[This is a slightly updated version of an article which appears in Noûs. Vol. 48. Footnotes that have been added or amended are followed by a *.]

Moral Epistemology: The Mathematics Analogy*

There is a long tradition of comparing moral knowledge to mathematical knowledge. Plato compared mathematical knowledge to knowledge of the Good.¹ In recent years, metaethicists have found the comparison to be illuminating.² Sometimes the comparison is supposed to show that moral realism is peculiarly problematic. For example, James Rachels writes,

[How do we know moral facts?....In mathematics there are proofs….But moral facts are not accessible by…these familiar methods [1998, p. 3].

But other times the comparison is supposed to show that moral realism is no more problematic than mathematical realism. For example, Hilary Putnam writes,

[Arguments for “antirealism” in ethics are virtually identical with arguments for antirealism in the philosophy of mathematics; yet philosophers who resist those arguments in the latter case often capitulate in the former [2004, p. 1].

---

¹ Thanks to Hartry Field, Toby Handfield, Brian Leiter, Joshua May, Thomas Nagel, Stephen Schiffer, Jeff Sebo, and Sharon Street for helpful feedback.

In this paper I discuss apparent similarities and differences between moral knowledge and mathematical knowledge, realistically conceived. I argue that many of these are only apparent, while others are less philosophically significant than might be thought. The picture that emerges is surprising. There are definitely differences between epistemological arguments in the two areas, contrary to what Putnam suggests. However, these differences, if anything, seem to increase the plausibility of moral realism as compared to mathematical realism, contrary to what Rachels suggests. It is hard to see how one might argue, on epistemological grounds, for moral antirealism while maintaining commitment to mathematical realism. But it may be possible to do the opposite.

1. Self-Evidence and Disagreement

One reason that moral knowledge invites comparisons with mathematical knowledge is that moral knowledge can appear a priori like mathematical knowledge. Sarah McGrath writes,

[W]e do not attempt to discover what people ought to do in particular circumstances by designing and performing crucial experiments; nor do we think that our moral beliefs are inductively confirmed by observation. Experience does not appear to play an evidential role in our moral knowledge. In these and other ways, moral knowledge seems to resemble mathematical knowledge more than it resembles the kind of knowledge that is delivered by the empirical sciences [2010, 108-9].

McGrath’s point is that we seem to arrive at some moral conclusions on the basis of reflection alone. Empirical evidence obviously bears on the question of what concrete things are good, bad, obligatory, and so on. For example, that David’s action was an instance of hitting one’s friend bears on the question of whether that action was wrong. But we seem to
arrive at some conclusions of the form, “if x is F, then x is M”, where ‘F’ is an intuitively
descriptive predicate, and ‘M’ is an intuitively moral predicate, independent of such
evidence. This is how we are often said to arrive at “pure” mathematical conclusions -- such
as that 2 is prime or that any set of real numbers with an upper bound has a least upper
bound.

But how might we acquire even defeasibly a priori justified moral beliefs? Again, as Rachels
writes, “In mathematics there are proofs….But moral facts are not accessible by…these
familiar methods [1998, p. 3].” That is, there appears to be an established method by which
we might acquire at least defeasibly a priori justified mathematical beliefs – namely,
mathematical proof. But it can easily appear that there is no such method in the moral case.

It is natural to object that Rachels simply confuses two notions of “proof”. In the sense in
which mathematicians “prove” mathematical theorems, we could equally “prove” moral
theorems. What is called a “proof” of a given mathematical proposition, p, is really just a
deduction (or deduction-sketch) of p from the relevant axioms. In other words, a
mathematical proof of p shows that if the relevant axioms are true, then so too is p.\(^3\) Moral
propositions are open to analogous “proof”. We could deem a set of moral propositions
“axioms”, and then show that the relevant propositions deductively follow from them.

But this response might miss Rachels’ point. Of course, for any proposition, p, we can find a
set of propositions from which p follows. But in mathematics, unlike morality, certain sets of

\(^3\) Michael Gill seems not to appreciate what mathematical proof accomplishes in his [2007]. On p. 19 he writes:
“No one disagrees about…basic arithmetic….But disagreement…do[es] characterize…high-level mathematics.
What, after all, is the job of a mathematician if not to try to prove or disprove theorems about which there is
disagreement and perplexity?” Disagreement that can be simply resolved by proof or disproof is not
mathematical disagreement. It is logical disagreement (disagreement about what follows from what).
propositions are thought to possess a privileged status. Mathematical “axioms”, it is commonly said, are also axioms in something like Descartes’ sense. They are self-evident. Thus, proving that p follows from the mathematical axioms is tantamount to proving that p is true. By contrast, there is no set of moral propositions that enjoys such a privileged status.

An immediate difficulty with this position is that, if there are self-evident mathematical propositions at all, they are not typically axioms, but theorems. Russell pointed out in his 1907 Cambridge lecture on the subject that axioms of mathematics are often more doubtful than the theorems that they imply. Consider, for a contemporary example, the Axiom Replacement of set theory. This axiom states that the image of any set under a definable function is itself a set. This does not seem to be self-evident in any interesting sense. It easily implies such dramatic results as that there is an ordinal greater than all f(x), where f(0) = Aleph_0 and f(x+1) = Aleph_f(x) for all natural numbers, x. Nevertheless, the Axiom of Replacement does seem plausible. In particular, it implies that there are ordinals as great or greater than omega + omega in a natural way, and this proposition is arguably self-evident in some sense. The Axiom of Replacement is plausible, not because it is itself self-evident, but rather because it allows one to prove propositions that are – much as laws of physics gain plausibility by allowing one to predict propositions that are observationally evident.

This means that the view that we arrive at a priori justified mathematical beliefs by proving them from self-evident mathematical axioms must be too simple. If there is a method by which we arrive at a priori justified mathematical beliefs, it rather resembles the “method” by which we are often said arrive at such moral beliefs -- reflective equilibrium. We begin with

---

4 See Russell [1907].
5 * Boolos seems not even to accept this consequence in his [1999], p. 121.
6 * For the locus classicus, see Rawls [1971]. This runs contrary to what Kelly and McGrath claim in their [2010]. (Since writing this, it has come to my attention that Charles Parsons also mentions Rawls’ notion of
particular propositions that we deem plausible and seek general principles which systematize those propositions. Such principles, in turn, often pressure us to reject the propositions with which we began as we seek optimum harmony between the two. Of course, proof can play a role in the process. With regard to the Axiom of Replacement, proof tells us that this axiom implies an array of plausible propositions (such as that there are ordinals as great as or greater than omega + omega), while arguably refuting some plausible propositions as well (such as, perhaps, that there do not exist ordinals greater than all f(x) above). However, whether such results show that we ought to endorse or reject the Axiom of Replacement is left open. That depends on which alternative would facilitate equilibrium among our mathematical beliefs.

Despite this complication, an analog to Rachels’ point may still hold. It may still be true that the particular mathematical propositions on which the method of reflective equilibrium – as opposed to the method of mathematical proof -- operates are self-evident in some sense in which corresponding moral propositions are not. Does the proposition that that there are ordinals as great as or greater than omega + omega possess a privileged status that, say, the proposition that murder is bad lacks? I assume that most of us believe that murder is bad, absent much contemplation, as most who understand the proposition that there are ordinals as great as or greater than omega + omega believe that proposition. However, it also seems that not everyone does. Maybe this is a relevant epistemological difference between morality and mathematics. Everyone – or everyone who understands – epistemically basic mathematical propositions believes them, but this is not so of epistemically basic moral propositions.

This cannot be right, however. There are obviously people who understand the relevant mathematical proposition that do not believe it either. Indeed, as Williamson has noted, this is apt to be the case for any proposition whatever due to variation in background philosophical commitment. For example, for any area of discourse, F, be it morality, physics, or mathematics, there are error-theorists with respect to F. These are philosophers who understand F-claims as well as the rest of us, but take them to involve commitments that are not satisfied. Hartry Field is an error-theorist about mathematics. As an error-theorist, he explicitly rejects the claim that there are ordinals as great as or greater than omega + omega.

Maybe the sense in which it is self-evident that there are ordinals as great as or greater than omega + omega, but not that murder is bad, is that, bracketing error-theorists, everyone who understands the former, but not the latter, proposition believes it. However, even this restricted principle is false. There are those who do not believe that there are ordinals as great as or greater than omega + omega, although they are not mathematical error-theorists.

Speaking of a variety of axioms, George Boolos writes,

I am by no means convinced that any of the axioms of infinity, union, or power [set]…force themselves upon us or that all the axioms of replacement that we can comprehend do…. [T]here is nothing unclear about the power set axiom [for example]…. But it does not seem to me unreasonable to think that… it is not the case that for every set, there is a set of all its subsets [1999, 130 – 131].

---

7 See Williamson [2006]. Williamson’s illustrations are somewhat different than the one I give presently.
8 See Field [1980] and [1989].
9 * Note that Boolos does not himself seem to reject all of these axioms. His point seems to be that those who do reject them are not necessarily irrational.
Indeed, for core claims in every area of mathematics – from set theory to analysis to arithmetic – there are some non-error-theorists who deny those claims. For example, Hermann Weyl famously rejects the fundamental principle of the calculus, the Least Upper Bound principle (according to which every non-empty set of real numbers with an upper bound has a least upper bound).10 Similarly, Edward Nelson even rejects the Axiom of Mathematical Induction (which states that if 0 has the property, F, and if n+1 has, F, whenever n has F, then all natural numbers have the property, F).11 Harvey Friedman contends, “I have seen some…go so far as to challenge the existence of 2^100...”12

Of course, many -- though not all -- such doubts are still evidently connected to intuitively philosophical considerations. The metaphysical doctrine that there do not exist (mere) sets

---

11 * See Nelson [1986]. In the published version of this paper, I assert that Nelson also rejects the Successor Axiom, according to which every natural number has a successor (see Nelson [1986], p. 176). It now seems to me more accurate to say that Nelson refuses to assert that for every actual (or “genetic”) x, there is an actual y, and y is the successor of x. He instead advances a weak inference rule. At MathOverflow (http://mathoverflow.net/questions/142669/illustrating-edward-nelsons-worldview-with-nonstandard-models-of-arithmetic) Nelson clarifies his position as follows.

To avoid vagueness, let Q* be Q with the usual relativization schemata adjoined. Construct a formal system F by adjoining an unary predicate symbol ψ, the axiom ψ(0), and the rule of inference: from ψ(a) infer ψ(Sa) (for any term a). I think this is an adequate formalization of the concept of an “actual number”. Is ψ(80^5000) a theorem of F? I see no reason to believe so. Of course, one can arithmetize F in various theories, even Q*, and prove a formula ∃p[p is an arithmetized proof in F of ‘ψ(80^5000)’], but to conclude from this that there is a proof in F itself of ψ(80^5000) appears to me to be unjustified.

Contrast F with the theory T obtained by adjoining to Q* a unary predicate symbol ϕ and the two axioms ϕ(0) and ϕ(0)&∀x′[ϕ(x′)→ϕ(Sx′)]→ϕ(x). Then one can easily prove in T ϕ(80^5000) or even ϕ(80^5000...5000). The ellipsis means that the iterated exponential term is actually written down.

In response to a question, he continues,

Q* proves that addition and multiplication are total, but does not prove that exponentiation is total. The situation is different with F and ψ; a rule of inference is far more restrictive than an implication, and F does [does not even] prove that if ψ(x) then ψ(Sx)

For a more explicit rejection of the Successor Axiom, see Doron Zielberger, “’Real” Analysis is a Degenerate Case of Discrete Analysis”, in Aulbach, Bernd, Saber N. Elaydi, and G. Ladus (eds.), Proceedings of the Sixth International Conference on Difference Equations, Augsburg, Germany: CRC Press, p. 33. (Thanks to Walter Dean for helpful discussion.)

12 See Friedman [2002], p. 4. For more on disagreement over axioms in mathematics, see Clarke-Doane [Forthcoming], Forster [2010], Fraenkel, Bar-Hillel, and Levy [1973], Maddy [1988a] and [1988b], Quine [1969], and Shapiro [2009].
in-extension informs the view of many of those who reject the Axiom of Choice (which states that corresponding to every nonempty sets of sets there is another containing exactly one member from each member of the first). Similarly, predicativism, the semantic doctrine that it is not coherent to define an object in terms of a set to which it belongs, informs the view of Edward Nelson who rejects Mathematical Induction. But such considerations inform one’s views with respect to morality as well. For example, any comprehensive moral outlook will be informed by metaphysical considerations concerning the existence God or by semantic considerations concerning the apparently action-guiding character of moral language.

It might be thought that I have focused on the wrong candidates for self-evident mathematical propositions. What if we consider the proposition that $1 + 1 = 2$ or that 7 is prime? I am not aware of any non-error-theorist who rejects these claims. But this still fails to establish a disanalogy between morality and mathematics. I am not aware of any non-error-theorist who rejects the claim that burning babies for fun is wrong, or that it is sometimes permissible for some people to stand up. I focus on interesting examples of apparently epistemically basic mathematical propositions because only they have the potential to illuminate our justification for believing in all of standard mathematics. Such propositions as that $1 + 1 = 2$ or that 7 is prime are comparably epistemically inert as the moral propositions just mentioned. For example, the propositions that $1 + 1 = 2$ and that 7 is prime are perfectly consistent with a radically heretical theory of arithmetic, such as Edward Nelson’s, which fails to validate even such fundamentals the Axiom of Mathematical Induction.

Perhaps self-evidence in the relevant sense does not require unanimity among non-error-theorists. What is required – at least if we are to be justified in regarding a given proposition as self-evident -- is general agreement. For any interesting proposition, there are those who
reject it. But that does not show that there are no interesting self-evident propositions. What would perhaps show this is widespread disagreement. There is such disagreement with respect to moral propositions, but not, it might be thought, with respect to mathematical ones. Brian Leiter writes,

[P]ersistent disagreement on foundational questions…distinguishes moral theory from inquiry in…mathematics…certainly in degree [2009, p. 1].

Even this suggestion, however, is doubtful. We may distinguish two ways in which disagreement from an area, D, may be widespread. First, it may be propositions-widespread, such that, for many kinds of D-propositions, p, there is a pair of people, P, such that P disagrees with respect to p (kinds of propositions, as opposed to raw numbers of them, is relevant because disagreement over p always involves disagreement over not-p, not-not-p, not-not-not-p, \textit{ad infinitum}). Second, disagreement may be people-widespread, such that, for many pairs of people, P, there is a proposition, p, such that P disagrees with respect to p.

I have already observed that mathematical disagreement is propositions-widespread just as surely as is moral disagreement, even bracketing error-theorists. In mathematics, there is disagreement over everything from set theory to mathematical analysis to arithmetic. However, it might still be thought that only moral disagreement is people-widespread.

It is obviously true that, in terms of raw numbers of people, there are not many pairs of people, P, such that there is a mathematical proposition, p, over which P disagrees. However, raw numbers of people who disagree over p cannot be what is relevant to the self-evidence of p. To illustrate, suppose that p is a proposition that only a small subset of the population
understands, but over which a significant proportion of that subset disagrees. Then surely p has no better claim to being self-evident than a claim that everyone understands and over which the same proportion disagrees. If any aspect of the actual distribution of opinion has bearing on the self-evidence of p, then it is the proportion of those with views as to whether p that disagree. The raw number of people that disagree over p is irrelevant.\textsuperscript{13}

With this clarification, mathematical disagreement seems to be people-widespread after all. Very few mathematicians, let alone lay people, seem to have a serious view as to what axioms in mathematics are true (as opposed to what follows from those axioms). However, among those who do, there is notorious disagreement. Bell and Hellman write,

Contrary to the popular (mis)conception of mathematics as a cut-and-dried body of universally agreed upon truths…as soon as one examines the foundations of mathematics [question of what axioms are true] one encounters divergences of viewpoint…that can easily remind one of religious, schismatic controversy [Bell and Hellman 2006, p. 64].\textsuperscript{14}

Of course, I do not suggest that mathematical disagreement is thoroughly analogous to moral disagreement. One obvious difference between mathematical disagreement and moral disagreement is that the latter tends to track with personal and religious investment in a way

\textsuperscript{13} McGrath fails to appreciate something like this point in her discussion of the relevance of philosophical disagreement to the question of whether we are justified in our common sense beliefs. See her [2007].

\textsuperscript{14} * Bell and Hellman qualify this passage with the following. While there is indeed universal agreement on a substantial body of mathematical results…as soon as one asks questions concerning fundamentals, such as…”What axioms can we accept as unproblematic?”…we find we have entered a mine-field of contentiousness [2006, 64]. If such “fundamentals” are contentious, then the “substantial body of mathematical results” to which Bell and Hellman refer may be limited to logical truths of the form “If T, then P”, where T is the finite subset of axioms used in the proof of P.
that the former does not. But this disanalogy only bolsters the suggestion that mathematical propositions have no better claim to being self-evident than moral ones. Mathematical disagreement typically occurs among the intellectually virtuous and seems to be largely independent of personal and religious investment. Such disagreement raises doubts about the supposed self-evidence of the relevant propositions far more effectively than paradigmatic moral disagreement. Unlike moral disagreement, mathematical disagreement cannot be explained away as reflecting the above distorting influences.

Perhaps the relevant sense of self-evidence is divorced from belief altogether. The sense in which epistemically basic mathematical propositions are self-evident is that anyone who considers them will find them to be plausible, even if philosophical or other commitments lead one astray. This is consistent with its being the case that a high proportion of those with views as to which mathematical propositions are true disagree.

But even this proposal is unlikely to distinguish morality from mathematics. It is doubtful that there is relevant consensus as to the plausibility of many epistemically basic mathematical propositions of interest. When Hermann Weyl proclaimed of the Least Upper Bound principle “in any wording [it] is false”, he did not seem to be registering a theoretical doubt.15 Or consider the “intuitions” that are traded in discussions of the backbone to standard set theory, the Axiom of Foundation (which states that every set occurs at some level of the cumulative hierarchy). The key questions here are whether it is plausible that there are sets that contain themselves or whether it is plausible that there are sets with infinitely descending chains of membership. Some seem to think that it is -- to banish such

sets would be unnaturally restrictive. But many others seem to think that it is not -- such sets are pathological. This seems to be a straightforward case of people disagreeing as to the plausibility -- not just the truth -- of epistemically basic mathematical propositions.

Of course, there remains a sense of “self-evident” in which it might be suggested that basic mathematical propositions are self-evident, and basic moral propositions are not. Rather than appealing to descriptive claims about what people do believe or find plausible, one might appeal directly to a normative claim about what people ought to believe or find plausible. The sense in which basic mathematical propositions are self-evident is that anyone – or anyone who understands – those propositions ought to believe them or find them plausible. The same is not true of basic moral propositions. But in the absence of an additional difference between basic moral and mathematical propositions (such as that only basic mathematical propositions are deemed plausible by all who understand them), this suggestion obviously just assumes what is at issue -- that epistemically basic mathematical propositions possess a privileged status that epistemically basic moral propositions lack.

2. The Quine-Putnam Indispensability Argument and Harman’s Objection

I have argued that, contrary to a common view, mathematical propositions do not seem to be “provable” or “self-evident” in any interesting sense in which moral propositions do not. This suggests that our mathematical beliefs have no better claim to being at least defeasibly a priori justified than our moral beliefs. Nevertheless, there remains the less commonly advanced possibility that our mathematical beliefs (perhaps additionally) enjoy empirical

---

16 See, for instance, Rieger [2011]. See also Quine’s discussion of the Axiom of “Regularity” in his [1969] and Forster’s discussion of the Axiom of Foundation in his [Forthcoming].

17 See Maddy’s discussion of the Axiom of Foundation in her [1988] or Boolos [1971], 491.
justification, while our moral beliefs do not. After all, mathematics, but not morality, appears to be indispensable to our best empirical scientific theories. Gilbert Harman writes,

In explaining the observations that support a physical theory, scientists typically appeal to mathematical principles. On the other hand, one never seems to need to appeal in this way to moral principles. Since an observation is evidence for what best explains it…there is indirect observational evidence for mathematics. There does not seem to be observational evidence…for basic moral principles [1977, pp. 9 - 10].

Harman is naturally read as sketching an argument for the view that we have empirical justification for believing in mathematical hypotheses, but not for believing in moral hypotheses. If Harman’s argument has any relevance to the realism-antirealism debate in the corresponding areas, then he must have intended a stronger conclusion. The stronger conclusion is that we have empirical justification for believing in mathematical, but not moral, hypotheses, realistically conceived – i.e., conceived, roughly, as being true or false, interpreted at face-value, relevantly independent of human minds and languages.18

Understood in this way, the argument that Harman sketches depends on three premises. First, mathematical hypotheses, realistically conceived, figure into our best empirical scientific theories. Second, if a hypothesis figures into our best empirical scientific theories, then we have at least defeasible empirical justification for believing that hypothesis. Third, moral hypotheses, realistically conceived, do not figure into our best empirical scientific theories.

---

18 For a detailed explication of realism in the relevant sense, see Section I of Clarke-Doane [2012].
The first two premises constitute what is known as the *Quine-Putnam Indispensability Argument* for mathematical realism.\(^{19}\) The third premise is often called *Harman’s Objection*.

The Quine-Putnam Indispensability Argument is problematic. Its most well-known problems involve the first premise. It may be possible to formulate attractive analogs to our empirical scientific theories that involve no commitment to mathematics, realistically conceived. There are a number of approaches.\(^{20}\) The most influential has been that of Field [1980] and [1989]. Field concedes that our empirical scientific theories currently involve commitment to mathematical hypotheses, realistically conceived. However, he argues that these theories ought to be regarded as convenient shorthand for better theories that involve no such commitment. Here is one of the most compelling reasons that Field offers for this view.

\[
\text{[I]t seems to me that…one wants to be able to explain the behavior of the physical system in terms of the intrinsic features of that system, without invoking extrinsic entities (whether mathematical or non-mathematical) whose properties are irrelevant to the behavior of the system being explained). If one cannot do this, then it seems rather like magic that the…[relevant] explanation works [1985, p. 193].}
\]

Field observes that mathematical hypotheses do not play the same role in empirical scientific explanations as hypotheses about such things as electrons. Electrons and their charges are *causally relevant* to the behavior of bodies. But this is not true of mathematical entities. Even if a given function’s taking on such and such a value at such and such an argument helps *explain* the moon’s orbiting Earth, no one thinks that that function or its taking on such

---

\(^{19}\) See Quine [1951] and Putnam [1971].

\(^{20}\) For a survey, see Rosen and Burgess [1997].
and such a value helps *cause* the moon to orbit Earth. This leaves it mysterious how the relevant explanation is supposed to work. If possible, we ought to avoid such mystery.

This suggests a problem with the second premise in the Quine-Putnam Indispensability Argument as well. Given the peculiar role that mathematical hypotheses play in empirical scientific theories, it is highly questionable whether they are justified by observation in the same way as the rest of those theories. If mathematical theories *were* justified in this way, then key elements of empirical scientific practice would be very mysterious. Empirical scientists do not seem to think twice about “postulating” the likes of functions. But postulations of new particles or forces are met with empirical scrutiny. Indeed, there is typically a great deal of arbitrariness as to *what* functions are to be “postulated” in a given instance. Many functions will often serve the relevant explanatory purpose equally. But scientists do not seem to be correspondingly reticent about asserting the resulting theory.  

Despite these difficulties with the Quine-Putnam Indispensability Argument, it does retain some advocates. What, then, are the prospects for undercutting Harman’s Objection? Compared to the prospects for establishing the corresponding premise of the Quine-Putnam Indispensability Argument, they do not seem to be very good. Canonical examples of moral explanations – such the well-known slavery example of Sturgeon [1985] – seem to be dramatically less impressive than canonical examples of mathematical explanations. The only question surrounding the former seems to be whether they are more plausible than the

---

21 One explanation of this practice, which rejects the Quine-Putnam Indispensability Argument, is that of Sober [1993]. Sober argues that theories are only confirmed relative to competitors. But since all of the competitors to empirical scientific explanations entail the same mathematical theories, the mathematical theories do not accrue confirmational support.

22 See, for instance, Colyvan [2001].
obvious non-moral explanations on offer. By contrast, in the mathematical case, there are
typically no apparent (remotely attractive) non-mathematical explanations on offer at all.23

To be sure, Harman’s Objection is not as easy to establish as it is sometimes said to be.
Defenders of Harman’s Objection sometimes suggest that it depends on the mere claim that
moral hypotheses fail to figure into our best empirical scientific explanations in a causal
capacity.24 They argue, for example, that the badness of slavery did not plausibly cause the
rise of abolitionism. However, what matters for Harman’s Objection is whether moral
hypotheses figure into our best empirical scientific explanations in any capacity. Again,
(virtually) everyone concedes that mathematical hypotheses fail to figure into our best
empirical scientific explanations in a causal capacity. Still, I suspect that little turns on this
complication. Compared with mathematical hypotheses, it does not seem plausible that
moral hypotheses figure into our best empirical scientific explanations in any capacity.

What does this show? It at most shows that some mathematical hypotheses enjoy stronger
empirical justification than any moral hypotheses. It does not show that the range of
mathematical hypotheses that we actually endorse do. Indeed, important parts of standard
mathematics are almost certainly not indispensable to our best empirical scientific theories.
For example, even fundamentals of modern set theory, such as the Axiom of Replacement
mentioned above, seem plausibly dispensable to our best empirical scientific theories. Quine
himself accepted this point and declared the higher reaches of set theory vacuous.25 But more

23 * It follows from Craig’s Theorem that there will always be some (recursively enumerable) non-mathematical
“explanation” available of a given empirical phenomenon, although it need not have any intuitive appeal. See
Craig [1953] and Field [1980], p. 8.
24 See Majors [2007].
25 Quine writes: “I recognize indenumerable infinities only because they are forced on me by the simplest
known systematizations of more welcome matters. Magnitudes in excess of such demands, e.g., or
inaccessible numbers, I look upon only as mathematical recreation and without ontological rights [1986, 400].”
might have to be rejected. It could be that while Field is incorrect that all mathematical hypotheses are dispensable in empirical science, any that, say, quantify over uncountable totalities are. If so, then we still lack empirical justification for believing most of standard mathematics. *Insofar as we are interested in the relative epistemological merits of the moral and mathematical hypotheses that we actually endorse*, the Quine-Putnam Indispensability Argument and Harman’s Objection seem, therefore, to be largely beside the point.

3. The Benacerraf-Field Problem and Mackie’s Problem

I have been discussing how we might be defeasibly justified in believing moral or mathematical hypotheses, realistically conceived. I have been arguing that the question does not seem to be more tractable in the mathematical case than it does in the moral. But assuming that we are so justified in believing mathematical or moral hypotheses, it is often argued that that our justification is *undermined* by an epistemological quandary. A classic statement of the quandary in the mathematical case is due to Paul Benacerraf. He writes,

> I find [mathematical realism] both encouraging and troubling…. [S]omething must be said to bridge the chasm, created by…[a] realistic…interpretation of mathematical propositions, between the entities that form the subject matter of mathematics and the human knower [1973, p. 675].

To appreciate Benacerraf’s worry, consider the sentence, “2 is prime”. Realistically conceived, this sentence is about a *number*. It predicates the property of being prime of it.

But how could we know this? Numbers and their properties would not be causally

---

In his [1969], he accepts the Axiom of Replacement applied to the von Neumann ordinals. Again, Quine is committed to the falsity of the Axiom of Foundation.  

26 See Feferman [1993].
efficacious. Of course, objects or properties may not strictly be the relata of causation anyway. Perhaps events or facts are. But whatever the literal relata of causation, it does not seem that mathematical objects or properties could participate in them. It does not seem that, say, the number 2, the property of being prime, 2’s being prime, or the fact that 2 is prime, could cause anything. These remarks make it hard to see how we could know that 2 is prime.

Hartry Field has transformed Benacerraf’s quandary into a clear challenge for the mathematical realist. His idea is to set aside the question of what it would take for us to know mathematical truths such as that 2 is prime. On any reasonable view of justification, Field suggests, our justification for believing in them would be undermined if it appeared in principle impossible to explain the reliability of our corresponding beliefs. Field writes,

We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in [mathematical] entities. These positive reasons might involve…initial plausibility…[or] that the postulation of these entities appears to be indispensable…. But Benacerraf’s challenge…is to…explain how our beliefs about these remote entities can so well reflect the facts about them…. [I]f it appears in principle impossible to explain this, then that tends to undermine the belief in mathematical entities, despite whatever reason we might have for believing in them [1989, p. 26].

Field’s challenge should not be confused with the challenge to actually explain the reliability of our mathematical beliefs. That challenge would clearly be too stringent. Consider our perceptual beliefs. People were presumably justified in holding those before anything like an explanation of their reliability became available. Even today we have no more than a sketch
of such an explanation. But it is less plausible that people would have been justified in holding their perceptual beliefs if it appeared to them in principle impossible to explain the reliability of their corresponding beliefs.

It is widely supposed that a simple analog of the Benacerraf-Field challenge arises for moral realism. Indeed, John Mackie’s *Ethics: Inventing Right and Wrong* contains what is commonly taken to be a direct application of Benacerraf’s reasoning. Mackie writes,

> It would make a radical difference to our metaphysics if we had to find room for objective values -- perhaps something like Plato’s Forms -- somewhere in our picture of the world. It would similarly make a difference to our epistemology if it had to explain how...objective values are or can be known, and to our philosophical psychology to allow such knowledge [1977, p. 24].

Mackie is naturally read as echoing Benacerraf -- “something must be said to bridge the chasm, created by...[a] realistic...interpretation of [moral] propositions, between the entities that form the subject matter of [moral]ity and the human knower.” Following Field, the challenge for the moral realist is to explain the reliability of our beliefs about these “remote entities”.

However, there is a difference between the moral case and the mathematical that is obscured by Mackie’s challenge. Contrary to what Mackie suggests, moral claims are not about peculiarly moral entities, in the way that mathematical claims about mathematical objects. Moral claims are about the likes of people, actions, and events. For example, the sentence

---

27 See, for instance, Huemer [2005], Street [2010], Enoch [2010], or Schechter [2010].
“Osama Bin Laden is wicked” is not literally about wickedness in the way that “2 is prime” is about the number 2. The latter sentence refers to (or first-order quantifies over) the number 2. But the former sentence does not refer to (or first-order quantify over) wickedness. It refers to a man, Osama Bin Laden.\textsuperscript{28}

The point may seem pedantic, but it alters the dialectic. The force of the Benacerraf-Field challenge depends on the following plausible principle.

\textit{Principle #0:} It appears in principle impossible to explain the reliability of our beliefs that are \textit{both} about causally inert objects \textit{and} that predicate causally inert properties.

Note that the puzzle in the mathematical case is not \textit{just} that the relevant \textit{properties} fail to participate in causal relata. \textit{The objects of which they are predicated} fail to so participate as well. By contrast, moral truths are about objects that \textit{do} participate in causal relata. For example, Osama Bin Laden, the Holocaust, and the Lincoln’s freeing of the slaves, all so participate. At most, moral \textit{properties} fail to participate in causal relata. Whether there is a compelling principle that covers the moral case as well as the mathematical is not obvious.

Let us try to locate such a principle. The most obvious potential such principle is simply this:

\textit{Principle #1:} It appears in principle impossible to explain the reliability of our beliefs that predicate causally inert properties.

\textsuperscript{28} Similarly, the sentence “Killing is wrong” is not about wrongness, interpreted at face-value. It is about killings – somewhat as “Perfect numbers are sums of their proper positive divisors” is not about proper-positive-divisorhood, interpreted at face-value. It is about perfect numbers.
Principle #1 is doubtful. Imagining that mathematical beliefs merely predicated causally inert properties provides an immediate reason for doubt. Imagine that while the peculiarly mathematical properties of mathematical objects failed to figure into causal relata, many other of their properties did not. As with the objects of moral predication -- people, actions, and events -- our empirical scientific theories had application to the behavior of mathematical objects, and we could observe or be otherwise causally affected by them. In this scenario, the mystery surrounding mathematical knowledge would surely seem much less compelling. Indeed, if, as some have argued, such physical postulates as electrons are only describable in mathematical terms, such postulates may actually be analogous to mathematical objects as I have imagined them here. But whatever epistemological challenges surround such physical postulates as electrons do not seem to be at all analogous to the Benacerraf-Field challenge.29

The same point can be made with more homely examples. Consider the property of being a restaurant. Prima facie this property has as strong a claim to being superfluous in causal explanations as any moral property. This is not to deny that there are “common sense” explanations that invoke the property of being a restaurant in a causal capacity. As in the moral case, we folk may find it natural to explain various phenomena with reference to the causal powers of being a restaurant. But if any causal explanation that invokes moral properties can be replaced by a better one that does not, then something similar would certainly seem to be true of the property of being a restaurant. It is not as if the property of being a restaurant is the “postulate” of a special science, or that it supports counterfactuals in any straightforward sense in which moral properties do not (more on this below). And, yet, I

29 Thanks to Hartry Field for suggesting this example.
know of no philosopher that concludes on this basis that it is in principle impossible to explain the reliability of our beliefs that predicate the property of being a restaurant.\textsuperscript{30}

What is it about the property of being a restaurant that makes it seem possible to explain the reliability of our corresponding beliefs? Perhaps it is the fact this property \textit{supervenes} on causally efficacious properties.\textsuperscript{31} Had the property of being a restaurant been distributed differently, the causally efficacious properties on which it supervenes would have been distributed correspondingly so. Because causally efficacious properties shaped our beliefs, our beliefs about restaurants would have reflected the difference. For example, had the Diner not been a restaurant, people would not have paid money in order to be served food at the Diner. But had people not paid money in order to be served food at the Diner, George would not have believed that the Diner was a restaurant. This suggests that Principle #2 might instead be compelling.

\textit{Principle #2}: It appears in principle impossible to explain the reliability of our beliefs that predicate causally inert properties which fail to supervene on causally efficacious properties.

The problem with Principle #2 is not that it is not compelling. There \textit{does} seem to be something deeply mysterious about the claim that our beliefs that predicate causally inert properties which fail to supervene on causally efficacious ones are reliable. The problem

\textsuperscript{30} Notice that the claim is not that we ought to agree that the property of being a restaurant is instantiated (along with whatever properties get mentioned in our best sciences, for example). The claim is that \textit{assuming} that that our restaurant beliefs are defeasibly justified and true, it seems possible to explain their reliability. The Benacerraf-Field challenge \textit{assumes} for the sake of argument the defeasible justification and truth of our relevant beliefs. It purports to \textit{undermine} the justification of those beliefs on the grounds that, \textit{even given the aforementioned assumptions}, it must be an “inexplicable coincidence” that our relevant beliefs are reliable.

\textsuperscript{31} For a suggestion along these lines, see Sturgeon [1986].
with Principle #2 is that it surely fails to cover the moral and mathematical cases. It is a virtual datum that moral properties would supervene on causally efficacious properties. In the mathematical case the supervenience is trivial – there can be no change in the distribution of mathematical properties absent a change in the distribution of causally efficacious ones because there can be no change in the distribution of mathematical properties simpliciter.

Perhaps what makes the reliability of our restaurant beliefs comprehensible is that the distribution of the property of being a restaurant is a *conceptual consequence* of the distribution of causally efficacious ones. Given a distribution of causally efficacious properties, the distribution of restaurant properties follows as a matter of *conceptual necessity*. There cannot even be an *intelligible* – let alone metaphysically possible -- worry that the property of being a restaurant might have been distributed differently while the distribution of causally efficacious properties remained the same. These considerations suggest that rather than Principle #2, Principle #3 might be compelling.

*Principle #3:* It appears in principle impossible to explain the reliability of our beliefs that predicate causally inert properties whose distribution is not a conceptual consequence of the distribution of causally efficacious properties.

Unlike Principle #2, Principle #3 does plausibly cover both the moral and mathematical cases. It is commonly acknowledged that there are few if any conceptual constraints as to the distribution of moral properties, given a distribution of causally efficacious properties.32 Even if it is not *metaphysically* possible, it is *intelligible to imagine* the distribution of the

---

32 For suggestions along these lines, see Blackburn [1971], Horgan and Timmons [1992] and Gibbard [2003].
moral properties being very different, while the distribution of the causally efficacious properties remains the same.

It might be thought that such a view about mathematics is harder to defend. Certainly there is a longstanding tradition according to which mathematics is, in some sense, just a body of conceptual truths. But, first, even if it were a conceptual truth that, say, if there are (pure) sets, then they instantiate the property of being well-orderable, it is hard to see how it could be a conceptual truth that there are sets. As Hume and Kant underscored, how can it be a conceptual truth that something exists? Second, even if it were a conceptual truth that there are such things as sets, it is hard to imagine a non-question-begging argument for the view that it is unintelligible to imagine the distribution of mathematical properties being different (while the distribution of causally efficacious properties remains the same). That this distribution is actually so different is precisely what George Boolos alleges with respect to sets, what Hermann Weyl alleges with respect to real numbers, and what Edward Nelson alleges with respect to natural numbers. Of course, it may be unintelligible to imagine certain distributions of mathematical properties. For example, perhaps it is unintelligible to imagine that 4 has the property of being prime. However, as before, something similar can be said of certain moral truths. It is dubiously intelligible that a given instance of torturing children just for the fun of it has property of moral goodness (in a world with the same distribution of causally efficacious properties). As the variation of disagreement in the two areas suggests, certain claims are arguably definitive of the corresponding subjects.

I take Principle #3 to articulate the epistemological challenge for moral realism that lies behind recent work by Alan Gibbard, Sharon Street, and others. Their concern seems to be

---

33 See Clarke-Doane [2012].
that, given all the conceptually – even if not metaphysically -- possible ways that the moral properties might correspond to the causally efficacious ones, it seems inexplicable that “we just happened to land” on the true such correspondence. Sharon Street writes,

[A]s a purely conceptual matter…normative truths might be anything…. Noting this sense in which the normative truth might be anything, and noting the role of…[causal] forces in shaping the content of our basic normative tendencies, we may wonder whether…these forces would have led us to…the…normative truth…. [2008, p. 208]

According to Street,

[T]he realist must hold that an astonishing [inexplicable] coincidence took place -- claiming that as a matter of sheer luck…[causal] pressures affected our evaluative attitudes in such a way that they just happened to land on or near the true normative views among all the conceptually possible ones [2008, p. 208].³⁴

There is something prima facie plausible about this line of thought. One would like to rule out “from the inside” the worry that the distribution of the relevant properties might have been different while the distribution of causally efficacious properties remained the same. But I submit that not even Principle #3 is compelling. In an important sense, it overgeneralizes like Principle #1.

³⁴ See also Gibbard [2003]. Note that, like Field’s worry, Street’s is not that the moral realist’s moral beliefs are not (defeasibly) justified. Street grants the moral realist -- for the sake of argument -- both the truth and (defeasible) justification of her beliefs. Street purports to undermine the justification of those beliefs. Her idea seems to be that, even if our moral beliefs are (defeasibly) justified and actually true, realistically conceived, the truth of those beliefs could only be an inexplicable coincidence.
Consider the restaurant case again. Let us grant that it is unintelligible to imagine that, *given* that the property of being a restaurant is instantiated at all, its distribution could be different while the distribution of causally efficacious properties remained the same. Still, it seems intelligible to imagine *that the property of being a restaurant is not instantiated*. Indeed, metaontologists who argue that there are no “common sense” objects, but only elementary particles appropriately arranged, argue for precisely the claim that there are no such things as restaurants.\(^{35}\) Credible or not, such proposals seem to be *intelligible*. But if they are, then there is still a sense in which we can intelligibly worry that the property of being a restaurant could have been distributed differently while the distribution of causally efficacious properties remained the same. We can intelligibly worry that the property of being a restaurant *might not have been distributed at all*.

Is this worry relevantly different from the worry that we have been considering in the moral or the mathematical cases? It might seem to be. It is natural to think that one task is explaining how, *given* that properties of a kind, F, are instantiated, we are reliable detectors of their distribution, and that another task is explaining how we are reliable detectors of *the fact that F-properties are instantiated*. But the relevance of this distinction is dubious upon inspection.

Suppose that one became convinced that it was simply unintelligible to worry that, *given* that properties of a kind, F, are instantiated, their distribution could be different while the distribution of causally efficacious properties remained the same. Still, one may worry that *only F-like properties* – as opposed to F-properties – could be instantiated, while the distribution of causally efficacious properties remained the same. For example, suppose,

\[^{35}\] See, for instance, Merricks [2001] or Van Inwagen [1990].
contrary to the above, that it is unintelligible to imagine that there are sets that fail to instantiate the property of being well-orderable. If so, then there is no intelligible worry that certain sets might have failed to instantiate the property of being well-orderable while the distribution of causally efficacious properties remained the same. Nevertheless, there is an obvious worry in the same spirit as this one: there might have been no sets at all, but only set-like things – where set-like things are just like sets except some set-like things fail to instantiate the property of being well-orderable.

In general, there is a “translation scheme” between talk of the distribution of properties of a kind, F, being different (under the assumption that this is intelligible), and talk of there being only F-like properties distributed (under the assumption that it is not). I am not claiming that the translation preserves meaning. I am claiming that it preserves the relevant epistemic mystery. If one endorses Principle #3, one ought to also endorse Principle #4.

Principle #4: It appears in principle impossible to explain the reliability of our beliefs that predicate causally inert properties whose distribution or instantiation is not a conceptual consequence of the distribution of causally efficacious properties.

Because Principle #4 would again imply that it appears in principle impossible to explain the reliability of our relevantly uncontroversial beliefs – such as our beliefs about restaurants -- both Principle #3 and Principle #4 must be rejected.

Is there another potentially compelling reliability challenge that covers the moral case as well as the mathematical? I cannot think of one. Since all of those that seem to so apply turn out
not to be compelling, I conclude, tentatively, that there may simply be no challenge for moral realism that is on a par with Benacerraf’s and Field’s – contrary to what is widely assumed.

4. Conclusions

I have discussed a number of apparent similarities and differences between epistemological problems for moral realism and epistemological problems for mathematical realism. I have argued that some of these are merely apparent, while others are of little philosophical consequence. There are definitely differences between epistemological arguments in the two areas, but these differences, if anything, seem to increase the plausibility of moral realism as compared to mathematical realism. Contrary to what is commonly assumed, it may not be possible to reject moral realism on epistemological grounds while failing to reject mathematical realism. But it may be possible to do the opposite.

Of course, it does not follow that it is not possible to reject moral realism on any grounds while failing to reject mathematical realism. There are obviously non-epistemological problems for moral realism that at least seem not to be problems for mathematical realism – contrary to what Putnam suggests in the quote that began this paper. In particular, moral judgment seems to be tied to motivation in a way that mathematical judgment is not. Given a Humean conception of belief, this observation leaves moral realism vulnerable to the objection that moral judgment is not belief at all. My own view is that this objection is problematic. But I cannot defend this assessment here.
BIBLIOGRAPHY


----- [1985] “Can We Dispense with Spacetime?” Reprinted, with postscript, in [1989].


Forster, Thomas. [Forthcoming] *The Axioms of Set Theory*. Cambridge: Cambridge University Press. Draft available online at:

<http://www.dpmms.cam.ac.uk/~tf/axiomsofsettheory.pdf>


