit an implausible one, since it entails that there are infinitely many ghosts (or whatever other surrogates we chose). But if we then embed T^* in some appropriately restricted possibility operator, claiming merely that it is consistent with the facts entirely about the entities we really care about, the result will be much weaker, and more likely to be useful to the nominalist.[20]

Another strategy avoids the arbitrary choice of substitute predicates by using higher-order quantification to Ramsify all the problematic predicates

[18] Some theorists have given formal theories of fine-grained concepts that can distinguish different sources of metaphysical necessity. For example, Kit Fine uses an operator 'It is true in virtue of the nature of X that...', where X is a plural term. Metaphysical necessity is truth in virtue of *everything*; when X are some but not all of the things there are, 'It is true in virtue of the nature of X that...' is a necessity operator stronger than metaphysical necessity. While Fine's conception of the entities in virtue of whose nature typical metaphysical necessities hold is far from being acceptable to a nominalist, the general picture of a rich space of intelligible weakenings of metaphysical possibility is one that nominalists might embrace.

[19] Cf. Chihara 1990.

[20] Only on a rather expansive conception of our a priori access to facts about metaphysical possibility will it be plausible that T deductively entails the possibility of T^*. Otherwise, employing this strategy will mean giving up on desideratum (i).

out of T. Let the problematic predicates be $F_1, \ldots, F_n$; let $T(X_1, \ldots, X_n)$ be the result of substituting appropriate higher-order variables $X_1 \ldots X_n$ for these predicates; and let $T^\exists$ be $\exists X_1 \ldots \exists X_n T(X_1, \ldots, X_n)$. We can then, if we wish, apply our restricted possibility operator to $T^\exists$ instead of T. Provided they can understand the higher-order existential quantifications in $T^\exists$ in such a way that they do not deductively entail that there are mathematical entities, even those who accept the necessity thesis can take the resulting theory to be nominalistic. Given our assumption that the existence of mathematical entities is not a priori, it suffices to understand the second-order notation in such a way that the theorems of standard systems of second-order logic—like $\exists X \forall y(Xy)$—are in fact deductively valid. Some will be happy to grant that we can directly learn to understand the second-order notation in such a way as to render standard logic is deductively valid (cf. Williamson 2003); others will hold out for a translation of the notation into ordinary English (see Lewis 1991, Rayo and Yablo 2001).

If we can understand second-order quantification in one of these ways, $T^\exists$ is already a nominalistic theory. Is there any point in embedding it further within a restricted possibility operator? Yes, if T included some strong mathematical theory like *ZFCU*; for then $T^\exists$ will entail the existence of a huge infinite number of entities of unspecified sort, whereas the claim that $T^\exists$ is consistent with the truth about the concrete world plausibly won't.[21] On the other hand, if T already entails the existence of a fairly large infinity of concrete entities such as spacetime points, we may be able, with some ingenuity, to weaken the required mathematical commitments so much that the concrete entities themselves suffice to witness the truth of $T^\exists$; in that case, we will achieve no further weakening by introducing the possibility operator.[22]

[21] However, if—as Williamson (2002) maintains—the right logic for all modal operators is a 'constant-domain' logic on which claims about how many things there are are necessary if true, nothing is gained by embedding $T^\exists$ within a possibility operator. If constant-domain modal logic is correct, the modal strategies with which this paper is concerned are less interesting than they might otherwise seem, since they are of no use for avoiding IBE-style arguments for strong claims about the cardinality of the universe.

[22] There is a version of mathematical structuralism that takes something like $T^\exists$ to specify the 'real' content conveyed by an utterance whose face-value content is given by T. Whether or not this is true as a hermeneutic claim, we should be interested in comparing the explanatory strength of the existentially quantified theory with a stronger theory that purports to talk about a distinctive realm of mathematical objects structured by some *sui generis* natural relations. The discussion of the explanatory goodness of section 4 does apply, mutatis mutandis, to $T^\exists$.

Like the strategy that invokes conceptual possibility, the strategy using surrogates and the strategy using higher-order quantification only work when T is a 'rich' theory—one that contains theoretical definitions like (*) that relate all its mixed mathematico-physical vocabulary to purely concrete and narrowly mathematical vocabulary. Otherwise, too much of the content of T will be lost in the transition to T^* or $T^\exists$—in the worst case, these theories will say nothing about any aspect of the concrete realm other than its cardinality.

3 The adequacy of $T^\Box$

This time I will consider the three desiderata in reverse order.

(iii) Given our assumption that *some* theories do not deductively entail that there are mathematical entities, I see no reason to doubt that $T^\Box$ is among them. Even if the necessity thesis were true, the result would be that $T^\Box$ is (vacuously) deductively entailed by the non-existence of mathematical entities. This would indeed make it useless to the nominalist, but the problem would involve desideratum (ii), not (iii).

(ii) Like $T^\Diamond$, $T^\Box$ will certainly fail to be empirically equivalent to T if T does not deductively entail, for each E that is part of our evidence, that E is entirely about the concrete. Otherwise, we can at best have empirical equivalence *modulo* an auxiliary hypothesis that does deductively entail these claims. But even if these claims are all a priori, it is still not obvious that $T^\Box$ deductively entails E whenever T does. The problem comes from the formal possibility that $T^\Box$ is vacuously true because no M-world is concretely indiscernible from the actual world—in other words, that the actual world is in Region 1 of Figure 1. If this cannot be ruled out a priori, $T^\Box$ will not, taken on its own, satisfy desideratum (ii): the fact that E follows from T is no reason to think that E holds throughout Region 1. We will need to package $T^\Box$ together with an auxiliary hypothesis that rules out the possibility in question:

($M^\Diamond$) Possibly, the concrete realm is just as it in fact is, and M.

It is evident that $T^\Box$ and $M^\Diamond$ together entail $T^\Diamond$, and thus entail every proposition entirely about the concrete realm that T entails.[23]

Is $M^\Diamond$ a priori? If we are not moved by the considerations that support the necessity thesis, and take an optimistic view of the scope of deductive rationality, there is some appeal to the idea that it is. If one is less optimistic than this about the extent of our a priori knowledge about metaphysical possibility, there are some possible strategies for avoiding the need to rely on $M^\Diamond$, corresponding to those we considered in discussing the necessity thesis in section 2. One is to replace the metaphysical necessity in $T^\Box$ with some stronger notion of necessity, so that we only need a counterpart of $M^\Diamond$ using a correspondingly weaker notion of possibility. Another is to eliminate specifically mathematical vocabulary, either by using surrogates like ghosts, or by using higher-order quantification to formulate a theory like $T^{\Box\forall}$:

($T^{\Box\forall}$) Necessarily, if the concrete realm is just as it in fact is, then $\forall X_1 \ldots \forall X_n$ (if $M(X_1, \ldots, X_n)$ then $T(X_1, \ldots, X_n)$).[24]

Assuming T is 'rich' in the sense of the previous section, to get something empirically equivalent to T, we will only need to combine $T^{\Box\forall}$ with $M^{\Diamond\exists}$:

($M^{\Diamond\exists}$) Possibly, the concrete realm is just as it in fact is, and $\exists X_1 \ldots \exists X_n M(X_1, \ldots, X_n)$.

The case for the apriority of claims like $M^{\Diamond\exists}$, which essentially say nothing more than that the concrete realm's being just as it in fact is is consistent with the universe as a whole having an appropriately large infinite cardinality, is relatively strong. However, even this requires a fairly expansive conception of the scope of deductive rationality: the justification even of claims of narrowly logical possibility, let alone claims of metaphysical possibility, is certainly much more mysterious and problematic than the justification of elementary logical theorems. If we take a narrower view of the a priori, then there will be no way to avoid having to include something like $M^\Diamond$ (or $M^{\Diamond\exists}$) along with $T^\Box$ (or $T^{\Box\forall}$) as part of the package which we assess for explanatory goodness. Section 7 below will consider whether this makes a significant difference.

(i) There is no narrowly logical reason to think that T must deductively entail $T^\Box$. This is clear from Figure 1: *prima facie*, the actual world could be one of the T-worlds in Region 3—a T-world that is concretely indiscernible from some M-world that is not a T-world. If we can rule this out a priori, it must be because M plays a distinctive role within T: M must, in conjunction with the claim that T is consistent with the truth about the concrete realm ($T^\Diamond$), deductively entail T.

I think it is quite plausible that this is so, in the case of standard mathematical theories and typical scientific theories that rely on them. How could an M-world w' fail to be a T-world while being concretely indiscernible from a T-world w? I see four possibilities. The first is that the truth of T at w depends on the obtaining there of some fundamental relations between concrete and mathematical entities—relations which do not supervene on the totality of facts either about the concrete realm or about the narrowly mathematical relations with which M is concerned (like set-membership).[25] Perhaps some of the mixed mathematico-physical predicates of T, like 'r is the ratio of the length of x to that of y', stand for these non-supervenient mixed relations.[26] If T admits of such 'heavy duty platonist' interpretations, it does not deductively entail $T^\Box$: the fact that w' is an M-world that is concretely indiscernible from a T-world does not ensure that the non-supervenient mixed relations behave in such a way as to make T true at w'.

[23] $M^\Diamond$ is somewhat similar to the claim that M is 'conservative' in the sense of Field 1980: that is, such that the result of combining it with any consistent theory entirely about the concrete realm is itself consistent. The differences are that (a) Field's notion of consistency is that of narrowly logical consistency, whereas $M^\Diamond$ uses something closer to metaphysical possibility; (b) Field's notion quantifies only over recursively axiomatisable theories about the concrete realm, whereas $M^\Diamond$ requires consistency with the *complete* truth about the concrete realm, even if it is not recursively axiomatisable; and (c) for the parallel to go through, one would need either to add a 'Necessarily' in front of $M^\Diamond$, or to change the definition of 'conservative' to require consistency with *true* theories entirely about the concrete rather than *consistent* theories.

[24] Hellman's reconstructive project (Hellman 1989) involves replacing each T with something like $T^{\Box\forall}$.

[25] Cf. the view Field (1984: §5) calls 'heavy duty platonism'.

[26] If so, 'theoretical definitions' like (*) are either contingent, or necessarily true but only because facts about which things are equally long are themselves not entirely about the concrete realm.

However, I doubt we would lose any explanatorily important generality if we confined our attention to theories T that explicitly state, for each of their mixed mathematico-physical predicates, that its extension supervenes on facts about the concrete realm and narrowly mathematical relations.[27]

A second possibility is that the truth of T at w but not w' depends on some facts entirely about the mathematical realm that do not follow from M, and obtain at w but not w'. Perhaps, for example, some large cardinal axiom is true at w but not at w', and this, together with the facts about the concrete realm, prevents T from being true there. But ordinary theories in the sciences don't work like this: they simply help themselves to as much mathematics as they need.

The third possibility is that the relevant difference between w and w' involves some *impure* mathematical claims which need to be true for T to be true. For example, perhaps at w there is a bijection between the set of spacetime points and the set of real numbers, while at the concretely indiscernible w' there is no such bijection. This sort of thing will happen all the time if M as a first-order theory, say first-order ZFCU. For given any infinite set U, there are many different ways to extend U to a model of first-order ZFCU where U serves as the interpretation of 'urelement'; different set-theoretic claims about the cardinality of the set of urelements are true in these different models. If these models correspond to different possible worlds in the obvious way, then the facts entirely about the concrete world and the truth of first-order *ZFCU* do not settle which pure sets stand in one-one correspondence with the set of concrete objects. But this is one of the places where we seem to manifest an understanding of questions of cardinality that transcends first-order logic. Even if there are infinitely many concrete objects, we can make sense of different hypotheses about how many of them there are without bringing in numbers at all. For example, we can make sense of the hypothesis that there are at most $\aleph_1$ concrete objects.[28] Among worlds where first-order ZFCU is true, we can distinguish the 'well-behaved' ones, where the set-theoretic characterisation of the set of concreta corresponds to the intrinsic fact of the matter about how many concreta there are, from the rest. This suggests that we can legitimately take M to be a second-order (or plural) version of ZFCU, in which case we don't have to worry about this kind of difference between w and w'. For according to a theorem of Zermelo (1930), second-order ZFC is 'quasi-categorical': for any two of its models, one is isomorphic to an initial segment of the set-theoretic hierarchy of the other. The result extends to models of second-order ZFCU, provided that the sets that interpret 'urelement' are themselves isomorphic. This can plausibly be taken to show that two possible worlds at both of which the second-order or plural version of ZFCU is true, and where exactly the same urelements exist, can differ set-theoretically only as regards the height of the hierarchy.

The fourth possibility is that the relevant difference between w and w' involves some facts about abstract entities other than mathematical ones, which matter to the truth of T. For example, maybe w' contains abstract angels, and T entails that there are none.[29] If this is an issue, it is easily dealt with by expanding M slightly—in most cases, it should suffice to include the claim that the only abstract entities are mathematical ones.

4 The damning analogy

Let's suppose that $T^\Diamond$ and $T^\Box$ both meet desiderata (i)–(iii): they are deductively entailed by T, empirically equivalent to it, and nominalistic. If so, everything turns on desideratum (iv): the anti-nominalist needs an argument that $T^\Diamond$ and $T^\Box$ are worse—less plausible as stopping-places for explanation—than T, for some good theory T.

How could one go about arguing for a conclusion like that? An attractive idea is that we should look at the way scientists actually reason. It is scientists, not philosophers, who are most noted for their skill in inductive reasoning. If we want to determine what the *best* kind of inductive reason-

[27]A theory T could satisfy this condition either by containing a necessitated version of a (*)-style theoretical definition for each of its mixed predicates, or by merely making the supervenience claim while leaving it open exactly how the supervenience works.

[28]As follows: whenever there are infinitely many of the xs and infinitely many of the ys, and all of the xs and ys are concrete, either there are as many xs as concrete objects, or there are are as many ys as concrete objects, or there are as many xs as ys.

[29]Do ordinary scientific theories ever entail such things? Anna Mahtani pointed out to me that they may do so implicitly, by making claims about causal or counterfactual relations among concreta that would not obtain in certain kinds of angel-worlds.

ing requires in a given case, we should not just sit around thinking about how *we* are disposed to reason; instead, we should be guided by examples of good inductive reasoning in scientific practice.

There are more and less direct ways to apply this 'naturalistic' methodology in assessing the epistemic credentials of nominalistic replacements for standard scientific theories. The direct approach is to listen to what actual scientists say when we present them with the particular theories we are interested in. The indirect approach is to see how scientists reason in other domains, extract some general epistemological principles about what makes for theoretical goodness and badness, and apply these principles to the questions we are interested in.

Burgess and Rosen (1997: §III.C.1.a) have recently championed the direct approach. They wryly suggest that nominalistic alternatives to standard scientific theories should be tested by submitting them to scientific journals like the *Physical Review*. If they are rejected, we are invited to conclude that the theories in question are worse, epistemically speaking, than the originals, so that the original theories' claim on our credence remains undiluted.

There are several reasons to be uneasy with such a naked appeal to authority. Let me mention three. First, the standards for acceptance in a given scientific journal are evidently quite far from the notion of theoretical goodness we are interested in: for one thing, they reflect facts about the currently accepted demarcation between the different branches of science. Biologists obviously shouldn't be worried by the fact that their papers would be rejected by *Physical Review*; why should it be different for nominalistic philosophers? The central *professional* judgment that would lie behind the rejection from *Physical Review* is 'this belongs in a philosophy journal'. This obviously has nothing to do with theoretical goodness, and everything to do with the demarcation of different subject areas within the overall scientific enterprise. Since the question of nominalism happens not fall into the remit of any of the currently constituted departments of the science faculty, any scientists we might ask about it would be going beyond their sphere of distinctively professional expertise. The situation is the same with many debates in the philosophy of physics. Philosophers who aspire to learn from physicists soon realise that *most physicists just don't care* about the theoretical differences that seem so important to us. Unless the lesson we think we should learn from the physicists that we shouldn't care either, our learning is going to have to be indirect, guided by analogies between our questions and the ones in which physicists do have a professional interest.

Second, even if deference to scientists shows that we have good reason to believe in numbers, it is not clear how this bears on our limited question, which is whether we have good reason *of an inductive, topic-neutral sort* to believe in numbers. The way scientists actually reason about mathematical entities seems strikingly different from the way they reason about electrons and such like. One natural moral for 'naturalists' to draw from this is that good reasoning about numbers is subject to topic-specific standards very different from those that govern inductive reasoning about other kinds of theoretical posits.[30]

Third, even if I were to concede that 'there are numbers' is true in ordinary scientific English, in the context of ordinary scientific discourse, there is a further question which I think I understand, and which this concession would leave open—whether there are numbers *in the most fundamental sense*.[31] If there is a distinct intelligible question here, no amount of direct deference to scientists will help us resolve it. Nevertheless, if there is such a question, it is especially plausible that it cannot be resolved by deductive or perceptual or topic-specific considerations, so it is important to investigate what good inductive reasoning might have to say about it. Even if scientists almost never use the fundamental quantifiers themselves, scientific considerations do give us reason to form opinions about what there is in the most fundamental sense. For example, I think we have some reason to think that among the things that there are, fundamentally speaking, are spacetime points or regions. The epistemic situation *might* be similar for numbers or sets. But if we want to see whether it is, we will get nowhere by trying to cajole scientists into making pronouncements on the question.

For these reasons, I think the indirect version of the naturalistic methodology is more promising for our purposes. What we want, essentially, is an

[30]Maddy (1997) draws the latter conclusion.

[31]See Dorr 2007: §1 for an introduction to the question as I understand it, and Dorr 2005 for a more concerted attempt to explain it to those who don't.

argument by analogy, of the form 'Given scientific practice, the following must be explanatorily bad theories; the theory we are interested in is similar in relevant respects to them; therefore the theory we are interested in is also a bad theory'. In the case of $T^\lozenge$, a rather powerful argument of this form is implicit in much of the literature on 'the indispensability argument', for example in Putnam's discussion of 'fictionalism' (1972: pp. 350–56); Field (1988: pp. 260–61) articulates it especially clearly. I will state the argument in my own way.

Scientists—at least when they are reasoning unselfconsciously and not being led astray by bad philosophy—believe that there are lots of things much too small to be observed by anyone. So we should conclude that this belief is required by good inductive reasoning. This is so despite the fact—much dwelt on by scientific anti-realists like van Fraassen (1980), who deny that we ought to believe in unobservable entities—that for any theory T postulating unobservable entities, we can easily find a theory which shares T's consequences about the observable realm without entailing anything about unobservables. One such theory is

($T^\blacklozenge$) Possibly, all observable matters are just as they in fact are, and T.

$T^\blacklozenge$ says that T is 'observationally adequate': those of its consequences that are entirely about observable matters are true. Since our evidence gives us good inductive reason to believe claims about unobservable matters that do not follow from any theory of the form of $T^\blacklozenge$, these must be bad theories. Of course, we have reason to believe many such theories: $T^\blacklozenge$ is after all deductively entailed by T. But this reason derives from the reason we have to believe stronger theories which do have nontrivial consequences about the unobservable world.

What is it about these theories that makes them so bad? One factor that seems closely related to explanatory goodness is simplicity. Could it be the sheer complexity of the notion of 'observability' (and its precisifications) that is responsible for the badness of $T^\blacklozenge$ (and its precisifications)? No. For even when we replace the notion of observability with something precise and reasonably simple, we still end up with theories which must be bad:

> Possibly, the position of the centre of mass of each atom is just as it in fact is, and T.

> Possibly, the total mass contained in each region of space at each time is just what it in fact is, and T.

Since we do, or at least could, have empirical reason to believe some reasonably specific claims about subatomic structure, and about fields other than the mass-density field, these theories must be bad in the same way as $T^\blacklozenge$. In the light of such examples, it is natural to conclude that the badness-making feature all these theories share is their distinctive *logical structure*. But this structure is also shared by $T^\lozenge$. So we can conclude that $T^\lozenge$ is also a bad theory.

For this argument to be defensible, we need to be careful about the meaning of 'bad'. One thing it certainly cannot mean is 'a priori unlikely': since $T^\blacklozenge$ is a logical consequence of T, it must be at least as a priori likely as T. What is a priori unlikely is not that $T^\blacklozenge$ should be *true*, but that it should be true without its truth being entailed or probabilified by the truth of some better theory—T itself, or some other theory that entails quite a lot about the unobservable portions of the world. This suggests the following gloss on 'bad': a theory is bad just to the extent that it is a priori unlikely that it should be true *without its truth being explained by that of any better theory*. While this is no good as a definition, given its circularity, it does suffice to ground useful entailments between claims about theoretical goodness and claims about a priori likelihood.[32]

[32] This notion of theoretical badness should not be confused with the related notion of *crying out for explanation*. Roughly, for a claim to cry out for explanation is for it to be unlikely *conditional on its being true* that its truth is not explained by that of any better theory. Not all bad theories cry out for explanation. Even if it is very unlikely that T is true and unexplained, it might be similarly unlikely that T is true and explained, in which case the probability that T is explained conditional on its being true will not be high. It is not plausible that the extent to which a theory cries out for explanation is determined by facts about its logico-syntactic structure. White (2007) considers theories that are enormous conjunctions of claims about the position and size of each pebble on a certain beach. Even though these theories are all syntactically very similar, some—such as those that describe arrangements in which the pebbles compose pictures of faces—cry out for explanation far more urgently than others. That's because there is a fairly good theory—that the pebbles were deliberately arranged by someone aiming to make a picture of a face—which raises the probability of the conjunctions that describe face-like arrangements, while there is no comparably good theory which significantly raises the probability of the conjunctions that describe random-looking jumbles.

The argument that since $T^{\blacklozenge}$ and $T^{\lozenge}$ are similar in logical form, they must also be similar in being bad theories may seem too tenuous to carry so much weight. $T^{\blacklozenge}$ and $T^{\lozenge}$ are dissimilar in lots of ways. For example, the objects 'modalised away' by $T^{\lozenge}$ are abstract (non-spatiotemporal, causally inefficacious,. . .), whereas those 'modalised away' by $T^{\blacklozenge}$ are concrete. Why rest so much on the similarities? But then again, what can we go on, besides such analogies, in coming up with an evaluation of $T^{\lozenge}$? We have several millenia's worth of experience to show us that, once we start appealing directly to our *distinctively philosophical* intuitions—intuitions concerning the alleged difficulty of knowing anything about causally inefficacious objects, for example—the debate about the epistemological status of nominalism will end up hopelessly deadlocked. There is something deeply appealing about the 'naturalistic' methodology that tells us to form opinions about controversial questions in applied epistemology, such as that of the epistemic status of $T^{\lozenge}$, by starting with the large body of case-by-case epistemological judgements common to all scientific realists, looking for whichever epistemological theory does the best job of accounting for and systematising this data, and following this theory where it leads.

5 Extending the analogy to $T^{\square}$?

We certainly *could* extract from examples like $T^{\blacklozenge}$ some general principle that would impugn $T^{\square}$ as well as $T^{\lozenge}$. For example, we could conclude that the use of modal operators, or at least their use in stating theories about subject matters that don't themselves have any special connection to modality, is a general source of theoretical badness. Or we could conclude that 'parasitic' theories, which embed other, stronger, theories, are *ipso facto* bad. There is something compelling about such conclusions: they resonate with our moral judgment that theft is bad, and honest toil is better. But we should not mistake a resonant metaphor for a good argument. Is it actually *true* that modality and/or parasitism are general sources of theoretical badness? If it is, $T^{\square}$-style theories are tarred with the same brush as $T^{\lozenge}$-style theories, and we are thrown back upon Field's programme. And who knows whether we will be able to come up with nominalistic versions of general relativity or quantum mechanics that are as utterly free of any taint of similarity to theories like $T^{\blacklozenge}$ as is Field's version of Newtonian gravitation? If not, the question of the explanatory goodness of $T^{\square}$-style theories remains unresolved.

Can we construct an analogy that does for $T^{\square}$ what the analogy with $T^{\blacklozenge}$ did for $T^{\lozenge}$? To do so, we would have to find some theory of the form

($T^{\blacksquare}$) Necessarily, if the observable facts are just as they in fact are and BLAH, then T

which is empirically equivalent to T. But what could BLAH be? For $T^{\blacksquare}$ to be observationally equivalent to T, it needs to be filled in in such a way that (i) every BLAH-world that is observationally indiscernible from a T-world is itself a T-world, and (ii) every world (or anyway, every world that is not observationally indiscernible from any T-world) is observationally indiscernible from some BLAH-world.

One proposition with these properties is the material conditional $T^{\blacklozenge} \supset T$:

($T^{\blacksquare\supset}$) Necessarily, if the observable facts are just as they in fact are and ($T^{\blacklozenge} \supset T$), then T.

$T^{\blacksquare\supset}$ is a priori equivalent to $T^{\blacklozenge}$: it is true at w if the observable facts at w are consistent with T, and false at w if the observable facts at w are inconsistent with T. So we have the same reason to think that $T^{\blacksquare\supset}$ is a bad theory that we have for $T^{\blacklozenge}$. However, the structural parallel between $T^{\blacksquare\supset}$ and $T^{\square}$ is much weaker than that between $T^{\blacklozenge}$ and $T^{\lozenge}$—too weak for the analogy to carry any force. The mathematical theory M that features in $T^{\square}$ is a simple piece of theory—much simpler than T itself, of which it is a conjunct—whereas $T^{\blacklozenge} \supset T$ is at least as complex as T. Moreover, $T^{\blacksquare\supset}$ embeds $T^{\blacklozenge}$, which we already have reason to think of as a distinctively bad theory, whereas $T^{\square}$ does not embed anything that we have independent reason to regard as problematic.

To get a serious argument by analogy going, we would need to find some *simple*, *unified* claim to substitute for BLAH in $T^{\blacksquare}$—something that we could think of as exhausting the 'non-observational content' of T in the same way M exhausts its mathematical content. But ordinary scientific theories about unobservables don't contain anything like this; nor do ordinary scientific theories about subatomic particles, electric charges, and so on. If some empirically successful theory T did turn out to have a fragment which determined all the facts about subatomic particles as a function of facts about the locations of atoms, while being consistent with any consistent hypothesis about the locations of atoms, then this fragment could play the role of BLAH. But in the cases where we clearly have reason to believe in some unobservable structure, we don't have reason to believe in any simple formula whereby the facts about that structure can be read off from other facts. Once such 'reading off' comes into view, our belief in the hidden structure wavers. An example is the gravitational field in Newtonian gravitational theory. Some rather elegant versions of the theory take the gravitational field (or the gravitational potential) to be a genuine piece of extra intrinsic structure, in virtue of which geometrically indiscernible regions of empty space could fail to be duplicates, but governed by laws which fully determine the field at any time given the distribution of mass at that time.[33] But even if such a theory perfectly fit our evidence, there would be little pressure to believe in the gravitational field as a piece of additional intrinsic structure. Given that the theory has a simple fragment which lets us read off the facts about the field from distribution of masses, the proposed structure seems redundant. A view that use the fragment to define the gravitational field extrinsically, in terms of the distribution of masses, is an attractive alternative to the view that takes it to be intrinsic and fundamental.[34]

The prospects for establishing the badness of $T^{\Box}$ using an argument parallel to our argument by analogy for the badness of $T^{\Diamond}$ seem poor. But wait: do we need a new argument at all? Section 3 held out the hope that $T^{\Box}$ and $T^{\Diamond}$ are a priori equivalent, in the central cases of interest. If they are equivalent, doesn't that mean that they are really the same theory, or at least, that they are alike in respect of theoretical goodness? —No, it had better not mean that. If we individuated theories coarsely, so that logical equivalence was sufficient for identity, the the argument from the similarity of logical form between $T^{\blacklozenge}$ and $T^{\Diamond}$ to their similarity in respect of theoretical virtue would be a non-starter. For on this way of thinking of theories, a theory has many different logical forms. The mere fact that two theories *can* be given analogous logical forms tells us little, since it is compatible with there being some other logical form that only one of the theories admits. In particular, if $T^{\Diamond}$ admits a '$T^{\Box}$-style' logical form in addition to its '$T^{\Diamond}$-style' logical form, while $T^{\blacklozenge}$ admits no relevantly similar logical form, that might well be an epistemically relevant difference between $T^{\Diamond}$ and $T^{\blacklozenge}$. Indeed, on this way of talking, *every* theory admits a '$T^{\Diamond}$-style' logical form—if T is entirely about subject matter S, then T is a priori equivalent to 'Possibly, the S-facts are just as they in fact are, and T'. If one is going to try to characterise good inductive reasoning using an IBE-style framework, a fine-grained conception of theories, as something like structured propositions, is more useful for the purposes of formulating generalisations about theoretical goodness.

6 Existential and universal quantification

So far, then, the only arguments on the table for the badness of theories like $T^{\Box}$ involve deriving some very sweeping general principle, like 'Parasitic theories are bad', from examples like $T^{\blacklozenge}$. This already looks tendentious. The current section will introduce some new data which will cast further doubt on these arguments, by suggesting that the contrast between parasitic theories which use possibility operators (like $T^{\Diamond}$) and those which use necessity operators (like $T^{\Box}$) may matter a lot to the theories' epistemic status. The data will involve the related contrast between existential and universal quantification.

If science can tell us anything at all about the unobservable world, one thing it tells us is that objects that are alike in respect of shape, size, motion and mass need not be exactly alike in all respects. For example, objects that are exactly alike in all those respects can fail to be perfect duplicates by having different distributions of electric charge. We have—or at least *could* have—good empirical reason to believe this, in spite of the following fact: for any theory T that entails that things sometimes differ intrinsically by having different charge distributions, we can find a weaker theory that has

[33] Interestingly, Field's way of nominalising Newtonian gravity depends essentially on taking this realist attitude to the gravitational field.

[34] In electromagnetism, by contrast, the possibility of source-free radiation prevents any analogous 'reading off' of the facts about the electromagnetic field from the distribution of charged matter.

all the same consequences as T concerning the shapes, sizes, motions and masses of material bodies, while being consistent with the thesis that these are the only intrinsic respects in which things ever differ. One such theory uses a possibility operator: 'Possibly, the facts about shape, size, motion and mass are just as they in fact are, and T'. But when T contains enough mathematics, there is no need to use a modal operator to formulate the desired weakening: ordinary quantification is enough. We can proceed as follows: (i) Express T in such a way that all talk of charge is accomplished by a single expression—say a one-place functor 'charge(x)'. (ii) Replace each occurrence of 'charge(x) = n' with '$c(x) = n$', where c is some new variable ranging over functions from bodies to real numbers; call the result of this $T(c)$. (iii) Let our new theory $T^{\exists c}$ be $\exists c T(c)$. Since the availability of theories like $T^{\exists c}$ does *not* undermine our inductive reason to believe that charge is just as real and intrinsic as shape, size, motion and mass, we can conclude that $T^{\exists c}$ must be a much worse theory than T.

Although $T^{\exists c}$ does not explicitly mention electrical charge, there are ways to interpret 'charge' on which it entails that 'objects have charges' is true. We could, for example, interpret 'charge(x) = n' as meaning 'for some function c such that $T(c)$, $c(x) = n$' or 'for the unique function c such that $T(c)$, $c(x) = n$'. Or, if we wanted 'If objects have charges, then T(charge)' to come out expressing a contingent truth, we could get a bit fancier, interpreting 'n is the charge of x' to mean something like '$c(n) = x$ for the function c that plays the simplest sufficiently "charge-like" role in relation to the facts about shape, size, motion, and mass.'[35] On these interpretations of 'charge', the proposition expressed by 'T(charge)' may even be a priori equivalent to $T^{\exists c}$. But these propositions are very different from the original theory T, as we were imagining it. On the intended interpretation, the facts about charge were supposed to be intrinsic, at least in the sense that objects with different distributions of charge are never duplicates. By contrast, any interpretation of T on which it follows from $T^{\exists c}$ will require 'charge' to express something highly extrinsic.

We might think of extending the 'charitable interpretation' trick to predicates like 'duplicate' as well as predicates like 'charge'; then $T^{\exists c}$ would after all entail that 'objects with different charge distributions are never duplicates' is true. But this way lies Putnam's paradox (Lewis 1984). There must be more to correctness of interpretation than considerations of charity! Otherwise any theory we ended up accepting would be correctly interpreted as equivalent to the result of 'existentially quantifying out' all of its non-logical constants; the only way such a theory could be false would be for it to make some false claim about the cardinality of the universe. Predicates like 'duplicate' seem an especially good place to put a stop to charity run amok. While it is debatable whether $T^{\exists c}$ is a priori consistent with the hypothesis that nothing is charged, it is very clear that it is a priori consistent with the hypothesis that objects are duplicates whenever they are alike in respect of shape, size etc.

So $T^{\exists c}$ really is weaker than the original T, and weaker than what the empirical success of T would give us reason to believe; our conclusion that it is a bad theory stands. There is a general pattern here. If we want to weaken a theory so as to eliminate its commitment to some sort of hidden structure, we can often do so by replacing the vocabulary which purports to characterise this structure with variables of an appropriate sort, bound by initial existential quantifiers. Philosophers who are suspicious of particular putative bits of hidden structure keep on rediscovering this fact, and announcing that they have shown how to eliminate the structure in question. But once we have realised the complete generality of the trick, we should not be impressed by their achievements. Here are some more examples.

(i) Many ordinary physical theories that speak of fundamental particles are naturally understood as entailing that these particles come in several qualitatively different kinds. But we can get rid of this entailment, by replacing each predicate purporting to stand for a kind of particle with a new plural variable, bound by an initial existential quantifier. So our new theory will look something like this: 'there are some particles, the xs; and there are some particles, the ys, and… such that T(the xs, the ys,…)'. We can also attempt to reinterpret the original theory so that it is entailed by the new theory, by analysing the predicates the purport to stand for kinds of particles as expressing extrinsic properties that par-

[35] In working out such an interpretation, we might take inspiration from 'best system' analyses of lawhood (Lewis 1994).

ticles instantiate in virtue of their motions with respect to other particles.

(ii) Ordinary physical theories formulated in co-ordinate terms are naturally understood as claiming that spacetime has a geometric structure much richer than that of mere topology: regions of spacetime can differ in all sorts of intrinsic geometric respects even when they are topologically indiscernible. But any such theory can be weakened so as to remove this implication: we need only say that *there is some co-ordinate system* which respects the spacetime's topological structure, relative to which the given dynamical equations are true. This minimalistic way of thinking about the content of such theories is favoured by van Fraassen (1970). And, under the influence of a positivist philosophy that rejects the whole idea of intrinsic structure, similar ideas are still sometimes found in physics textbooks—it is common to formulate Newton's first law as the claim that 'there exist inertial frames', where these end up getting defined as frames in which Newton's laws hold.[36] But I have no interest in deferring to the opinions of physicists when these are manifestly influenced by the discredited anti-realist philosophies of the past. We *do* have good reason to ascribe to spacetime an intrinsic geometric structure that goes far beyond topology. Thus the existentially quantified theories which purport to explain all our observations without entailing that there is any such structure must be bad theories.

(iii) Similarly, Newtonian mechanics, by speaking about absolute accelerations, seems to require reality to have a geometric structure that fails to supervene on the history of distances between pairs of particles. This is generally agreed to be a major problem for Leibnizian relationalism, according to which there is no more to geometric structure than the history of these distances. But according to Huggett (2006), there is no problem. The Leibnizian relationalist can simply adopt a theory of the form 'there are some admissible co-ordinate systems in which Newton's laws hold', where 'admissible' co-ordinate systems are those that respect the history of inter-particle distances. Huggett goes on to suggest a 'best system' analysis of claims about absolute acceleration: roughly, to be accelerating is to be accelerating according to every 'best' co-ordinate system, where 'best' is understood in such a way that if there are any admissible co-ordinate systems in which Newton's laws hold, they are guaranteed to include all the best ones.

(iv) Imagine a theory of fluid mechanics that describes a world entirely filled with a continuous fluid, moving around in various ways, and instantiating different fundamental scalar quantities like mass-density and charge-density. To the extent that we had reason to take such a theory seriously as a fundamental theory, we would have reason to reject a stringent version of Humean supervenience on which there is nothing more to the world than points standing in spatiotemporal relations and instantiating intrinsic properties. But our fluid-mechanical theory can be weakened to make it compatible with this stringent Humeanism, by existentially Ramsifying the vocabulary that purports to characterise the velocity field, or the partition of spacetime into trajectories of matter-points. And provided that this existentially quantified theory is true, we will be able to find extrinsic 'best system' interpretations of expressions like 'trajectory of a matter-point' or 'velocity field vector' under which the original theory is true. Sider (2002) develops such analyses, and concludes that the Humean has nothing to fear from the celebrated Leibniz-Russell-Broad-Kripke-Armstrong spinning disc/spinning sphere/infinite river objection. But he is wrong. If we are ever justified in positing hidden structure, sufficient empirical successes by a fundamental fluid-mechanical theory would justify us in positing hidden structure of a non-Humean sort.

(v) Once we have enough mathematics on board, we need not have recourse to possibility operators, as in $T^{\blacklozenge}$, if we want to weaken any theory so as to eliminate the implication that there are unob-

[36]See, e.g., Woodhouse 2003.

> servable objects: we can use existential quantification over *models* to achieve the same effect. We need only define up some notion of what it is for a model to 'accurately represent the observable facts': then our new theory can simply say that the old theory is true in some model that accurately represents the observable facts.

The rule seems to be this: when we modify a theory by replacing an expression that purported to stand for some aspect of the intrinsic structure of the world with a variable bound by an initial existential quantifier, the result is generally much worse than the original theory, even when it is empirically equivalent. The other important observation is that these bad existentially quantified theories are closely akin to bad theories like $T^{\blacklozenge}$. Indeed, the work done by existential quantifiers could in each case be done by an appropriate possibility-operator. For example, the Leibnizian relationalist could offer a theory of the form 'Possibly, the facts about the intrinsic properties of particles and the inter-particle distances are just as they in fact are, and T', where T is some orthodox version of Newtonian mechanics that entails the existence of rich geometric structure. We should thus expect that the true explanation of the badness of $T^{\blacklozenge}$ and $T^{\lozenge}$ involves some feature which they share with the bad existentially quantified theories we have just been considering. And there is no mystery about what that could be. Even those who reject the ontology of possible worlds can recognise the logical parallels between possibility operators and existential quantifiers which form the basis for possible-worlds model theory. For example, the inferences $\Diamond(\phi \vee \psi) \vdash \Diamond\phi \vee \Diamond\psi$ and $\exists x(\phi \vee \psi) \vdash \exists x\phi \vee \exists x\psi$ are both valid, while the inferences $\Diamond\phi \wedge \Diamond\psi \vdash \Diamond(\phi \wedge \psi)$ and $\exists x\phi \wedge \exists x\psi \vdash \exists x(\phi \wedge \psi)$ are not. It would thus not be at all surprising if the true canons of theoretical virtue turned out to group possibility operators and existential quantifiers together, as distinctive sources of theoretical badness.

This is already enough to cast doubt on the claim that the badness of $T^{\blacklozenge}$ is due to some feature it shares with $T^{\Box}$, such as relying on a modal operator or being parasitic. For our existentially-quantified theories use no modal operators, and are not in the relevant sense parasitic.[37]

If the canons of theoretical virtue are sensitive to the logical parallel between possibility-operators and existential quantifiers, as they seem to be, it stands to reason that they should also be sensitive to the logical parallel between necessity operators and universal quantifiers. So we can support the claim that there is an epistemologically important difference between necessity operators and possibility operators by arguing for an epistemologically important difference between universal and existential quantification. And in fact, *prima facie*, there is such a difference. It is utterly standard for a theory to consist of a universal quantification, or a conjunction of universal quantifications. By contrast, my attempts to imagine scientifically interesting theories that are conjunctions of existential quantifications all have something of the flavour of the bad theories considered above. Consider for example a theory that says that there is a point of space towards which all bodies accelerate (in certain specified ways). It seems to me that if we found out that there was such a point, we would have reason to think that it was intrinsically special, or at least that it could be distinguished by some structural role simpler than that of being a point towards which bodies accelerate in the specified ways. We should thus not be satisfied with the existentially quantified theory as a stopping place for explanations—we should hold out for some stronger theory which gives a substantive characterisation of the attractive point.

So things seem to be going well for $T^{\Box}$. We could further bolster the case for its theoretical goodness if we found some class of universally-quantified theories that stand to $T^{\Box}$ as the bad existentially-quantified theories considered above stand to $T^{\lozenge}$, and those theories turned out to be theoretically good. What could these theories be like? The logical structure we are looking for is fairly distinctive: just as it is only in special cases that one

[37] Perhaps there is a *historical* sense in which the existentially quantified theories can be said to be 'parasitic': the process whereby they in fact came to our attention involved our first thinking of a strong, hidden-structure-positing theory, and then noticing that we could weaken it by existentially quantifying out the structure-characterising predicates. But I doubt that such merely historical features of theories should matter when we're considering *ideal* inductive reasoning, as opposed to heuristics and rules of thumb. Anyway, the history of science is full of cases where a good theory was arrived at by weakening some worse, stronger theory: for example, special relativity was derived in this way from Lorentzian mechanics. The history of such episodes shows that a general suspicion of such theories is not much good even as a heuristic.

can *weaken* a theory by embedding it within a restricted necessity operator, likewise it is only in special cases that one can weaken a theory by replacing one of its primitive expressions with a variable bound by a restricted universal quantifier.

It seems to me that ordinary physical theories stated in co-ordinate terms fit the bill. When physicists write down equations about the rates of change of physically interesting quantities with respect to the x, y, z and t co-ordinates, they don't mean to suggest that there is a distinguished, intrinsically privileged co-ordinate system, concerning which we could sensibly ask questions like 'how far are we from the origin?'. Rather, they are—sometimes explicitly, sometimes implicitly—making universally quantified claims, to the effect that the equations in question hold true for *every* co-ordinate system that is 'admissible', in the sense of fitting in the right way with the intrinsic structure of the space in question.[38] Invariably, however, the claim that the equations hold for all admissible co-ordinate systems is a consequence of the claim that they hold in any one admissible co-ordinate system: if this were not true, we would know that we were working with an unduly impoverished conception of the intrinsic structure of the space, and thus an unduly generous definition of 'admissible'. The universally quantified theory that physicists actually take seriously is thus deductively entailed by the silly theory that posits a unique privileged co-ordinate system. The transition from the latter to the former stands to the transition from T to $T^{\Box}$ as the transition from T to $T^{\exists}c$ stands to the transition from T to $T^{\Diamond}$.

I don't suppose anyone has ever believed in the 'one true x-axis'. But actual physics does provide examples of transitions of the kind in question. One is the transition from Newton's version of Newtonian mechanics, with absolute motion and rest, to a version that does away with absolute motion and rest while keeping absolute acceleration. The history of this transition is complicated by the fact that the formal machinery required for a rigorous formulation of the latter, namely neo-Newtonian spacetime (Sklar 1974), was developed only after Newtonian mechanics had already been rejected for independent reasons. Nevertheless, the expert consensus is that the banishment of absolute motion and rest is a major theoretical improvement. And the standard way of formulating the neo-Newtonian theory is to claim that the equations of Newtonian mechanics hold in every inertial frame—that is, in every co-ordinate system that fits the intrinsic geometry of the spacetime in a specified way.

(An even more dramatic version of this sort of transition occurs with gauge-symmetric field theories. As standardly formulated, these theories speak of certain fields, understood as functions from points to geometric objects of some sort; but it is understood that by changing these functions in certain ways—for example, by adding any divergence-free vector field to the electromagnetic potential—one gets something that can be understood as 'an equivalent description of the same situation', analogous to a change of co-ordinates. Sometimes, in dealing with a particular problem, one will fix on a particular gauge, just as one might fix on particular co-ordinates. But much of the time one does not do this; in these cases, talk about a gauge-symmetric field like 'the electromagnetic potential 4-vector' is understood as governed by a tacit initial universal quantifier over 'admissible' fields of the relevant kind.)

Admittedly, in all these examples, it has turned out not to be necessary to formulate the theory as a universal quantification over co-ordinate systems or gauges. Modern differential geometry makes possible 'co-ordinate free' statements of basic physical laws; co-ordinate systems still play a role, but only at the foundations, where the distinction between admissible and inadmissible co-ordinate systems features in the definitions of differential manifolds, smooth functions, vector fields, fibre bundles, etc. This is undoubtedly an important theoretical advance. But we must not overstate its significance. The physics community was happy with co-ordinate-based formulations when these were the only ones available; and they are still dominant, outside some specialised contexts where foundational questions

[38] Standard theories in physics don't normally answer the question *what it is* for a co-ordinate system to be 'admissible': instead, they merely place constraints on the answer to this question, by characterising a relation between co-ordinate systems such that all and only those co-ordinate systems that bear that relation to some *other* admissible co-ordinate system are themselves admissible. We *could* take a 'heavy duty platonist' view of admissibility as a primitive, non-supervenient relation between physical and mathematical objects. But this view is *prima facie* unappealing, and no more compulsory than the corresponding view about any of the other mixed mathematico-physical predicates that crop up throughout physics.

loom large. A crucial property of the notation used in co-ordinate-free statements is the fact that equations involving co-ordinates equations can be read transparently off the co-ordinate-free ones. Even the most fervent apostles of the co-ordinate-free approach do not think that it saved us from having to believe in a single privileged co-ordinate system.

These data help to confirm the hypothesis that existential quantification *as such* is a distinctive source of badness. Weakening a theory by 'existentially quantifying out' some putatively structure-characterising predicates makes it worse. By contrast, in those special cases where one can weaken a theory by 'universally quantifying out' some putatively structure-characterising predicates, the result is often an improvement on the original. Together with the observation that the operations of 'existentially quantifying out' and 'embedding within a possibility operator' seem to make for badness *in the same way*, this provides some reason to think that the operation of 'embedding within a necessity operator' is like the operation 'universally quantifying out' in *not* making for theoretical badness, even in the special cases where its effect is to weaken the original theory.

Here is a toy theory of how these asymmetries might work. The badness of a theory increases with the number of symbols it takes to express the theory, in an appropriately canonical language. But the rate of increase is much greater within formulae governed by existential quantifiers and possibility operators. Or to be more precise: it is greater within formulae governed by existential quantifiers and possibility operators which occur in positive contexts, and within formulae governed by universal quantifiers and necessity operators which occur in negative contexts. (We don't want to be able to make a theory better by replacing $\exists$ with $\sim\forall\sim$, or $\diamond$ with $\sim\Box\sim$.) Thus, in general one can improve a theory by replacing a long existential quantification '$\exists x\phi(x)$' with a conjunction of the form '$\exists x\psi(x) \wedge \forall x(\psi(x) \supset \phi(x))$', where ψ is considerably shorter than ϕ. This toy theory inherits most of the defects of symbol-counting as a measure of the epistemically important notion of simplicity. But I hope that a more plausible measure of simplicity could be tweaked in a similar way, so as to make complexity within the scope of existential quantifiers contribute more to badness than complexity elsewhere.

7 What about $M^\diamond$?

As noted in section 3, the claim that $T^\Box$ is empirically equivalent to T depends on the assumption that $M^\diamond$ is a priori:

($M^\diamond$) Possibly, the concrete realm is just as it in fact is, and M.

If $M^\diamond$ is a priori, we can help ourselves to it for free in deriving empirical consequences from $T^\Box$, just as we can help ourselves to theorems of classical logic, even very complex ones; it does not have to be counted as part of the total package to which we apply our syntactic tests for theoretical badness. But is it a priori? I quite like the idea that that for such claims we can know a priori that there is no relevant gap between logical and metaphysical possibility, or between metaphysical possibility and metaphysical consistency with the truth about the concrete world, so that if we could know a priori that M is *logically consistent*, we could deduce $M^\diamond$. But even if I am right about this, and even if the logical consistency of M is in some sense a logical truth, one might well balk at the idea that our justification for believing it is a matter of deductive rationality alone. The logical consistency of a theory like ZFCU is clearly epistemologically problematic in a way that theorems of predicate logic, even complex ones, are not. Indeed, if we are justified in believing ZFCU to be consistent at all, part of the story about why we are so justified involves the empirical fact that so far no-one has succeeded in deriving a contradiction from it. This seems to be within the sphere of inductive rationality: the hypothesis that ZFCU is consistent is supported by its constituting a good explanation of our failure to derive contradictions from it. If so, we do not get to help ourselves to $M^\diamond$ for free. And that is a worry, because the logical form of $M^\diamond$ is the very same as that of $T^\diamond$ and $T^\blacklozenge$, which we found to be a source of badness. If we need to posit $M^\diamond$ as part of an explanation of some empirical phenomena—either the wide range of empirical phenomena putatively explained by T, or merely our failure to derive contradictions from M—doesn't the analogy with $T^\blacklozenge$ show that the explanation in question is a bad one?

Two responses to this objection seem promising to me. The first response concedes that $M^\diamond$ needs to be included along with $T^\Box$ in the total package of theory for the purposes of applying IBE, but insists that, despite sharing a logical form with $T^\lozenge$ and $T^\blacklozenge$, $M^\lozenge$ is nevertheless much better than them, since M is so much simpler than the total physical theory T to which $T^\lozenge$ and $T^\blacklozenge$ apply their possibility operators. The contrast in simplicity is genuine: while laws in physics sometimes admit of very compact statements, these invariably involve many expressions whose definitions have been carefully crafted to allow for such compactness. And the toy theory of section 6 predicts that simplicity matters: the idea is that complexity is worse within the scope of an existential quantifier or possibility operator, so that one does better the more of the meat of one's theory one manages to exclude from such contexts.

The example of co-ordinate systems is helpful here. If a theory that begins 'For every admissible co-ordinate system. . . ' is to have any empirically interesting consequences, it will need to be combined with something that entails that there is at least one admissible co-ordinate system. In some especially nice cases, like Euclidean geometry, we can find geometric axioms which do entail this (given an appropriate definition of 'admissible') without being at all syntactically analogous to bad existentially quantified theories. But in other cases, we are still left with a residual existential quantification governing something somewhat complex, albeit much less complex than the original theory. For example, in standard treatments of differential geometry it is a basic axiom that for every point there is *some* admissible assignment of co-ordinates (in R^n) to points in a neighbourhood of that point. And unless we go with an implausible conception of admissibility as a primitive property of co-ordinate systems, its definition in terms of intrinsic geometric relations holding among points of space is bound to introduce a certain amount of complexity. Nevertheless, these theories are explanatorily good: no one thinks that there are any epistemic advantages to be gained by positing a new piece of fundamental structure that assigns to each point a unique privileged local co-ordinate system around that point. So however we end up implementing the idea that existential quantification is a distinctive source of badness, the existential quantification in 'for each point, there is an admissible co-ordinate system around that point' had better not turn out to make for too much badness. If our grouping of possibility operators together with existential quantifiers is on the right lines, then, we shouldn't just throw up our hands whenever we see something with the logical form $M^\lozenge$ shares with $T^\lozenge$ and $T^\blacklozenge$. Rather, we should try to take considerations of simplicity into account, being guided as much as we can by analogies such as the one with co-ordinate systems.[39]

The second response is less concessive. Even if it is recognised that empirical considerations sometimes play a role in explaining why it is reasonable for us to believe logical truths, perhaps we should not expect the rules that seem to govern good inductive reasoning about logically contingent matters to carry over to the realm of the logically necessary. For there is something odd about the thought that we could be justified in believing something logically contingent (for instance, that there are models of ZFCU) in virtue of its *recognisably* entails something (for instance, that there are no valid derivations of contradictions from the axioms of ZFCU) which is in fact a logical truth, hence entailed by everything. Our limited ability to recognise certain kinds of logical truths as such seems like a reason for diffidence and caution: it would be strange if it rationally required us to be *more* opinionated about some logically contingent matters. Perhaps, then, we should understand our conclusions about explanatory goodness as telling us in the first instance about the standards an extremely idealised kind of inductive reasoning, whose prerequisites include being fully confident of all logical truths, including truths about logical consistency. If so, we should not expect them to play the same role in an account of good inductive reasoning at a more humanly attainable level—about how we should best accommodate our limitations, such as our inability reliably to distinguish logical truths from nonlogical ones.

8 Other theoretical virtues

Burgess and Rosen (1997: §III.C.1.a) put forth the following argument for the theoretical badness of a wide range of nominalistic theories. *Famil-*

[39] It may be too that such examples show that theories that require the initial quantifier order $\forall\exists$, as in 'for every point, there is an admissible co-ordinate system around that point', are *ceteris paribus* better than theories with an initial $\exists$. If there is anything true in this vicinity, it is good news for us: to the extent $M^\lozenge$ is plausible, the stronger claim that *necessarily* it is possible for the concrete realm to be just as it in fact is while M is true—in other words, that *every* world is concretely indiscernible from some M-world—is also plausible, and could be used instead of $M^\lozenge$ in our total theoretical package.

iarity, *perspicuity* and *fruitfulness* are features that make scientists favour a theory, other things being equal. So they are theoretical virtues: inductive reasoning that favours theories that have them is, other things being equal, good reasoning. But the nominalistic theories under consideration lack these virtues. And other things are close enough to being equal. Hence the nominalistic theories are worse than the theories they aim to supplant.

For various reasons, I am not close to being convinced by this argument, at least as applied to theories like $T^{\Box}$. Here are some of the reasons.

(i) I find it implausible that considerations like 'familiarity' play any role in the theory of *ideal* inductive reasoning. Normally, the facts about the order in which theories are invented have no evidential bearing on which theories are true; a perfect reasoner, whose degrees of belief fit the evidence, would thus treat these facts as irrelevant. Of course, real human beings, including scientists, are prone to favour the theories they encounter first over newly invented alternatives. And this conservative bias is not a mere defect, but makes sense given our other limitations. Since we don't have the time or ability to think through each theory for ourselves, we can legitimately use our knowledge of a theory's origins and history as a shortcut, and hold new theories in suspicion even when we haven't yet been able to uncover any intrinsic problems with them. But this won't matter if we are concerned with questions about how ideal inductive reasoning works.[40]

What's the *point* of investigating ideal rationality, if the claim that ideal inductive reasoning does not require belief in the existence of mathematical entities is consistent with the claim that the best kind of inductive reasoning available to human beings does require such a belief? —Well, what's the point in ever being concerned with questions about normative ideals? The answer, I think, lies in some principle like the following: if we ought$_{\text{non-ideal}}$ to believe that we ought$_{\text{ideal}}$ to believe that P, then we ought$_{\text{non-ideal}}$ to believe that P.[41]

[40]Similar remarks apply to perspicuity and fruitfulness, if these are construed in a way that ties them closely to the contingent capacities of human beings.

(ii) There may be some good sense in which Field-style nominalistic theories are less 'fruitful'—less capable of being patched up or extended to account for new phenomena—than the platonistic theories they replace; there is the risk that a small-looking change to the platonistic theory will generate a completely new set of difficulties for the reconstruction project. But it is hard to think of a sense of 'fruitfulness' on which $T^{\Box}$ could be said to be less fruitful than T, since there is such an obvious one-to-one correspondence between modifications we might perform on T and the corresponding modifications on $T^{\Box}$.

(iii) Of course, if we decided to believe only $T^{\Box}$ rather than some platonistic T, there would be no need for us to go around pronouncing the 'Necessarily, if the concrete realm is just as it in fact is and M, then. . . ' all the time, or forming conscious mental representations of it. It would naturally fade into the background, and the detailed business of scientific theory-construction and communication could take place in exactly the same way as before. This sort of thing is common in scientific practice: long stretches of thought and discourse are implicitly governed by assumptions which many competent practitioners would not readily be able to make explicit. All of this makes it hard to get the charge of lack of 'perspicuity' to stick.

(iv) Suppose I am right in thinking that I understand *fundamental* meanings for quantifiers which are different from the meanings they express in ordinary scientific contexts. No matter how seriously we take virtues like familiarity, they are not going to be much help in drawing epistemic distinctions between theories expressed using fundamental quantifiers—all such theories are pretty unfamiliar, awkward to work with, etc.

[41]And likewise if we ought$_{\text{non-ideal}}$ to believe that we ought$_{\text{ideal}}$ *not* to believe that P, then we ought$_{\text{non-ideal}}$ not to believe that P.

Another putative virtue that might be thought to favour T over $T^{\Box}$ is that of *ideological economy*. This pair of theories might seem perfectly to exemplify Quine's famous tradeoff between ontology and ideology; and if one thinks in these terms, trading in $T^{\Box}$'s distinctive ideology (modal operators, including whatever is required to make sense of restrictors like 'the concrete realm is just as it in fact is') for T's ontology (mathematics) might seem a benefit. But this can't be the right way to think. We just do understand modal operators: a blank rejection of all modal claims is not a serious theoretical option. And because of this, we can just see, at least in the central cases of interest, that if T is true, $T^{\Box}$ must also be true.[42] It is mysterious how the metaphor of economy could apply in such a case: its natural use is in making comparisons between *competing* theories.

The notion of ideological economy does feature in an important class of arguments from $T^{\Box}$ to the claim that there are abstract objects of some sort. These arguments work by exhibiting some general *metaphysical analysis* of modal operators like those used in $T^{\Box}$, under which $T^{\Box}$ and/or $M^{\Diamond}$ turn out to require the existence of abstract objects, such as possible worlds. It is claimed that the unless the analysis in question is accepted, we will have to be committed to the ideologically uneconomical view that the modal operators are 'metaphysically primitive', a brute addition to the overall structure of reality on a par with the structures investigated by fundamental physics. Evaluating this kind of argument is a big task. For one thing, it is far from clear how to understand the notion of metaphysical primitiveness in such a way as to make room for a debate whether metaphysical possibility and necessity are primitive. (Nominalists can easily accept that modal facts *supervene* on the non-modally-specified facts; is this enough for them not to be primitive?) For another thing, it is far from obvious that invoking abstract ontology lets one avoid taking modality as primitive.[43] I won't attempt this task here. I just want to point out that this kind of argument is quite different from the argument we have been concerned with, which purports to establish the existence of abstract entities using an inductive inference that essentially involves our specific, contingent evidence, as opposed to general considerations that would apply in the same way no matter what our evidence had been like.[44]

9 Objections

(i) *Theories about simplicity*

Section 5 claimed that there is no way to fill in the BLAH in $T^{\blacksquare}$ with something comparable with M in terms of simplicity, etc., in such a way as to do for subatomic particles or some other category of physical unobservables what $T^{\Box}$ does for numbers. What about something like $T^{\Box s}$?

($T^{\Box s}$) Necessarily, if the laws are as simple as they could be given the facts about observables, T.

This has the same consequences for observables as T. And typically, when T is itself simple, it will be reasonable to be confident that $T^{\Box s}$ is true if T is.

By using simplicity in this way within the theory, one can also formulate universally quantified replacements for the existentially quantified theories we considered in section 6:

[42] I am abstracting away here from the worries discussed in section 7

[43] While metaphysical possibility and necessity are tricky, it is easier to see how abstract ontologies containing facts, propositions, or possible worlds might help with the analysis of the expression 'the concrete realm is just as it in fact is'. But nominalists are not forced to take this expression as metaphysically primitive. One strategy is to metaphysically analyse it as a conjunction, whose first conjunct says that everything exists that in fact exists, and whose remaining conjuncts are of the form $\forall x_1 \dots x_n(Rx_1 \dots x_n \leftrightarrow \downarrow Rx_1 \dots x_n)$, where R is some metaphysically primitive physical predicate. ($\downarrow$ is the Hodes 'backspace' operator: see note 13 above). The first conjunct is tricky to make sense of in a way that does not make it redundant; one way to do so would require another new operator that 'undoes' the effect of 'in fact'. Must nominalists take *these* operators as metaphysically primitive? Not if they are the kind of nominalists who can tolerate a bit of second-order quantification—then they can analyse a claim like '$\Diamond(\forall x_1 \forall x_2(Rx_1x_2 \leftrightarrow \downarrow Rx_1x_2) \wedge T)$' as $\exists X(\forall x_1x_2(Rx_1x_2 \leftrightarrow Xx_1x_2) \wedge \Diamond(\forall x_1x_2(Rx_1x_2 \leftrightarrow Xx_1x_2) \wedge T))$, where the second-order variables are understood as 'rigid'. Rigid second-order quantification likewise lets us analyse 'Possibly, everything exists that in fact exists and T', as '$\exists X(\forall x(x = x \leftrightarrow Xx) \wedge \Diamond\exists Y(\Box\forall x(Xx \leftrightarrow Yx) \wedge T))$'. (See Hodes 1984.)

[44] Sider (MS) develops a worked-out way of talking about ideological economy in terms of what I was calling 'metaphysical primitiveness', and uses it to argue for the existence of sets.

($T^{\forall s}$) On every topologically admissible co-ordinate system $\langle x, y, z, t\rangle$ which permits a maximally simple statement of the laws, $T(x, y, z, t)$.[45]

The challenge for me is to explain why $T^{\Box s}$ and $T^{\forall s}$ are bad in a way that doesn't impugn $T^{\Box}$.

The challenge isn't so hard to meet. The first thing to observe is that the notion of simplicity that features in these theories is itself quite complex (on each of its precisifications), and needs to be for $T^{\Box s}$ and $T^{\forall s}$ to be at all plausible. And the second thing to observe is that this complexity is itself embedded in the scope of a an existential quantifier or possibility-operator, in a positive context, thanks to the quantification implicit in the notion of *maximal* simplicity. This can be seen easily in the case of $T^{\forall s}$, which can be spelled out as follows:

> For every topologically admissible co-ordinate system $\langle x, y, z, t\rangle$: either there is a topologically admissible $\langle x', y', z', t'\rangle$ which permits a simpler statement of the laws than $\langle x, y, z, t\rangle$ does, or else $T(x, y, z, t)$.

The same structure can be seen in $T^{\Box s}$ if we paraphrase it in terms of possible worlds:

> For every world w where the observable realm is just as it in fact is: either some world w' where the observable realm is just as it in fact is has simpler laws than w, or T is true at w.

Thus I doubt that explaining the badness of $T^{\Box s}$ and $T^{\forall s}$ requires any new insights beyond those contained in the toy theory of section 6.

(ii) Semantic ascent

Another suggestion for doing without possibility-operators or existential quantifiers is to use the resources of proof-theory. We could replace $T^{\blacklozenge}$ with something like this:

($T^{\vdash}$) Whenever there is a valid derivation whose only premise is 'T' and whose conclusion is a sentence S that is entirely about observable matters, S is true.

Making this precise is a big task; but if we do it properly $T^{\vdash}$ will entail every claim about the observable world that can derived from T.[46] Thus $T^{\vdash}$ must be a bad theory, for the same reasons as $T^{\blacklozenge}$.[47] How are we to explain the badness of $T^{\vdash}$? Filling in the definitions of derivation and truth, and replacing the quote-name 'T' with a syntactic description of a sentence expressing T will leave us with something quite complex by any reasonable standard. But if we are careful, it will be possible to keep much of this complexity—in particular, the complexity required for the syntactic description of 'T'—out of the scope of existential quantifiers.

The same kind of problem arises in a more straightforward way for existentially quantified theories like those considered in section 6. Surely on any sensible account of theoretical goodness, the transition from T to a theory that says that 'T' is true should never count as an improvement. But while a statement of the latter theory in fundamental terms will be long and intuitively quite complex, its general logical form will be much the same irrespective of whether the existential quantifications in the T we started with were complicated or simple.

There is some temptation to give up the high level of abstraction I have been looking for in an account of theoretical goodness, and admit special-purpose principles that apply only to theories involving semantic ascent. But, given the multifarious forms that a 'syntax' might take, it is hard to see how such principles could be stated in a way that would give them sufficient generality. So I hope that these cases can be dealt without introducing anything beyond general considerations of complexity into an account of theoretical badness. (I note, for example, that any theory that mentions some particular sentence by means of a description of its syntax

[45] Indeed, under a 'best system' analysis of geometric predicates that go beyond the favoured minimal base (in this case, topology), T itself will turn out to be equivalent to something like $T^{\forall s}$.

[46] I don't think we will need to worry about the semantic paradoxes: since the conclusion is only concerned with truths entirely about the observable realm, we can understand 'S is true' to mean that S is true in a set-sized model whose domain is the set of observable objects, which interprets 'red' as standing for the set of observable red things, and so on.

[47] Unless T is first-order, $T^{\vdash}$ will not be quite as strong as $T^{\blacklozenge}$; but I can't see why this would matter.

will have to contain many more quantifiers than that sentence itself does.) Given the point in the previous paragraph, this means that the transition from T to the theory that 'T' is true will not always be equally deleterious: we lose less if we start with a bad, existentially quantified theory than we do when we start with a universally quantified theory of equal syntactic complexity. But I don't see that that should matter, provided that what we end up with is always worse than what we started with.

In any case, while these cases may be problematic for the toy theory of section 6 and some of its elaborations, they do not threaten the central point of my defence of $T^{\Box}$. However we end up explains the badness of $T^{\vdash}$, there is no reason at all to fear that the explanation will in any way impugn $T^{\Box}$.

(iii) Differential equations

Does my hypothesis that complexity is always worse when it is in the scope of existential quantifiers stand up in the light of examples from actual science? I wish I knew. One worry I have thought about arises from the fact that our most basic physical theories are, or centrally involve, differential equations. Given the familiar epsilon-delta definition of the derivative operator, any differential equation will reduce under analysis to some rather complex formula with an initial string of universal quantifiers followed by a string of existential quantifiers. Won't my suggestion have the absurd consequence that these theories are no good?

Well, this consequence wouldn't be absurd if we could show these theories to be logically equivalent to some other theories with less complexity in the scope of existential quantifiers. For the notion of badness I am working with, insofar as it cares about fine-grained distinctions between logically equivalent theories, seems to go beyond any concept of explanatory quality that we antecedently understand. I will be happy if I can get the right results about what it would be reasonable to believe given this or that evidence.

In the case of differential equations, it does seem possible to restate the theories so as to eliminate the problem. Instead of replacing each use of a differential operator with its epsilon-delta analysis, we can regard the derivative operator as a variable bound by a universal quantifier: 'for each operator d on such-and-such space of functions which is a derivative operator according to the epsilon-delta definition, . . . d . . .'. By expressing things this way, we can keep the meat of the differential equations outside the scope of any existential quantifier.

(iv) Fundamental properties

When we state physical theories, we give names to the physically fundamental properties and relations that feature in them: 'mass', 'charge', 'electronhood', etc. But according to Ramsey, Carnap and Lewis, these names are disguised descriptions. When we state the theory, we are really saying nothing more than that *there are* some properties that (uniquely?) play such-and-such structural roles. (The roles need not be specified entirely in observational terms: for Lewis, for example, it will include the specification that the properties and relations that play it are *natural* ones.) So, on this view, even our best theories in physics will be revealed, under analysis, to have the allegedly problematic structure of 'one big existential quantification'.[48]

One response to the objection is simply to deny that theoretical terms are disguised descriptions. Kripke's arguments against the view that proper names are disguised descriptions seem to work just as well against the corresponding view about theoretical terms.

The problem with this response is that it is hard to see how it could matter. Suppose the practice of introducing names for properties and relations had never occurred to us. Instead of introducing names for properties like *electronhood*, we might confined ourselves to describing them as the occupants of some theoretical role. While such a practice might be inconvenient in various ways, it is implausible to think that it would severely diminish our ability to provide good explanations in physics.

If the ability to introduce a name to refer to something instead of denoting it using quantifiers did matter in the way it would have to for this response to work, couldn't those who want to avoid positing some bit of hidden structure use the same strategy to avoid having to rely on theories with

[48]This will also be true under 'structural strategy' for giving nominalistic analyses of predicates like 'electron' discussed in Dorr 2007: §4.ii.

complicated existential quantifications? But clearly there is no explanatory progress to be made by replacing the theory that there is *some* function from objects to numbers that uniquely plays the 'charge' role with the theory that *Clyde* is such a function, where the name 'Clyde' was introduced by the stipulation that it refers to the function that uniquely plays the charge role. The introduction of the name only seems like progress if we think that our ability to name the function derives from its representing some kind of natural, intrinsic structure.

The following response seems more promising to me. Even if we think of theories in fundamental physics as involving existential quantification over natural properties, if these theories are good candidates to be the best explanation of our evidence, they will include clauses which state that the natural properties they talk about are the only natural properties there are. Because of this, the theories will be equivalent to conjunctions with one relatively simple, existentially quantified conjunct, which merely says how many natural properties and relations there are, and one more complicated universally quantified conjunct, which describes the structural roles played any natural properties and relations there might be. In the simplest case, imagine a theory according to which there is only one natural property and one natural binary relation. We could represent it as a single existential quantification:

> $\exists p \exists r$(p is the only natural property $\wedge$ $r$is the only natural binary relation$\wedge T(p,r)$).

But we can also state it as a conjunction:

> $\exists p \exists r$(p is the only natural property $\wedge$ $r$is the only natural binary relation) $\wedge$
>
> $\forall p \forall r$((p is a natural property $\wedge$ r is a natural relation) $\supset T(p,r)$).

In the more general case, the first conjunct may say something like 'there are two natural properties and two natural binary relations'. Then the second conjunct will need to be more complex:

> Whenever p_1 and p_2 are distinct natural properties and r_1 andr_2 are distinct natural relations, either $T(p_1, p_2, r_1, r_2)$, or $T(p_2, p_1, r_1, r_2)$, or $T(p_1, p_2, r_2, r_1)$, or $T(p_2, p_1, r_2, r_1)$.

As the number of properties and relations increases, the number of disjuncts we need will increase factorially. If we are taking length of formulae as a measure of complexity, this will seem very worrying. But it is plausible that in this case symbol-counting fails quite badly as a measure of complexity. Intuitively, a long disjunction whose disjuncts are all and only the formulae generated from some simple combinatorial principle seems far less complex than a much shorter disjunction with miscellaneous, unrelated disjuncts. Consider how long the first-order translations of sentences involving numerical quantifiers quickly get, while remaining intuitively quite simple.[49]

In fact, this example brings out the surprising explanatory power of the idea that big existential quantifications are a distinctive source of theoretical badness. It is a commonplace—a version of Ockham's Razor—that we should not attribute more structure to the world than we require for our explanations. If we find that we only need to posit eight natural properties and three natural relations to explain all our evidence, we should be pretty confident that there are not any additional 'junk' natural properties or relations, marking out joints in nature that play no role in explaining anything we know about. But why should a theory that rules out 'junk structure' be better than a theory that simply leaves the question open? This is not explained by the thought that simpler theories are better. A simple theory doesn't have to entail that the *world* is in any sense simple: it could leave it open how simple the world is. But the idea that big existential quantification are bad provides a neat answer. A theory that rules out junk structure will be logically equivalent to a conjunction in which the existentially quantified conjunct is relatively simple, whereas a theory that leaves it open how much junk structure there is will only be statable as a big existential quantification. Other applications of Ockham's razor to rule out other kinds of 'junk' can be accounted for in a similar way.[50]

[49] Moreover, the best theories will, I think, tend to posit few natural properties of each adicity; they will have lots of symmetries of the kind that would make many of the disjuncts logically redundant; and where the posited natural properties are not related by symmetries, they may be related by simple asymmetries that we can use to restrict the universal quantifier in the second conjunct, doing away with the need for the disjunction of permutations.

[50] Thanks to Antony Eagle, Hartry Field, Anandi Hattiangadi, John Hawthorne, Anna

References

Balaguer, M. 1996: 'Towards a Nominalization of Quantum Mechanics'. *Mind*, 105, pp. 209–226.

—— 1998: *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.

Burgess, J. P. and Rosen, G. 1997: *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford: Oxford University Press.

Chihara, C. 1990: *Constructibility and Mathematical Existence*. Oxford: Oxford University Press.

Dorr, C. 2005: 'What We Disagree About When We Disagree About Ontology'. In Kalderon (2005), pp. 234–286.

—— 2007: 'There Are No Abstract Objects'. In T. Sider, J. Hawthorne, and D. Zimmerman (eds.), *Contemporary Debates in Metaphysics*, Malden, Mass: Wiley-Blackwell, pp. 32–64.

Field, H. 1980: *Science Without Numbers*. Oxford: Blackwell.

—— 1984: 'Can We Dispense With Spacetime?' *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1984, pp. 33–90. Reprinted in Field 1989, pp. 171–226.

—— 1988: 'Realism, Mathematics and Modality'. *Philosophical Topics*, 19, pp. 57–107. Reprinted in Field 1989, pp. 227–281.

—— 1989: *Realism, Mathematics and Modality*. Oxford: Blackwell.

van Fraassen, B. 1970: *An Introduction to the Philosophy of Time and Space*. New York: Columbia University Press.

—— 1980: *The Scientific Image*. Oxford: Oxford University Press.

Hellman, G. 1989: *Mathematics Without Numbers*. Oxford: Oxford University Press.

Hodes, H. 1984: 'On Modal Logics Which Enrich First-order S5'. *Journal of Philosophical Logic*, 13, pp. 423–454.

Huggett, N. 2006: 'The Regularity Account of Relational Spacetime'. *Mind*, 115, pp. 41–73.

Mahtani, Kieran Setiya, Ted Sider, Gabriel Uzquiano, and audiences at Texas, MIT, the Eastern APA, London, and Cardiff.

Kalderon, M. (ed.) 2005: *Fictionalist Approaches to Metaphysics*. Oxford: Oxford University Press.

Kripke, S. 1972: *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, revised edition.

Lewis, D. 1984: 'Putnam's Paradox'. *Australasian Journal of Philosophy*, 62, pp. 221–236. Reprinted in his *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999, pp. 56–57.

—— 1991: *Parts of Classes*. Oxford: Blackwell.

—— 1994: 'Humean Supervenience Debugged'. *Mind*, 103, pp. 473–490.

Maddy, P. 1997: *Naturalism in Mathematics*. Oxford: Oxford University Press.

Melia, J. 1995: 'On What There's Not'. *Analysis*, 55, pp. 223–229.

Putnam, H. 1972: *Philosophy of Logic*. Harper and Row. Reprinted in his *Mathematics, Matter and Method: Philosophical Papers, vol. 1*, pp. 323–357.

Quine, W. v. O. 1948: 'On What There Is'. *Review of Metaphysics*, 2, pp. 21–38.

—— 1976: 'Whither Physical Objects?' In R. Cohen, P. Feyerabend, and M. Wartofsky (eds.), *Essays in memory of Imre Lakatos*, Dordrecht: D. Reidel, pp. 497–504.

Rayo, A. and Yablo, S. 2001: 'Nominalism through De-Nominalisation'. *Noûs*, 35, pp. 74–92.

Sider, T. 2002: *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.

—— MS: 'Writing the Book of the World'.

Sklar, L. 1974: *Space, Time and Spacetime*. Berkeley: University of California Press.

Tegmark, M. 2008: 'The Mathematical Universe'. *Foundations of Physics*, 38, pp. 101–150.

Vaihinger, H. 1924: *The Philosophy of 'As If': a System of the Theoretical, Practical and Religious Fictions of Mankind*. London: Routledge. Translated by C. K. Ogden.

White, R. 2005: 'Explanation as a Guide to Induction'. *Philosopher's Imprint*, 5, pp. 1–29.

—— 2007: 'Does Origins of Life Research Rest on a Mistake?' *Noûs*, 41, pp. 453–477.

Williamson, T. 2002: 'Necessary Existents'. In A. O'Hear (ed.), *Logic, Thought and Language*, Cambridge: Cambridge University Press, Royal Institute of Philosophy Supplement: 51, pp. 233–251.

—— 2003: 'Everything'. In J. Hawthorne and D. Zimmerman (eds.), *Philosophical Perspectives 17: Language and Philosophical Linguistics*, Oxford: Blackwell, pp. 415–465.

Woodhouse, N. M. J. 2003: *Special Relativity*. London: Springer.

Wright, C. 1983: *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

Yablo, S. 2001: 'Go Figure: A Path Through Fictionalism'. In P. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy XXV: Figurative Language*, Oxford: Blackwell, pp. 72–102.

—— 2005: 'The Myth of the Seven'. In Kalderon (2005), pp. 88–115.

Zermelo, E. 1930: 'Über Grenzzahlen und Mengenbereiche: neue Untersuchungen über die Grundlagen der Mengenlehre'. *Fundamenta Mathematicae*, 16, pp. 29–47.