

Of Numbers and Electrons

Cian Dorr

Draft of January 18, 2010

No comment too big or too small

1 The indispensability argument

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 the analogy as unhelpful because they think that we can know that there are mathematical objects 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 see* that there are sets—e.g. sets of dogs standing in the street in front of us—just as we can directly see that there are red things; no inductive inference of any sort is required. Still others will regard the analogy as unhelpful because they

¹The tradition in question thinks of itself as upholding “the indispensability argument for the existence of mathematical entities”. Since I don’t want to get hung up on the question what the key term ‘indispensable’ could sensibly mean in this context, I will not be discussing this argument as such. I hope that the ideas that lie behind it will nevertheless get a fair hearing.

²See, e.g., Wright 1983.

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—it is very hard to say anything substantive about the distinctive standards that apply to inductive reasoning without doing this!—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 assume that our evidence does not deductively entail that there are mathematical entities.

Having internalised this supposition, 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 is a realm of entities as drastically unlike those objects as numbers or sets would have to be is eyebrow-raisingly audacious, as inductive inferences go. The hypothesis that there just aren’t any such entities certainly doesn’t *feel* very like the sceptical hypotheses that we get used to dismissing. Nor, for that matter, does it feel at all like the hypothesis that there are no electrons—there is nothing in the history of experimental sci-

³See, e.g., Maddy 1997.

ence that stands 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 and praiseworthy 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 P becomes the claim that some theory that entails P is better *qua* explanation of E than any of those that do not, by a wide enough margin. 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 (i.e. non-platonistic) theory that is as good as, or at least not much worse than, all of these platonistic theories, *qua* explanation of our evidence. The challenge, then, is to actually produce such a theory, or at least to give enough of a sketch 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 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 re-

place.⁴ 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.⁵

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, carrying out the programme for theories like general relativity and quantum mechanics poses distinctive challenges which have yet to be tackled convincingly.⁶ It is too early to say whether we can nominalistic theories in these domains which do as good a job as Field's Newtonian theory at providing the intuitive sense of explanatory satisfaction and avoiding anything which might strike someone as a cheap or cheesy trick. And even if everything goes as well as could be hoped in physics, it is a deep and 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 genuine epistemic advantages, if any, all this honest toil has over theft. There are mechanical

⁴This claim is true only for the version of Field's theory that uses something like second-order or plural 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 and §III.B.1.b. However, the claims about the physical world that follow from the platonistic theory and don't follow from the first-order version of Field's theory are quite esoteric; certainly none of them are part of our evidence. So even if we denied the legitimacy of second order or plural logic, it would be hard to see how the additional concrete consequences of the platonistic theory would do anything to improve its claim to be well-supported by our evidence: if anything, the reverse seems more likely to be true.

⁵On the role of explanatory satisfaction in giving nontrivial content to IBE, see White 2005.

⁶In the case of quantum mechanics, the process of constructing anything as rigorous as a Field-style theory will inevitably incorporate some particular approach to the measurement problem. For this reason I am not convinced by the treatment in Balaguer 1996, which assumes a propensity-based "solution" to the measurement problem that I find quite unsatisfactory.

methods which take arbitrary platonistic theories as input and output replacement theories which at least seem to be nominalistic and empirically equivalent. The outputs of these algorithms will certainly feel like cheats to those who have internalised the constraints that make Field’s programme so challenging. 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 of 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 strategies. 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 .

In other words, T is *nominalistically adequate*: it is consistent with the totality of truths about the concrete realm, or equivalently those of its consequences which are entirely about the concrete realm are true. ‘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)—it is *as if* T were true, as far as the concrete realm is concerned.⁷

⁷Strictly speaking, the ‘as if’ claim appears to be a counterfactual conditional: if T were true, the concrete realm would be just as it in fact is. Formally, this does not follow from T^\diamond : ‘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 are *ipso facto* closer to the actual world than any other T -worlds.

The second method requires us 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 bit of artificiality we can take M to be something like real analysis or ZFCU (= ZFC + ‘there is a set of all non-sets’). 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 the contemporary fictionalists like Yablo (2001, 2005), something roughly like T^\square is 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 .⁸

Thinking of T^\diamond and T^\square in terms of possible worlds can help to elucidate their content, although of course people who don’t believe in abstract entities will not want to take these glosses fully seriously. In these terms, T^\diamond is true at a possible world w iff some T -world is *concretely indiscernible* from w , and T^\square is true at w iff every M -world that is concretely indiscernible from w is a T -world.⁹ This is represented in

⁸A closely related approach 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 . This is formally weaker than T^\square —in possible-worlds terms, even if not all M -and- C -worlds are T -worlds, the closest ones could be—and arguably does a better job than T^\square at capturing the intuitive content of the ‘according to the fiction’ claim. However, the counterfactual formulations introduce new complications, and it’s hard in practice to think of any case where the alleged logical differences between counterfactuals and strict conditionals could matter.

⁹We 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^\square would, like ‘actually P ’ be necessary if true. Rather, ‘in fact’ works like the ‘backspace’ operator in Hodes 1984—it “undoes” just one modal operator.

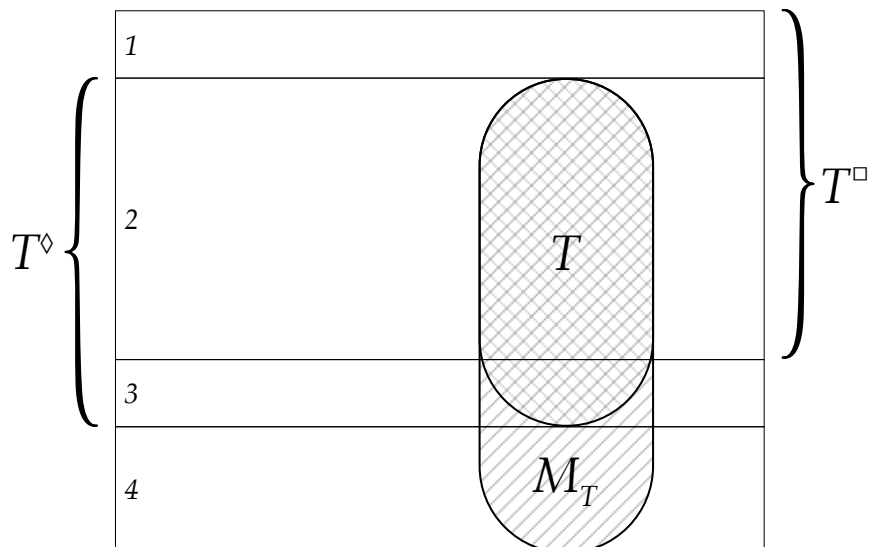


Figure 1. Possible-worlds representations of T^\diamond and T^\square . Points on the same horizontal line represent concretely indiscernible worlds.

figure 1. As the figure makes clear, there is no formal guarantee that T^\diamond and T^\square are true in the same worlds—for this to be the case, the sets of worlds labelled ‘1’ and ‘3’ would need to be empty. But as we will see in section 3, there is reason in many of the central cases of interest to think that these sets *are* empty, so that T^\square and T^\diamond are 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) weaker than, i.e. deductively entailed by, T ;
- (ii) *empirically equivalent* to T , at least in the minimal sense that it entails every proposition E that is both in fact part of our evidence and deductively entailed by T ; and
- (iii) *nominalistic*, i.e. does not deductively entail that there are mathematical entities;
- (iv) Explanatorily better than, or at least only slightly 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^\square yield good theories as outputs, when given good theories as inputs. I will take up this question in section 4. But first, I should say at least something about how T^\diamond and T^\square fare with respect to the first three desiderata. Section 2 will sketchily address these questions for T^\diamond ;

section 3 will address them for T^\square . These sections will also introduce some more sophisticated variants of T^\diamond and T^\square that one might turn to if one thought that T^\diamond and T^\square themselves failed in some serious way to satisfy the desiderata. Those who want to get straight to the epistemology should feel free to skip ahead to section 4.

2 The adequacy of T^\diamond

(i) That T deductively entails T^\diamond seems obvious: the inference from ϕ to ‘Possibly ϕ ’ seems manifestly deductively valid, as does the inference from anything to ‘the concrete realm is just as it in fact is’.¹⁰

(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 . And since T deductively 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 or all of the properties of concrete objects that feature in our evidence (redness, squareness, etc.) in fact consist in certain relations holding between concrete objects and non-concrete ones, and thus 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 the real number that is the ratio of their lengths to be close to 1,

¹⁰One might, I suppose, regard it a consistent hypothesis there is something radically defective about modal concepts; but even on the hypothesis of such defectiveness, it seems better to regard ‘Possibly’ as inert (so that ‘Possibly ϕ ’ is equivalent to ϕ) than to reject all possibility-claims across the board.

so that it is metaphysically impossible for any objects to be roughly equally long in a world without numbers. 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).

One would need a fairly expansive conception of the scope of “deductive entailment” to regard such hypotheses as deductively inconsistent. 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’s a very serious limitation. The auxiliary hypothesis is quite plausible; and as far as I can see, none of the standard platonistic scientific theories we have reason to take seriously derives any explanatory power from being consistent with its negation. The only exceptions are the hyper-Pythagorean theories, which achieve a somewhat attractive economy by entailing that there is nothing more to reality than the mathematical realm; this vision will eventually have to be compared on other grounds with the more familiar kinds of economical world-views available to the nominalist.

(iii) The most important reason to doubt that T^\diamond is nominalistically acceptable 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, and we interpret the modal operator in T^\diamond as expressing metaphysical possibility, then T^\diamond does after all deductively entail the existence of mathematical entities, and is thus useless to the nominalist.

Why believe the necessity thesis? One reason to believe it would be a commitment to the stronger claim that the existence of mathematical entities is itself a priori (i.e. deductively entailed by everything). But we have agreed to set this view aside for the purposes of evaluating the force of the inductive argument for their existence. Apart from this, the most promising argument I know of is one that parallels Kripke’s argument, in the appendix to *Naming and Necessity*, for the claim that if there are no unicorns, it is metaphysically impossible for there to be unicorns. But while Kripke’s argument has been influential, it is not all

that strong. It is not clear why a failed attempt to introduce a word as a natural kind term should confer on it an empty intension rather than, say, 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’. Thus, it seems to me that the necessity thesis would be a weak reed indeed upon which to rest the case against nominalism.

Even if we accept the necessity thesis, we may hope to find a nominalistically acceptable reading of T^\diamond by reading ‘possibly’ as expressing something weaker than metaphysical possibility. What other modalities could we appeal to? One idea is to invoke some kind of conceptual possibility, perhaps something 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 in terms of purely physical predicates (like ‘equal in length’) and narrowly mathematical ones (like ‘is a member of’). For example

(*) $r =$ 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

$$f(a)/f(b) = r.^{11}$$

It is quite plausible that something like (*) is *metaphysically* necessary. But it is not so clear that there is any conceptual necessity in the vicinity. The case would be even less clear if we were dealing with some less familiar notions than that of length, e.g. charge-density. If these theoretical definitions are not conceptually necessary, and the original theory T makes no explicit use of “purely physical” predicates

¹¹This is similar to the definition of ‘The distance from x to y is r ’ endorsed by Putnam (1972: 340).

like ‘equal in length’, then it will be much too easy for the conceptual reading of T^\diamond to be true: since 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, it’s hard to see this lack of specificity as bringing any distinctive explanatory benefits. *Sometimes* a less specific theory, e.g. one describing some general dynamical law, is explanatorily better than all of its extensions, e.g. extensions that add various piecemeal facts about the positions and velocities of particular particles. But leaving it open how one’s mixed predicates relate to purely physical and narrowly mathematical ones doesn’t seem to achieve anything like this kind of explanatorily powerful generality. 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^\square . 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: we can then use this class of propositions in specifying interpretations of T^\diamond and T^\square on which they are neither too weak nor too strong to be useful to the nominalist.¹²

¹²Some 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.

But suppose that we both accept the necessity thesis and reject modalities more fine-grained than metaphysical necessity and possibility. Then, while T^\diamond is no use as it stands, we may still be able to modify it so as to achieve the desired effect. One strategy for doing this is to substitute some nominalistically unproblematic predicates for the problematic ones (those whose meaning allows some of their argument places to be filled only by mathematical entities). The idea is to replace T in T^\diamond with an isomorphic theory T^* , in which all the work that was done in T by problematic predicates is done by unproblematic ones, e.g. predicates applicable to concrete inscriptions (as in Chihara 1990) or predicates applicable to angels. 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 an angel 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 angel that contemplates them in some distinctive way. Strictly speaking, T^* already looks to be a nominalistically acceptable theory; but it is probably a very implausible one, that deductively entails the existence of an infinity of angels or inscriptions or whatever. But if one then embeds T^* in some appropriately restricted possibility operator, claiming it to be consistent with the facts entirely about the entities we really care about (or with the facts entirely about the entities that in fact exist), the result will be much weaker; indeed, it will at least be arguable that we know enough a priori about metaphysical possibility for the claim that T^* is possible in the relevant restricted sense to be deductively entailed by T .

Another strategy avoids the arbitrary choice of substitute predicates by using higher-order logic to “Ramsify out” all the problematic predicates in T . Let the problematic predicates be F_1, \dots, F_n ; let

Metaphysical necessity is truth in virtue of *everything*, so that when X are some but not all of the things there are, ‘It is true in virtue of the nature of X that . . .’ is an operator with the logic of a necessity operator that is stronger than metaphysical necessity. While Fine’s conception of the entities in virtue of whose nature typical metaphysical necessities hold is anything but acceptable to a nominalist, the general picture of a rich space of intelligible operators corresponding to strengthenings of metaphysical necessity or weakenings of metaphysical possibilities is one that nominalists could embrace.

$T(X_1, \dots, X_n)$ be the result of substituting appropriate higher-order variables $X_1 \dots X_n$ for these predicates; and let T_*^\diamond be the result of first existentially quantifying these new variables and then applying an appropriate restricted possibility operator:

(T_*^\diamond) Possibly, the concrete realm is just as it in fact is and
 $\exists X_1 \dots \exists X_n T(X_1, \dots, X_n)$.

Assuming the necessity thesis, for T_*^\diamond to be nominalistically acceptable we must be able to interpret the second-order existential quantifications here in such a way that they do not require there to be mathematical entities. Given our assumptions, it is sufficient to do this if we can understand the second-order notation in such a way that the theorems of standard systems of second-order logic are in fact deductively valid—e.g. $\exists y(y = y)$ must deductively entail $\exists X \exists y(Xy)$. For *ex hypothesi*, $\exists y(y = y)$ does not deductively entail that there are any sets. Some will be happy to grant that we can directly learn to understand the notation in such a way that the standard logic is deductively valid; others will hold out for a translation of the notation into ordinary English, e.g. one in terms of parthood and plurals, as in Lewis 1991. (Again, $\exists X_1 \dots \exists X_n T(X_1, \dots, X_n)$ already looks to be a nominalistic theory, strictly speaking; but it entails the existence of a huge infinite number of entities of some unspecified sort, whereas T_*^\diamond seems to avoid this implication as well as the implication that there are mathematical entities.)¹³

Like the strategy that invokes conceptual possibility, these strategies only have the desired effects when T is a “rich” theory that contains theoretical definitions like (*), relating all its mixed mathematico-physical vocabulary to purely concrete and narrowly mathematical vocabulary. Otherwise, too much of the content of T will get “Ram-

¹³However, if—as Williamson (2002) maintains—the right logic for all modal operators is a “constant-domain” logic on which it’s impossible for there to be more or fewer things than there in fact are, this apparent benefit of T_*^\diamond is lost: T_*^\diamond also requires a huge infinite supply of entities of unspecified sort, and indeed would seem to be equivalent to the bare $\exists X_1 \dots \exists X_n T(X_1, \dots, X_n)$. More generally, if constant-domain modal logic is correct, the modal strategies with which this paper is concerned lose much of their interest, since they will work only given strong assumptions about the infinity of the universe.

sified away”, and T_*^\diamond will entail little or nothing about the concrete realm (perhaps only something about its cardinality).

3 The adequacy of T^\square

(i) There is no narrowly logical reason to think that T must deductively entail T^\square . 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 -worlds that are not T -worlds. 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 concrete realm—i.e. T^\diamond —deductively entail T .

I think it’s fairly 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 three possibilities. One 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 that feature in M (e.g. set-membership). 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 such non-supervenient mixed relations, so that “definitions” like (*) are either contingently true, or necessarily true but only because facts about which objects are equally long are not really entirely about the concrete realm. If the negations of such “Heavy Duty Platonist” hypotheses are not a priori, and are not deductively entailed by T , then T does not deductively entail T^\square . For all we can tell a priori, T may be true at the actual world while being false at certain concretely indiscernible M -worlds where the non-supervenient mixed relations don’t behave in the right way. But I can’t see how we could lose any explanatory power if we confine 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 together with the narrowly mathematical facts. (T may go on, like the “rich” theories of section 2, to explicitly state the theoretical

definitions that ensure this supervenience, or it may leave this open.)

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 at w' , 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 w' there is no such bijection. This sort of thing may 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 there are concretely indiscernible worlds where first-order ZFCU is true and in which the cardinality of the set of concrete objects is different. But this is one of the places where we seem to manifest an understanding of questions of cardinality that transcends first-order logic. Whether or not there are numbers, we can make sense of different hypotheses about how many concrete objects there are. For example, we can make sense of the hypothesis that there are at most \aleph_1 concrete objects.¹⁴ There is a fact of the matter about how many concrete entities there are: 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 truth about how many concreta there are. 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

¹⁴As follows: whenever there are infinitely many of the x s and infinitely many of the y s, and all of the x s and y s are concrete, either there are as many x s as concrete objects, or there are as many y s as concrete objects, or there are as many x s as there are y s.

this kind of difference between w and w' . For according to a theorem of Zermelo, second-order ZFC is “almost” categorical: for any two of its models, one is isomorphic to an initial portion 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. One does not need to be super-serious about the possible-worlds semantics for modal claims to think that these theorems justify the claim that two concretely indiscernible worlds at both of which the second-order or plural version of ZFCU is true can differ set-theoretically only as regards the height of the hierarchy.

(ii) As with T^\diamond , T^\square 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 entails these claims. But even the claims are all a priori, it is still not obvious that T^\square deductively entails E whenever T does. The problem comes from the formal possibility that T^\square is vacuously true because no M -world is concretely indiscernible from the actual world—that the actual world is somewhere in region 1 of figure 1. If this scenario cannot be ruled out a priori, T^\square will not satisfy desideratum (ii), since the fact that E follows from T provides no reason to think that E holds throughout region 1.

But perhaps this scenario *can* be ruled out a priori. If we are not moved by the considerations that support the necessity thesis, and don’t mind a fairly expansive view of the scope of deductive rationality, there is some appeal to the idea that it is a priori that it *is* possible not only for there to be mathematical entities, but for them to conform to our standard theories of them, and for them to do so while the facts entirely about the concrete world remain the same. The claim, in other words, is that M^\diamond is a priori:

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

If this is a priori, then obviously T^\square deductively entails T^\diamond ; so if T^\diamond deductively entails E , T^\square does too.

If one is less optimistic than this about our ability to have a priori knowledge of metaphysical possibilities, there are some possible

fallback routes, corresponding to those we considered in discussing the necessity thesis. One is to replace the metaphysical necessity in T^\square with some stronger notion of necessity, so that we only need a version of M^\diamond using a correspondingly weaker notion of possibility. Another is to eliminate specifically mathematical vocabulary, e.g. by using higher-order quantification to formulate a theory like T_*^\square :

(T_*^\square) Necessarily, if the concrete realm is just as it in fact is, then $\forall X_1 \dots \forall X_n$ (if $M(X_1, \dots, X_n)$ then $T(X_1, \dots, X_n)$).¹⁵

For T_*^\square to be empirically equivalent to T , T will need to be “rich” in the sense of section 2, and M_*^\diamond will need to be a priori:

(M_*^\diamond) Possibly, the concrete realm is just as it in fact is, and $\exists X_1 \dots \exists X_n M(X_1, \dots, X_n)$.

The case for the apriority of claims like M_*^\diamond , 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 quite strong—considerably stronger than the corresponding case for M^\diamond .

However, even this requires a somewhat expansive conception of the realm of deductive rationality. The epistemology even of narrowly logical possibility is quite mysterious—at least part of the story about why it’s OK for us to be fairly confident that, say, ZFCU is consistent involves our empirically derived knowledge that so far no-one has succeeded in deriving a contradiction from it. If one excludes claims like M^\diamond or M_*^\diamond from the realm of the a priori on these kinds of grounds, then one will need to add M^\diamond or M_*^\diamond as a further auxiliary hypothesis to T^\square or T_*^\square in order to get something empirically equivalent to T . Section 7 below will consider whether this might make a significant difference to the explanatory goodness of the total package.

(iii) Once one grants the assumption that *some* theories do not deductively entail that there are mathematical entities, I can’t think of any reason to doubt that T^\square -style theories are among them. Even if the necessity thesis were true, the result would be that T^\square is (vacuously)

¹⁵Hellman’s reconstructive project (Hellman 1989) essentially amounts to replacing each T with something like T_*^\square .

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).

4 The crucial analogy

Let's suppose, then, where T is the theory that, by the anti-nominalist's lights, best explains our total evidence, T^\diamond and T^\square are nominalistic, deductively entailed by T , and empirically equivalent to it. If so, the anti-nominalist needs an argument that T^\diamond and T^\square are worse—less plausible as stopping-places for explanation—than 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 reasoning requires in a given case, we should not just sit around introspecting; instead, we should be guided by examples of good inductive reasoning in scientific practice.

There are more and less direct ways to apply this kind of “naturalistic” methodology in assessing the epistemic credentials of nominalistic replacements for standard scientific theories. The direct way is to look at what actual scientists say about the particular theories we are interested in. Burgess and Rosen (1997) have recently championed this sort of 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 to conclude that the theories in question are worse, epistemically speaking, than the originals, and that so far, the original theories' claim on our credence remains undiluted.

There are several reasons one might have for being dissatisfied 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. 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 just the same with many debates in the philosophy of physics. Philosophers who aspire to learn from physicists quickly realise that *most physicists just don't care* about the theoretical differences that seem so important to us; unless the lesson we draw 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 take a professional interest.

Second, even if we do learn from scientists that we have good reason to believe in numbers, it is not clear how this bears on my question, which is whether we have good reason *of an inductive, topic-neutral sort* to believe in numbers. The way scientists actually reason about mathematics seems strikingly different from the way they reason about electrons and so forth. One natural moral for “naturalists” to draw from this is that we have *deductive* reason to believe in numbers, or at least 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.¹⁶

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*.¹⁷ If we understand this question, no amount of direct deference to scientists will help us resolve it.¹⁸ Nev-

¹⁶Maddy (1997) draws the latter conclusion.

¹⁷See 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 think they don't.

¹⁸Not that Burgess and Rosen think otherwise—it is clear from their recent paper (2005) that if they had any objection to the view that there are no numbers in the

ertheless, 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 behooves us to try to investigate what good inductive reasoning has 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, though developments in quantum gravity have made the case less strong than it used to be, there is still some reason to think that among the things that there are, fundamentally speaking, are spacetime points or regions. The epistemic situation *might* be the same for numbers or sets. But if we want to see whether it is, we're not going to get anywhere by trying to convince scientists to take a professional interest in the question. We will have to proceed indirectly, by trying to extract from scientific practice generalisations about what makes for theoretical goodness, and applying them to the questions we are interested in.

What we are looking for is an argument by analogy, of the form 'X is an explanatorily bad theory; T^\diamond/T^\square is similar in relevant respects to X; therefore T^\diamond/T^\square is also a bad theory'. In the case of T^\diamond , a rather powerful argument of this form is implicit in much of the literature on "the indispensability argument", for example Putnam's discussion of "fictionalism" in Putnam 1972; Field (1988: 260–61) articulates it especially clearly. Let me state the argument in my own way.

We can and do have excellent empirical reason to believe that there are lots of things much too small to be observed by anyone. This is so despite the fact—much dwelt on by scientific anti-realists like van Fraassen (1980), according to whom we have no such reason—that for any theory T postulating unobservable entities, we can easily find a theory which shares all T 's consequences about the observable world without entailing anything about unobservables—viz., the claim that T has only true consequences for the observable portion of the world, or equivalently:

(T^\diamond) Possibly, all observable matters are just as they are, and T .

Since our evidence gives us good inductive reason to believe claims

most fundamental sense, it would be on the grounds that this claim is unintelligible.

about unobservable matters that do not follow from any theory of the form of T^\diamond , these must be bad theories. While we have reason to believe many such theories— T^\diamond is after all deductively entailed by T —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^\diamond (and its precisifications)? No. Even when we replace the notion of observability with something precise and reasonably simple, we still end up with bad theories:

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 within ϵ of its actual value, and T .

Since (given scientific realism) 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, whatever the badness-making feature of T^\diamond might be, it must be a feature these theories share. In the light of such examples, it is natural to conclude that the badness-making feature is the distinctive *logical structure* all these theories share. But this structure is also shared by T^\diamond . So we can conclude that T^\diamond is also a bad theory.

What could 'bad' mean, for this conclusion to be a reasonable one? One thing it certainly can't mean is 'a priori implausible': since T^\diamond is a logical consequence of T , it must be at least a priori plausible as T . What is a priori unlikely is not that T^\diamond 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 implausible 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 many useful entailments between claims about theoretical goodness and claims about a priori plausibility.¹⁹

The argument that since T^\blacklozenge and T^\diamond 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^\diamond are dissimilar in lots of ways. For example, the objects “modalised away” by T^\diamond 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^\diamond ? 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^\diamond , 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

¹⁹This 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 a priori 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, so that the probability that T is explained given that it is true is not high. It would not be plausible to think that the extent to which a theory cries out for explanation is determined by facts about its logico-syntactic structure. Roger White (2007) considers theories that are enormous conjunctions of thousands of sentences, each of which specifies the position and size of a pebble on a certain beach. Even though these theories are all syntactically very similar, some—e.g. 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 are relatively good theories—e.g. that the pebbles were deliberately arranged by someone aiming to make a picture of a face—which raise the probability of the the conjunction that describes a face-like arrangement, whereas there are no comparably good theories which raise the probability of the conjunction describing the random jumble.

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^\diamond . We could, for instance, conclude that the use of modal operators 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^\diamond -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 with 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^\diamond ? To do so, we would have to find some theory of the form

(T^\square) 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^\square 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 that is not observationally indiscernible from any T -world is observationally indiscernible from some BLAH-world.

One proposition with the desired properties is the material conditional $T^\blacklozenge \supset T$:

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

T^\blacksquare is clearly 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 . Given this, we have the same reason to think that T^\blacksquare is a bad theory that we have for T^\blacklozenge . However, the structural parallel between T^\blacksquare and T^\square is much weaker than that between T^\blacklozenge and T^\diamond —too weak for the analogy to carry much force. The mathematical theory M that features in T^\square is a simple piece of theory—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 embeds T^\blacklozenge , which we already have reason to think of as a distinctively bad theory; 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* fragment of T to substitute for BLAH in T^\square —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, then the claim that the facts about subatomic particles are determined in the relevant way might be able to play the role of BLAH. (Though even then there would be worries if the fragment all by itself ruled out certain metaphysically possible patterns of facts about atoms.) But in the cases where we most clearly have good 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 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.²⁰ But even this theory perfectly fit our evidence, there would be little pressure to believe in the gravitational field as a piece of additional fundamental structure. Given that the theory has a simple fragment that determines it as a function of the distribution of masses, the proposed additional structure seems redundant: a view that uses this 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.²¹

I conclude that the prospects for establishing the badness of T^\square using an argument parallel to our argument by analogy for the badness of T^\diamond are poor. But wait: do we need a new argument at all? In the central cases where theories like T^\square are live options, section 3 argued that they are deductively equivalent (perhaps modulo some auxiliary hypotheses) to the corresponding T^\diamond . Doesn’t that mean that T^\diamond and T^\square are really the same theory, or at least 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 has a T^\square -style logical form in addition to its T^\diamond -style logical form, and T^\blacklozenge admits of no logical form that is relevantly analogous to that of T^\square , that might well turn out to be an epistemically relevant difference between T^\diamond and T^\blacklozenge . (Indeed, on this way of individuating theories, just about every theory *can* be expressed using a T^\diamond -like logical form—if T talks 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

²⁰Field’s approach to nominalising of Newtonian gravity depends essentially on taking this attitude to the field.

²¹In electromagnetism, by contrast, the possibility of source-free radiation prevents any analogous “reading off” of the facts about the electromagnetic field.

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 T^\square -style theories involve deriving some very sweeping general principle from examples like T^\diamond , e.g. to the effect that “parasitic” theories are bad. This already looks pretty tendentious. The current section will introduce some new data which will both cast further doubt on this argument, and suggest that the contrast between parasitic theories which use possibility operators (like T^\diamond) and those which use necessity operators (like T^\square) 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 still 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 all the same consequences as T concerning the shapes, size, 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 the shapes, sizes, motions and masses of material bodies are just as they in fact are, and T ’. But provided T already contains enough mathematics, there is no need to use a modal operator in formulating the desired weakening of T . We can proceed as follows: (i) Formulate T in such a way that all talk of charge is accomplished by a single type of expression, say ‘ n is the electric charge of x ’. (ii) Replace each occurrence of the formula ‘ n is the electric charge of x ’ in T with ‘ $f(x) = n$ ’, where f is some new

variable ranging over functions from bodies to real numbers; call the result of this $T(f)$. (iii) Let our new theory T^\exists be $\exists fT(f)$. Since the existence of theories like T^\exists *doesn’t* undermine our reason to believe that charge is just as real and intrinsic as shape, size, motion and mass, we can conclude that T^\exists must be a much worse theory than T .

Although T^\exists does not explicitly mention electrical charge, if it is true, there are ways to interpret ‘charge’ on which it is true to say that objects have charges. We could, for example, interpret ‘ n is the charge of x ’ as meaning ‘ $f(x) = n$ for some function f such that $T(f)$ ’, or ‘ $f(x) = n$ for the unique function f such that $T(f)$ is true’. Or, if we wanted ‘If objects have charges, then T ’ to come out expressing a contingent truth, we could get a bit fancier, interpreting ‘ n is the charge of x ’ as meaning something like ‘ $f(n) = x$ for the function f that plays the simplest sufficiently “charge-like” role with respect to the facts about shape, size, motion, etc.’²² On these interpretations, T may even turn out to be deductively entailed by T^\exists . But these interpretations of T are very different from the intended interpretation of T , as we were imagining it. On the intended interpretation, the facts about charge were supposed to be intrinsic, 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 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 would after all entail that objects with different charge distributions are never duplicates. But this way lies Putnam’s paradox. 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. And predicates like ‘duplicate’ seem especially good place to put a stop to charity run amok: while it is not entirely clear that T^\exists is consistent with the hypothesis that nothing is charged, it is

²²In working out the details of such an interpretation, we might take inspiration from Best System analyses of lawhood as in Lewis 1994.

as clear as anything that it is consistent with the hypothesis that all objects alike in respect of shape, etc., are duplicates.

So T^3 really is weaker than the original T , and weaker than what our evidence (we are supposing) makes it reasonable to believe; our conclusion that it is a bad theory stands. And 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 very often do so by replacing the vocabulary which purports to characterise this structure with variables of an appropriate sort, bound by initial existential quantifiers. People who are suspicious of particular bits of putative 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 x s; and there are some particles, the y s, and. . . and the x s repel one another and attract the y s, and. . .’. 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 particles instantiate in virtue of their motions with respect to other particles.²³
- (ii) Ordinary physical theories formulated in co-ordinate terms are naturally understood as entailing that spacetime has a geometric structure much richer than that of mere topology: regions of spacetime can differ in all sorts of intrinsic geomet-

²³Depending on the form taken by the original theory, we might or might not end up having to say that it is sometimes indefinite what kind a given particle belongs to.

ric respects even if 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* on which the given dynamical equations come out true (and which respects the intrinsic, topological structure of the space). This minimalistic way of thinking about the content of ordinary theories is favoured by van Fraassen (1970). And under the influence of a discredited positivist philosophy that rejects the ideology 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.²⁴ But I at least have no interest in deferring to the opinions of physicists when these are manifestly shaped by past anti-realist philosophies. We *do* have good reason to posit a spacetime whose intrinsic geometric structure goes far beyond topology; and 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 decisive problem for “Leibnizian” relationalism, which claims 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 interparticle distances. (Huggett goes on to suggest a “best-system” style analysis of claims about inertial motion, absolute acceleration, etc.: roughly, to be moving in-

²⁴See, e.g., Woodhouse 2003.

ertially is to be moving inertially according to every “best” co-ordinate system, where “best” is understood in such a way that if there are some 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 quantifying out the vocabulary that purports to characterise the velocity field, or of the partition of spacetime into trajectories of matter-points. And provided that this existentially quantified theory is true, we will be able to give some extrinsic, best-system-style analysis of expressions like ‘trajectory of a matter-point’ or ‘velocity field vector’ that make the original theory 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^\diamond , if we want to weaken any theory so as to eliminate the implication that there are unobservable objects; we can use existential quantification over models to achieve the same effects. Our new

theory will simply say that there is some model of the old theory that accurately represents the observable facts (or the centres of mass of atoms, or whatever.)

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^\diamond . Indeed, the work done by the existential quantifier 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 interparticle 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^\diamond and T^\diamond 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 it turned out that the true canons of theoretical virtue 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^\diamond is due to some feature it shares with T^\square —e.g. 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.²⁵

²⁵Perhaps there is a *historical* sense in which these our existentially-quantified theories are “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

If the canons of theoretical virtue care about the logical parallel between possibility-operators and existential quantifiers, as they seem to, it stands to reason that they should also care about 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, and between T^\square and T^\diamond , 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 place which all bodies tend to move towards (in such-and-such specified ways). It seems to me that if we found out that there was such a place, 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 place towards which bodies move 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 place.

So things seem to be going well for T^\square . We could further bolster the case for its theoretical goodness if we found some class of universally-quantified theories that stand to T^\square as our existentially-quantified theories stand to T^\diamond , and those theories turned out to be theoretically good. What could these theories be like? The structure we are looking for is fairly distinctive: just as it is only in special cases that the result of embedding a theory within a restricted necessity operator is weaker than the original, likewise it is only in special cases that the result

predicates. But I find it implausible that such merely historical features of theories should matter when we're considering *ideal* inductive reasoning, as opposed to heuristics and rules of thumb. In any case, the history of science is full of cases where a good theory (e.g. special relativity) was arrived at by weakening some worse, stronger theory (e.g. Lorentzian mechanics)—the history of such episodes shows that a general suspicion of such theories is not much good even as a heuristic.

of replacing one of a theory's primitive expressions with a variable bound by a restricted universal quantifier is weaker than the original.

It seems to me that ordinary physical theories stated in co-ordinate terms fit the bill. When physicists state equations about the rates of change of physically interesting quantities with respect to the x , y , z and t co-ordinates, they normally 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 a universally quantified claim, 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. 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. The universally quantified theory that physicists take seriously is thus a consequence of the theory positing a privileged co-ordinate system. The transition from the latter to the former stands to the transition from T to T^\square as the transitions using existential quantifiers stand to the transition from T to T^\diamond .²⁶

I don't suppose anyone has ever believed in the "one true x -axis". But the history of physics does provide real 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

²⁶Standard theories in physics don't normally directly address the question *what it is* for a co-ordinate system to be "admissible": rather, they merely place constraints on what this could amount to, by characterising some 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 such theories to be talking about a primitive, non-supervenient relation between physical and mathematical objects. If we are looking for nominalistic reconstructions, we will of course not want to take this view.

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 such-and-such equations hold in every inertial frame, i.e. in every co-ordinate system that fits the intrinsic geometry of the spacetime in a certain 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, e.g. 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 a lot of the time one does not do this; in these cases, talk about a gauge-symmetric field like “the electromagnetic potential” (understood as a vector field) is tacitly understood as governed by an initial universal quantifier over gauges, i.e. over vector fields that fit in the right way with the intrinsic structure that underlies the gauge field, whatever that might be.)

Now, in all these examples, it has turned out that the formulation of the theory as a universal quantification over co-ordinate systems or gauge fields is not obligatory. 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 a primitive 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 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 enthusiastic advocates 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 “weakening by embedding within a necessity operator” is like the operation “weakening by universally quantifying out” in *not* making for theoretical badness.

To fix ideas, 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 *in positive contexts*, and within formulae governed by universal quantifiers and necessity operators *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 complex 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 The consistency of M with the facts about the concrete world

As noted in section 3, the claim that T^\square 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 presumably help ourselves to it for free in deriving empirical consequences from T^\square , 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 think it is quite plausible that we can know a priori in this domain there is no relevant gap between logical and metaphysical possibility, so that if we could know a priori that M is *logically conservative* (logically consistent with any consistent theory entirely about the concrete), we could deduce M^\diamond . But even if this is right, and even if the logical conservativeness 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 ZFC, for example, is clearly epistemologically problematic in a way that theorems of predicate logic, even complex ones, are not. Indeed, if we are justified in believing ZFC to be consistent at all, at least 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 ZFC 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. But 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 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 possible lines of response to this objection seem plausible to me. One is to concede that M^\diamond does need to be included along with T^\square in the total package of theory for the purposes of IBE, but to insist that,

despite sharing a logical form with T^\diamond and T^\blacklozenge , M^\diamond is nevertheless much better than them, in virtue of M ’s being a much simpler theory than the total physical theory T to which T^\diamond and T^\blacklozenge apply their possibility operators. The contrast in simplicity is striking: 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 this matters: the thought 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. For a theory that begins ‘For every admissible co-ordinate system. . .’ to have any empirically interesting consequences, it will need to be combined with something that entails that at least one co-ordinate system is “admissible” in the relevant sense. In some especially nice cases, like Euclidean geometry, we can find geometric axioms which entail this and aren’t at all syntactically analogous to our paradigmatically bad existentially quantified theories. But in other cases, we are still left with a residual existential quantification over co-ordinate system, albeit one that is much simpler than the claim that embeds the whole theory within an existential quantification over co-ordinate systems. In standard treatments of differential geometry, for example, it is taken as a basic axiom that for every point there is *some* assignment of co-ordinates (in R^n) to points in a neighbourhood of that point that is “admissible” in the sense that it respects the intrinsic topological and differential structure of the space. And this seems tolerable: no one thinks that there are any explanatory benefits to be gained by positing a new piece of fundamental structure that assigns to each point a 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 so, and if our grouping of possibility operators together with existential quantifiers is on the right lines, we shouldn’t throw up our hands

whenever we see something with the logical form M^\diamond shares with T^\diamond and T^\blacklozenge —rather, we should try to take considerations of simplicity into account, being guided as much as possible by analogies such as the one with co-ordinate systems.²⁷

The other 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 characterise 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 (e.g. that there are inaccessibly many objects) in virtue of the fact that it *recognisably* entails something (e.g. that there are no valid deductions of contradictions from the axioms of ZFC) 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 is odd to suppose that it could rationally require us to be *more* opinionated about some logically contingent matters than we would otherwise be. Perhaps, then, being an *ideally* good inductive reasoner, in the sense that we have been concerned with, requires being fully confident of all logical truths, including true claims of logical consistency. If our conclusions about theoretical goodness are telling us in the first instance about the nature of this kind of ideal reasoning, there is no reason to expect them to play the same role in the complete story about non-ideal rationality—about how human beings should best accommodate their limitations, e.g. by having certain non-extreme degrees of belief in claims of logical consistency.

²⁷It 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, since to the extent M^\diamond 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—i.e. that *every* world is concretely indiscernible from some M -world—is also plausible, and could be used instead of M^\diamond in our total theoretical package.

8 Other theoretical virtues

Burgess and Rosen (1997: §III.C.1.a) suggest the following style of argument for the theoretical badness of a wide range of nominalistic theories. *Familiarity*, *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 sufficiently close to being equal. So the nominalistic theories are worse than the theories they aim to supplant.

Here are a few rather disconnected thoughts which indicate why I’m not close to being convinced by this argument, at least as applied to T^\square -style theories.

(i) I don’t think it’s plausible that considerations like “familiarity” have any role to play in an account 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 would have degrees of belief that fit with the evidence, and would thus treat these facts as irrelevant. Of course, real human beings, including scientists, surely are prone to favour the theories they encounter first over newly invented competitors. And this 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 newly invented 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, as I have been.

What’s the point of being concerned with those questions, 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 ideal rationality? The answer, I think, lies in some principle like this: if we ought_{non-ideal} to believe that we

ought_{ideal} to believe [/not to believe] that P , then we ought_{non-ideal} to believe [/not to believe] that P .

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’s hard to think of a sense of “fruitfulness” on which T^\square 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^\square .

(iii) Of course, if we decided to believe only T^\square 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. Given this, it’s hard to get the charge of lack of “perspicuity” to stick.

(iv) Suppose I’m right in thinking that I understand a “fundamental” sense of quantification which is different from the sense quantifiers bear in typical 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.

Another putative virtue that might be thought to favour T over T^\square 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^\square ’s distinctive ideology (modal operators) 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 (in the central cases of interest) that if T is true, T^\square must also be true. It’s mysterious how the metaphor of economy could apply in this case: it seems to be adapted for making comparisons between *competing* theories.

The notion of ideological economy does feature in an important class of arguments from T^\square to the claim that there are abstract objects of some sort. These arguments work by exhibiting some general *analysis* of the modal operators used in T^\square under which T^\square turns out to require the existence of abstract objects, e.g. possible worlds.²⁸ 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 the world. Evaluating this kind of argument is a big task—most of the extant analyses of modality using abstract ontology seem to bring in some new fundamental ideology of their own; and in any case, it is by no means clear how to understand the notions of metaphysical primitiveness and analysis in such a way that the argument carries weight. I won’t attempt the task here. I just want to point out that this argument is quite different from the argument we have been concerned with, which purports to show that abstract objects are required for a good explanation of our evidence, even on the assumption that that evidence itself consists of truths entirely about the concrete world.

9 Objections

9.1 Theories *about* simplicity

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

²⁸More plausibly, the existence of such objects would be required by a possibility claim like M^\diamond .

(T^{\square_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^{\square_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:

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

Indeed, if we give “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^{\vee_s} . The challenge for me is to explain why T^{\square_s} and T^{\vee_s} are bad in a way that doesn’t impugn T^{\square} .

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^{\square_s} and T^{\vee_s} to be at all plausible. And the second thing to observe is that this complexity is itself embedded in the scope of 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^{\vee_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 $\langle x, y, z, t \rangle$, or else $T(x, y, z, t)$.

The same structure can be seen in T^{\square_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^{\square_s} and T^{\vee_s} requires any new insights beyond those contained in the toy theory of section 6.

9.2 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

(T^+) Whenever D is a valid derivation [in such-and-such system] whose only premise is ‘ T ’ and whose conclusion is a sentence S that is entirely about observable matters, S is true [in a certain big-enough model, e.g. one whose domain is the set of all concreta].

If we choose the model well, the conclusion that S is true in it will entail S ; thus T^+ will entail every claim about the observable world that can be derived from T , and thus T^+ must be a bad theory, for the same reasons as T^{\blacklozenge} . (Unless T is first-order, T^+ will not be quite as strong as T^{\blacklozenge} ; but I can’t see why this would matter.) How are we to explain the badness of T^+ ? Filling in the definitions of derivation and truth in the relevant model and replacing the quote-name ‘ T ’ with a description capturing all the relevant syntactic detail 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 of 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 ‘“ T ”’ is true [in such-and-such big-enough model]’ 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 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, when one reflects on 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 it 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 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 ‘“ 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 an equally complex, but better, universally quantified theory. 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, I see nothing here to threaten the central point of my defence of T^\square . Somehow, an account of theoretical badness needs to explain the badness of theories like T^+ ; but we have seen no shadow of a reason to think that whatever explains this will also impugn T^\square .

9.3 Differential equations

Does my suggestion 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 a little bit 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, it seems that 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, it 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 ex-

planatory 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, . . .’. (That there is exactly one such operator will follow from the background mathematics and does not need to be explicitly specified.) By doing it that way, one can keep the meat of the differential equation from being embedded in the scope of the existential quantifier.

9.4 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”.²⁹

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

²⁹The same will be true on the family of nominalistic analyses of predicates like ‘electron’ considered in Dorr 2007: §4.ii under the heading “the structural strategy”.

properties like *electronhood*, we might confine 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 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 clause, 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 by any natural properties and relations there might be. Thus, for example, we can replace the existentially quantified theory

(N) There are 8 natural properties $p_1 \dots p_8$ and 3 natural binary relations $r_1 \dots r_3$ such that every natural property is one of $p_1 \dots p_8$ and every natural relation is one of $r_1 \dots r_3$, and $T(p_1 \dots p_8, r_1, \dots, r_3)$.

with the conjunction of

(N₁) There are 8 natural properties $p_1 \dots p_8$ and 3 natural binary relations $r_1 \dots r_3$ such that every natural property is one of $p_1 \dots p_8$ and every natural relation is one of $r_1 \dots r_3$

and

(N₂) Whenever $p_1 \dots p_8$ and $r_1 \dots r_3$ are some distinct natural properties and relations, either $T(p_1 \dots p_8, r_1, \dots, r_3)$ or $T(p_2, p_1, \dots, p_8, r_1, \dots, r_3)$ or ... [and so on for all $8! \times 3!$ permutations].

Of course, $8! \times 3!$ is a very big number; if we are taking length of formulae as a measure of complexity, N_2 will look quite bad. But this is one of those cases where symbol-counting fails quite badly as a measure of complexity. Intuitively, a disjunction like the one in N_2 , 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.³⁰

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

³⁰Moreover, *really* good theories will, I think, posit fewer natural properties of each adicity; they will have lots of symmetries of the kind that would make many of the disjuncts in N_2 unnecessary; and where the posited natural properties are not related by symmetries, they may be related by simple asymmetries that do away with the need for permutations altogether.

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 stable 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.

References

- Balaguer, Mark (1996). 'Towards a Nominalization of Quantum Mechanics.' *Mind* 105: 209–226.
- (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.
- Burgess, John P. and Gideon Rosen (1997). *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford: Oxford University Press.
- (2005). 'Nominalism Reconsidered.' In Stewart Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, 515–535. Oxford: Oxford University Press.
- Chihara, Charles (1990). *Constructibility and Mathematical Existence*. Oxford: Oxford University Press.
- Dorr, Cian (2005). 'What We Disagree About When We Disagree About Ontology.' In Kalderon (2005), 234–286.
- (2007). 'There Are No Abstract Objects.' In Theodore Sider, John Hawthorne and Dean Zimmerman (eds.), *Contemporary Debates in Metaphysics*, 32–64. Malden, Mass: Wiley-Blackwell.
- Field, Hartry (1980). *Science Without Numbers*. Oxford: Blackwell.
- (1988). 'Realism, Mathematics and Modality.' *Philosophical Topics* 19: 57–107. Reprinted in Field 1989: 227–281.
- (1989). *Realism, Mathematics and Modality*. Oxford: Blackwell.
- Hellman, Geoffrey (1989). *Mathematics Without Numbers*. Oxford: Oxford University Press.
- Hodes, Harold (1984). 'On Modal Logics Which Enrich First-order S5.' *Journal of Philosophical Logic* 13: 423–454.
- Huggett, Nick (2006). 'The Regularity Account of Relational Space-time.' *Mind* 115: 41–73.
- Kalderon, Mark (ed.) (2005). *Fictionalist Approaches to Metaphysics*. Oxford: Oxford University Press.
- Lewis, David (1991). *Parts of Classes*. Oxford: Blackwell.
- (1994). 'Humean Supervenience Debugged.' *Mind* 103: 473–490. Reprinted in Lewis 1999: 224–247.
- (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Maddy, Penelope (1997). *Naturalism in Mathematics*. Oxford: Oxford University Press.
- Putnam, Hilary (1972). *Philosophy of Logic*. Harper and Row. Reprinted in Putnam 1975: 323–357.
- (1975). *Mathematics, Matter and Method: Philosophical Papers, vol. 1*. Cambridge: Cambridge University Press.
- Quine, Willard van Orman (1948). 'On What There Is.' *Review of Metaphysics* 2: 21–38. Reprinted in Quine 1953: 1–19.
- (1953). *From a Logical Point of View*. Second, revised edition. Cambridge, Mass.: Harvard University Press, 1980.
- (1976). 'Whither Physical Objects?' In R.S. Cohen, P.K. Feyerabend and M.W. Wartofsky (eds.), *Essays in memory of Imre Lakatos*, 497–504. Dordrecht: D. Reidel.

- Sider, Theodore (2002). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Sklar, Lawrence (1974). *Space, Time and Spacetime*. Berkeley: University of California Press.
- Tegmark, Max (2008). 'The Mathematical Universe.' *Foundations of Physics* 38: 101–150. Available online at <http://arxiv.org/abs/0704.0646>.
- Vaihinger, Hans (1924). *The Philosophy of 'As If': a System of the Theoretical, Practical and Religious Fictions of Mankind*. London: Routledge. Translated by C. K. Ogden.
- van Fraassen, Bas (1970). *An Introduction to the Philosophy of Time and Space*. New York: Columbia University Press.
- (1980). *The Scientific Image*. Oxford: Oxford University Press.
- White, Roger (2005). 'Explanation as a Guide to Induction.' *Philosopher's Imprint* 5: 1–29.
- (2007). 'Does Origins of Life Research Rest on a Mistake?' *Noûs* 41.
- Williamson, Timothy (2002). 'Necessary Existents.' In Anthony O'Hear (ed.), *Logic, Thought and Language*, Royal Institute of Philosophy Supplement: 51, 233–251. Cambridge: Cambridge University Press.
- Woodhouse, Nicholas Michael John (2003). *Special Relativity*. London: Springer.
- Wright, Crispin (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Yablo, Stephen (2001). 'Go Figure: A Path Through Fictionalism.' In Peter French and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy XXV: Figurative Language*, 72–102. Oxford: Blackwell.
- (2005). 'The Myth of the Seven.' In Kalderon (2005), 88–115.