The Marriage of Rationalism and Empiricism: A Naturalistic Account of A Priori Knowledge

Sharon Berry

March 31, 2010

Contents

Chapter 1

Introduction

1.1 A Classic Question

This thesis aims to show how completely naturalistic processes could have given us a faculty of rational insight, which reliably produces knowledge about mathematics and logic - even on the most realistic and objective construals of these subjects. I say 'could', because it's not up to philosophy to fill in the details of how humans actually did acquire the ability to do mathematics. If I succeed, I will have done three important things:

1. I will have shown how to reconcile rationalism with naturalism. Rational intuition need not be an occult faculty; perfectly ordinary processes -of the kind already being studied by psychologists, historians and biologists - already suffice to explain it. As we will see later on, this, in effect, amounts to a marriage of empiricism and rationalism.

2. I will have simultaneously refuted a range of transcendental arguments which have the following familiar form. We have mathematical knowledge, it would be impossible (miraculous) that we had mathematical knowledge if $\sim$P, therefore P. Here are some of the most famous conclusions which philosophers have argued in this way:

- our souls observed mathematical objects before we were born (Plato),
- mathematics concerns the most general form of sensible intuition (Kant)
- mathematical statements are about fictions (Hartry Field)

3. I will have solved the so called 'access problem' for Platonism in a very general way. Although the access problem strictly concerns our access to abstract objects alternative theories of mathematics which deny that there are any such abstracta to need accessing often find themselves facing an analogous problem. Fictionalists need to explain how our beliefs can track what would be true in a certain fiction. And even the formalist - who thinks of mathematical reasoning as nothing more than symbol manipulation devoid of factual

content- faces a problem of explaining how we manage to get the kind of symbol manipulations which help us form reliable empirical beliefs. If it works, my proposal dissolves all these relatives of the access problem as well.

In addition to this, giving an intuitively acceptable, positive, story about how our a priori access to mathematics *could* work, might well be the only really decisive way to resolve what have otherwise been ancient and intractable debates about the ontology of mathematics. Many philosophers find the idea of rational insight into mathematics spooky, or feel that on the Platonist picture our knowledge of mathematics looks like a miracle. Those who feel this worry have tried to articulate what they find unacceptable by proposing various principles (e.g. that we couldn't know things about platonic mathematical objects because in order to know about something you need to have causal contact with it). Their rationalist or Platonist opponents then argue against these particular principles. But, the initial feeling of mystery precedes all these attempts to articulate it - and it remains even when the particular principles have been refuted. The result is an intractable debate, in which those who have concerns about rational insight into mathematics are temporarily silenced (until they manage to formulate a new principle), but never satisfied. So, I propose to cut the Gordian knot: to show there's nothing spooky about rational intuition or knowledge of mathematical objects, by *actually exhibiting* a story about how we could have gotten a faculty of rational intuition which yields knowledge of platonically-construed mathematical objects - one which even the strictest objector will admit, doesn't posit any miracles or otherwise violate the requirements of naturalism.

1.2 The Proposal

My proposed account breaks down into two big steps. The first step, is to explore what it would take for some creature's armchair reasoning to *not* be reliable (either because it very frequently leads to false conclusions, or because it is so incoherent as to fail to allow any interpretation at all) and show that the relevant scenarios, nearly all involve dispositions towards something like *practical failure*. The second step, is to argue that perfectly ordinary naturalistic processes can be seen as systematically working to remove such dispositions to fail. The relevant processes include straightforwardly psychological ones, like belief revision, and perhaps also natural selection (more on this in chapter 3). One important upshot of this combined strategy- if it works- will be that the explanation of how we manage to get into states which count as having true beliefs about math and logic, turns out to be fully continuous with our most plausible explanations of our ability to form true beliefs about straightforwardly empirical matters (astronomy, psychology etc). More specifically, I will invoke causal interactions with the world in conjunction with broadly Davidsonain principles of charitable explanation.

However, at first, it's not at all clear how the kinds of causal interactions we have with the world *could* systematically correct beliefs about *the abstract* *objects of realistically-construed mathematics*. In the case of empirical beliefs about donkeys, we have some idea of how causal interactions could systematically improve the accuracy of our beliefs - but it's not as if the empty set will come down and kick me if I start to form false beliefs about it!

Thus, two more ideas - the main novelties of my thesis - are needed. **First**, I claim that (once we suitably distinguish between causation and justification), revisions of logical and mathematical beliefs which are *causally prompted* by their leading to false predictions about experience, turn out to be completely ubiquitous. For example: Frege's inspection of the ink marks in Russell's letter causally prompted him to revise his mathematical beliefs in a way that removed the clash between theoretical prediction ('My system is consistent, and I'm skilled at using it, so I won't see what strikes me as an inscription of a valid proof of contradiction from my system, seen under good lighting, on a day when I've had my coffee') and experience.

In this way, there turn out to be completely common-place, naturalistic, processes (like belief-revision at the mathematician's desk) which systematically exert pressure towards our accepting logical and mathematical theories which are internally coherent, and succeed in whatever practical applications we expect them to have. But, coherence and practical success are one thing, and truth is another. The truth of many mathematical claims seems to require existence of certain objects. And yet, if e.g. set-theoretic-reductive Platonism is right, there is only one kind of mathematical object - the sets. So, it still looks like the Platonist might have to say that we got terribly lucky in choosing (from among the many equally coherent and practically successful mathematical objects), axioms which correctly describe the structure of the unique kind of mathematical objects which actually exist.

So, **second**, I will argue that even the most hardcore Platonist already has the resources to allow for 'multiple targets' for mathematical knowledge in a way that defuses this problem. What is this?

Well, no Platonist wants to deny that, say, Euler knew a bunch of truths about the numbers. And yet, he didn't posit any objects with the structure of the hierarchy of sets (or whatever else the one really fundamental kind of mathematical objects turns out to be). So how does the reductive Platonist reconcile this, with his view that the sets are the only fundamental mathematical objects that exist? He (traditionally) says that numbers are, in fact, sets. Thus, even though Euler did not manage to correctly guess the fundamental mathematical objects, **he could still make true existential claims - because (unbeknownst to him) his talk about numbers referred to and correctly described a certain collection of sets**. And the same things goes, more generally, for other kinds of objects mentioned in pre-20th century mathematics(real numbers, functions, points). Indeed, what makes the sets an appealing candidate for the fundamental mathematical objects is that so many coherent, intuitively mathematical, practices can be understood as implicitly talking about the sets.

But, if this is right, then shouldn't the same thing hold for alternative possible mathematics-like practices which were not in fact developed? Thus (when he remembers what he wants to say about past mathematicians) even the most conventional Platonist can consistently hold that making true existential statements doesn't require correctly guessing what the fundamental kind of mathematical objects are. Rather, you just need to engage in a sufficiently coherent mathematical practice and, (like Euler and Greek geometers and modern set theorists -should sets turn out not the fundamental objects of mathematics), you'll count as speaking about these fundamental objects and, indeed, be in a good position to assert truths about them. And, an analogous point also applies to less reductionistic Platonists, who acknowledge a wider domain of mathematical objects. (Less strenuously reductive Platonists have an analogous but simpler story available)

Putting all this together, we get a story about how perfectly naturalistic processes might have given us a faculty of rational intuition which reliably delivers truths about math and logic - one that works even when these subjects are construed as realistically as possible. Naturalistic processes, like theory-revision caused by recalcitrant experience at the blackboard, systematically press us towards accepting mathematical theories and methods of armchair reasoning which are internally coherent, and succeed at whatever practical applications we take them to have. But then it turns out that, in doing armchair reasoning which is internally coherent and practically benign in the way mentioned above, we are likely to also succeed in saying largely true things - even when make existential claims. And this holds, *even if the most conventional Platonist is right* about the ontologically committing force of such claims and the paucity of mathematical objects. Finally, when we survey the whole process just described, it turns out to be easy to see how such a process could not only count as yielding true beliefs about mathematics, but as yielding knowledge as well.

1.3 Outline

The chapters to come will fill out, and argue for this proposal as follows. In Chapter One, I will try to motivate and sharpen the intuitive problem of naturalistically accounting for mathematical knowledge by showing how existing theories fail to plausibly address it.

In Chapters Two and Three I will try to show how a completely naturalistic account of our getting logico-mathematical beliefs which are at least practically helpful. Chapter Two invokes mechanisms of Quinean/Millian revision, and shows how these can be invoked without taking on any unappealing Quinean baggage. Chapter Three considers possible evolutionary mechanisms to supplement this story.

Then, in Chapters Four and Five I turn to questions of semantics and epistemology, to complete the story. In Chapter Four I propose a Platonist story about how, in getting experience-matching logic and mathematics like practices from the mechanisms above, it is no miracle that we also wound up with substantial true beliefs about logic and mathematics. And in Chapter Five I argue that the reliable true beliefs generated of Chapter Four should count as nothing less than mathematical knowledge.

Chapter 2

Realism, Naturalism and the Problem of a Priori Knowledge

2.1 Introduction

In this chapter, I will try to a) clarify the intuitive problem of a priori knowledge, and b) say why existing answers to it don't suffice. Thus, I'll start by considering *what kind* of story about our access to mathematics would be needed to do the three things this work aims to do: reconcile rationalism with naturalism (by showing that there need be nothing spooky about rational intuition), refute transcendental arguments that mathematical knowledge would be impossible or miraculous unless some anti-realist thesis X holds, and provide a general solution to the access problem in philosophy of mathematics. Then I'll say why the major existing philosophies of math, and accounts of the a priori don't suffice.

In a nutshell, the problem with existing *philosophies of math* is that they just trade one access problem (how could we get knowledge about sets?) for another (how could we get knowledge about consistency/what would be true in the fiction/what's a second order logical consequence/what's a conceptual truth) which is not obviously any less puzzling.

And the problem with existing (non-supernatural) *general accounts of the a priori*, is that they tend to only consider questions about justification. So, if we distinguish the question 'how do we manage to get *true beliefs* about logic/math/what's contained in a concept?' from 'how, given that we do somehow systematically form true beliefs about logic/math/what's contained in a concept, can we count as having *knowledge* of these things?', they answer the latter without saying anything about the former. This is fine so far as it goes, and makes sense insofar as these theories typically arise from epistemology, and are intended to address concerns about the structure of justification, how there can be basic knowledge etc.

But it creates a problem for using these theories in the present context. The problem is this: worries that rational intuition is incompatible with naturalism, or that Platonism has an access problem, arise just as much when we restrict our attention to the question of how we manage to reliably form *true beliefs* from the armchair/about platonic mathematical objects, as when we ask about how *a priori knowledge* is possible. Thus, theories which talk only about justification, while leaving our ability to systematically get the right answer a mystery, are, at best, only part of the story, and at worst, totally inadequate as responses to the empiricist or the anti-realist. (Actually, insofar as the hypothesis that we have reliable intuitions about mathematics still seems as hard to cohere with the rest of our beliefs as the hypothesis that we have reliable *clairvoyance*, one might worry that these accounts of rational intuition which brush aside reliability worries, ultimately fail as accounts of justification as well!).

2.2 The Problem

The intuitive puzzle about a priori knowledge can be roughly stated as follows 'How can we account for apparent possession of methods of armchair reasoning which reliably yield mathematical and logical knowledge?' But what, exactly, is mysterious here? And what kind of account would be needed to show that access to mathematics doesn't require any kind of spooky faculty, refute transcendental arguments etc? Let me note a few points about how I understand the question at hand.

2.2.1 An explanatory problem not one of possibility

Firstly, this question is (as suggested in the preface) closely related to Kant's question 'how is a priori knowledge possible?' but, most directly, the puzzle isn't that it seems *metaphysically impossible* for creatures like us to have a priori knowledge, but rather that it seems *miraculous* that we do. In particular, I don't mean to suggest that being the beneficiary of some kind of reliable mechanism, which pushes your mathematical beliefs towards accurately matching the relevant mathematical facts (the kind I aim to produce in this thesis) is a necessary condition for having mathematical knowledge. Perhaps merely having true beliefs about math together with a certain kind of phenomenology would suffice to grant one mathematical knowledge- and it's clear how a benevolent deity or even chance could *at least in principle* endow a creature with that.

The feeling that there's something spooky about a faculty of rational intuition, which leads us to right answers, isn't a feeling that it would somehow be *impossible* for a creature's mind to be stocked with true beliefs about mathematics, logic, objective aesthetic and moral facts etc. Rather, one feels uneasy because it seems it would take a *massive miracle* or stroke of luck to achieve this. We have some notion of how contingent facts about the physical world should 'kick back' in such a way as to correct our false impressions and lead us to form true beliefs about them - but no such account seems possible, with regard to the more abstract subject matter of math, logic, realistically construed ethics etc.

2.2.2 Explaining a regularity

Secondly, note that in asking for an explanation for how we could have got in a position to know things like math or logic by armchair reasoning, we aren't asking **of particular things which we accept** (e.g. 2+2=4), why it is true, or why experience appears to conform to it. Rather, we are asking for an explanation of **why we tend to accept statements which true/which experience appears to conform to to**, as opposed to others, which would not have this property.

To put the same point a different way, we are asking for a **systematic** explanation of the undeniable correlation between: a) which mathematical statements people are psychologically disposed to accept b) which mathematical statements are true/correct/such that experience appears to conform to them.

I take the same intuitive question/desire for explanation to be expressed in many other places. Hartry Field asks for an explanation of why it's reliably true that if mathematicians believe that P then P. And Linnebo asks for an explanation of the fact that if "2+2=4" hadn't expressed a truth then you wouldn't have believed it. [1] If we had a explanation of how mathematicians reliably managed to get true beliefs (rather than false or incoherent ones) we would be able to answer both questions above.

2.2.3 The question has an explanatory and a justificatory component

Thirdly, we can distinguish explanatory and a justificatory components within to the question, and it's important to note that a theory which addresses one component might not thereby to address the other.

How can we **manage to get** true mathematical/logical beliefs? (e.g. how do we wind up finding true mathematical claims 2+2=4 obvious rather than false ones like 2+2=5)

How can we count as having **knowledge** of these claims given that we have somehow managed to get true beliefs about them? (e.g. how come someone who finds it obvious that 2+2=4 counts as having knowledge rather than mere true belief)

Now, one might think that all the serious philosophical problems concern justification - we can just wait for science to answer the explanatory question, just like we can wait for it to explain how we manage to systematically get true beliefs from visual perception or how we manage to successfully do face

[1] Perhaps Carrie Jenkins' requirement that if we are to count as knowing that P, the fact that P must figure suitably in the explanation for why we believe that P can also be understood as raising my descriptive question about how mathematical knowledge is possible.

recognition. But there's a difference between these cases (at least in how they relate to the current state of philosophical understanding). We are ignorant of how face recognition works, but there's no a priori feeling of impossibility, which is being used to drive a philosophical argument. No one is claiming that adequate scientific explanation of face recognition will be impossible, unless we accept some controversial philosophical claim C about the nature of faces or recognition.

In contrast, the apparent nature of math and logic (as traditionally understood) can make a naturalistic account of how we manage to align our beliefs with such apparently abstract and remote but objective facts seem *impossible*. [2] And this feeling of impossibility has then significantly motivated some very radical doctrines in the history of philosophy: Plato's doctrine of recollection, Kant's idea that mathematical knowledge is 'the general form of sensible intuition', formalist idea that mathematics is a chess game, or the fictionalist idea that it's a game of make-believe or Mill, Kitcher, Maddy and perhaps also Quine's[3] infamous infamous empiricism about mathematics. Now, if my thesis succeeds, I will show that (contra Plato, Kant and Hartry Field) a perfectly ordinary naturalistic explantation can be given for our ability to reliably form true beliefs about mathematics - even when these are construed in the most straight forward realist/platonist way possible.

Furthermore, I think that once we have this kind of account of how we might systematically wind up with true beliefs about math, the further task of accounting for the fact that we have lots of logical and mathematical *knowledge* becomes a lot easier - and I'll show how it can be answered for almost all mainstream theories of knowledge. This connection between issues of explanation and justification might seem strange, but questions about whether we are reliable, or whether we can fit the idea of reliable access into the rest of our beliefs about how the world works are highly relevant to the question of whether we have mathematical knowledge on a number of different accounts. For the reliablist, it's straghtforward how having an answer to the explanatory question is relevant. But, more interestingly, the internalist who thinks that in order to be jusitified in trusting one's apparent perceptions of certain states of affairs one needs (unlike the person who is suddenly made clairvoyant) to lack reasons for thinking that it would be a miracle if this sense were reliable, will similarly want to diffuse the appearance that mathematical knowledge conflicts with our scientific beliefs of the world just as much as mind-reading does.

Thus, a satisfying account of a priori knowledge would address both questions. And it's important to note the difference between them since an answer to one question may entirely fail to answer the other. A theory of default warrant whereby anyone who found certain 'basic' mathematical truths obvious thereby automatically counted as knowing them -and any conclusions they might deduce from them- would do nothing to explain how we manage to find

[2]Note that what seems impossible is not our having substantial true beliefs about math and logic itself - a benevolent deity might ensure this- but the existence a satisfying naturalistic account of how we manage to do so

[3]I have in mind the Quinean doctrine that mathematics is responsive to new results in physics.

true statements obvious. Conversely, if we have a satisfactory causal + semantic story about how people systematically wind up forming true beliefs when they reason from mathematical principles they find obvious, one could still ask questions about whether people who arrived at true mathematical beliefs like that would count as having knowledge.

But note: in saying we need both an account of how we could have come to reliably form true beliefs about math and logic, and an account of how these true beliefs could count as knowledge, I don't mean to imply that the answers to these two questions must somehow be related. In fact, these questions might turn out to be totally separate if, for example, it turns out that anyone who is brutely inclined to believe that 2+2=4 thereby immediately counts as being justified. In this case, the task of accounting for our having mathematical knowledge would factor cleanly: there would be one task of explaining how we managed to get true beliefs about math, and then a story about why anyone who had the right kind of true beliefs about math would count as having knowledge.

2.2.4 A question about how we got the right methods

Fourthly, it's clear that part of the story about how how we manage to learn things about math and logic goes as follows. We accept some true general principles and valid methods of inference. Then, we combine these to form proofs, which lead us to additional mathematical and logical knowledge.

There may be important philosophical questions to be asked about how making these valid inferences preserves knowledge (for example, there may be some cases where knowledge is not preserved by logical inference e.g. a person might know that they see a zebra, but not that they don't see a cleverly painted mule). However, the worries about how we can have mathematical knowledge without needing any kind of spooky rational insight arise long before these problems come up. For, even if we had a perfect account of this, we would still lack an important piece of the answer to the explanatory question: namely, an explanation for how we could have gotten knowledge of the above-mentioned, true, general principles and valid methods of inference in the first place.

More generally, I think the really hard part of naturalistically accounting for a priori knowledge of mathematics (the part that looks so hard that it motivates transcendental arguments that we must be making some false assumption, either about math or about nature), starts when you push back through all the cases where you can explain *one* kind of a priori knowledge, by assuming *another* kind of a priori knowledge is already in place. So, that is where I will focus my efforts.

2.2.5 Math as a special test case

Finally, I will focus on mathematical knowledge as a primary test case, even though the question I hope to answer arises for all forms of a priori knowledge.

Within discussions about a priori knowledge, the special case of mathematical knowledge has a privileged role. On the one hand, mathematical knowledge seems more 'substantive' [4] than logical knowledge or knowledge of analytic truths - and hence, perhaps, it's especially puzzling that we can know it a priori. But, on the other hand, it isn't a live option to simply deny that we have anything like mathematical knowledge [5] (in the way that some people do deny that we have anything like aesthetic or moral knowledge). These two features of mathematics make it a sort of "dark tower" of rationalism, a nice hard test case which any fledgling account of a priori knowledge must be able to deal with.

2.3 Three traditional approaches

Now let's turn to the answers in the literature. To my knowledge, three basic kinds of answers to the problem of a priori knowledge have been proposed:

2.3.1 Supernatural accounts

What I will call **Supernatural accounts** appeal to supernatural forces or encounters, which shape our psychology to match relevant logical and mathematical facts, and thereby explain how we could have gotten a faculty which allows us rational insight into these matters.

Supernatural accounts (as I will use the term) differ from other rationalist theories of mathematics, which simply posit a faculty of rational intuition, and do not try to explain how humans could have got such a thing. Philosophers like Frege, Russell, and Godel are (as I understand them) rationalists, but not proponents of any supernatural account. They posited that we have some faculty which allows us to reliably form true beliefs about certain logical and mathematical facts, in a way that does not appeal to experience. And the first two, at least, were interested in reducing the extent to which distinct rational faculties would be needed (if logicism worked, then a faculty that let you form true beliefs about logic, would thereby let you form true beliefs about math as well). But, they did not attempt to answer the psychological/explanatory question about how we could have gotten such a faculty of rational insight.

[4] Personally, I'm not sure there *is* a distinctive sense in which mathematical truths are 'substantive' but knowledge of logical or analytic truths is not. Many mathematical statements are ontologically comital. But statements like 'If there are finitely many people at Jane's party, each of whom has finitely many cars, there is one person such that no one has more cars than him' However, for our purposes we can avoid debates about the notion of substantiveness by simply asking: How is a priori knowledge of *mathematics* possible? In answering this question we will have explained how "substantive a priori knowledge" is possible - in whatever sense the example of mathematics shows us that substantive a priori knowledge **is** possible.

[5] Though, of course, the Fictionalist and the Platonist will have serious disagreements about what this knowledge amounts to (knowledge about what's true in a fiction, or knowledge about objects called the numbers), everyone will agree that in learning what an ordinary person would express by saying 'there are infinitely many primes' we are learning something

In contrast, supernatural accounts supplement rationalism with a positive, supernatural, theory how we got the relevant faculty of rational insight. Descartes' idea of a benevolent creator putting correct inference dispositions into our minds like sparks into a tinder, the Leibnitzian theory of pre-established harmony, and Plato's doctrine of recollection are all examples of supernatural accounts. I won't try to argue against these accounts here, but rather take their intuitive unacceptability for granted.

2.3.2 Empiricist accounts

Next, **empiricist accounts** of our access to math and logic, take math and logic to be merely a part of science - denying that these domains have any special a priori status, which could require a different account of how we know them. Thus, this style of account appeals to causal interaction with the world to account for the match-up between psychology and logico-mathematical fact - the same kinds of causal interactions which the world which help us form true beliefs about what would traditionally be distinguished as a posteriori matters like biology and physics. Quine gives the most famous version of this story.

Such empiricist accounts have (in principle) a nice, perfectly naturalistic story for how agreement could be produced between our logical and mathematical beliefs and the subject matter which they purport to represent. If we take systematic revision in response to bad predictions about experience to be capable of producing true beliefs about *physics* or *medicine*, why not math and logic as well - given that these are just part of science?

However (as noted above) the empiricist buys this appealing explanatory strategy, at the cost of denying the common intuition that many claims in mathematics and logic admit of a special kind of justification which is (in some relevant and distinctive sense) independent of experience. Thus in terms of the distinction above, between the explanatory question and the justificatory one, one might feel that the empiricist explains how we get the right answers about math, in a way that makes it impossible to say that we have the kind of justification for mathematical claims (a priori justification) which ordinarily take ourselves to have.

But, even at the level of explaining how we manage to get the right answers, traditional empiricism faces some problems. For, this empiricist story fits awkwardly with some apparent empirical facts about how humans do math and logic. As Charles Parsons has pointed out: we don't seem to be willing to revise our basic mathematical or logical beliefs in response to say, new observations of supercolliders in the way that we would revise our beliefs about physics or chemistry.[6] And some basic facts of logic and mathematics feel psychologically obvious to us in a way that no scientific facts do - which one would not expect on the empiricist account.[7] ADD QUOTES

See the next chapter for more detailed discussion of Quine - whose theory I

[6] Quine on the philosophy of mathematics
[7] Quine on the philosophy of mathematics

take as a starting point- and the last chapter of this thesis for criticisms of more recent empiricist views (Kitcher, Maddy and Jenkins).

2.3.3 Meta-semantic accounts

Finally, **meta-semantic accounts** appeal to facts about how the way we use a word helps determine what that word means in our language, to explain the match between our logico-mathematical beliefs and reality. The basic thought goes like this. How come we are disposed to believe true things like 2+2=4 rather than falsehoods like 2+2=5? Perhaps the answer is just this. It's arbitrary what we are disposed to say. But if we had been disposed to say that "2+2=5", our word "+" would not have expressed the plus function, but something different (e.g. the plus 2 function), such that this statement would have expressed a truth. Thus, a match between what claims we are disposed to accept (on the one hand), and the logical and mathematical facts (on the other), is ensured by something like Davidson's principle of charity. We can get true beliefs about math and logic just by adopting arbitrary principles for how to reason a priori. For, whatever we are disposed to say in the armchair will determine the meaning of our relevant words - in such a way as to ensure that whatever we are disposed to say when following these dispositions counts as true.

More generally: the meta-semantic strategy tries to account for our match between our logico-mathematical beliefs, and the logical and mathematical facts which are relevant to the truths of these beliefs, without positing any kind of mechanism that effects what sentences we are disposed to assent to (or what we are disposed to practically respond to seeing certain inscriptions on a blackboard) in the direction of fitting logical or mathematical facts. Instead, the match between what we believe a priori and what's actually the case, is explained by **metasemantic facts**: facts about how our (and our community's) dispositions to use a word, effect what that word winds up meaning in our language. On this view, our willingness to make certain kinds of inferences and assertions with a feeling of obviousness, to not demand any justification for them etc., amounts to a kind of stipulative definition. And the result is that any assertion made in accordance with this practice will count as true. Thus, there's no need (the meta-semantic strategy says) to posit explanations for how we come to do (what amounts to) asserting largely mathematical truths, rather than mathematical falsehoods or nonsense, when we reason a priori, because *nothing we could do would count as asserting largely falsehoods or nonsense*.

One *might* interpret Carnap as giving this kind of account (if one thinks he was concerned to account for our access to math and logic at all) [8]. For, Carnap's principle of tolerance makes one free to adopt arbitrary meaning-giving stipulations: the framework conventions for our language. All the statements which logically follow from the statements on the list, or are otherwise suitably indicated by the framework conventions, will then count as expressing truths

[8] Ricketts has plausibly argued in X that he was *not* concerned to answer this question

in our language. Thus, we can arrive at true beliefs a priori, just by following out the consequences of our own meaning-giving stipulations. Now, if we take these meaning-giving stipulations to be explicit statements (e.g. posted on the laboratory door) there's a problem - as Quine pointed out. One would need to *already* have logical knowledge to work out whether a given new sentence was a logical consequence of the sentences stipulated to be true or otherwise indicated on the list. Thus, meaning-stipulations could (even in the best case) only give us access to a priori truths, if we already had some other form of access. A natural way of responding to this concern would be to say that the relevant stipulations aren't explicit statements, which we would need to consciously apply, but rather these conventions are implicit in our use of words. Thus, one winds up saying that we use logical and mathematical words, in such a way as to make certain kinds of statements and inferences count as framework stipulations for the meaning of these words - and hence to ensure the truth of these statements and the truth-preservingness of the inferences. And this is just (a version of) the meta-semantic story.

Now, I claim that (contra the meta-semantic story) that it's simply not true that, communities who embrace arbitrary patterns of assent as obviously compelling meaning-giving truths, will thereby fix the meaning of their terms in such a way as to ensure that they count as saying largely true things a priori (in the sense that we *do* take ourselves to be in a position to say largely true things about math and logic a priori). For example, consider people who say what "a+b" is in the same way that we would answer questions about sums mod 17, *but* make (verbally) the same inferences about 'how many apples or oranges there are all together in a basket', or 'how many male or female rhymes there are in a poem', and draw the same practical consequences from these statements as we would from statements about the regular plus function. These people (if embodied in a world like ours) would be constantly running into practical problems, and making practical mistakes - in ways that make it reasonable to interpret them as having false beliefs about sums (or the a priori consequences of facts about sums for facts about numbers of objects). Also, consider people who accept syntactically inconsistent axioms, together with the same principles of logical inference that we have (including the principle that everything follows from a contradiction). Here we are inclined to say that either, they are wrong about some of their basic axiomatic beliefs, or their practice is so incoherent as to fail to express anything at all. Thus, it's simply not true, that nothing would count as having a practice of armchair assertion which failed to get things largely right. Despite occasional failures (like accepting inconsistent axioms for set theory, or taking ourselves to know a priori that parallel lines in physical space must meet at most once), human knowledge of math and logic looks *very different* from what one would expect an arbitrary practice of making meaning-giving stipulations to yield.

I think the meta-semantic strategy can look more appealing than it is for two reasons. Firstly, we can wind up considering particular patterns of armchair reasoning on their own, rather than thinking about how they might fit into a person's overall way of life (which is, of course, what's relevant to assigning meanings to their words). When we think about particular pieces of mathematical practice (e.g. the behavior we exhibit when calculating sums mod 17 at the blackboard), we can see how these behaviors might be part of an overall pattern of use - of drawing certain verbal conclusions from claims about what '2+7' is, and acting on these conclusions in certain ways - which we would interpret as saying largely true things. This can make it seem like we are free to add such behaviors to our own practice at will, and hence to make arbitrary meaning-giving stipulations for our words. But, the fact that for each pattern of a priori 'reasoning' from one sentence to another there's *some* web of assent dispositions and actions in which it could count as leading one to assert truths, doesn't entail that arbitrarily adding new patterns of a priori reasoning to *your current* web will amount to adding new true beliefs. (For example, the very same pattern of blackboard behavior which amounts to inferring new truths about arithmetic mod 17 when we do it, would count inferring falsehoods about sums or not having any meaning at all when people in the community mentioned above do it.) If you are going to keep old conventions of use - such as the laws of logic you already have and scientific uses of language for prediction etc. which depend on it not being possible to derive every sentence in your language - you will need, at the very least, to avoid making logically inconsistent stipulations. Thus, in order to reliably get to new true beliefs by making 'arbitrary' stipulations you already need to be sensitive to consistency and conservativity facts - will adding these axioms allow me to prove contradiction? will it let me is already to suppose that we have some kind of access to/sensitivity to the logical and mathematical facts?

So, at this point, the meta-semanticist faces a choice. Getting math right seems to require more than an ability to make stipulations. You need some kind of ability to make the right kind of stipulations (e.g. first order consistent stipulations, if the system of stipulations you have already made contains first order logic). Someone who just pulled sentences out of a hat and stipulated that they were to express truths, would be quite unlikely to arrive at anything like the mathematical knowledge which we have. If the meta-semanticist denies this, his view looks implausible. But if he agrees that in addition to the ability to make stipulations we also have something like ability to recognize the right kind of stipulations, and both are needed to account for our possessing substantial true beliefs about math, he has dodged this bullet, but now needs to explain how we could have got this ability. Are we guided by a separate kind of mathematical insight that lets us see that certain kinds of stipulations will always be consistent or conservative? If so how could we have gotten *that?* Do we get to good stipulations by brute trial and error? Empirical experiments with attempts to produce strings of certain kinds? So what he has given us is, at best, a reduction of the question, rather than a satisfactory answer.

As a brief glance at the beginning of a typical paper will show, stipulative definitions obviously play an important role in mathematics. But mere consideration of our ability stipulate can't suffice to explain our mathematical knowledge, in the absence of some account of how we are able to recognize acceptable vs. unacceptable stipulations.

This brings us to the second reason the meta-semantic story can look unduely attractive. Certain patterns of use - such as introducing 'bachelor' as an abbreviation for 'unmarried man' - are obviously safe and unlikely to introduce contradiction to our general pattern of making inferences and acting on our conclusions about the world. Thus, one might think that the same easy 'process' of knowledge generation, which is at work when I stipulate "'let 'bachelor' mean 'unmarried man'", and then infer from "all unmarried men are unmarried men" to "all unmarried men are bachelors", could account for all my a priori knowledge. But to say this ignores the substantive capacities which are at work even in this process - that I start out with a consistent system of logical inferences, which I used to get 'all unmarried men are unmarried men', and I systematically add terms (along with their associated new patterns of inference) to it, only when (as is particularly clear in the case of introducing a term which functions as a mere abbreviation for the other), these new patterns of inference don't allow me to infer contradiction or - any new statements which not involving the term which might express falsehoods. To presuppose that our psychological dispositions to assent to or reject new definitions/ways of using a word are sensitive to consistency facts, is already to presuppose the kind of beneficial match-up between our psychology and logico-mathematical facts which it is our task to account for.

Now, in rejecting the pure meta-semantic strategy, I don't mean to say that the close relationship between use and meaning can do nothing to explain how we wind up asserting such a large proportion of truths from the blackboard or in the armchair. I merely want to point out two things. Firstly, the right explanation for our access to logic and math a priori can't be that we get math right because nothing would count as systematically failing to get it right because. We've just seen some things that would count as failing to get it right. And, secondly, to say that we reliably get true beliefs by making 'arbitrary' meaning-giving stipulations and then following out their consequences presupposes an account of how we are sensitive to facts about consistency and conservativity so as to make *the right kind* of 'arbitrary' stipulations.

2.3.4 Non-answers in the literature

Finally, there are also some other philosophical theses in the literature which are quite relevant to the task at hand, but which I want to flag as *not* providing an answer to the question 'How can we account for apparent possession of methods of armchair reasoning which reliably yields true beliefs about math and logic' [9], **on their own**.

- Theories of justification on which certain basic statements of math and logic are 'hinge propositions', which need no justification, or count as prima facie justified for anyone who believes them - provided only that this person has a certain phenomenological feeling of obviousness or insight. Even if one accepts a theory of this kind, one still doesn't have an

[9] in the sense that's relevant to this thesis

account for the striking fact that we systematically manage to get true beliefs i.e. how we manage to get the right answer in the first place (it just says why we will count as knowing the right answer whenever we are so lucky as to get it). That is, a theory like this can only provide an answer to the explanatory question, not the justificatory one.

- Theories in philosophy of math (like Fictionalism or If-Then-ism), which deny that mathematical objects literally exist. Even if doing mathematics is just a matter of working out what follows from arbitrary hypotheses - where 'follows' means either, what follows logically, or what 'would have to be true' in a fiction where these hypotheses held - there's still a question of how we manage to get accurate beliefs about what's a logical or necessary consequence of what. Also, most Fictionalists and If-thenists would want to say that we very often manage to form *consistent* hypotheses and one needs an explanation for that. [10]

- Theories on which the apparent conformity of experience to mathematics, is just an artifact of the way our minds 'construct' experience. Here the idea is, we can correctly predict we'll never see two parallel lines that meet more than once, because our minds process sensory input in such a way that, whatever comes in, they will always represent it with an experience in which parallel lines meet at most once. This might seem like it answers the problem - we can predict certain features of experience in advance of experience because these are features which our mind will always add. But note that a piece of very substantive modal knowledge is required - knowledge that necessarily my mind will not represent a scenario in which X - and no explanation has been given of how we could manage to get this knowledge (for example, how do we distinguish cases where we have just never yet had an experience that represents X, from cases where our minds couldn't represent a scenario in which X?). It's one thing to have a mind which can't represent scenarios in which P, and another thing to realize that you have such a mind. Also, of course, insofar as one doesn't take the same story to apply to logic, nothing has been said about how we manage to get logic right.

- Theories (a la Peacocke) on which basic logical and mathematical statements are true 'in virtue of our concepts' and realizing that something is true in virtue of one's concepts justifies us in believing it. If we construe this kind of story as an attempt to, not only answer the justificatory question but also the explanatory one (i.e. to account for our ability to systematically form true beliefs about math and logic) then it's missing something crucial. For, consider what you might call incoherent concepts

[10] I take Michael Potter to be making this point in the essay 'What is Mathematical Knowledge?', though he also seems to be making the stronger claim that Fictionalists and If-thenists are committed to view that we manage to systematically get consistent hypotheses in order to make their explanations of the applicability of our mathematical doctrines - a view which I argue against in 'Berkeley and the Bridges: what the applications of math don't commit us to'. See appendix

or pseudo-concepts, like Frege's notion of extensions (which falls victim to Russell's paradox) or the connective 'tonk', which has the introduction rules of 'or' and the elimination rules of 'and', and hence allows you to derive any sentence from any other. [11]. If our armchair access to math and logic is a matter of forming concepts, and making the claims and inferences that feel like obvious truths in virtue of these concepts, one needs an explanation for how we avoid forming pseudo-concepts like the above - how we manage to detect when a possible pattern of language use doesn't really correspond to a concept/corresponds to an incoherent concept or otherwise systematically avoid adopting adopting such patterns.

- Theories of meaning on which people that didn't have largely correct beliefs about math and logic wouldn't count as thinking anything about math, or wouldn't count as thinking at all. Relabeling the apparently possible, and worrying, scenarios in which we (for example) find syntactically inconsistent axioms obvious, or make the inferences characteristic of tonk, as ones in which we aren't thinking at all, doesn't do anything to help explain how we manage to avoid being in such bad scenarios.

2.4 Conclusion

Thus, I take it, the task of providing a plausible, non-spooky, epistemology for mathematics has not yet been satisfactorily addressed.

[11] You simply reason from A to 'A tonk B' to B

Chapter 3

Math as Science with Stockholm Syndrome

3.1 Introduction

Kant noted that we come to opinions by a priori reasoning (in advance of experience), which experience then appears to conform to. Reasoning in the armchair leads us to conclusions like 'everything is self-identical', which our limited experience seems to bear out -or at least not conflict with- as opposed to conclusions like 'everything is lime green'. The fact that our inclinations to a priori (logical and mathematical) reasoning are practically benign in this way is interesting in itself, and this chapter and the next one will be devoted to showing how we can naturalistically account for it.

In this chapter, I'll start by trying to get clear on exactly how our mathematical and logical practices are practically benign(/helpful in a certain limited sense). Then I'll turn to some roughly Quinean mechanisms of theory revision, which can account for how we got practices with this fortunate feature. I will try to show how we can appeal to these quinean mechanisms without being committed to various unattractive features of Quine's overall theory.

3.2 Practical Helpfulness

All philosophers of math will agree that people do something they call "math", and that this activity is practically helpful, in a certain sense. This is often put pretty loosely by saying 'Math helps us build bridges that stand up'. But I think we can say something much clearer than that - as follows.

Our grasp of math (such as it is) has at least three aspects:

- We can follow proofs. You will accept certain kinds of transitions from one mathematical sentence to another, (or between mathematical sentences and non-mathematical ones) when these are suggested to you.

- We can come up with proofs. You have a certain probability of coming up with chains of inference like this on your own.
- Proofs can create expectations in us. Accepting certain sentences makes you disposed to react with surprise and dismay should you come to accept other sentences. (e.g. if you accept "n is prime" you will react with surprise and dismay to a situation where you are also inclined to accept "n has p, q, and r as factors").

Now, the sense in which our mathematical practices are **practically helpful** is this:

First, our reasoning about math fits into our overall web of beliefs in such a way as to *create additional expectations*. That is, there are many situations S with the following feature: People in situation S who realize their dispositions to make/accept certain mathematical inferences, arrive in a state where they will be surprised by more things than people in the same situation who don't realize these disposition.

For example, plonk a bunch of people down in front of a bowl of red and yellow lentils. Make each person count the red lentils and the yellow lentils. Now, give them some tasty sandwiches and half an hour. Some of the people will add the two numbers resulting from their count. Others will just eat their sandwiches. Now, note that the people ho did the math have formed extra expectations, in the following sense. If we now have our subjects count the lentils all together, the people who did the sum will be surprised if they get anything but one particular number, whereas those who didn't do the math will only be surprised if they get anything outside of a certain given range.

Secondly, the extra expectations raised by doing math are *very often correct*. When doing mathematical reasoning about your situation puts you in a state where (now) you'd be surprised if a certain observation/reasoning yields anything but P, applying this process tends to actually yield P. (This is especially true for certain kinds of specially direct feeling applications, which we feel extreme confidence about e.g. applications of reasoning about proof theory to predict what inscriptions of symbol strings we will or won't find). Thus, 'composing' a process of mathematical reasoning M with some other reasoning processes A (i.e. going through A and then applying mathematical reasoning M to whatever results you got from doing A), frequently yields correct expectations about the result of going through a different process B, if it yields any expectations at all.

And finally, this is (potentially) helpful, because it means that - not only do we acquire the disposition to be surprised if B yields something different - but any further inferences/actions which would get triggered by doing B, happen *immediately* after doing A and M - you don't have to spend time and energy doing B. For example, in the case from the previous post: if we imagine that all of our sample population have inductively associated counting 1567 lentils in total with having enough to make soup, the people who did the addition after counting the lentils separately, would start cooking earlier than those who did something else instead.

To summarize:

Doing math is practically helpful in the sense that: spending time doing math raises extra expectations (relative to spending that time eating sandwiches) about the results of certain other processes, and these expectations are generally correct. Thus, mathematical reasoning constitutes a reliable shortcut, leading us to take whatever actions would be triggered by going through some other process B without actually going through B.

3.2.1 Examples of Practical Helpfulness

Here are some examples of how mathematical/logical reasoning is practically helpful in the sense above. In each case we have a combination of some (potentially empty) collection of observational procedures A, combining with mathematical procedure M, to yield expectations about (which I have symbolized with a →) the results of some distinct procedure B (usually another observational procedure).

1. Observe a computer (wiring looks solid, seems to be running program p etc.), derive that program it's running doesn't halt, expect it to still be running after first 1/2hour → observe computer after 1/2 hour

2. Observe cannonballs, form general beliefs about trajectory of a cannonball launched at various angles, observe angle of launch for a particular cannonball, derive where trajectory lands → measure where this ball does land.

3. Prove a general statement, expect 177 not to be a counterexample → (directly) check whether 177 is a counterexample.

4. Conclude that some system formalizes valid reasoning about some math truths, expect that you aren't looking at an inscription of a proof of "0=1" in that system → check what you have to see if it's an inscription a proof in the system, if it ends in "0=1".

5. Count male rhymes in poem, count female rhymes, then add → count total rhymes

The case of reasoning about the numbers, also provides a bunch of nice examples of practical helpfulness. Each of the following procedures allows us to correctly anticipate certain results of applying the procedure below it.

- general reasoning about the numbers: $\forall x \forall y \forall z\ ((x+y)+z) = (x+(y+z))$
- calculations of particular sums: 22+23=45
- assertions of modal intuition: whenever there are 2 apples and 2 oranges the must are 4 fruit
- counting procedures: there are two "e"s in "there"

Hopefully the above section has made the sense in which mathematics can be practically helpful more clear. But this only serves to dramatize the remaining question. How could creatures like us have gotten a practice like mathematical and logical reasoning which has this good features?

3.3 A Quinean answer, and three objections to it

Quine has (at least part of) an answer: we revise our beliefs when they lead us to make predictions that fail. If my expectation that the computer would not halt in the next 5 minutes had lead me to expect to not see it spitting out a certain line of text, but then I *did* see it spitting out a certain line of text I would come to revise some of my beliefs, my observational practices or my behaviors in response to observing certain things. If we suppose ourselves and our scientific community to constantly be undergoing such practical testing and revising our beliefs, observational practices, and ways of acting on our beliefs in response to practical failure, this would easily explain how we might wind up with a total system that fits together in such a way as to *not* lead to practical failure.

But wait! Traditionally philosophers have objected that this picture of rational belief revision can't be right since a) mathematical beliefs are not rationally revisable in response to experience in the way scientific beliefs are b) some mathematical principles are specially obvious in ways that no scientific principles c) we find the falsehood of our mathematical beliefs inconceivable. Now I claim that, even if each of these three points is right, **none of these facts undermines the Quinean explanation of how we could have managed to get practically successful beliefs/beliefs that appear to be supported by all future experience**. Thus, we can accept the Quinean explanation of practical success -and hence get the explanatory benefits of Quine's well-known account of mathematics - without accepting the epistemological and psychological claims (there is no a priori knowledge, math is just another branch or science) which people have found so implausible.

3.4 How to get the Quinean explanation of our having a practically successful mathematics without the Quinean epistemological, phenomenological baggage

Why? Let's start with a) mathematical beliefs are not rationally revisable in response to experience. Now note, all the Quinean explanation above requires is that revisions in mathematical beliefs are causally responsive to the experience of using these beliefs to generate false predictions i.e. recalcitrant experience can reliably **cause** you to change which mathematical beliefs you hold. Now this is indubitably true as regards particular mathematical beliefs. If I calculate in my head that 23+27=41 and then punch this sum into a calculator and get 50 (i.e. not the result that my total web of mathematical and electrical beliefs prompted me to expect), this will prompt me to change some of my beliefs, in particular my mathematical beliefs. And presumably in making this change I am not doing anything irrational. Hearing testimony, or seeing scientific experiments go astray can also send us back to the chalkboard to check our calculations, and thus reliably prompt theory revision.

Is this enough? You might worry that the only way experience can prompt a revision is when we have made what we would (antecedently) have considered a calculating error. But this is not right. Firstly, observations do make people reject certain methods of a priori reasoning as fallatious (e.g. seeing light bend lead people to reject the principle that space necessarily had a certain geometry, and repeated experience with computer simulations of the Monty Haul problem leads people to reject the fallatious but tempting chain of reasoning which leads the conclusion that switching doors can't help). These are cases where people changed their very rules for how to reason a priori (not just see that something was what they would already have recognized as an error had they been more careful).

Secondly, note that the kinds of recalcitrant experience which prompt revision need not look like scientific experiments. They will often happen in the math department. Finding out that a certain set of axioms is inconsistent involves a kind of recalcitrant experience. You expect never to find any string of symbols on paper which you see ends in "0=1" and you check (or your computer checks) line by line to have certain syntactic properties corresponding to being a proof in FOL from the axioms in question. But, one day, you do find yourself seeing such a string. Here the scientific elements in question are quite minimal (you think you/your computer's output co-varies in certain ways with the syntactic properties of the string). And we probably wouldn't count our evidence for these simple scientific claims as part of our justification for believing the string is inconsistent or rejecting the axioms. But this makes no difference to the Quine-style explanation above.

If mathematical principles that lead to practical failure when combined with the rest of our beliefs, observation practices etc. reliably get revised we have our explanation of how we could have gotten mathematical principles that don't lead to practical failures [1]. It doesn't matter what we say about this revision.

Usually we will say that we should have known all along, so that the recalcitrant experience (even if it were causally necessary and sufficient for producing the revision) doesn't figure in the justification for the new beliefs (we have rationally revised our beliefs as a result of recalcitrant experience but not "in response to it"). Now it's easy to see how b) (the special obviousness of some mathematical claims) and c) (the feeling that whatever mathematical claims we accept are necessary truths) don't pose a problem for the Quinean explanation either. So long as mathematical beliefs do get revised when we have certain

[1] assuming there were any reasonably simple such principles out there to find

recalcitrant experiences like seeing an inscription of a proof of contradiction from certain axioms or getting an unexpected output from a calculator (which they clearly do), it doesn't matter whether they have a special phenomenology. The feelings of special certainty and inconceivability of falsehood [2] which once attended the idea that for any property there must be a extension i.e. a set containing all and only the things that have that property in no way prevented us from revising this principle once contradiction was derived.

[We should note here, a simple fallacy which confuses not being ready to revise a belief in response to evidence with thinking it inconceivable that one could encounter evidence which would warrant revising it. If I found a proof of contradiction in ZF I would stop believing it. Thus, it's quite clear that there's one kind of experience which I would count as evidence against ZF, namely, seeing a proof of "0=1" from ZF. However, I now do believe ZF and think it's consistent. Hence, I also think the one situation which would prompt me to give up ZF is impossible. Given this, it would be (true but) very misleading for me to say that no observation could possibly count as evidence against ZF. The observation of seeing a string of inkblots with a certain combination of syntactic properties (namely those which correspond to being a proof of "0=1" from ZF) would be good evidence against it. But, of course, I think there can be no such string so this "observation" is impossible.]

3.5 Wait, but what about innate inclination towards certain kinds of good mathematical beliefs?

We now have a semi-Quinean explanation for the fact that when you come to some conclusions about math a priori, and them apply them in building bridges the bridges you build tend to stand up. Mathematical principles which lead to bridge-failures (let's say, false predictions about whether you'll see a given bridge fall) get revised just like principles of physics or beliefs about building materials or the reliability of contractors. Mathematical beliefs have lots of special features - they can feel specially obvious, we take them to be necessary truths whose falsehood is inconceivable. And, when we revise them, we don't just form a new the mathematical belief (e.g. 'There is *not* an extension corresponding to every property') but we are also likely to form a new normative belief ('e.g. I shouldn't have ever assumed that there are extensions corresponding to every property, that is a fallacy.') and our notions of what's conceivable (we now find it conceivable that the property of being a property that doesn't belong to it's own extension doesn't have an extension). [In the more interesting cases like that of Euclidean Geometry, we change our whole notion of what descriptions count as *conceiving of* a scenario in which P as opposed to merely describing it]

But there's another kind of objection. This story all takes place at the level of

[2]Frege might well have said to himself 'it's just clearly inconceivable that there could be some property that applied to certain things, but no extension containing the things which it applied to'

adult language users revising their mathematical beliefs (causally) in response to recalcitrant experience. But what if one could show that some portion of mathematical belief is innate, i.e. that we are somehow inclined to prefer the kind of mathematical beliefs that combine in a practically successful way with the rest of our beliefs *literally in advance of any relevant experience* (e.g. when we are babies)?

There's one worry about simply accepting the above mentioned story about the revisability of logical and mathematical practice in response to experience as a full answer to the our question about the harmony between the conclusions we arrive at in the armchair and experience is that experience **couldn't** correct math. I have tried to argue against this view in this chapter. But another possible worry is that experience **doesn't need to** do much correcting because we seem to have an innate cognitive set up which already steers us towards the kinds of mathematical hypotheses which would conform to experience - and it's inexplicable how we could have got *that*.

So, in the next chapter I'll turn to the question of how one can naturalistically explain our having an innate inclination towards certain kinds of good (experience-matching) logical and mathematical hypotheses rather than others.

Chapter 4

A Nudge from Natural Selection?

4.1 Introduction

We left off in the previous section with a need to account for how little Quinean Revision seems to have to do in getting us to accept good (i.e. experience-matching) logical and mathematical beliefs. We seem to have a kind of nudge in the right direction, even in advance of modifying our views to fit any potentially recalcitrant experience. What do I mean by a nudge? Well, there are two phenomena which it looks like Quinean revision alone might be insufficient to account for.

Firstly, animals and pre-linguistic babies (who can't even pose scientific theories, much less engage in Quinean revision) have something suggestively like mathematical knowledge.

For example, Spelkie's looking-time experiments show that if you put three dolls behind a cover, and then take one away, babies and apes will look much longer if they see 1 or 3 or 4 remaining dolls rather than 2 when you lift the cover. Some people would describe this by saying that the creatures in question know that 3-1=2. (The same results persist when you change the numbers, but fall apart when distinction between numbers greater than 3 is required.) Others, who take there to be a much more intimate relationship between language and thought, would deny that these creatures have any such belief. I won't take a stand on this interpretive issue, because the only thing that matters now is whether we can account for animals and babies doing what they do (call it using mathematical knowledge or not) naturalistically.

Secondly, there are certain kinds of bad (experience-resisting) mathematical theories which (we somehow expect) no one would ever pose. For example, consider the hypothesis 'Addition is symmetric, but only for numbers up to 500, then a+b=b+a+1 whenever either a or b is ¿ 500'. If someone did pose this theory - while engaging in the kinds of ordinary computational practices for computing individual sums which might make us want to attribute the meaning plus to their word ''+' - they would run into recalcitrant experiences when doing particular sums like 245+688, and likely either give up this piece of doctrine, or change their algorithm for computing ''+' in such a way as to reestablish fit. However, it seems plausible that no one who learned the ordinary addition algorithm would ever even *propose* something like this as a hypothesis. Thus Quinean Revision, can't explain why we don't have the kind of experience-conflicting mathematical theory described above.

To put this second point positively: we seem to have a kind of attraction towards good (experience- matching) logico-mathematical theories in advance of experience and Quinean Revision. I use the word 'nudge' here to stress how weak the phenomenon in question is. It's not as if we are irrevocably hardwired to accept certain mathematical theories (as I argued above history shows that we are quite willing to change mathematical doctrines, and reject arguments that intuitively seemed appealing as fallacious), nor are the kinds of theories that we are drawn to invariably correct (cf. the Wason card-flipping experiments, and the gambler's fallacy).

Also, it's very important to note, that this inclination to avoid forming bad theories doesn't entail that we are innately inclined to form any *particular* good theory. Compare the claim that *we are innately inclined to prefer they hypothesis that addition is symmetric over the hypothesis that it is not symmetric in the way outlined above* with the claim that *we are innately inclined to believe that addition is symmetric* simpliciter. The former seems very plausible, while the latter is possibly also true but much more controversial. The issue at stake here is just the same as in the following empirical case. Contrast: 'We are innately inclined to prefer the hypothesis that the surface of the international space station is grey all over to the hypothesis that the space station is grey in places where we can see it an pink everywhere else' with 'We are innately inclined to believe that the international space station is grey all over'. It would be crazy to think that we are hardwired to form this (or any) particular belief about the ISS, but not to think that we are innately disposed to prefer the former hypothesis to the latter, *should we wind up forming any beliefs about the ISS at all*.

So, in this chapter I'm going to show how we can naturalistically account for these two phenomena: suggestively mathematical good behavior in babies and animals, and a nudge towards experience-matching mathematical doctrines, even in advance of experience.

4.2 Accounting for suggestively good mathematical behavior

I'll start with the task of accounting for what I'm calling 'suggestively mathematical good behavior' observed in pre-linguistic (and hence pre-Quinean theory revising) infants and animals (e.g. Spelkie's differential looking times). Personally, it seems obvious to me that there's no philosophical problem in accounting for this kind of well observed natural phenomenon - why wouldn't evolution be able to build animals with these kinds of practically useful dispositions to behavior?

But for thoroughness, I'll now go through an example story of how this kind of suggestively good mathematical behavior *could* have arisen. I'm only doing this to show that we don't need to posit any kind of spooky faculty of Godelian concept inspection or platonic recollection of the forms to explain it. There are still interesting questions to be asked about how various animals *actually* got the particular pieces of suggestively mathematical good behavior which they have. But I think we philosophers should be inspired by Hegel's success with deducing the number of planets, and Kant's in deducing the necessary truth of Newtonian Mechanics to leave those questions to the evolutionary biologists and psychologists who will go actually make relevant observations.

4.2.1 Mathematically-shaped problems + adaptations to solve them

The story starts with mathematical challenges and natural selection's ability to meet them. Nature poses certain kinds of mathematically structured problems. For example: If you go north for n minutes, then east for m minutes then north again for o minutes west for p minutes, all traveling at a constant speed, what angle do you wind up from home? The Tunesian Desert Ant faces this problem when it needs to go home after foraging in the desert while the wind blows around visually recognizable landmarks. and it's nerve system has evolved to correctly calculate what angle to turn to go home using "deductive reckoning" i.e. the correct trigonometry algorithm.

People and monkeys don't face that problem, but they do face problems like the following. If 2 stable objects like apples or dolls are put behind a curtain and 1 is removed, and then the curtain is raised, how many objects should you expect to see, i.e. when should you go looking for some aspect of the situation you expected to miss? Also, if Jim always gathers berries according to such and such a pattern, are there places in this berry field which she would systematically miss?

And language-using animals living in communities where everyone has some a priori reasoning ability face challenges where having more (so that you can reason better about what they can vs. can't get using their reasoning) would be helpful e.g. if Jim reasons according to such-and-such logical laws, can he combine pieces of information which he has to discover that both: I have an basket of tasty fruit in my hut, and no one is now guarding the house? (though it may make sense to think about this example more later when we have the story about a nudge towards good logico-mathematical theories on the table)

Now, it's perfectly clear that nature can build the nervous systems that realize the right kinds of algorithms to deal with these problems. Since it so happens that the right kind of algorithm to decide what angle to turn around if you are Tunesian Deseart Ant is also the right algorithm to calculate the angle between the start and end points of a certain path in a nearly Euclidean space, it's clear that nature can build something that realizes that algorithm (at least for a range of suitably small inputs - perhaps the ant would eventually loose track if you made it make 1000s of turns). In the case of differential looking times, the needed algorithm: if you seem to see n added and m taken away, expect n-m is much simpler, but also clearly physically realizable. (Interestingly though, we don't realize that algorithm for slightly larger numbers, but a more permissive one which makes you look longer if you see any number of objects outside of a certain expected *range*. Perhaps this is an adaptation to the fact that we have visual-processing inputs which very reliably co-vary with whether there are 3 things, but which only less reliably co-vary with whether there are, say, 24 things - so it pays not to be too suprized if your optical processing is off by 1 for larger numbers)

And, at least in simple cases like the differential-looking-time behavior, (I take it) there's nothing mysterious about how natural selection could have lead us to the right kind of behavior. The right kinds of algorithm is simple and easy to physically realize, and fact that when you see 1 apple removed from 3 apples leaving you with 1 apple you HAVE missed something generates selective pressure to be disposed to *look for what you've missed in such circumstances*. So, we have natural selection for certain kinds of behavior which are suggestively similar to the behaviors which would flow from mathematical knowledge in an adult language speaker.

Now let's add two things.

4.2.2 Getting more good behavior than you need

First, let's note that something very general about the world (I don't know whether the same thing would apply to any other possible world as well). Often the physically easiest way of building something which succeeds in a certain desired range of cases is to build something which can also succeed in a wider range of cases. Suppose I give you a home electronics kit and ask you to build a machine that displays the right answer when any one of 1000 different addition problems I give you are punched in. You could build a machine that works like a look-up table and just does that. But it would be cheaper and quicker to build a different kind of machine - a calculator which correctly returns the right answer to *any* sum below the same amount. And the same thing applies even if there's no designer. If you knew that random shaking up of a home electronics kit had produced something that got these 1000 sums right, you'd be surprised if it didn't get other sums right too.

The same thing seems to apply to physical organisms as well. Selection for eyes that could e.g. let us distinguish different kinds of food and preditors and whatnot also gave us eyes that could distinguish different patterns of ink on a page. Selection for reflexes that e.g. help us dodge falling rocks also gave us reflexes which help us dodge flying tennis balls.

So, it wouldn't be very surprising if it turned out that the Tunesian Desert Ant is able to successfully compute the angle home after trips longer than it ever actually takes in the wild, or if our differential-looking-time behavior lead us to expect a range of numbers including the right answer even for very large sums.

4.2.3 From input-output to proto-belief

A very simple creature like the Tunesian Desert Ant's suggestively good mathematical behavior might just be a matter of always responding to one particular sensory input (e.g. from the legs and the jaws when the pick up food) with one particular behavioral output (turning around). But as we get to more complicated animals, we find that their behavior patterns can interact with one another. So we might imagine a more sophisticated ancestor of the ant which still turns around in accordance with the same algorithm (it might even be realized in a very similar way) but this turning behavior also takes account of:

- multiple sensory inputs (the creature might not rely on this calculation if it's felt itself being blown by the wind, or go home earlier if it sees a predator, or not bother calculating at all if it sees its home)
- internal state inputs (it might go home earlier if it gets too hungry, and there is food stored at home)
- multiple motor outputs (having calculated that home is at angle a, it might go straight home or stop and hide if it has seen a preditor and is in a region where it can camoflogue)
- real time modification to inputs and outputs: if the creature sees an animal which it had previously assumed to be innocuous eat one of its own kind maybe it will 'learn that animal is a predator', in the sense of acquiring the disposition to walk home

Kim Sterelney has proposed how certain kinds of environmental complexity might select for these more complicated behavior patterns [1], but whether or not he's right, surely there's no puzzle about how there could have come to be systems which integrate stored information from different sources like this to determine behavior.

I am calling these internal states with complex relationships to input and output pseudo-beliefs because of their suggestive similarity to beliefs

- multiple sensory inputs: many different observations may be relevant to forming a belief
- internal state inputs: feelings of hunger/pain/tiredness etc. may be relevant to forming a belief

[1] and I take the helpful terminology above from him

- multiple motor outputs: the same belief may lead to different actions depending on what other sensory and internal input you have
- real time modification to inputs and outputs: how you act on a given belief, or when you are willing to form it may change as you come to form other beliefs in response to experience.

We could also think about pre-linguistic-animals as having pseudo-theory-preferences in virtue of being disposed to alter their proto-beliefs in some ways rather than others e.g. they will be hesitant to eat melons that smell like melons that made them sick in the past, but not melons that are the same size as melons that made them sick in the past (just like a person who prefers the hypothesis 'all melons that smell like this are sick-making' over 'all melons this size are sick-making').

Let me emphasize here that I'm not suggesting any of these things have psychological reality. I'm just using the most convenient vocabluary we have for describing the complex conditional behavior patterns which even simple animals. All we need for our purposes is that pre-linguistic animals' brains have *some* kind of storage can play this kind of complex role in producing behavior, and that heritable mechanisms influence how this storage is used.

So, a more sophisticated way of dealing with mathematically shaped environments is to form proto-beliefs and proto-belief forming behaviors. Actual differential-looking behavior might be an example of one of these more subtly sensitive behavior patterns: If the monkey is hungry it might look less long at everything. If it has watched the things in question split or fuse frequently in the past, this may gradually decrease difference in looking time when it sees 'the wrong number'.

4.3 A nudge towards experience-matching theories

So now we see how natural selection could have produced behavior that looks suggestively like that of someone putting their mathematical knowledge to work. The baby looks longer when it sees two dolls added, one removed, and then anything but two left. The adult sees two dolls added, and one removed, forms a bunch of beliefs 'There were two dolls added and then one was removed. If there were 2 dolls behind the curtain at time t0, and only one was removed and no dolls were added or split or fused, there is now one doll behind the curtain. I didn't see any other dolls added or removed, and I don't think dolls spontaniously split and fuse, so I should be very surprised if there were anything other than 2 dolls left!' and then looks longer if he sees anything but 2 dolls.

But there's still an important gap to bridge if we want to explain away our second phenomenon: the nudge towards certain kinds of good (experience-matching) theories. The problem is to see how we can connect up (in a naturalistically acceptable manner) pre-linguistic creatures proto-beliefs, with language users' full-on beliefs.

How do heritable tendencies to behave a certain way before learning language wind up correlating with/producing tendencies tendencies to form combinations of beliefs and desires which yield the same patterns of behavior in the language using adult? How, for example, does a baby's liking for grapes (it tends to eat objects that look a certain way x) relate to the the toddler-it-becomes's liking for grapes (it believes that things which look like x are grapes, and it desires to eat grapes)?

This seems like a potentially important question for for cognitive science and perhaps philosophy of language, and I won't attempt to answer it. If this is a problem it is a very general one, and apparently quite independent from issues about the ontology or epistemology of mathematics. Thus, I take it I can assume, for present purposes, that some satisfactory answer can be found.

4.4 Summary

Now, putting this all together with the previous chapter, we have an overarching story about how we could have managed to get the kinds of beliefs which experience appears to bear out.

Mathematically shaped problems in nature produce creatures suitable behavior patterns. This then leads (in creatures with language) to prefer certain kinds of logico-mathematical theories over others (namely, the ones that would lead to these good behavior patterns). And then (if the creatures wind up having the leisure and inclination to posit general theories) Quinean theory revision irons out the kinks. Where evolution just gave us a fast and frugal heuristic rather than an exceptionlesly accurate law (e.g. inclination to accept the gambler's fallacy, failures at the wason card-selection task) and this inclination has lead us to propose a theory that conflicts with experience, recalcitrant experience leads us to revise. The end result is that we wind up with the kinds of beliefs about math and logic which experience largely seems to match.

The situation is much the same as with folk-physics. Evolution may have given us, along with other animals the dispositions to (behave like we expect) objects in the world around us to behave in certain ways. Not all these expectations are correct, but there's systematic pressure for us to get physical intuitions that are correct enough, in cases that might have been important in the context of evolution. This innate folk physics then influences what actual physical theories we posit, giving them a nudge in (usually) good directions. Now in the progress of science we test and elaborate these theories, extending them to contexts which would not have been relevant to survival on the savannah and correcting them when they fail.

Chapter 5

From practically benign beliefs to true beliefs

5.1 Introduction

In the past two chapters, I have tried to show that a naturalist can account for our having *some* practice, with the distinctive psychological and sociological features, that philosophers have traditionally been interested to note in mathematics. Specifically, (if what I say there works, it shows) there's no problem reconciling naturalism with our being inclined to do the kind of armchair reasoning which leads us to build bridges that stand up, nor with the fact that this reasoning has a distinctive phenomenology, and is much less susceptible to revision than typical scientific reasoning.

If the particular kind of naturalistic account suggested in chapters 2 and 3 works, this would be interesting in its own right.[1]

[1] Why? Note that if we look over the mechanisms involved in chapters 2 and 3, none of them appeal to any kind of philosophically controversial claims about what (if anything) mathematical statements *mean*. For example: it doesn't matter to Quinean revision, whether mathematical statements *actually* say anything about the course of future experience. All that matters, is that you are inclined to expect certain things from experience (e.g. to expect programs to halt, or to expect cannon balls to land in certain places) when you have calculated out certain results, which you would not have expected before you got them. For, if you have the disposition to go from certain results at the blackboard, to expectations about the course of future experience in some systematic way, (and the disposition to revise your methods at the blackboard and/or your tendencies to let these shape your future expectations in the face of recalcitrant experience) Quinean revision will lead you towards a combination of mathematics + tendencies to apply mathematics, which fits with experience.

And the upshot of this, is that even the most austere formalist (who thinks all mathematics is a matter of formal game-playing) can point to the results in chapters 2 and 3 to explain the (otherwise suspicious!) - See Godel's criticisms of Carnap in 'Is mathematics syntax of language?', in [Gdel, 1995a].- fact that the particular kind of formal games we are inclined to play, fit together with with our psychological tendencies to go from strings produced in mathematical games to expectations about future experience, in such a way as to yield correct predictions about computers, what will happen when you count apples and oranges etc.

But the aim of this thesis is more ambitious. As stated in chapter 1, I want to reconcile naturalism with standard Platonic realism about mathematics, by showing that realists don't need to appeal to any kind of spooky (i.e. non-naturalistic) faculty to account for our possessing substantial mathematical knowledge. In order to do this, I will need to show, not only that one can naturalistically account for our having a *practically good mathematics-like practice*, but that the realist can plausibly claim that engaging in such a practice would allow us to gain *knowledge of platonic facts about mathematics*.

That is: if you accept everything I've said up to this point, you might still think it looks miraculous

- that we wound up with a practically good practice which also happens to counts as one of saying largely true things about some realm of abstract objects
- that the largely true things we are inclined to assert about mathematics, also happen to be things that we are justified in asserting.

The next two chapters will be devoted to the first of these issues - to connecting up mere practical goodness with truth. (We will consider issues about knowledge and justification in the chapter after that). I will propose a realist answer to the access problem, which accounts for our possession of substantial true beliefs about mathematics without positing any causal contact with abstract objects, or spooky pre-established harmony.

5.2 The proposal

5.2.1 General picture of meaning and reference

My story starts with a vague, but popular, Lewisian picture of how words meaning and reference work in general (the next chapter addresses alternative theories of meaning). The picture is this: our words acquire the 'most natural' meaning compatible with the way that we use them. So, for example, our uses of the word "plus" could be interpreted either as saying largely true things[2] about the plus function, or as saying largely true things about Kripkie's quus function - if all you cared about was having an economical theory. But, the plus function is the more natural target, so people who have the pattern of linguistic behavior which is compatible with both plus and quus, count as meaning quus. The same story goes both for properties (like the three place relation "...plus...is...", or the property of "...is red") and objects (like llamas).

[2]There's some disagreement about whether we translate so as to maximize truth, reliability, similarity to the translator etc. This difference won't make a difference to what follows, but I do want to suggest that that if Quine's critics are right and there is a distinction between the sentences that feel analytic to us vs. those that don't (never-mind whether this distinction tracks anything of philosophical interest), then meaning may be determined in such a way that sentences which a speaker feels are analytic are more likely to count as expressing truths

Practical helpfulness fits into this picture insofar as it's a fact about linguistic practice. Actually, it's a very important one since it has a big effect on what we intuitively think it would take for a given potential interpretation to 'fit with' the way a speaker uses a word. Think about the following example. If a native is disposed to say "gavagai" when pointing to a bunch of boxes, and there are in fact rabbits in the boxes, this aspect of his practice counts in favor of interpreting gavagai as meaning rabbit. *But* if he is also disposed to be surprised and go back on this verdict, should he open the boxes, this counts (a little bit) against translating "gavagai" as meaning rabbit. More generally, when a pattern of reasoning fails to be practically good, it leads speakers to form expectations that a certain other line of reasoning or observation must come out a certain way, but things don't actually come out that way. (This is just how we defined practical helpfulness/harmfulness). So, failures of practical helpfulness *reduce* the extent to which we aim to interpret someone as believing truths. When someone is disposed to make inferences/observations that lead to their getting surprised, we want, ceterus paribus, to say that this is because one of their current beliefs is false, and going through these inferences/observations would lead them to realize this.

Now, since the controversies about Platonism focus on reference to objects, let's zoom in and note three things about how reference to objects works on this picture. Firstly, if there is no sufficiently natural object that fits well enough with our usage (e.g. allows for an assignment of meaning to our words that makes enough of what we say true) our expressions can fail to refer. So, for example, the name "Santa Claus", or the putatively referring expression "the fastest speed" will fail to refer. Secondly, there may be a range of different equally -or almost equally- natural objects, compatible with the use of a given expression. For example, some utterance of "that cloud up there", might be equally plausibly assigned reference to one of a number of different meriological sums of water molecules, because of the similar naturalness of these objects, and similar fit with our usage. In this case, reference is vague[3] between the different candidate reference.

These two points matter because standard Platonism says that we refer to abstract objects, and that we do so definitely enough that many questions about mathematical objects have determinate (i.e. non-vague) answers.

Thirdly, this standard picture of reference allows that the genuine referential structure of an expression need not be the same as the surface grammatical structure of that expressions. So, to choose an extreme example, ordinary people can be speaking truely when they say "there's a hole in the wall" even if there are no such things as holes. Expressions like this will be understood by appeal to some kind of paraphrase e.g. as really referring only to chunks of wall, and saying that they are arranged a certain way. The relevant complex paraphrase may not match the grammatical or linguistic structure of the expression very well - or even with the speaker's own expectations about the physics and/or metaphysics of the objects he is talking about.

[3]Whatever vagueness turns out to be

But, fourthly, people can disclaim the paraphrase when engaged in philosophical discussion. e.g. a philosopher who asserts that "there really are holes", can be saying something that's false, even if the mechanic's "there's a hole in your sink" is true. That is: if nihilism about holes is correct, the benevolent paraphrase that removes commitment to holes only gets deployed in ordinary contexts and not in philosophical ones.

I hope these points have outlined a familiar picture of how reference to objects works in ordinary cases. The Platonist claims that there are abstract objects, and thinks our talk of mathematical objects works in this ordinary way, much like our talk of tables and chairs. [4] So, now our challenge will be to see whether he can indeed make our reliable judgements about mathematics look like anything other than magic.

5.2.2 Abstract objects provide plentiful targets for reference

Well, let's start by unpacking what we need to show. In terms of the conventional picture of reference outlined above, what does it take for a variant on our mathematical practice to count as expressing largely truths about abstract objects? What we need is for there to be some collection of abstract objects which correspond closely enough to the usage in that mathematical practice, for expressions in it to be correctly paraphrasable as referring to those objects. But this means that, in order to make it not seem miraculous that we wound up with a mathematical practice that counts as expressing largely truths about mathematical objects, we need there to be (in the relevant sense) *enough* different patterns of abstract objects.

Let me say more. It needs to be the case that, if we had gotten a different mathematical practice, we would likely have gotten another one that counted as expressing largely truths. Now, we've already (in the discussion of Logicism) seen that some very bad, tonk-like practices can't plausibly be interpreted as expressing the truth about anything. But the work in chapters 2 and 3, provides a naturalistic explanation for why, if we had gotten a different mathematical practice, it wouldn't have been one of *those*. Tonk-like practices would be practically harmful and, mechanisms of natural selection and Quinean revision lead us towards practices that are largely practically helpful. So it's not surprising that we don't employ reasoning that's tonk-like.

But even once we toss out the practically harmful variants of mathematical practice, we are left with a wide swath of possibilities. If only, say, two of these possible variants on mathematical practice would count as expressing truths about mathematical objects, knowledge of platonic facts would still seem miraculous. 'How did we just happen to get one of the few mathematical practices which have suitable abstract objects corresponding to them?' the critic of Platonism might well ask. Hence what we need is for there to be

[4] This smooth semantics for mathematical language, which makes it work just the same as not mathematical language, is arguably one of the main selling points of Platonism - see Benaceraff's problem article CITE

'enough' mathematical objects for all (or most) of the different possible variants of mathematical practice which we might have gotten, to have suitable objects corresponding to it. That is: there need to be enough different structures and patterns of mathematical objects for each of these mathematical practices to be interpreted as talking about some of these objects.

Putting this all together: what we need for mathematical knowledge to not look mysterious, is for there to be **enough differently structured combinations of abstract objects to provide target domains for each of the different practically benign alternative mathematical practices that we could have gotten**.

5.2.3 Contrast with Less Pedantic Formulations

In a sense, this is just an overly pedantic way of arriving at an intuition which is commonly used to motivate both Plenetuidnous platonism and Structuralism. Namely: 'Knowledge of mathematical objects would not be naturalistically puzzling if there were mathematical objects 'corresponding' to every possible 'coherent' mathematical theory we could have gotten. If all possible mathematical objects/structures are real, then there's no mystery about how we managed to get substantial mathematical knowledge, beyond the mystery of how we managed to come up with a coherent theory. Whichever coherent theory we came up with, there would have been some suitable mathematical objects conforming to that theory, which we could have counted as talking about. Hence it might be arbitrary which coherent mathematical theories we developed, but whichever ones we developed would have counted as largely correct.'

However, pedantry pays off in this case, because there are well-known problems for the view naively stated above, which I will argue do not apply to the more limited and pedantic view stated above. For one thing, if 'coherent' means logically consistent, then it can't be the case that all logically consistent mathematical theories are true. For, there are logically consistent theories whose conjunction is not logically consistent (consider: "∃x ∃y - x=y" and "-∃x ∃y - x=y").[5] But if 'coherent' doesn't mean logically consistent, it's not at all clear what it does mean (for example, the statement "all possible mathematical objects exist" is something that *everyone* would agree to, so this can hardly be the content of plenetudinous platonism). [6]. Also, there's a natural worry on the picture above about how there can ever be a right answer to logically independent questions: wouldn't the theories you get by adding P vs. adding -P be equally 'coherent', and hence equally true? So, how can there be a right answer?

Furthermore, we will see that the profusion of abstract objects needed is actually much smaller than you might think. In particular:

- The existence of the hierachy of sets will be enough for the reductive platonist.

[5]CITE boolos
[6]CITE Eklund

- The existence of distinct arrangements of fundamental objects with each of the first order structures modelable in the sets will be enough for the non-reductive platonist.

Finally, note that we are not claiming that all consistent mathematical theories are true, but rather all/most sufficiently practically good practices are correctly interprable as being *largely* true.

In a sense this should be obvious: what we are trying to give a Platonist-Naturalist account for is human access to mathematics, and our beliefs about mathematics are clearly not completely correct. So, if I proposed a story on which all mathematical theories people were likely to generate would be completely true, it would be clear that I was saying something wrong. Rather, what we need to keep in mind is the sense in which it feels intuitively obvious that Euler had mathematical knowledge. This is what the Platonist must accept, and this knowledge of Euler's is what the Platonist must account for.

In particular, note that Euler arguably had false beliefs:

- about the philosophical/physical significance of geometry (he might well have thought that space necessarily obeyed Euclid's axioms)
- about various particular mathematical questions (even the greatest minds can make mistakes, this is why they wait four years before giving out the Fields Medal[7])
- about foundational issues in philosophy of math (I don't know much about Euler's particular philosophy of math, but if formalism were true and he was a platonist, or if platonism were true and he was a formalist, neither of these would be a reason to deny the intuitive claim that Euler clearly knew lots of mathematics)

My task in this chapter, and the work as a whole, will be to show that Platonist doesn't need to invoke anything miraculous to account for this this kind of limited, but still impressive, mathematical knowledge: almost any kind of practically benign variant on our current mathematical practice would do equally well.

5.2.4 From practical goodness to accuracy

Now, here's the argument. Suppose that some Platonist ontology is correct: either there is a single fundamental kind of mathematical object like the hierarchy of sets (you might call this view classic, or reductive, Platonism), or there are distinct fundamental objects with all possible first order structures (as per Plenetudinous Platonism). [8]

[7] This is the equivalent of the Nobel Prize in mathematics

[8] Logically, a third possibility would be to hold that there are a few distinct fundamental objects. For example, someone might think there are natural numbers and sets as distinct fundamental objects, but claims other mathematical objects like real numbers or topological spaces are really

1. **Practical Goodness** The story in Chapters 2 and 3 provides mechanisms of attraction towards practically good theories. Thus, no miracle is required to account for our getting (putative) logico-mathematical intuitions which are at least practically benign.

2. **Logic** This explains why we would wind up with patterns of logical inference like those for 'and' and 'or' rather than those for pseudo-connectives like 'tonk'. For, as mentioned above, reasoning with 'tonk' allows one to get any conclusion from any premises, and hence has a good chance of leading one from the true to the observably (or otherwise inferably) false. Thus, these naturalistic forces which lead to practical benignness can account for our winding up with some (largely) correct logical system, like first order logic (rather than a bad system which allowed). For example, we can see every day how Quinean correction leads people away from bad patterns like inferring the consequent.

   Now, admittedly, such forces cannot be expected to lead to a *unique* logic, in the sense that a formal system which substituted the Sheffer stroke for the usual "and","or", and "not" of standard first order logic would be a different logic. But such differences are surely irrelevant to the task of explaining how we managed to get substantial mathematical knowledge. People who used the Sheffer stroke instead of our familiar connectives would surely count as having mathematical knowledge just as much as we do. More interesting variations on classical logic are also possible. So this kind of variability poses no problem that mere naturalistic correctives could have lead us to a logic like standard First Order Logic.

   A potentially more threatening kind of variability concerns alternative patterns of logical reasoning which would be largely practically good, but are not mere notational variants. Does our access to mathematics depend on our being eerily lucky enough to have gotten classical logic as opposed to one of these? Well, what would we say about the mathematical knowledge of a population that accepted intuitionistic logic, or dialethic logic? It seems to me that intuitionist could perfectly well count as knowing lots of math, even if they are missing out on opportunities to learn various things by proofs from contradiction (since they don't accept the law of the excluded middle). Even a community of dialethic logicians (i.e. people who hold that there can be true contradictions) could count as having substantial mathematical knowledge provided that their deployment of contradiction was disciplined enough not to lead to tonk-like conclusions. Hence, I propose that practical benignness considerations make it unsurprising that we got *some* largely truth preserving system of logical inferences, and it's arbitrary for our purposes that Classical First Order Logic is the particular one that become dominant in the actual world.

about either the numbers or the sets. Since I don't know anyone who holds this position, I won't explicitly consider whether it allows for an analogous account of mathematical knowledge. But I would tentatively guess it does.

[From this point on, I'll assume in what follows that we are talking about mathematical practices which include classical first order logic, simply because this lets us *relate* questions of practically benign practices to mathematical notions like (first order) consistency or having a (first order logical) model. The popularity of first order logic means that we have relatively good mathematical grip on how theories that contain FOL can relate to patterns of objects, and various ways in which reference might fail.]

3. **Consistent or Prune-able to be Consistent** If a practice containing first order logic allowed for the *easy* derivation of first order logical contradiction, then it would not be practically benign. So, if some practically good practice is first order inconsistent, and hence allows for the derivation of contradiction, then this derivation must not be 'easy'. More specifically, the paths that lead to this contradiction must not be part of the core common usages of the practice (otherwise the mechanisms of Quinean revision etc would be able to weed them out).

   But if this is true, then (I claim) then such a practice can be plausibly understood as formed by a central fragment which is first order consistent, together with some additional false beliefs. Examples of this kind are easy to find in history. Consider people who made inconsistent arguments about continuity and limit, before the epsilon delta definition of continuity. Do we understand them as completely failing to assert true things, or talk about mathematical objects? Certainly not. Rather, we say that they knew many true things about numbers, and even about limits. We cannot interpret them as talking about mathematical objects which have all the properties they expect, for their beliefs allow for the derivation of contradiction. However, we can interpret them as talking about objects (the numbers) whose properties were correctly reflected by a large, central, fragment of their mathematical practice.

   Hence, we can assume that we are dealing with a practice that either is first-order consistent, or has some central fragment which is.

4. **Enough Candidates for Reference: May The Best One Win** Finally, the completeness theorem for first order logic assures us that there are models for every first order consistent theory in the heirarchy of sets, and our plenetudinous assumption yields that there are fundamental objects. Hence, we are assured that there's an object that *at least* matches up with the pattern of first order logical inferences that makes up the practice. The assumption of practical goodness also insures us that the facts about this target object will also match up with the pattern of inferences about concrete objects, and practical applications commonly made by people with that practice.

   However, this doesn't mean that there is a unique object that matches the first order aspects of our use. Except for the case where we already accept a complete first order theory for a given area of mathematics[9], there will generally be *many different* arrangements of mathematical objects, which pass this bar.

Nor does it mean that the practice will count as referring to some such object that perfectly conforms to the first-order features of the practice. For, as noted above, interpretation generally takes into account **both** the naturalness of the target and the extent to which our assertions come out true of that target. We are perfectly happy to interpret a community as thinking falsely that crystals cure cancer, rather than truly ascribing the property of being-a-crystal-or-curing cancer to crystals, insofar as the latter property is both a more natural target, and fits better with other aspects of their use.

Rather, what we have is an assurance that the class of mathematical ensembles that match the practice 'reasonably well' is non-empty. We have a collection which at least makes the first order inferences mathematicians are likely to draw come out to be exactly correct. Within this class, reference will accrue to the most natural collection of objects that best fits the practice (all aspects of it, not just the first order ones), with ties leading to vagueness - just as happens in the general case above.

If all the candidates are equally natural, then claims which are true of some candidates but false of others will be vague in their truth value. A plausible example of this would be some expression like 'the set of polygons' as it occurs in the ongoing dialog of Lakatosh's Proofs and Refutations.

In contrast, if some candidates are more natural than others (or if there's a class of maximally natural candidates), then there will definite facts about all sentences which have the same truth-value when interpreted as referring to any of the most natural objects, but the correct theory of vagueness (whatever it is) will apply to sentences which differ between equally natural objects. This fits with the widespread intuition that there are definite facts about all questions in number theory (where we think that the smallest models of first order axiomatizations of number theory are especially natural candidates[10], and must agree on all first order questions).

Thus, mere practical helpfulness, leads to (a certain degree of) truth, even on the Platonist picture. The crucial points at work are -on one hand- the profusion of mathematical objects, and -on the other- the structural focus of mathematics.

[9] That is, except for cases like algebra (?) for the reals, where mathematical practice tells us to accept statements that already allow one to derive the truth or falsehood of any statement

[10] most natural both intrinsically, and most natural in the interpretation they allow one to give to expressions like "smallest", which we use both to characterize the numbers "the numbers are the smallest collection containing 0 and closed under successor" and in more concrete scenarios "that is the smallest list of books which contains one work despised by each of my advisors"

Hence, (putting steps 1-4 together), it is not miraculous on the Platonist picture that, in getting a mathematical practice that's practically helpful, via the mechanisms of chapters 2 and 3, we also wound up with one that leads us to so much true belief.

5.3 Filling out the picture: proofs, definitions, and general mathematical intuition

In the previous section, I tried to answer puzzlement about how creatures like us could possibly have gotten the ability to form reliable true beliefs about realist mathematical objects. However, I will admit that the practice described above doesn't look very much like the mature mathematics we know and love. So, now I want to add some complications.

In the basic scenario we have informal (unarticulated) intuitions about how mathematical facts should be related to observable facts about certain physical set ups. A process of expanding and revising these expectations then either fills in correct beliefs about what the-abstract-objects-which-allow-these-applications are like, or gets one to give up the expectation (maybe the applications you expect are incompatible with each other). Then eventually our reasoning about these objects (e.g. "3" or the 'the triangle in platonic heaven') extends far enough beyond the initial application, that the initial applications are no longer central to the sense of the term. In this way we get the current situation where we expect certain mathematical facts to match up with the results of certain physical processes of counting, measuring etc. but we don't think this match is necessary.

Now there are three extremely central aspects of mathematics which are missing (or at least not explicitly mentioned) in this story. The first is proof, the second is stipulative definition, and the third is general mathematical intuition.

5.3.1 Proof

Proof already has a place in the picture above, but given the obvious importance of the notion I want to emphasize it. A notion of proof (in partiuclar, dispositions to accept certain inferences and not others in the course of a mathematical proof) are one half of what the methods of revision mentioned in our first stage acts on. On the one hand, we have informal expectations about how facts about "3" will relate to facts about the result of concrete counting procedures. On the other hand, we have ways of forming and rejecting beliefs about what the facts about "3" are. These can come into conflict, and they are what get revised.

I should stress here that the critical items to consider (on this account) are *dispositions to accept certain putative proofs*, not *beliefs* about what's a proof. Similarly, when I talk about implicit expectations about how certain mathematical facts should mirror facts about concrete physical objects, I mean dispositions to go from accepting some claim about the mathematical objects, to believing some claim about the physical objects. This is important because of Lewis Carrol's point in 'What the Tortuse said to Achillies'. Mere beliefs on their own don't generate new beliefs without some procedure of inference. So, the fact that we have dispositions to make mathematical/logical inferences -not just nod yes to some finite list of mathematical beliefs- is a crucial part of how we do mathematics. Furthermore, insofar as we are trying to account for how creatures like us can get math and logic right, it won't do to just take inferences for granted (any potentially puzzling mathematical belief can always be redescribed as a rule of inference that allows one to infer that belief from empty premises).

This way of thinking about things also fits nicely with common sense intuitions about history. An ancient Greek who expected the results of measuring physical triangles to be similar to those calculated for abstract, platonic triangles, need not have had any kind of explicit theory about what the relationship between facts about abstract triangles and concrete ones that went beyond vague talk of similarity. It seems bizzare to say that he implicitly or subconsciously knew Tarski's (?) elegant theory relating facts about geometry to claims about equidistance and betweenness relations on points in space. But what such a Greek *would* have is very definite dispositions to go from making certain measurements to expecting certain calculations to come out a certain way, and vice versa. Hence it's attractive from the point of view of common sense, as well as rigor, to think of our expectations for how math should apply to the world in terms of inference dispositions, not just beliefs.

[maybe this should go earlier]

Thus proofs are really central to the story told above. Giving putative proofs (i.e. mathematical arguments one is disposed to be compelled by) is what leads to conflict and revision. And the result of this revision is - if the story works- an expansion and reshaping of what inferences a person will and won't accept, towards increasing soundness and completeness with respect to the facts about the relevant mathematical objects. I have only focused on the issue of accounting for how we get good reliable notions of what's a good proof, because this the most intuitively puzzling part of the story. Once we have got our hands on such a good (informal) system of proof, all that remains is to actually give proofs in that system (i.e. proofs that we are disposed to accept as compelling) and we can thereby mill out much mathematical knowledge.

5.3.2 Stipulative definitions

In contrast, stipulative definitions are a core part of current (mature?) mathematical practice which is genuinely missing from the picture above. At no point in the story above was "3" or "triangle" given a stipulative definition. But stipulative definitions are ubiquitous and play an extremely important, and perhaps even distinctive role in mathematics, so it will be important to see how they fit in.

Stipulative definitions function much as Kenny Eswaran has argued axioms do, to focus disagreement. Rather than each person working with their own intuitive notion limit or straight line, a stipulative definition of limit or line lays down a key fact which all participants will then agree to. In this way, stipulative definitions can prevent people from talking past each other, and increase our confidence in mathematical results which can be shown to follow from the stipulative definition (plus a limited range of inference procedures which we are already very confident about). This increased confidence in a stipulative definition tracks genuine facts about reliability, because the very act of accepting something as a stipulative definition can make it more likely for that sentence to express a truth.

However, it's important to notice certain limits to the power of stipulative definition. Given our wide freedom to make many different (potentially unmotivated and absurd, but mathematically completely acceptable) stipulative definitions, you might think that we can just accept stipulative definitions at random, or that these can be a source of mathematical knowledge. However, as we saw in chapter 1, this is not the case. Accepting arbitrary combinations of claims as an implicit definition of some new term is not a generally safe procedure. Burge's example of tonk demonstrated this. Instead, our ability to reliably form true beliefs by making definitions turns on a kind of antecedent mathematical insight. We can reliably judge that definitions of a certain form (e.g. something is a fi-prime if and only if it is a prime and greater than five) will not allow for the derivation of any new false consequence when combined with our other mathematical and logical reasoning, and hence accepting this stipulative definition does pose a danger of leading us into false beliefs. And when confronted with more complex forms of implicit definition we can make the same kind of judgement, though sometimes with more explicit reasoning (e.g. showing that some function on objects in a domain genuinely induces a corresponding function on equivalence classes of that domain), or with less confidence (e.g. complex implicit definitions motivated by an informal description).

Thus, accounting for our ability to make successful implicit definition is something that any adequate platonist account of our ability to get math right must do.

I propose that that we get to successful (i.e. non-tonk-like) implicit definitions in three ways.

The first is to start, with an informal practice (like the practice of talking about "3" mentioned in the section above) and then to articulate some of the central principles and methods of inference that formed part of the practice - and stipulate that these are to count as a definition from now on. In this way we get a systematic explanation for the two phenomena which doggedly raised problems for old-fashioned stipulative-definition based accounts of mathematics in chapter 1. We will tend -though certainly not infallibly, as history shows - to make coherent (e.g. logically consistent) stipulations because they are fragments from a practically coherent practice, which was revised whenever any of its inferences lead to conflict. And our newly minted stipulative definitions will come with concrete and largely correct expectations about how facts about the items/properties from the stipulative definition should relate to the observable world, because the stipulations inherit them from the preceding practice.

But not all successful implicit definitions are inspired by some concrete scenario, or intended applications.

The second way of getting to acceptable stipulative definitions involves explicit mathematical reasoning. If we already accept principles that allow us to prove there is a unique function or set corresponding to a certain definition, then we can explicitly define some term as the name for this object or relation. I take it this part of the story is uncontroversial, since it's an explicit part of ordinary mathematical practice.

More interestingly, the third way of getting an acceptable stipulative definition involves informal mathematical reasoning, and sometimes mental or physical pictures. Here the phenomenology may be that you seem to be visualizing (there will be many different methods of visualization) how a function could behave the way your stipulative definition requires, or it may simply be a direct inclination to judge that some implicit definition is acceptable. Here we have arrived at something that looks the most like the infamous 'spooky' rationalist faculty to just see that some definition is consistent (and indeed corresponds to some aspect of platonic reality), or that facts about an object introduced by mere stipulation will then have certain definite consequences for facts about concrete objects. But, notice that the same mechanisms which discipline and correct our judgements about mathematical objects posited for some specific reason, can apply -indirectly- at this higher level as well!

For consider a) someone's methods for accepting stipulative definitions and b) their methods for going from claims about newly minted objects with stipulative definitions to claims about concrete objects (e.g. from facts about Turing machines, to facts about the human female computers of Turing's day). These *methods* have a possibility for leading to conflict or agreement, and these methods are subjected to correction. We can say things like 'I thought there couldn't be multiple functions satisfying this description because I was assuming that every function was something you can draw like this...' and then modify this bad method of inference.

So again we have an explanation for the facts that troubled old positivist accounts. It's no surprise that we manage to get consistent stipulative definitions, and correct expectations for how facts about the objects stipulatively defined will relate to the world, even when using general mathematical intuitions to posit novelties, because these methods of general mathematical reasoning (implicit heuristics for deciding when definitions are consistent, and when some applications will succeed) are themselves corrected by experience and revision. (Note that - as usual - this doesn't mean that you need to cite experience when justifying you in proposing some intuitively acceptable stipulative definition. Indeed, insofar as we pick up methods of mathematical reasoning from our teachers, notation and general culture, the relevent experiences may be ones that didn't even happen to you.)

5.4 Objections

The remainder of this chapter will now be devoted to addressing some natural objections to the claim above.

5.4.1 Contrast with Science: Philogiston

The first objection I'll consider is arguably a bit of a straw man, but I think it may be helpful to just say what's going on in this case. A standard objection to plenetudinous platonism is that consistency doesn't imply truth in the physical realm, so it would be odd if it does in the mathematical realm. As Micheal Potter puts it, we can perfectly coherently post a winning American rugby team, or a northeast passage, but that coherence doesn't make the posit true.

So, you might wonder, how can I say that mere practical helpfulness (which is a kindof coherence insofar as it involves raising expectations that then get satisfied) is enough to lead to substantial correctness? Doesn't this lead to the view that claims about philogiston or winning American rugby teams are true since these posits are perfectly consistent?

The key point is that practical goodness doesn't just mean inconsistency. I can have a perfectly *internally* logically consistent system which proves statements about what numbers I should get when I count rhymes or oranges, which yields massive practical failure at every step. The sense in which it's consistent or 'coherent' to posit a winning american rugby team is just this kind of internal consistency. There's no contradiction to be derived from that sentence alone. But when we ask about practical goodness, which is a matter of how the expectations raised by one reasoning subsystem fit into the rest of a creature's pattern of thinking, we see that the practice assuming there's a winning american rugby team is quite practically bad. Someone who adopted this, while keeping the rest of their reasoning largely the same (I take it this is what it would mean to posit a winning american rugby team), would be disposed to have all kinds of experiences that surprised them. They would, for example, reason to the conclusion that this team should occur in sports almanacs, and then when they deployed their normal procedures for checking whether a team is an an alamanac they would get a surprise.

On the other hand, if we try to keep fixed the assertion of the sentence "there's a winning American rugby team", while changing our subject's methods for checking sports almanacs, ways of forming verdicts about what teams are playing when they attend rugby matches etc. it becomes less plausible that the sentence they are using means the same thing it does in standard English. So it's hard to see this as a case of positing that there's a winning American rugby team, in a way that's practically good, as opposed to positing some other (likely true) statement.

5.4.2 Contrast with Science: Empirically Adequate Physical Theories

A better place to look for problems in how this theory applies to science might be empirically equivalent (but incompatible) theories in physics. It's plausible that, whatever the right theory of physics is, there will be variations on it which lead to all the same conclusions about what people can observe, but which posit a different and incompatible fundamental physical structure to the world. Take some normal physics and the equivalent theory in terms of brains in vats, or take special relativity and an equivalent theory in terms of Lorrenz Contractions. Such pairs of theories lead to the same "observable" conclusions, and will be similar in practical goodness.

My theory yields the (possibly unattractive) conclusion that people proposing these empirically adequate but ultimately wrong theories would indeed have substantial physical knowledge. Obviously, how much of a bullet this is to bite depends on what substatial means, so lets recall the sense in which this theory purports to explain our having 'substantial' mathematical knowledge. Euler was our example of the kind of imperfect but impressive mathematical knowledge we needed to explain. So here's what we get: in the same sense in which Euler counted as knowing many things about number theory, even while he was (perhaps) wrong about fundamental philosophical questions about the nature of the numbers, or the physical significance of geometory, proponents of Lorentz Contraction or Brain-In-Vat physics count as having substantial physical knowledge, even though they are wrong about how to understand the apparent shrinking of physical objects moving near the speed of light, or the microphyical goings on that give rise to the patterns in sensory experience which they know very well.

Note also that the sense of substantially here isn't the (interesting, but different) sense in which we might say Newtonian Mechanics was largely right, but strictly wrong. It would be awkward to compare the imperfections in Euler's beliefs with that. For (surely we want to say) most of Euler's beleiefs about number theory were literally and completely true. It's not like $e^{i\pi}$ is only *approximately* -1! But the same thing plausibly holds for our proponents of empirically equivalent but wrong physics. The proponent of Lorrentz contracting knows many many literally true things about how various measurements will turn out. They know exactly what various clocks should read, and how to build machines to do various things etc. Similarly the proponent of the Brain-In-Vat theory knows lots about patterns in how their experience should unfold, they are just wrong in thinking that the "apples" whose behavior they know lots of literal truths about are data structure in the Vat rather than physical objects.

Hence, I don't think this is much of a bullet to bite.

5.4.3 Boolos and plenetudinus platonism

As noted above, the theory advocated above may seem reminiscent of ontological maximalism (the view that there are mathematical objects corresponding to all consistent theories) or neo-Fregeian views, on which new kinds of abstract objects can be introduced by merely laying down sufficiently clear, consistent, 'abstraction principles' saying when two such objects are supposed to be different from one another. So it's worth checking whether George Boolos' famous objection to these views applies to my theory as well.

Boolos points out that there can be consistent mathematical principles which are not consistent when combined [11]. His most famous example presents abstraction principles for both cardinalities and "parities". While each principle is consistent on its own, the principle for cardinalities requires one consider infinite collections while the principle for parities precludes this. Thus, accepting both principles leads to contradiction. This might seem to show there are two practically good variants on matheamtical practice which can't both yield (largely) true conclusions.

But note that anyone with a practically good practice matching the abstraction principle for parities, can't be inclined take any property to apply to infinitely many things. If they did, they could actually encounter the Boolos' problem and hence be surprised. EXPAND? Thus, while one possible interpretation of their behavior would be to say that they accepted the full abstraction principle for parities (which requires that there are only finitely many objects), an (arguably superior) interpretation is available:

Why not interpret those who appear to accept the abstraction principle for parities as having some kind of implicit quantifier restriction to the finite collections of objects which they are actually willing to make claims about. That is, one might think that people in this situation would not really count as accepting the abstraction principle, but rather a weaker theory, which only says that there are parities associated with *certain* properties (e.g. properties which apply to objects within *within the relevant restricted domain*).

5.4.4 A priori access to morals and theology - does this story prove too much?

Second, the appeal I have made to suitable facts about meaning/charitable interpretation (if we had some other coherent math-like practice, we would count as having true beliefs about some other part of the range of fundamental objects) can seem like black magic. Isn't this all a bit too easy? Why couldn't you tell the same story about a priori access to moral facts, or theological ones?

The reason why the same story won't work for morals or theology is this. Mathematics allows for, and indeed involves actively seeking out, new kinds of structures, and ways that objects could (in principle) be related to one another. So, even if we posited, and reasoned coherently about, objects that looked totally different from the ones that we have now, we could easily still be getting at mathematical truths. Thus, it's plausible to say that if our reasoning at the

[11] The Standard of Equality of Numbers. In George Boolos (ed.), Meaning and Method: Essays in honor of Hilary Putnam, 26178. Cambridge: Cambridge University Press. Reprinted in Boolos 1998: 20219.

blackboard could had a totally different (but consistent and practically benign) shape then it now does, we would still count as having the same degree of access to mathematical knowledge.

In contrast, in the moral case, when we imagine different practices of writing certain consistent things down on blackboards and then coming to, say, aspire to, praise or blame various actions, we don't feel that people engaged in these would count as having equal insight into the moral facts. People whose patterns of praise, blame, aspiration etc tracked totally different things than our own, would count as either not having any moral insight at all, or as being deeply mistaken about morality. People with radically different (coherent) mathematical practices strike us as learning about something equally interesting but different, whereas people with radically different (coherent) moral practices strike us (or, at least, strike the moral realist) as actively getting things wrong.

Contrast what an ancient Greek's reaction would be to discovering arabic mathematics (we've been doing geometry, they are doing this totally different thing, algebra - cool, let's go learn about it!) vs. discovering regions of arabic ethical reasoning that differed as widely from their own (they think plays aren't pious, but actually immoral - that's crazy!).

Perhaps the proponent of rational intuitions about morality could argue that all the physically easy ways of getting a practice that looked like it involved praise and blame, and aspiration etc. would substantially agree, and embody equal moral insight. There might, for example, be some evolutionary benefits to cooperation, and cheat detection etc. But even if they could argue in this way that, given that the moral facts are a certain way, it's no surprise that our moral intuitions correctly track them, it will still look quite miraculous that the most evolutionary beneficial patterns of cooperation, and the morally best ones, should happen to be the same. In contrast, it does not seem miraculous that our world behaves in ways that can be helpfully described by comparison to some (of the vastly many different) structures available to mathematical objects.

Similarly, in the case of theology, it's hard to imagine that structurally very different variants of current reasoning about theology, would all embody equal degrees of insight into the nature of the divine. Even in the actual world, there are many different, largely internally coherent, theological practices which we interpret to be making radically incompatible claims - hence we can't claim that any different, coherent, way of doing theological reasoning would embody an equal insight into theological facts. (Though perhaps the ancient Romans, and modern Unitarians would disagree with this claim?)

To repeat, (I claim) one can't imagine an activity which a) felt just like our reasoning about math and logic, only with different sentences feeling obviously compelling and b) was practically benign - so e.g. it didn't lead to any unexpected contradictions or false empirical predictions - but was hopeless at getting at the truth. In contrast, one *can* imagine and activity which felt just like our reasoning about morals and god, only with different sentences feeling obviously compelling, but was practically benign.

5.4.5 Independent questions

A point that is worth reiterating here, is that this account still allows us make sense of the intuitive notion that it's possible to get mathematical questions which are independent of all currently accepted axioms *wrong*. It might seem troubling that once we have assured ourselves that a potential new axiom is independent of our current axioms, the reductive platonist must admit that adding the axiom would lead us to state where most of our mathematical practice still count as giving us knowledge, and so would adding its negation! So, how can there be a right answer about whether the axiom is true?

Note that the challenge here, isn't to explain why all our mathematical views are correct (mathematicians can make mistakes), but rather to explain how our mathematical beliefs are largely correct, and our methods of forming them are reliable, to whatever extent is required for these methods to count as giving us mathematical knowledge.

So, what happens if a proposed new axiom is false, but my community comes to accept it? Well, there are two possibilities. One option is that adding the axiom makes a small difference to their resulting mathematical practice. In this case there's no change in meaning, and that particular axiom expresses a falsehood, but the result is still a community that's reliable enough to count as having substantial mathematical knowledge. Alternatively the mathematical practice might be so radically altered that, on the old interpretation, its conclusions would be largely unjustified. But in this case, the radical change in practice will also suitably change the meaning of the relevant mathematical terms, yielding a new interpretation on which the conclusions condoned by the practice would (largely) count as mathematical knowledge.

5.5 Conclusion

Putting all of the above together, we get an answer to the access problem which satisfies our initial characterization of realism about mathematics, and fits other common intuitions as well. As required by my initial stipulation: there are mathematical objects, mathematical claims can be true or false independent of any human activity, and we can meaningfully wonder about questions that are independent of our currently accepted axioms and formal methods of reasoning. But adopting this view also yields some other intuitive consequences: People are reliable, but not infallible about mathematics. It's easier to count as being right about math, than about philosophy of math (e.g. claims about which mathematical objects are most fundamental). And, in fact, the strength and prevalence of a mathematical judgement correlates with its reliability.

We get the last result, that mathematical judgments that feel obvious, or are widely shared, are more likely to be correct than weaker or more controversial mathematical intuitions, because the former carry more weight in our interpretive practice. That is, the correct interpretation of a mathematical practice will (ceterus paribus) take more pains to make wide-spread or obvious-feeling judgements come out as true than tentative or controversial ones. Thus, while we don't get the kind of perfect certainty regarding mathematical intuitions about, say, large cardinals, which has struck many as deeply counterintuitive, we can still justify our extreme confidence regarding less controversial mathematical claims.

Chapter 6

More about the Meta-Semantics

6.1 Introduction

In the previous chapter, I tried to put together results from chapters 2 and 3 with common particular intuitions, and a popular/orthodox Lewisian view about how the meanings of our words are determined, to yield an account of mathematical knowledge according to Platonism. However, almost nothing is really orthodox in philosophy, so in this chapter I will try to show how the limited meta-semantic claims needed for the argument in the last chapter relate to some conflicting ideas about meaning and reference.

The biggest intuitive source of worry concerns the possibility of talking about abstract objects.

One reaction you might have to the previous chapter, is that if Lewisian Orthodoxy allows for such an easy account of mathematical knowledge, this shows Lewisian Orthodoxy must be wrong. Is it really plausible, even assuming that mathematical objects exist, that merely engaging in a practice with the kind of anthropological features discussed in chapters 2 and 3 would enable you to talk about such objects?

In this chapter I will consider this question from the point of view of some general theories in philosophy of language. I'll start by addressing the intuition that reference requires causal contact, so nothing could possibly count as talking about the Platonist's abstract objects.

Then, in light of the considerations which tell against this causal constraint I'll define a notably liberal notion of what it is for a mental state, or utterance to "carry information" about a state of affairs. Then (making occasional use of this notion of information carriage), I will sketch a positive picture of why engaging in a coherent practice of the kind discussed in chapters 2 and 3, might qualify as talking about abstract objects - according to the standards imposed by a variety of different contemporary theories of meaning.

Finally I'll argue that the story in the previous chapters actually allows us to 'see how' reference to abstract objects is possible, on views like Linnebo's and Field's where reference requires certain kinds of reliable covariation.

6.2 The causal constraint

So, let's start with the causal constraint. The Platonist wants to say that claims about "3" are claims about a certain abstract object, 3. But, we have never had any causal contact with this abstract object, and indeed we never could. Thus the Platonist is committed to denying that one can only have names for objects with which one has had causal contact. Is this a problem? I am going to argue that it isn't, because there are sufficient independent grounds for thinking that the causal constraint can't be right.

The shallow argument for this conclusion, is that we can introduce names by stipulating that a given name is to refer to the unique object (if there is one) which satisfies a given description, but such names form a counterexample to the claim that we cannot have names for objects which we have not causally interacted with. So, for example, we might say "Let 'Blargo' name the unique fattest sentient being in the universe, if there is one". Then, if there is such a creature, it would seem that our word "Blargo" refers to him/her/it.

However, one might say that the implicit use of a definite description shows that what's going on isn't 'really' a name (even if subsequent time erases all memory of the verbal stipulation by which the name was introduced). Perhaps there are two ways of talking about objects - you can name them (and this only works for objects you have causally interacted with), or you can 'talk about them' in a looser sense, by using definite descriptions which are logical combinations of names of things you have causally interacted with, together with property terms and logical vocabulary. Thus, one might argue that, we cannot employ names for mathematical objects, and (unless we can provide a definite description which uniquely picks out the numbers or the sets), we cannot talk about them indirectly either. This brings us to a deeper objection to the causal constraint on names.

The deeper objection attacks the motivation for the causal constraint, as follows. If you think that a name like "Jim" can only come to make a definite contribution to the truth conditions for sentences containing it, by way of a speaker having causal contact with the person named, then how can you account for the apparent meaningfulness of property words and logical constants? Presumably, phrases like ''...is red' or '$\exists$ x' also make a definite contribution to the truth conditions of sentences which contain them. But there are no objects associated with these terms (except perhaps properties like redness) for us to have causal contact with. So, it's hard to see how causal contact would be needed, before a term like "3" could come to make a definite contribution to the sentences it occurs in.

To put things more positively: if you think that property words like "... is red" can come to stand for e.g. the most natural/simple/eligible-for-reference, *property* that applies sufficiently often when your community uses the word "red", then why can't object-words like "three" come to refer to the most eligible-for-reference *object*, whose features largely accord with the way your community uses the word "three"? That is - why can't whatever factors you think fix the contribution which words that *aren't* names make to the truth conditions of sentences, also sometimes work for words that *are* names, like "3"?

6.3 Reference to mathematical objects, on three major theories of meaning

So much, by way of trying to dispel the notion that we can't have names for objects which we can't causally interact with. Now let's turn to the more positive part of the story. Why would you think that engaging in a useful, coherent, practice (the kind that might be forged by the forces in chapters 2 and 3) could count as forming beliefs about abstract objects?

To my knowledge, the three major contemporary theories of meaning are: causal (Fodor and Dretske), teleofunctional (Millikan), and interpretational (Davidson).

6.3.1 Information Carriage

Both causal and teleo-functional accounts appeal to (something like) a notion of a mental state (or linguistic item) 'carrying information'.

Fodor defines 'information carriage' in terms of causation and laws: a mental state B carries information about an object A, if and only if it's a law that objects of kind A typically cause mental states of type B. So, for example, the sound "cow" carries information about cows, because cows typically cause "cow"s.

But, if we keep this causal definition of information carriage (and say that atomic terms in our language gain their meaning in part by carrying information about something as causal and teleofunctional theories are going to), we face immediate problems trying to understand how predicate expressions, or logical expressions could ever be meaningful. It's immediately nonsensical to require that ...is reds cause "is red"s or that ands cause "and"s.

Given this, it seems only reasonable (and charitable to the views in question) to require, not that it be a law that some *object* causes some mental/linguistic object, but rather that it be a law that some *state of affairs* covaries with the production of that object. Thus, we won't focus on it being a law that cows cause "cow"s (maybe there's a sizable fraction of cases where the conspicuous *absence* of a cow causes the utterance "drat, there aren't any cows here"), but rather the fact that reliably, *the facts* about whether there's a cow nearby covary with the facts about whether "there's a cow nearby" gets asserted. So: a sentence carries information about a state of affairs, if reliably/it's a law that utterances of that sentence co-vary with with the facts about whether that state of affairs obtains.

One can then imagine pulling facts about how various atomic terms contribute to determining truth conditions of whole sentences out of these facts about information carriage, by looking at what (if anything) sentences containing those terms carry information about (and satisfy certain other requirements with regard to).

Now, if we can use this broader notion of information carriage, we wind up getting both a Millikan-style teleofunctional account and a Fodor-style asymmetric dependence relationship between a) facts about the Platonist's abstract mathematical objects and b) sentences of a human language and/or mental states associated with these sentences.

6.3.2 The Story for Causal and Teleo-functional Accounts

Consider, for example, your disposition to accept or reject statements of the form "a+b=c". These carry information about the sums of various numbers, in the above-mentioned sense. There's a reliable covariation between the facts about whether someone will make a claim of the form "a+b=c" and the logical facts how many F-or-Gs there are when there are a Fs and b Gs (i.e. the facts about what a+b is).

But your dispositions to accept or reject "a+b=c" statements also carry information about a number of other things. For example, these dispositions reliably track facts about how many apples and oranges you will *count* if you count a apples and b oranges in some particular way (e.g. by listing number-names as you point to them one by one), and then count the apples and oranges together, or what number you will get if you try to mentally count how many rhymes there are in a poem that you have previously counted to have a male rhymes and b female rhymes. Think of all the different applications which we might make of the computation 24+27=51, to predict the results of counting some kind of object, via some particular counting procedure. The facts about when you assent to claims of the form "a+b=c" will also co-vary with these facts. So, how come "24+27=51" expresses the abstract, mathematical, fact that that 24+27=51, rather than any of these more concrete claims about what will happen if you count certain things? This problem of cutting down the possible candidate meanings for a given sentence, (assertions of a single word or sentence will covary with many different things) is already well known, and the major contemporary accounts have tried to deal with it in a few ways.

Millikan appeals to the 'consumers' of a given information carrying state, to try to solve this problem. Even though a given frog mental state covaries with the presence of: flies, black dots on its visual field, smallish flying black objects and many other things, the 'consumers' of this mental state (e.g. the mental machinery which which lead to the frog flicking out its tongue whenever it's brain gets into this state) only 'use' the fact that this state carries information about flies. In order to determine what function a given information carrying mental state has, we are to look to the consumers of this mental state. What (if anything) do the various other processes tied to this mental state use it to carry information about?

Fodor, in contrast, appeals to 'asymmetric dependence' for a solution (he takes the notion of biological function which Millikan appeals to, to be insufficiently sharp) Here, the idea is that utterances of "fly" mean fly, rather than (say) small moving black object, because the fact that small moving black objects cause "fly"s, depends on the fact that flies cause "fly"s. The closest possible worlds in which it stops being a law that moving black objects cause "fly"s, are ones in which we get better at distinguishing flies from little black pellets, so it's still a law that "fly"s cause flies. But, conversely, (Fodor thinks) if it stopped being a law that flies caused "fly"s, little black pellets would not cause utterances of "fly" either. Thus (to put things in terms of information carriage) the meaning of a sentence M is the (unique?) state of affairs which that sentence carries information about, *in such a way that the fact that the sentence carries information about any* ***other*** *thing counterfactually depends on that sentence's carrying information about M.*

I won't try to comment on whether either of these strategies ultimately succeeds in whittling down the range of possible interpretations, to match up with our intuitive judgments about the meanings of utterances. But they do, at least, help us eliminate the more concrete interpretations suggested above. For, our judgements about "a+b=c" facts will have a number of different concrete 'consumers' (e.g. judgments about how many fruit one should expect to find when counting a apples that have been put into a basket with b oranges, or how many rhymes one will find in a poem that contains a male rhymes and b female rhymes). As Frege emphasized, we are willing to count a wide range of different kinds of objects - from gingerbread cookies to numbers. Thus, if we are looking (a la Millikan) for what the mental state we associated with assenting to "24+27=51", is *used* by these diverse and changing consumers, to information about, the general mathematical fact about the sum of 24 and 27 seems a better candidate than one of the contingent empirical facts about what applying various counting procedures is likely to produce.

And, if we are looking (a la Fodor) for asymmetric dependence, one can plausibly claim that the closest possible worlds in which our statements of the form "a+b=c" stop carrying information about the results of counting some particular kind of object, some particular way (maybe your 'accuracy' at counting by pointing decreases as your eyesight gets worse, so that facts about what answer you are likely to get no longer mirror facts about how many fruit of a given kind there are), are ones in which most of the other applications we expect still seem to work, so our dispositions to assert statements of the form "a+b=c" would remain the same - and hence would still carry information about the fact that a+b=c. In contrast, plausibly, the closest possible worlds where we stop using "a+b=c" claims in a way that matches up with the facts about sums (scenarios where e.g. we suddenly embrace a different algorithm for calculating "addition"), are not worlds in which most of these particular applications are still likely to work. If we started computing "+" totally differently, the things we said about this new functions would no longer carry information about the results we were likely to get when counting apples and oranges. Thus, the fact that dispositions to assent to "a+b=c" claims carry information about what happens when you successfully count apples and oranges, seems to depend on these utterances matching up with/carrying information about the mathematical facts, but not there is not a converse dependence.

Another, more properly mathematical example, is the case of predictions about what (possible) chess games or lines of logical deduction there "are". We use claims of the form ""$\forall$x if x is a number then F(x)", to make predictions about what you could and couldn't get by going in accordance with certain definite rules. A sophisticated / explicit way of doing this would be to use numbers to code up sequences of board positions in chess, or sequences of lines in an argument. A less fancy way of doing this, is to reason about stages, (if there is some chess game which ends like this, there must be some least stage n, where the board looks like that). Nonetheless our statements in arithmetic don't *mean* claims about e.g. what chess games are physically possible, even though they may carry information about this. For, (a la Millikan) what all the different consumers of mental states corresponding to assertions about "all the numbers" are using these state to get information about, isn't any such particular application (which would only be directly relevant to one of them) but rather general facts about the numbers. (Or at least, this seems like the most attractive thing to say, if we assume that Millikan's notion of what information consumers are 'using' a given information source for is well defined in the first place.) And (a la Fodor) if claims about "all the numbers" stopped carrying information about what chess games one can vs. can't physically create, they would still carry information about the numbers, whereas if we started using claims about "all the numbers" in a way that no longer carried information about what's true of all the integers, (say, we started using "number" the way we now use "real number") these practical applications of claims about no number coding a proof would (generally) also fail.

So, what I want to claim is that, if Fodor and Millikan are right about meaning, (and if the platonist is right about the nature of mathematical facts) there is a plausible story to tell about how we wind up being able to assert things about abstract mathematical facts. In a nutshell, these abstract mathematical facts (anything like *this* must also be like *that*) are what the range of otherwise very different practical applications - which drive and maintain our use of mathematical terms- have in common. Thus the Platonist can plausibly say, what we have seen above that he wants to say, about Euler.

But what about the more ambitious claim that if we had gotten a different practically benign math-and-logic-like practice, we would still have counted as asserting largely truths? Presumably the same thing applies to other possible structures which we might have posited to systematically help predict experience. We have already seen that (even the most reductive of platonists) as a wide range of mathematical structures available to be talked about, so that alternative practices of blackboard reasoning which might be used to systematize different kinds of experience could (for just the same reasons here rehearsed in the actual case of addition and the integers) count as talking about these abstract objects.

6.3.3 The story for the Interpretationalist

The same kind of story applies to interpretationalism.

Interpretationalism (as I will use the term in what follows), takes facts about what people say and believe to be true in virtue of some very complicated combination of physical (and perhaps qualia) facts, just as facts about where there are chairs or livers or dearths of water supervene on these facts. Thus, it has the same naturalistic status as the more ambitious attempts to naturalize meaning above. *But*, unlike these the more ambitious causal and teleofunctional accounts above, interpretationalism does not propose any nice, purely phyiscal, necessary and sufficient conditions for the truth of claims about what a given creature means or believes. According to the interpretationalist, we are taught to recognize instances of saying that it's snowing [1] just as we are taught to recognize instances of lamps. And, there's no reason to expect that there will be short necessary and sufficient conditions (statable in purely physical + qualia language) for saying that it's raining, any more than there's reason to expect similar necessary and sufficient conditions for being a lamp. [In case this isn't obvious, let me stress that interpretationalism is in no way committed to the (anti-realist, and apparently regress-generating) view that people only count as having beliefs, insofar as there are other people to interpret them as having certain beliefs (i.e. form the belief that they have certain beliefs). Just as there could be lamps in a world without people, there could (so far as interpretationalism is concerned) be a single believer without anyone to interpret him (though, of course, there might be other private-language-based reasons why this turns out to be imposible.]

Since interpretationalism doesn't lay down explicit constraints about what physical/dispositional/qualia states count as believing what, we can't give as explicit an argument to the interpretationalist as to the proponents of the other two major views. Instead, the interpreationalist will evaluate claims of the form 'any practically benign math and logic like practice would count as getting things largely right' in the same way as he evaluates claims like 'knowledge is justified true belief'. That is, he will imagine a range of cases, and try to bring his trained ability to spot instances of belief that P to bear on these imaginary cases. Here again, the range of uses to which (what are apparently) the same computations can be put, pull in favor of interpreting these computations as making some, suitably abstract, claim about structural features which are shared between all the intended applications (whenever there are 2 objects of any kind and 2 objects of any other kind, there are 4 objects in total). So, if you are an interpretationalist (as I am), just consider the kind of cases described above, where arithmetic is applied, and consider whether (if platonism is right) you'd be willing to interpret someone exhibiting these dispositions to behavior, as making a claim about these platonic mathematical objects, rather than about any of the particular applications. And, in order to evaluate the more ambitious claim, (that other practically benign math-and-logic-like methods of reasoning would equally well count as reliably leading us to form true beliefs)

[1] even, to some extent, in the situation of radical translation

try to imagine one that wouldn't. The easy cases of logic and math-like practices that seem totally benighted (tonk-like reasoning that allows practitioners to infer any statement from any other, or applying the kind of computations we use to calculate the plus mod 27 function in the way that we actually apply computations of the plus function) are *not* practically benign.

6.4 Reference and Reliability: The Best Answer

In the last two sections, I have tried to show that a simple causal constraint on reference is ill-motivated, and that my story about reference to abstracta is compatible with the major big-picture theories of reference. These arguments all have a somewhat negative defensive character, but I think now that we have cleared the ground in this way, a much more positive answer to worries about reference to abstracta is available.

Here's what I have in mind:

Arguably, what's intuitively worrying about the idea that we refer to mathematical objects is that there doesn't seem to be the right kind of *connection* between the human mathematician and the things she is supposedly talking about. Initially, Benacarraf phrased this worry in terms of reference to abstract objects, but that proposal was rejected, partly for the reasons covered above. However, (at least for many people who find reference to mathematical objects problematic) the effect of these criticisms of the causal criterion were to silence criticism without really answering it. Perhaps the lack of causal connection wasn't the problem with platonism, but reference to abstracta still felt problematic.

Thus (as we saw in chapter one) Hartry Field and Oystein Linnebo have each proposed their own updated version of the problem.

One can see both of these challenges as worries about something like reliable connection our (putative) beliefs, and the domain that we are talking about. When we have beliefs about physical objects like llamas, there's an obvious way that experience can kick back and correct us. But with mathematics it seems like there is no such corrective force. There's no mechanism of correction.

Naively one might try to express this worry by saying something like this "Look, your belief that 2+3=5 has nothing to do with these abstract objects, the numbers, which you claim it is about. All you do is say what feels plausible to you for whatever psychological reasons, and this is never corrected by experience or anything else. So, even if you happen to have all true beliefs about the numbers, these beliefs are highly unreliable. Whether or not 2+3=5 has nothing to do with explaining why you believe what you do. **Even if 2+3 weren't 5 you would still believe it**". However, the latter formulation runs into another problem which threatens to silence without convincing. The problem is that 2+3=5 is a necessary truth, so we can't really make sense of this counterfactual. For example, we can't consider the closest possible world in which 2+3 isn't 5, and ask whether the mathematician shows himself to be unreliable by continuing to believe that 2+3=5. Hence a little more ingenuity is required.

Field asked for an explanation of why, 'Reliably, if mathematicians believe that P, then P'. And Linnebo asked for an explanation of why 'If S had meant something that wasn't true, mathematicians wouldn't have believed that S'. The idea here, is that mathematicians accept the sentence "2+3=5" and "there are infinitely many primes" and many other statements which express truths, and -intuitively- it's not just a fluke that this is the case. Rather acceptance by the mathematical community and expressing a truth are related. If "2+3=5" hadn't expressed a truth (e.g. because "+" became the symbol for multiplication), then mathematicians wouldn't have accepted it, and so on for various other sentences.

But now recall that these are the very challenges that my story was proposed to answer! If all the stuff about the wideness of the mathematical universe, and Quinean revision and a nudge from nature is correct, what we have is exactly a mechanism that provides the desired and apparently missing kind of kick-back.

- insofar as your judgements about "the numbers" stray just **a little bit** from the facts about the numbers, they are likely to lead to conclusion that conflict with your other judgements about "the numbers", and hence be corrected.

- insofar as your judgements about "the numbers" stray **a lot** from the facts about the numbers, they will be steered by Quinean Revision, and nudge from nature etc, towards consistency and correct empirical applications, and hence (given the profusion of targets for mathematical knowledge) wind up expressing the truth about some other mathematical structure (e.g. the integers rather than the natural numbers).

Hence, I claim, absent some **other** reason to think we can't refer to mathematical objects, we have a nice explicit mechanism that explains the phenomena Field and Linnebo want explained. The points above directly explain why reliably mathematicians assert truths. For, when we consider close possible worlds in which mathematics had gone differently, the same mechanisms of Quinean selection and nudges from mathematically shaped problems in nature will be effective in these worlds as well.

Perhaps it seems like I'm assuming the last phase of my story works (practical benignness is likely to yield largely true beliefs), at exactly the point where I need to prove that. But I don't think that's true, so let me stress something about the dialectic at this point.

The Linnebo and Field objection is that if we were to interpret people as referring to mathematical objects, they would come out as being unreliable (in a certain relevant sense) about them. It is for *this* reason that they claim (in Field's case), or worry that (in Linnebo's case), we cannot count as referring to mathematical objects. The idea is that *even if you initially suppose that the platonist is quite right about how reference to mathematical objects works*, the Platonist cannot account for a particular kind of connection between mathematical objects on the one hand, and beliefs on the other. But (you might think) such a connection is required for reference. So we have something like a proof by contradiction. If we suppose that the platonist is right about reference, we arrive at an unacceptable conclusion (the fit between mathematical facts and mathematicians beliefs about them is - in a certain sense - just lucky a fluke). But this unacceptable conclusion is false. So, having seen that we can derive falsehood, and thence contradiction from the platonic theory of reference, we can conclude that it is false.

Hence, Linnebo and Field are exactly conceding -for the sake of argument- that the Platonist can have their intuitive semantic theory (e.g. Orthodox Lewisianism). What the Platonist then needs to show is that it comes out to be no fluke that if mathematicians assert P then P/mathematicians tend to assert sentences that are true. I have provided a mechanism that purports to do just that. It would be viciously circular for them to object to this mechanism on the grounds that reference to mathematical objects is impossible. If the reason why reference to mathematical objects impossible is because no acceptable story about kick-back can be told, the reason why my story about kick-back is unacceptable can't be that it requires reference to mathematical objects.

More importantly than this merely dialectical victory, I believe that my account can intuitively satisfy those who worry about Platonism because of concerns about our apparently inadequate relationship to Platonic objects. By actually giving an explicit and fairly detailed story I hope to convince, where previous arguments have merely silenced. The notions of adequate explanation, or naturalistic acceptability are famously easier to know when you see them (not that this is trivial) than to give formal criteria for. Putting together the previous chapter we get a detailed picture of how familiar mechanisms could work to reliably line up the beliefs of creatures like us with the kind of mathematical objects which the Platonist claims these beliefs are about.

When we look back over the past few chapters, and the panorama which stretches from evolved differential looking times in babies, to Quinean revision in the development of number theory around arithmetic (always preserving correct concrete applications!), we can actually see how ordinary factors could work to produce an alignment between abstract facts about mathematical objects on the one had, and human psychology and practice on the other.

I submit, and hope, that many who were initially puzzled by reference to abstract objects will find this an intuitively satisfying demonstration of how the right kind of connection can be achieved by completely ordinary and familiar mechanisms. Our having reliable true beliefs about mathematics initially looks like magic, but now we have laid out the combination of common-place springs and leavers that suffice to achieve this spooky looking trick.

Chapter 7

From true belief to knowledge

7.1 Introduction

In the previous chapters I have outlined a descriptive story about how we could have come to reliably use the kinds of methods of a priori reasoning which reliably yield true beliefs about math and logic. If what I have said works, I have answered the initial naturalistic question about how we wind up accepting the kinds of things in the armchair which experience then appears to bear out. I have also answered a more ambitious question about how we reliably wind up with true beliefs as a result of a priori reasoning about math and logic. But perhaps we still haven't accomplished our over-arching goal - to account for mathematical knowledge.

For, unless you are an extreme reliablist, reliable true beliefs about math and logic do not mathematical and logical knowledge make.

Thus, in this chapter I aim to close the gap.

7.2 A brute appeal to intuition

In my thesis I aim for catholic appeal. So, before I go on to make some substantive claims about the nature of knowledge which will likely be controversial I want to make an argument which even the harshest opponents of my epistemology might accept. The argument uses what might be claimed to be the most successful methodology ever used in a philosophy paper [1]. It goes like this:

Consider creatures who did acquire reliable true beliefs about mathematics in one of the ways sketched in the preceding three chapters.

Would they count as having mathematical knowledge? If not, what more could they possibly need? If you think they *would* count as knowledge (or it's

[1] i.e. the method of presenting individual cases for intuitive judgment which Gettier used to argue that knowledge isn't justified true belief

easy to fill in certain details of the case in such a way that they would), I claim that I have given you what I promised - a solution to the intuitive mystery of how a priori knowledge is possible. So let me rest my case there, before I start saying things that you will disagree with.

But perhaps you are not satisfied with the above simple argument. It isn't obvious to you that the creatures above are in a situation which lacks some essential ingredient for knowledge, but on the other hand you don't feel quite comfortable saying that they *do* have knowledge either. Since I aim to clear up any puzzlement about how a priori knowledge is possible, my work isn't yet done.

So, what I'd like to do for you, is to advocate a very general picture of the nature of knowledge, and what it takes for someone to count as having knowledge. Using such a picture we can see more explicitly how the naturalistic processes I have proposed would have given us not only reliable true belief but mathematical knowledge.

7.2.1 Good methods working as intended: a unified theory of knowledge

For a creature to have knowledge is a matter of it having a certain kind of systematic connection to the world/the facts which it purports to know about. Pessimistic induction over the history of Gettierology suggests that exactly what kind of systematic connection is require is either very difficult or impossible to spell out in suitably precise and informatively different terms. But -in terms which are neither as precise nor as informative as one might wish for - here's what I want to propose about the connection.

We naturally see people as forming beliefs via certain methods/mechanisms. So, for example, we think that normally a single methods generates both my belief that 2+2=4 and that 2+3=5. And we criticize people for forming certain beliefs by connecting these beliefs to certain methods that putatively yielded these beliefs, saying e.g. 'by that method of reasoning you could have come to the obviously false conclusion that x'. Different methods, such as mathematical induction, or observations about what furniture there is made in good lighting can be more or less reliable in producing true beliefs/not producing false beliefs/improving the extent of the true beliefs which the method user already has.

Good methods bring those who use them to form true beliefs in certain systematic ways, which we see as either working as intended or not working as intended in various cases (for example, if you believe the true testimony of someone who takes themselves to be lying to you, the method of believing testimony is not working as intended, even if it leads you to a true belief). Now I claim that: S knows that P iff their belief that P was caused or preserved by a reliable method, working as intended. Or actually (for reasons having to do with the fragmentation of knowledge which I'll get into in the next section) I claim that S has **knowledge-light** that P.

There are three kinds of good methods, if we divide methods up by how they systematically produce true belief when working as intended. Perceptual methods involve a causal interaction between a creature and their environment, which systematically brings the creature's beliefs to match up to the facts by changing the creatures mental state. We might say they yield 'perceptual knowledge' when working as intended. Active methods systematically bring a creatures beliefs that match up with the environment by changing the environment. People sometimes call the resulting knowledge 'maker's knowledge' and I think this is the kind of knowledge which Elizabeth Anscombe wanted to say characteristically comes along with action. Thirdly, there are mechanisms which systematically produce a fit between a creature's beliefs and the environment because they produce - in a way that's totally causally insensitive to the environment - beliefs that happen to be true of the environment. We can call this 'innate knowledge'

Here's a little example. Jim is a little fish who has two different digestive systems: one that works well for digesting food with a high PH and one that works well for digesting food with a low PH. He has magnetic sensors which let him detect his distance from a certain acid geisure in the north pole. He also has a PH sensor which he can poke out at will, and an ability to squirt acid. One day, Jim runs into a bit of food. Now he needs to know whether the food has a high or low PH, in order to put it into the correct digestive system. There are three methods he could use to do this. First, if he is near the north pole he will eat the food with his low PH mouth. Here he is using (what corresponds to) innate knowledge. Or, if the food is not near the pole he might squirt out some acid, so that now (whatever the food's original PH was) it has a low PH and then digest it with his low PH mouth. Here he is using maker's knowledge. Or still, if he is out of acid, he may use his PH sensor so that if the food has a low PH, it will put the PH sensor in one state which will lead to him eating it with his low PH mouth, and conversely if the food has a high PH, this will change the state of the sensor in a way that leads to him eating it with a high PH mouth. This corresponds to perceptual knowledge.

In all three cases, note, we have mechanisms at work which systematically lead Jim's beliefs about the food in front of him to match up with the facts, so there's a similarity between what might seem like bewilderingly different kinds of knowledge (perceptual knowledge, maker's knowledge, knowledge that comes from internal reasoning). And when we attribute someone knowledge, we do so in virtue of taking them to be employing a good method of one of these kinds.

7.2.2 What methods? Why knowledge attribution is deeply unprincipled

I have just given a loose sketch of a theory of knowledge which unifies different kinds of a knowledge in a way that you might find attractive. Now I will fill in the picture by adding something which, I am sorry to say, you are almost certain to find quite unattractive.

I proposed that knowledge is a matter of forming true beliefs as a result of good methods working as intended. But I did not say what exactly determines what method a person counts as using, or how good a given method has to be. One might simply propose these topics as open questions for future research. (And I suggest you do, if you like the above story, but don't like what follows).

But I think this would get things completely wrong. There are no philosophically interesting facts about what determines what method a person is using, or how reliable a method has to be (and in what ways) to count as sufficiently good to qualify for knowledge.

For example, consider my current belief that 22+7=29. What method did I use to form it? There are infinitely many different methods which yield this particular beliefs (for example, there are methods corresponding to every recursive function that takes (22,7) to 29). Also, don't forget very specific methods like 'believe 22+7=29' or 'believe n+7=29 for all n' - which come in varying degrees of reliability.

Does what I say when asked determine this? Often I can say nothing other than that a certain claim seemed obviously true, or obviously plausible in light of some other particular claim. Thus we can't equate my using a given method with my being disposed to say that I am using that method.

Perhaps if neuroscientists study my brain there will be some combination of mental mechanisms whose joint product is the sum total of the actual physical processing going on in the brain. But there's the same kind of ambiguity in deciding how to carve up what the brain actually does when I form beliefs into a joint product of attempts to realize different methods (and whether or not this particular belief counts as being formed on the basis of a good method depends completely on how we do this cutting up).

Thus, I want to suggest that there isn't anything really principled behind our inclination to say that certain beliefs are formed via one particular method. We have some general interpretive practices (perhaps not entirely stable ones), and make some intuitive judgments like 'he formed that belief by the method of consulting a fortune teller, so its unreliable'. And so, just as we take people's actual -perhaps not entirely stable- dispositions to count certain lumps of metal as forks to fix an extension for the word 'fork' this usage presumably determines facts (in at least some cases) about what method a given person counts as employing. But this is not to suggest that the boundary between forks and non-forks is of any intrinsic interest.

And I strongly suspect the same thing applies to our judgments about whether a given method is good enough. For example, most people would allow that the method of assuming a+b=b+a is sufficiently good, but what about the equally reliable method of assuming all instances of the Goldbach conjecture? Does our feeling that it's epistemically Ok to just the one method without trying to independently justify it, but not to use the other really track anything more than psychological facts about what kinds of true mathematical statements actual people tend to find obvious?

In any case, the distinction between knowledge and mere justified true belief seems to turn on the question of whether the subject was using a sufficiently good method, and this method was working as intended. But if there's something unprincipled and arbitrary in the facts about what method a given person counts as using, this same lack of principle will infect our judgments about whether what they have counts as knowledge.

And this brings us to the reason why I say you will almost certainly find this further specification of my theory of knowledge unattractive. For, if the distinction between knowledge and mere true belief turns out to depend on unprincipled distinctions of this kind, it plausibly can't support the normative weight many people would intuitively like to put on the notion of knowledge.

Now, note that this account of knowledge makes it plausible that the processes described above would wind up giving us knowledge as well. For (though the lack of any principled statement of how facts about what method someone is using supervene on physical facts precludes us from giving an explicit proof) there's nothing mysterious in the idea that in getting mechanisms which reliably produced true beliefs, we would wind up in states that can be interpreted as using good methods.

7.3 Conclusion

In this chapter I have tried to show how my account for how we could have gotten reliable true beliefs about math winds up also accounting for how we could have gotten mathematical knowledge. I started by reviewing the kind of faculty which my naturalistic processes could plausibly have given to us, and making a brute appeal (a la Gettier) to intuitions about what counts as knowledge. Then I sketched a positive theory of the nature of knowledge, on which being able to get knowledge is in essence a matter of having some faculty which produces a systematic match between what you believe and the truth. Not all such systematic true-belief creating processes are equal, though. Our concept of 'knowledge' and 'justification' gives a special thumbs-up to some such systems, and not others (you can get knowledge by taking your sense perception at face value, but not by taking equally well functioning ESP for granted). Roughly speaking, we give this thumbs up to those true-belief-producing-processes which can be found in most/all normal people. Whatever physical etc. systems produce the kinds of basic logical and mathematical statements and inferences which we allow to figure in proofs without further justification are among these widely shared faculties.

I don't know whether starting evolution again would have granted us the same faculties, so there's a sense in which it's lucky that we have mathematical "knowledge" (in our actual sense of the term) rather than mere true belief. Maybe we might have wound up finding certain inferences brutly compelling which we would say are valid but need multi-stage justification. And, if so perhaps we wouldn't count as having mathematical knowledge (because we "keep assuming things" and "can't give rigorous gap free proofs"). But if this the only sense in which luck is required to explain our having mathematical knowledge, there's no naturalistic puzzle about how we could have gotten mathematical knowledge left.

Chapter 8

Conclusion

In this chapter, I will now conclude, by showing how my naturalistic account of mathematical knowledge compares to simialr theories in the literature (Philip Kitcher's Empiricism, Penelope Maddy's early and late Naturalism and Carrie Jenkins' Empirical Grounding). Then I'll make some admittedly ambitious claims about how it relates to the classic Logicist and neo-Logicist projects of reducing math to logic, and to the old battle between Rationalism and Empiricism. In particular:

I'll argue that my account solves the epistemic worries that motivated logicism, without being subject to the technical problems that beset logicism - indeed, that it answers these worries in a deeper and more satisfying way than even a fully successful logicist reduction would. Logicism would only have reduced questions about mathematical knowledge to questions about logical knowledge which have turned out, in the intervening century, to be anything but trivial. My story presumes nothing about logic, and instead shows how perfectly ordinary processes (which biologists, psychologists and historians are studying right now) can account for both logical and mathematical knowledge.

I will propose that historical empiricist objections to a priori knowledge, can be understood as motivated by the kind of naturalistic worries about how to account for a priori knowledge which are the main topic of this thesis. Hence, if the story you have followed in these pages is correct, it amounts to what might be called (after Blake's 'Marriage of Heaven and Hell') a 'Marriage of Rationalism and Empiricism'. We have seen how one can maintain the purest rationalism about the justification of mathematical propositions, and the objectivity (indeed the platonic existence) of mathematical facts, while satisfying and addressing the naturalistic worries that have made rationalism seem unacceptable for four (?) centuries.

8.1 Kitcher's Empiricism

Philip Kitcher's neo-Millian The Nature of Mathematical Knowledge provides another natural contrast with the view advocated here. Verbally at least, Kitcher's account couldn't be more different from mine. Where I aim to reconcile naturalism with rationalism and platonism, Kitcher explicitly denies that we know anything a priori, and that there are mathematical objects like sets numbers etc. However, as we will see, there are also some important similarities. Both theories make central use of mathematical revision, and the concrete applications of mathematics in accounting for human mathematical knowledge. And, in one sense, both have the effect of diminishing the distance between a priori and a posteriori knowledge (though, of course mine does this much less than Kitcher's).

Kitcher begins with an argument that nothing is known a priori. A priori knowledge (he suggests) would have to be knowledge that anyone possessing the relevant concepts could have, regardless of the course of their experience. But, what about the experience of getting evidence that your mathematical reasoning abilities are impaired? Imagine someone who gets strong evidence that their ability to add is seriously impaired: they remember drinking heavily, they have just done 100 addition problems while in this stated and then checked them over and found many apparent errors in each. Now they do one more addition problem - correctly, as it happens. But they seem to see a number of respected mathematical experts disagreeing with their answer, and perhaps the calculator gives a different answer. Kitcher proposes that a person with this course of experience, although fully in possession of all the relevant concepts would not be justified in believing that eg. 3795+476= 4271. Hence, even this proposition does not count as a priori.

Against this argument, I take the standard line CITE that a priori justification is defeasable. Just as a normal person looking at a chair in good lighting gets justification for believing that there's a chair in front of them (but this justification can be undercut if they acquire evidence to the effect that they have taken a chair-hallucination causing drug), our intoxicated adder has (default) justification for believing each premise and making each inference as they work out the sum, but this justification is undercut by the evidence to the effect that their capacities aren't working correctly. Thus, although Kitcher has shown an important and previously under-appreciated interaction between a priori and a posteriori justification (a posteriori evidence can be undercut justification secured by a priori reasoning), he is wrong to conclude that there is no a priori truths.

In Kitcher's terms: a priori propositions are ones which anyone with the relevant concepts can acquire justification for believing, but some courses of experience may provide enough empirical defeaters to prevent propositions believed with this justification from counting as knowledge.

Then he argues briefly against Platonism, largely on epistemic grounds. The Platonist, he argues, needs a story about how 'rational insight' into facts about mathematical objects is supposed to work, whereas all they seem to provide is metaphors about seeing. Also, given the fact that apparent mathematical insight can be wrong (e.g. naif set theory, contradictory reasoning about limits), he suggests that they needs some story about how one can tell when it's working. He also mentions Benacarraf's second worry - that there seems to be no principled way of saying which set the number 3 is (given the existence of multiple reductions).

This whole thesis is, in effect, an attempt to thoroughly answer the first challenge. If it succeeds, I have shown how the Platonist can give a very concrete picture of what rational intuition is (a moderately reliable tendency to make correct mathematical and logical judgements), and how we could have got such a thing. As regards the second, Kitcher seems to be assuming that a genuine faculty of rational intuition would have to be infallible. If one thought that (as, perhaps, historically many Platonists have) there would be genuine question of how we manage to tell when we are using this faculty as opposed to doing whatever alternative activity leads to false mathematical beliefs. However, I and most other contemporary rationalist explicitly deny this claim. So, on my view, naive set theorists and reasoners who reached contradictory beliefs about set theory were using their faculty of rational intuition just as much as anyone else. Using your faculty of rational intuition requires nothing more impressive than trying to reason about math and logic i.e. starting with premises that feel compelling to you, and engaging in inferences that feel equally compelling, and subjecting the whole process to revisions when you run into conflict.

(The hard part is just to explain how creatures like us could have got into a situation where doing this was e.g. reliable enough to count as a 'faculty of rational intuition' rather than just a 'faculty of randomly making stuff up')

We have also already addressed the second Benacarraf worry in chapter 4. To review: A non-reductive platonist will say that numbers are distinct from sets, so the correct answer to "which set is 3?", is neither. And a reductive platonist who does think the numbers *are* sets can say that the expression "3" is vague in its reference (whatever the correct account of vagueness turns out to be), except in contexts where explicit definitions in math textbooks fix that the word is to refer to one specific set (in the context of that discovery). Such contextual specifications are already ubiquitous in other areas (e.g. 'for the purposes of this survey of plant-life "the foot of mount everest" will mean all of the mountain between n and m meters in elevation').

Now for Kitcher's positive view. Mathematics is the study of idealized possibilities for action. So, for example, claims about sums are true in virtue of facts about how an idealized creature could physically 'collect' objects - where physically moving objects together is one example of such collection. In particular, the fact that "if one performs the collective operation called 'making two', then performs on different objects the collective operation called 'making three', the total oportation is an operation of 'making five'". And we can learn such claims empirically as children by actually physically moving together pieces of gingerbread or whatnot. Claims about geometry wind up involving more elaborate idealization: they concern the possibilities for action (e.g. extending lines, dropping parallels) of an idealized agent on idealized objects in idealized space. But there still substantial similarity between actual human behavior and the mathematical facts about ideal behavior, which allows us to learn mathematical facts from operations that we can actually perform, and to learn about actual operations by considering facts about ideal operations. And set theory is done in a way that quantifies not over objects (the set) but over operations of collecting. The overall result is a picture on which that there are objective mathematical facts, but no mathematical objects. Mathematical facts are true in virtue of modal claims about idealized actions: every combination of actions of this kind, would amount/allow to an action of that kind. As Kitcher light-heartedly puts it, mathematics is about "permanent possibilities of manipulation".

Zooming in on epistemology we get the following three stage picture:

First, there are actual human activities of collection and ordering. This can involve physically moving Millian gingerbread around, or moving representations of various objects around. Importantly, the latter category includes the activity of moving around ink marks on paper which themselves represent possibilities for moving around cookies.

Secondly, experience with these activities of collection and ordering justifies us in making stipulative definition of facts about ideal agents, objects, spaces. So, for example, experience *actually* collecting stuff, suggests a notion of how ideal collection might work, and this leads us to make a suitable definition of an ideal agent in terms of how they *could* collect stuff. In the case of geometry, physical interactions may also suggest a notion of ideal geometrical objects and space - as well as an ideal geometrical agent to manipulate these things in.

Finally, similarity between these facts about the ideal agent and facts about the actual world, helps justify applications of these mathematical facts (facts about ideal agents) to actual real world scenarios. We can infer things about actual activites of collecting cookies/moving around shapes from mathematical reasoning about how an ideal agent could collect things/manipulate ideal objects in an ideal space, because the latter is similar to the former.

If you haven't read Kitcher's book, this view can seem very strange. Why do you need empirical justification for making a stipulative definition? And what do we mean by similar here? But the analogy that Kitcher makes, which we should keep in mind, is to scientific reasoning about 'ideal' systems - in particular the ideal gas law.

In this case it seems much more natural to say that interaction with actual gasses lead to, and helped justify the stipulative definition of an ideal gas, as one for which pv=nrt. And also: the resemblance between actual gasses and ideal gasses helps account for the application of facts about ideal gasses (derived from the stipulative definition) to conclusions about how actual gasses will behave.

So much for setting up the positive view. Now I'd like to make some objections.

The main point I want to start with is dialectical. Kitcher explicitly admits that the normal platonic view of mathematical objects would be preferable if all things were equal, but thinks this has epistemic problems which lead to the conclusion that we should prefer his view. In particular, he seems to admit Benaceraf's point that a straightforward platonic understanding of mathematical objects allows for an appealingly smooth semantics, which fits better with the semantics we give to ordinary talk about physical objects which seems to have -at least superficially- the same structure. In contrast, Kitcher's reconstructions of mathematics as reasoning about possibilities of action for an ideal agent fit more awkwardly with ordinary mathematical speech than the straightforward platonic understanding of the claim that 2 is prime as making a claim about an object called 2.

Given this, the burden is on Kitcher to show how understanding mathematics as reasoning about ideal agents, and permanent possibilities of combination, somehow does better -epistemically or otherwise- than the standard platonist understanding of mathematical talk of numbers as being about objects called numbers. So, why is Kitcher's view better?

First, we should note that we can't motivate this distinction by general worries about abstract objects. In order to capture the practice of modern set theory Kitcher winds up quantifying over possibilities of collection. (In a nutshell, sets at a higher level of the set theoretic heirarchy correspond to possible ways of collecting sets at the lower levels). So by Quine's criterion Kitcher is equally well committed to 'possibilities of collection', just as much as the standard platonist. And, presumably, these are abstract objects. Hence, intuitions that a) there can't be abstract objects or b) we couldn't possibly know anything about abstract objects cannot be used to support his semantics over my more straightforward semantics.

Instead, I take it, Kitchers epistemic worries are of a more subtle and local nature. Bringing mathematical facts closer to physical ones by understanding math as talking about ideal possibilities of collection is supposed to help

- a) Justify the transition from physical interactions to making stipulations
- b) Account for the transition from conclusion about abstracta to expectations about concreta, based on reasoning about similarity.

However, I shall now argue that Kitcher's proposal does a worse job than standard platonic semantics when combined with my proposal, in both of these respects.

Let's start with a). I think Kitcher is right to point out that there are epistemic issues surrounding giving stipulative definitions. We have seen that accepting certain kinds of stipulative definitions (like those corresponding to tonk) can lead to the derivation of contradiction, or false empirical premises. However, it is surely going too far to say that in order to be justified in accepting a stipulative definition we need to be justified in thinking some actual physical system has the properties mentioned. Wouldn't we be equally justified in defining a shmideal gas to be one for which $pv=nrt^2$? Claims about shmideal gasses derived from this definition seem equally correct and knowable, even if they are much less scientifically useful.

My proposal about stipulative definitions (above) aims to the epistemic issues related to stipulative definitions in a more nuanced way. On that view (as we saw), stipulative definitions can be motivated either by some kind of intended applications (Let the plus function the one that mirrors *these* logical facts about how many Fs and Gs there are in *that* way. Then the plus function seems to satisfy this recursive defintion. Now, stipulatively re-define the plus function to be that which satisfies this recursive definition). But they can also be motivated by general mathematical reasoning whereby it is proved that there is some property/ object corresponding to the intended definition (e.g. stipulating that one term is to be a shorthand for some combination of others never leads one to falsehood). Or stipulative definitions can be motivated by a more general, informal, sense of when stipulative definitions are practically benign, and hence likely to lead us to largely true beliefs - although such stipulations will often initially be somewhat tentative.

Hence, I claim that the right view of the epistemic challenges that face us when making stipulative definitions does not require the kind of direct similarity between the abstract objects or properties being defined and some real one, which Kitcher's non-standard semantics provides.

There are also problems with part b). Firstly the notion of similarity is infamously vague and ambiguous. Things can be similar in one respect, and different in another. For this reason, even accounting for scientific reasoning about ideal gasses in terms of similarity judgements seems a little crude. When we introduce a model for some physical system, we expect the physical system to behave like the model in some ways but not others - if we model the ocean waves as waves in a liquid of infinite depths, we expect to be able to carry over some consequences for the ideal system to conclusions about the actual world (e.g. about patterns of waves) but not others (e.g. about what would happen if we went deep sea diving). We might treat this by saying that scientific models come with specific intended rules of application saying what their predictions are for real physical systems. However, even this seems a little crude, insofar as a given model might explicitly be introduced to make one kind of prediction, but naturally suggest to scientists a range of other applications (with decreasing degrees of confidence in different kinds of applications).

For this reason, I want to suggest that my model for the application of mathematics works better Kitcher's - even when applied to the scientific models that inspired his theory. For recall that on this model we have, associated with a given mathematical theory, a range of expected applications - where these are understood as dispositions to go from conclusions in the mathematical system to expectations (i.e. dispositions to be surprised by) observations of the concrete world, and vice versa. These expectations can be of various strengths, and can be modified somewhat separately from one another. In the wave example above, I might very confidently draw conclusions about short term patterns in waves from my model of the ocean as a liquid with infinite depth, but only have a modest inclination to wonder whether the model will predict large-scale seasonal changes in the behavior of the ocean as well. And experience might lead me to give up the latter expectations without giving up the former.

If we just thought about models using an undifferentiated notion of similarity (drawing conclusions about the actual world as motivated by considerations of similarity) this would not apply we cannot make such of this kind of learning the ways in which your models applicability are limited.

This contrast between different applications of claims about abstract objects becomes even starker when we consider mathematics. In the absence of further chemical analysis, our expectations that novel actual gasses should behave like ideal ones are relatively fuzzy (exactly how similar should we expect the actual measurements to the results of ideal calculations?) - as it happens, the degree of divergence from the behavior of ideal gasses systematically varies as you go from left to right across the periodic table. And these expectations are relatively weak, in contrast to some expectations we have for mathematical objects. For, many core applications of mathematics are as certain as we can about anything, and extremely precise. For example: If "a+b=c" there will never be a apples and b oranges and only c fruit. If a certain program doesn't halt, then no computers which behaves in such-and-such way (fill in a specification that the computer must go from physical state a to physical state b whenever it reads a memory location in physical state c corresponding to each of the n-tupples in the turing machine for this program), will end up in so-and-so state (where this is the physical state associated with the turing machine's halting state).

Thus, to summarize, we don't need to reconstrue mathematical talk to make the objects of mathematics 'similar' to physical activities like collecting, in order to make sense of applications, because the applications of mathematical and scientific models are both better cashed out in terms of profiles of expected applications, than in terms of some undifferentiated similarities. The numbers don't need to be 'similar to' my activity of collecting cookies in any substantial sense, in order to account for the helpfulness of arithmetical reasoning in my dealings with concrete objects like cookies. All we need is the clear and concrete fact that all instances schematic inference from 'a+b=c' to 'if there are a gingersnaps and b oreos, and no gingersnaps are oreos, there are at least c cookies' are truth preserving.

If we want to describe this more vaguely and heuristically, we can say that the behavior of the plus function is 'structurally similar' to certain patterns in which sentences are logical truths. But we shouldn't forget that this notion of similarity cashes out in terms of concrete claims about certain *particular transitions* from sentences about the one subject matter to sentences to sentences about the other subject matter being truth preserving (these are my 'expected applications').

So, I've just tried to argue that Kitcher admits he has the burden of proof with regard to a standard platonist understanding of mathematics, but then fails to provide any reason for preferring his (superficially less natural) semantics. One might suspect that he saw how concrete applications of mathematics could be used to account for our systematically getting the right verdict about one kind of abstracta (possibilities of collection), but not how these techniques could be generalized to work for the standard platonist as well. Hence it seemed like he had a story about our access to mathematics, while the platonist has none, so his view was to be preferred on epistemic grounds. In fact, if the view in this thesis is correct, he has a story, and the standard platonist can tell a similar - indeed slightly more attractive- one.

I will end this section by noting one active objection to Kitcher's view, which concerns his treatment of incompleteness. When you think of mathematics in terms of stipulative defintions (whether the stipulations concern kitcher's ideal agents or something else), problems of incompleteness arise as follows. No finite set of stipulations you make can have, as first order logical consequences, all the truths of number theory. Intuitively, there's a fact of the matter about every sentence of the form ∀x Fx, where F is a simple predicate of arithmetic, that can be checked to hold or fail to hold of each particular number. However, no recursively axiomatizable theory can capture all these facts (hence in particular no theory that applies first order logic to some finite set of stipulations).

Kitcher is quite aware of this fact, and gets around it by saying that he allows infinitely many stipulations with regard to his ideal agents. But it seems to me that saying this leads to a dilemma. Intuitively, what I express when I state the Goldbach conjecture *right now* is true. Maybe I will never know, but I can meaningfully wonder which one it is. Now if the infinitely many stipulations are supposed to have already been made (somehow, when I was first learning mathematics) then it's hard to see how anything I have done could count as making infinitely many stipulative definitions.

I mean, arguably the grip we have on stipulative definition is already straining at the leash, when we consider Kitcher's proposal. I do recall making a few stipulative definitions, where I said aloud of wrote on a problem sent something like "let a good number be one such that...". In that sense, I don't recall making even one stipulative definition about an ideal agent. So, it seems like the stipulative definitions involved would have to be unconscious ones. Insofar as people can unconsciouls try to undermine loved ones, perhaps they can unconsciously make stipulative definitions about ideal mathematical agenthood. But when we are now required to say that I have made infinitely many situplative definitions, the total collection of which suffices to first order logically entail all truths about number theory, (at least I personally find) the leash snaps, and I loose all notion of what these implicit stipulations are supposed to be. I mean, presumably Kitcher isn't proposing that we could discover the answer to the collatz conjecture by undergoing some kind of regress therapy, which takes us back to gradeschool and allows us to remember a few more of the infinite number of stipulative definitions of ideal agency which we made then!

So it seems Kitcher must mean the alternative to this proposal: that there are infinitely many stipulations in the sense that we could make infinitely many stipulations. But this simply doesn't solve the problem. For the intuition is not that I could stipulate new definitions for some of my words such that "Every integer greater than 2 can be expressed as the sum of two primes." expressed a truth or a falsehood. *Every* string of symbols has the property that we could use stipulative definitions to get it to express either a truth or a falsehood. Rather the intuition is that when I ask myself *right now* whether "Is every integer greater than two expressable as the sum of two primes?" that question has an answer. Short of saying that future implicit definitions can somehow retroactively change the truth conditions for my claim (and humanity will exist forever so as to make each of the relevant infinite list of stipulations), it is hard to see how Kitcher can account for this fact.

In contrast, my view takes mathematical talk of "X"s to apply to the most natural target(s) in the wide platonic universe, which best satisfy our claims and expectations about the Xs (weighted to so as to give more weight to things that we have given literal stipulative definitions for). In this way, we can say that there is a fact of the matters as to whether my *current* utterance of "Every integer greater than 2 can be expressed as the sum of two primes." without requiring some kind of infinite chain of stipulations.

8.2 Penelope Maddy's Naturalism

Another contrast I'd like to consider is between the theory advocated here and Penelope Maddy's early and late naturalistic accounts of mathematical knowledge.

8.2.1 Early Maddy: Seeing the Sets

- I'm not claiming you can see sets (the only sense in which this is plausible is one in which you can see anything that's real, because the theory/observation distinction is not well defined)

8.2.2 Second Nature

general notion of second philosophy

-I'm not saying logic is contingent

-I'm not advocating some simple formula for what set theory to add, there's a fact of the matter and we may never know

8.3 Carrie Jenkins' Empirical Grounding

Carrie Jenkins' 2009 book 'Grounding Concepts: an Empirical Basis for Arithmetical Knowledge', proposes a theory that has a lot in common with the project of this thesis.

Jenkins' essential task in this work is the same as mine, as is her view of the major roadblocks. She too, aims to give a naturalistic account of mathematical knowledge (though a vastly less ambitious one, insofar as she limits herself to CHECK particular facts about arithmetic, which many take to be logical rather than mathematical in character). And she is equally motivated by the 'tonk' problem which faces stipulative-definition based accounts. On her view, as well as on mine, a core challenge in accounting for mathematical knowledge is to explain how humans can have managed get "good" combination of inference patterns that count as thinking true things about some domain of mathematical objects/having a coherent conception of what those objects must be like, rather than "bad", tonk-like patterns of reasoning. Finally, in a sense, we both appeal to causal interactions with the world, to explain how we wind up with such combinations of inference dispositions.

However, there are some significant differences, and I will argue that Jenkins' theory helps itself to some faculties and notions that seems almost as problematic as mathematical knowledge in the first place.

Jenkins' proposal is that experience has "non-conceptual content" which 'grounds' our acquisition of concepts, so as to help us form coherent ones. This is what leads us towards good (non-tonk-like) connectives, and coherent concepts in general. Then, when we have a coherent concept of something like the numbers, we inspect it to see what what must be true of the numbers and reason correctly about them. Between these two factors, arithmetic winds up being - paradoxical though it sounds - both empirical and a prori. It is empirical because it is empirically grounded, and a priori because the chains of reasoning that justify arithmetical reasoning only appeal to conceptual truths, not any facts about experience.

The big obvious worry here, is how concept inspection is supposed to work. Presumably Jenkins isn't proposing that we fly to the third realm to literally see concepts there. Indeed, if concepts are abstracta (as most philosophers now would seem to accept that they are), then why couldn't we use whatever faculty allows us to 'inspect' concepts, to 'inspect' numbers themselves directly! On the other hand, if concepts are supposed to be something psychological like brain structures, it's still not clear how you would inspect them. (Self-performed brain surgery?) Also, there are well known issues raised by Frege in taking concepts to be something psychological. EXPAND? Jenkins admits that how concept-inspection exactly works is an open question. QUOTE. But you might worry that explaining how we can see this kind of necessary abstract truths (X concept requires Y) isn't much less of a problem than explaining how we can see abstract truths about other things, like arithmetic. Indeed CHECK Jenkins says nothing about why conceptual knowledge would be less puzzling.

Another potential worry concerns the dependence on non-conceptual content. By the time things are conceptualized, Jenkins thinks, we already have the basic logical principles for arithmetic going. Once I am able to see things as two apples and three oranges, I have already accepted the principles of individuation and identity which yield arithmetical knowledge, so one cannot appeal to something like counting objects one sees to explain how we get those concepts. Rather, the experience that leads me to these good concepts (and hence, eventually, arithmetical knowledge) must be non-conceptual.

As Jenkins is aware, there's a huge literature on the very possibility of non-conceptual experiential content. Main worries about it include...

Finally there's the grounding relation Jenkins appeals to, and which she argues shows that our knowledge of arithmetic is empirical. What exactly does it mean to say that experience grounds arithemetical concepts? Insofar as people born into a sensory deprivation chamber are supposed to have extremely stunted development, it is plausible that experience is causally necessary for *all* concepts. So, what exactly is the contrast supposed to be? Clearly she holds that experience plays a crucial role in explaining how we got the good concepts of arithmetic as opposed to other tonk-like ones.

So, perhaps the distinction is between concepts where innate mental mechanisms make it likely that if we develop any concept of p at all we will develop a good one vs. concepts for which experience with the world is required to thin down the ranks of similar but incoherent concepts that we would be innately willing to accept. But this exposition is itself highly problematic. For one thing, there's the innate vs. non-innate distinction, which is generally rejected in philosophy of biology EXPAND?. For another thing, there's the notion of bad conceptual understudies (similar but incoherent concepts) which experience is supposed to be required to eliminate. To make sense of this we need a notion of conceptual similarity which is not obvious how to spell out in suitable detail. For example: experience is clearly supposed to be required to get us to not form the concept tonk. But which (if any) good concepts is tonk a version of, such that we would then know that these concepts are grounded empirically? Finally, there's the fact that, if this *is* what Jenkins means it would be bizzare to offer a book of philosophical arguments to that effect, instead of a collection of experiments in early childhood psychology. If having a bad tonk-like concept, is a matter of being inclined to accept certain verbal inferences (what else would it be?), then surely it is a *psychological* matter which bad concepts we are innately disposed to accept, and hence need experience to prune away. Even if Jenkins accepted that all concepts were empirically grounded, this doesn't seem like the kind of thing one could learn a priori (surely it would be possible to build a robot which simply refused to accept tonk-like inferences whatever its training, and ditto for bad versions of the concepts of arithmetic).

Thus I'm pretty confident Jenkins doesn't have this in mind. This train of thought above suggests that she must means something about the justification rather than merely the causal explanation for our winding up with good concepts. But remember that experience is not supposed to figure in the justification for propositions in arithmetic. Perhaps experience is supposed to justify us in having using the concepts? If so, it acts in a odd way since no one can remember this experience...

So, to summarize:

- The idea that there's non-conceptual content is a controversial point in philosophy of perception.
- The idea that experience can "ground" concept acquisition without playing a justificatory role in the conclusions drawn is not at all clear. What is this not-justificatory, but presumably not just causal relationship of grounding supposed to be?

- What is concept inspection, and how is it supposed to work? Jenkins admits that this is an open question for further research.

In contrast, my account of mathematical knowledge doesn't require any of this controversial, and under-explained philosophical machinery.

Putting things into Jenkin's vocabulary, my story becomes the following:

People 'grasp concepts' and 'inspect' them *simply in the sense that they are disposed to go from seeing things, to saying things, to being surprised if we then see other things, in certain ways*. Call these dispositions inference dispositions. No voyage to the third realm is necessary, nor self-surgery, nor a deep philosophical research project.

When these inference dispositions lead us to be surprised, (i.e. when our reasoning is practically unhelpful) we tend to modify them. Evolution may also play a role in making our psychology such that we aren't likely to form certain kinds of inference procedures which would then lead to us being surprised, or making us likely to modify these inference dispositions in helpful ways.[Note the latter tendency to modify behavior in responce to experience one way rather than another is something you already need to explain how rats can learn that pushing a lever releases food, so I don't think this is very controversial.]

Thus, it's not surprising that we should have wound up with the kind of combination of arithmetical inference dispositions, observational practices, and ways of applying arithmetic to the actual world, which makes our expected applications of arithmetic work out.

Note that no special new kind of justificatory status (grounding) is needed. We have an explanatory question: how can people manage to avoid bad concepts like tonk, and make true arithmetical judgements? And we have a clear (if boringly minimal to historians or psychologists - but it's their job to fill in the interesting contingent details) explanation. People just go from one sentence to another in a way that feels natural them (whether or not they are so fortunate as to be working with coherent concepts like +, rather than doing reasoning like Frege did about extensions). And when this natural-feeling reasoning leads to a surprise, they revise. Between this kind of correction, and a helpful nudge from nature with regard to relatively innate cognitive architecture, we manage to get methods of reasoning about arithmetic and particular counting procedures that correctly match up with each other.

Note also that my story doesn't need to appeal to any kind of special experience. Correcting your concepts is just a matter of having a bad concept and revising it. We know perfectly well what that's like. Many people can even remember having the intuitions associated with the naive conception of set and then being surprised by Russell's paradox and giving some of them up. The same goes for extremely intuitive but paradoxical reasoning about truth or heaps. That's what letting "experience"(i.e. causal interactions with the world, particularly blackboard and quiet room with a cup of coffee shaped regions of it) correct your experience feels like. You just feel strongly inclined to make some inferences, and turn out to be actually bad inferences, and conflict with other chains of reasoning you accept, so you stop making them. No psychedelic odyssey is required.

The core underlying problem here, I think, is an assumption that the right *explanation* of how we are able to know so much about mathematics must have some impact at the level of *justification*. This, arguably, is what moves Jenkins to posit a special kind of justification relation just for this purpose, and a special kind of experience which is stripped of contents so that it can play a non-question begging role in this justification. But, in fact, I think there's very little relationship. At most, the correct naturalistic account of how we are able to make correct judgements about mathematics figures as a defeater defeater. For, you might argue that anyone who finds it puzzling how a certain putative faculty for getting knowledge could work, has thereby some reason to distrust the results of that faculty. If so, the story told here defeats that reason for distrust, if it succeeds in addressing and removing the puzzlement.

The task of explaining how mathematical knowledge is compatible with naturalism, (or platonism, or mathematical facts being objective), to someone who finds it obvious that we have mathematical knowledge but has doubts about the latter positions, is quite different from the task of answering a skeptic about mathematical knowledge(as a whole) by providing some extra layer of justification for mathematical knowledge (as a whole). The former is my task, and Jenkins' task, as I understand her. The latter is, arguably, both unnecessary and impossible. (Most people think that the task of justifying logic to someone who rejects logic is hopeless, and plausibly there's no sharp distinction between mathematics and logic that would be relevant for these purposes.)

Now Jenkins accepts this point at the level of justification for mathematical beliefs, but seems to reject it at the level of grounding. If Jenkins wants to posit this new kind of justification I see no problem with that, but if the impression that grounding in experience is needed leads us to assume a priori what should be empirical matters of the relevant importance of evolutionary selection vs. individual learning on our developing certain patterns of inference rather than others, this is highly problematic.

extra Jenkins objections: - assumes analytic/synthetic distinction, obviously, for conceptual truths - ties useablitiy of concepts to indespencibility, but this is a bad feature of quine's account, surely we are justified in believing set theory even when it goes beyond physics, also cant be pick concepts that cut things up in rather arbitrary ways e.g. yuppie.

8.4 A new approach to logicism

Let me now say how the story I have proposed here relates to Logicism. I will argue that problems for contemporary logicism naturally lead us to ask questions about the basic motivations for the logicist project, which then lead to the key question in this thesis. Furthermore, if my proposal is correct, it can be seen as answering the questions which initially motivated logicism in a deeper way than even a the most successful logicist reduction could hope to.

So, let's start with the question, what is the point of Logicism? Where would the philosophical interest be in reducing mathematics to logic? I submit that the interest of Logicism comes from the idea that we take logic to have some special epistemic or metaphysical properties. If we could show mathematics to be really just logic, we could infer that mathematics has these properties as well.

8.4.1 Motivations for Logicism

In particular, consider the contrast between Kant's attitude towards logic and mathematics. As we discussed in the first chapter

8.4.2 Problems for Current Logicism

Remember that the logicist project is to show mathematical facts to be logical facts by producing a reduction of mathematics. However, this project turns out to face both technical and conceptual problems. So far as the technical problem of producing a reduction is concerned the problem is to find a system of 'logic' as a reduction base. The logical system which Frege originally proposed for a foundations of arithmetic turned out to lead to logical contradiction. And the new, consistent, systems which neo-logicists proposed are likely inadequate to account for more than arithmetic. [fill in, quotes]

Conceptually, there are problems about how to tell whether a given formal system counts as logic. Even if we had a perfectly acceptable translation of mathematics into some formal theory, there would be a question about whether this theory counted as logical. For example, there are some reasons to think that one can state reduce set theory to second order logic. But, then, it's a very controversial question whether second order logic counts as logic. There's no generally agreed on definition of logic, so it's not clear when the logicist can declare victory.

[add section on Frege's generality criterion?]

8.4.3 Stepping back

Putting these ideas about the epistemic motivation for logicism with the problems besetting neo-logicism above suggests a natural proposal. If you are interested in logicism because you think that logic is somehow less epistemologically mysterious, and a reduction of mathematics to logic would allow one to extend the same epistemology to mathematics, and hence explain how mathematical knowledge is possible, then why not start by figuring out how logical knowledge is possible.

If we had a clear picture of how creatures like us are able to get logical knowledge (which is supposedly a much easier problem), then we could look at this picture and see whether it applies to mathematical knowledge. Rather than arguing vainly about whether some particular formal system (e.g. second order logic) counts as logic, we can cut to the chase by asking directly whether this formal system is knowable in whatever easy way logic is supposed to be knowable.

once we do this (on my account) it turns out that a precisely analogous story can be told about mathematical knowledge, hence we get the desired explanation, and a picture where logic and math are continuous (this was intuitively appealing) without needing to provide a reduction to some stereotypical instance of logic - which seems impossible in the case of set theory.

-also without saying that math literally is logic, in case you want to reserve the term for things that are tidy a la Quine, or general a la Gila Sher

8.5 The Marriage of Rationalism and Empiricism

preserves what we like about platonism:

a)rationalism

b)possibility of unknowable truth

c) ordinary semantics has been preserved

historical motivations for empiricism are satisfied.