Chapter 2
What Is the Benacerraf Problem?

Justin Clarke-Doane

In "Mathematical Truth," Paul Benacerraf presented an epistemological problem for mathematical realism. "[S]omething must be said to bridge the chasm, created by […] [a] realistic […] interpretation of mathematical propositions… and the human knower," he writes.[1] For *prima facie* "the connection between the truth conditions for the statements of [our mathematical theories] and […] the people who are supposed to have mathematical knowledge cannot be made out."[2]

The problem presented by Benacerraf—variously called "the Benacerraf Problem" the "Access Problem," the "Reliability Challenge," and the "Benacerraf-Field Challenge"—has largely shaped the philosophy of mathematics. Realist and antirealist views have been defined in reaction to it. But the influence of the Benacerraf Problem is not remotely limited to the philosophy of mathematics. The problem is now thought to arise in a host of other areas, including meta-philosophy. The following quotations are representative.

> The challenge for the moral realist […] is to explain how it would be anything more than chance if my moral beliefs were true, given that I do not interact with moral properties. […] [T]his problem is not specific to moral knowledge. […] Paul Benacerraf originally raised it as a problem about mathematics.
>
> Huemer (2005: 99)[3]

> It is a familiar objection to […] modal realism that if it were true, then it would not be possible to know any of the facts about what is […] possible […]. This epistemological

Thanks to Dan Baras, David James Barnett, Michael Bergmann, John Bigelow, John Bengson, Sinan Dogramaci, Hartry Field, Toby Handfield, Lloyd Humberstone, Colin Marshall, Josh May, Jennifer McDonald, Jim Pryor, Juha Saatsi, Josh Schechter, and to audiences at Australian National University, Monash University, La Trobe University, UCLA, the University of Melbourne, the University of Nottingham, the Institut d'Histoire et de Philosophie des Sciences et des Techniques and the University of Sydney for helpful comments.

J. Clarke-Doane (✉)
Columbia University, New York, USA
e-mail: justin.clarkedoane@gmail.com

F. Pataut (ed.), *Truth, Objects, Infinity*, Logic, Epistemology,
and the Unity of Science 28, DOI 10.1007/978-3-319-45980-6_2

> objection […] may […] parallel […] Benacerraf's dilemma about mathematical […] knowledge.
>
> Stalnaker (1996: 39–40)[4]

> We are reliable about logic. […] This is a striking fact about us, one that stands in need of explanation. But it is not at all clear how to explain it. […] This puzzle is akin to the well-known Benacerraf-Field problem […].
>
> Schechter: (2013: 1)[5]

> Benacerraf's argument, if cogent, establishes that knowledge of necessary truths is not possible.
>
> Casullo (2002: 97)

> The lack of […] an explanation [of our reliability] in the case of intuitions makes a number of people worry about relying on [philosophical] intuitions. (This really is just Benacerraf's worry about mathematical knowledge.)
>
> Bealer (1999: 52n22)

> [W]hat Benacerraf […] asserts about mathematical truth applies to any subject matter. The concept of truth, as it is explicated for any given subject matter, must fit into an overall account of knowledge in a way that makes it intelligible how we have the knowledge in that domain that we do have.
>
> Peacocke (1999: 1–2)[6]

One upshot of the discussion below is that even the above understates the case. An important class of influential but *prima facie* independent epistemological problems are, in relevant respects, restatements of the Benacerraf Problem. These include so-called "Evolutionary Debunking Arguments," associated with such authors as Richard Joyce and Sharon Street.

The Benacerraf Problem is, thus, of central importance. It threatens our knowledge across philosophically significant domains. But what exactly is the problem? In this paper, I argue that there is not a satisfying answer to this question. It is hard to see how there could be a problem that satisfies all of the constraints that have been placed on the Benacerraf Problem. If a condition on undermining, which I will call "Modal Security," is true, then there could not be such a problem. The obscurities surrounding the Benacerraf Problem infect all arguments with the structure of that problem aimed at realism about a domain meeting two conditions. Such arguments include Evolutionary Debunking Arguments. I conclude with some remarks on the relevance of the Benacerraf Problem to the Gettier Problem.

2.1 Benacerraf's Formulation

Benacerraf's "Mathematical Truth" has been deeply influential—but more for its theme than for its detail. The theme of the article is that there is a tension between the "standard" realist interpretation of mathematics and our claim to mathematical knowledge. Benacerraf writes,

> [O]n a realist (i.e., standard) account of mathematical truth our explanation of how we know the basic postulates must be suitably connected with how we interpret the referential apparatus of the theory. […] [But] what is missing is *precisely* […] an account of the link between our cognitive faculties and the objects known. […] We accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties.
>
> Benacerraf (1973: 674)

Benacerraf is skeptical that such an account exists. Thus, he thinks, we must either endorse a "non-standard" antirealist interpretation of mathematics or settle for an epistemic mystery.

What is Benacerraf's reason for being skeptical that our mathematical beliefs can be "appropriately related" to our cognitive faculties, if those beliefs are construed realistically? His reason is the causal theory of knowledge. He writes,

> I favor a causal account of knowledge on which for *X* to know that *S* is true requires some causal relation to obtain between *X* and the referents of the names, predicates, and quantifiers of *S*. [….] [But] […] combining *this* view of knowledge with the "standard" view of mathematical truth makes it difficult to see how mathematical knowledge is possible. […] [T]he connection between the truth conditions for the statements of number theory and any relevant events connected with the people who are supposed to have mathematical knowledge cannot be made out.
>
> Benacerraf (1973: 671–673)

There is a natural response to this argument. Even if the causal theory of knowledge were plausible in other cases, it seems inappropriate in the case of mathematics. As Øystein Linnebo remarks,

> By asking for a causal connection between the epistemic agent and the object of knowledge, Benacerraf treats […] mathematics […] like physics and the other […] empirical sciences. But […] [s]ince mathematics does not purport to discover contingent empirical truths, it deserves to be treated differently.
>
> Linnebo (2006: 546)

Indeed, not even the originator of the causal theory of knowledge, Alvin Goldman, intended that theory to apply to mathematics. In the article to which Benacerraf refers, Goldman begins,

> My concern will be with knowledge of empirical propositions only, since I think that the traditional [justified true belief] analysis is adequate for knowledge of non empirical truths.
>
> Goldman (1967: 357)

But the causal theory of knowledge is not even plausible in other cases. It does not seem to work with respect to knowledge of general truths or with respect to knowledge of truths about spatio-temporally distant events. Indeed, it has been almost universally rejected for reasons that are independent of the Benacerraf Problem (Goldman rejected the theory long ago).[7]

For these reasons, Benacerraf's own formulation of the problem no longer carries much weight. But it is widely agreed that Benacerraf was onto a genuine

epistemological problem for mathematical realism nevertheless. W. D. Hart summarizes the prevailing opinion nicely.

> [I]t is a crime against the intellect to try to mask the [Benacerraf] problem […] with philosophical razzle-dazzle. Superficial worries about […] causal theories of knowledge are irrelevant to and misleading from this problem, for the problem is not so much about causality as about the very possibility of natural knowledge of abstract objects.
>
> Hart (1977: 125–126)

The genuine problem to which Benacerraf was pointing is commonly thought to have been identified by Hartry Field. I turn to his formulation of the problem now.

2.2 Field's Improvement

Field's presentation of the Benacerraf Problem is the starting point for nearly all contemporary discussion of the issue[8]. It has a number of virtues which will occupy me below. Field writes,

> Benacerraf formulated the problem in such a way that it depended on a causal theory of knowledge. The [following] formulation does not depend on *any* theory of knowledge in the sense in which the causal theory is a theory of knowledge: that is, it does not depend on any assumption about necessary and sufficient conditions for knowledge.
>
> Field (1989: 232–233)

In particular,

> 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 those 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.
>
> Field op. cit.: 26

Three observations about Field's formulation of the Benacerraf Problem are in order. First, as Field emphasizes, his formulation of the problem does not assume a view as to the necessary and sufficient conditions for knowledge. Field's formulation *does* assume a view as to the conditions that are (merely) necessary for *justification* (if justification is taken to be necessary for knowledge, then of course Field's formulation implies a necessary condition for knowledge too). According to Field, if one's beliefs from a domain F are justified then it does not appear to her in principle impossible to explain the reliability of her F-beliefs.

Note that the claim is not—or, anyway, should not be—that if one's F-beliefs are justified, then one *can now explain* the reliability of one's F-beliefs. That would clearly be too stringent. Consider our perceptual beliefs. People's perceptual beliefs

were presumably justified 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's perceptual beliefs would have been justified if it appeared to them in principle impossible to explain the reliability of those beliefs.

The second observation is that Field's formulation of the Benacerraf Problem is *non-skeptical*. Field does not merely claim that it appears (to us realists)[9] in principle impossible to offer an explanation of the reliability of our mathematical beliefs *that would convince a mathematical skeptic*—one who doubts that there are any (non-vacuous) mathematical truths at all. Field *grants* for the sake of argument that our mathematical beliefs are both (actually) true and (defeasibly) justified, realistically conceived.[10] Field claims that, *even granted these assumptions*, it appears in principle impossible for us to explain the reliability of our mathematical beliefs.

This is important. In granting these assumptions, Field can draw a contrast between the likes of perceptual realism—realism about the objects of ordinary perception—and mathematical realism. Notoriously, it appears in principle impossible (for us realists) to offer an explanation of the reliability of our perceptual beliefs that would convince a perceptual skeptic. What we can arguably offer is an explanation of the reliability of our perceptual beliefs that assumes the reliability of our perceptual beliefs. We can arguably offer an evolutionary explanation of how we came to have reliable cognitive mechanisms for perceptual belief, and a neurophysical explanation of how those mechanisms work such that they are reliable.[11] Clearly, neither of these explanations would convince someone who was worried that we were brains in vats. The arguments for evolutionary theory and neurophysics blatantly presuppose the reliability of our perceptual beliefs. Still, these arguments do seem to afford our perceptual beliefs a kind of intellectual security. The question is whether analogous arguments are available in the mathematical case.[12]

The final observation is that, while Field does not seem to recognize it, his formulation of the Benacerraf Problem does not obviously depend on the view *that mathematics has a peculiar ontology*. *Prima facie*, his challenge merely depends on the view that mathematical truths are causally, counterfactually, and constitutively independent of human minds and languages.[13] The converging opinion that there is no epistemological gain to "trading" ontology for ideology in the philosophy of mathematics reflects this point. But the point is often misconstrued.[14] The point is not that the explication of the ideological "primitives" will still somehow make reference to abstract objects, so the apparent loss of ontology is illusory. The point is that abstract objects are not what give rise to the Benacerraf Problem in the first place.

This point should not be surprising. As was mentioned in the introduction, something similar to the Benacerraf Problem is commonly thought to arise for realism about domains like morality, modality, and logic. But none of these domains, at least obviously, has a peculiar ontology. In the context of nominalism about universals, morality merely carries with it additional ideology ("is good," "is bad," "is obligatory," and so on). Similarly, for one who takes modal operators as

primitive, the same is true of modality. Finally, it is certainly not mandatory to think that there are peculiar objects corresponding to (first order) logical (as opposed to metalogical) truths. However, in none of these cases does the existence of something like a Benacerraf Problem seem to depend on the plausibility of an ontologically innocent interpretation of the corresponding domain.[15]

The canonical formulation of the Benacerraf Problem, due to Field, is, thus, appealing. It does not rely on a theory of knowledge, much less a causal one. It does not simply raise a general skeptical problem for mathematical realism that has an analog in the perceptual case. Finally, it does not, in any obvious way, rely on an ontologically committal interpretation of mathematics. Nevertheless, it is unclear at a crucial juncture. It is unclear what it would take to *explain the reliability* of our mathematical beliefs in the relevant sense. In what sense of "explain the reliability" is it plausible *both* that it appears in principle impossible (for us realists) to explain the reliability of our mathematical beliefs, *and* that the apparent in principle impossibility of explaining the reliability of those beliefs undermines them?

2.3 Safety

In addressing this question, it will be helpful to begin by considering an account of mathematical truth that even Field believes meets his challenge. We can then look for the sense in which this view can "explain the reliability" of our mathematical beliefs and in which the apparent in principle impossibility of explaining their reliability would undermine them.

The view in question is a version of mathematical pluralism. The key idea to this view is that consistency suffices for truth in mathematics. This contrasts with "standard" mathematical realism, according to which the overwhelming majority of consistent (foundational) mathematical theories are false (just as the overwhelming majority of consistent physical theories are false). Although this view takes many forms [see Carnap [1950] (1983); Putnam [1980] (1983); Linsky and Zalta (1995); and Hamkins (2012)], Field has been clearest about the merits of Mark Balaguer's "Full Blooded Platonism" (FBP). According to FBP, consistent mathematical theories are automatically about the class of objects of which they are true, and there is always such a class (where consistency is a primitive notion, and the notion of truth is a standard Tarskian one).[16]

FBP was specifically advanced as a solution to Field's formulation of the Benacerraf Problem. Mark Balaguer writes,

> The most important advantage that FBP has over non-full-blooded versions of platonism […] is that all of the latter fall prey to [Field's formulation of] Benacerraf's epistemological argument.
>
> Balaguer (1995: 317)

Balaguer explains,

> [FBP] eliminates the mystery of how human beings could attain knowledge of mathematical objects. For if FBP is correct, then all we have to do in order to attain such knowledge is conceptualize, or think about, or even "dream up," a mathematical object. Whatever we come up with, so long as it is consistent, we will have formed an accurate representation of some mathematical object, because, according to FBP, all [logically] possible mathematical objects exist.
>
> Balaguer loc. cit.

Field agrees. He writes,

> [Some philosophers] (Balaguer (1995); Putnam [1980] (1983); perhaps Carnap [1950] (1983) solve the [Benacerraf] problem by articulating views on which though mathematical objects are mind-independent, any view we had had of them would have been correct […]. [T]hese views allow for […] knowledge in mathematics, and unlike more standard Platonist views, they seem to give an intelligible explanation of it.
>
> Field (2005: 78)

In what relevant sense of "explain the reliability" does FBP explain the reliability of our mathematical beliefs? The quotations above suggest that FBP explains the reliability of our mathematical beliefs in the sense that it shows *that had our mathematical beliefs been different (but still consistent), they still would have been true*. However, FBP only shows this assuming that the mathematical truths are the same in the nearest worlds in which our mathematical beliefs are different.[17] If there are no (existentially quantified) mathematical truths in the nearest worlds in which we have different mathematical beliefs, for instance, then had our mathematical beliefs been (appropriately) different, they would have been false. I shall assume, then, that Field intends to grant not just the (actual) truth of our mathematical beliefs, but also that the mathematical truths are the same in all nearby worlds.

Would an explanation in the relevant sense need to show that had our mathematical beliefs been different, they still would have been true? Surely not. Consider the perceptual case. Had our perceptual beliefs been (sufficiently) different (but still consistent), they would have been false. The closest worlds in which we have (consistent) perceptual beliefs as of goblins, say, is a world in which we are deluded somehow. Perhaps it is true that had our perceptual beliefs been different, but "similar," they still would have been true (assuming that there is some independent way to explicate the notion of similarity). But it is still doubtful that an explanation in the relevant sense would need to show this. The observation that had our beliefs of a kind F been different, they would have been false only seems undermining *to the extent that they could have easily been different*. If the closest worlds in which our F-beliefs are different but "similar" are remote, then it is hard to see how the observation that had we been in those worlds, our F-beliefs would have been false, could undermine them.

The reasonable explanatory demand in the neighborhood, which FBP does seem to address (if the mathematical truths are the same in all nearby worlds), is to show that our mathematical beliefs are *safe*—i.e. to show that we could not have *easily*

had false mathematical beliefs (using the method we used to form ours).[18] For typical contingent truths F, our F-beliefs can fail to be safe in two ways (assuming their actual truth). They can fail to be safe, first, if it could have easily happened that the *F-truths* were different while our F-beliefs failed to be correspondingly different. They can fail to be safe, second, if it could have easily happened that our *F-beliefs* were different while the F-truths failed to be correspondingly different. If, however, the F-truths could not have easily been different, then our F-beliefs cannot fail to be safe in the first way. If, moreover, the F-truths are "full-blooded," in the sense that every (consistent) F-theory is equally true, then our F-beliefs cannot fail to be safe in the second way (assuming the "safety" of our inferential practices). Thus, if the mathematical truths are the same in all nearby worlds, and we are granted the (actual) truth of our mathematical beliefs, then FBP shows that our mathematical beliefs are safe.

But if the mathematical truths are the same in all nearby worlds, and we are granted the (actual) truth of our mathematical beliefs, then our mathematical beliefs may well be safe *even if standard mathematical realism is true*. Again, if the mathematical truths could not have easily been different (whether or not they are full-blooded), then our mathematical beliefs cannot fail to be safe in the first way (assuming their actual truth). Moreover, *if we could not have easily had different mathematical beliefs* (even if, had we, they would not still have been true), then they cannot fail to be safe in the second way (again, assuming their actual truth). But there are reasons to think that we could not have easily had different mathematical beliefs. Our "core" mathematical beliefs might be thought to be evolutionarily inevitable.[19] Given that our mathematical theories best systematize those beliefs, there is a "bootstrapping" argument for the safety of our belief in those theories. Our "core" mathematical beliefs are safe; our mathematical theories "abductively follow" from those; our abductive practices are "safe" (something Field, as a scientific realist, would presumably concede); so, our belief in our mathematical theories is safe.

Of course, to the extent that our mathematical theories do not best systematize our "core" mathematical beliefs, this argument is not compelling. But, then, to that extent, the standard realist should not believe in those theories anyway. The reason that standard realists typically refuse to endorse either the Continuum Hypothesis (CH) or its negation is precisely that neither CH nor not-CH seems to figure into the (uniquely) best systematization of our mathematical beliefs.

This argument for the safety of our mathematical beliefs obviously turns on speculative empirical hypotheses. In particular, what evolutionary considerations most clearly suggest is that we would have certain quantificational and geometrical beliefs about things in our environments in certain situations. It does not seem that it was evolutionarily inevitable for us to have the "core" pure mathematical beliefs that we have. One way to address this problem would be to argue that our pure mathematical theories plus bridge laws linking pure mathematical truths to quantificational and geometric truths about things in our environments "abductively follow" from the latter. But the point is that there is a promising argument here—one that it does not appear "in principle impossible" to make.

It appears, then, that the standard mathematical realist may also be able to show that our mathematical beliefs are safe. Is there some *other* relevant sense of "explain the reliability" in which FBP can explain the reliability of our mathematical beliefs? I cannot think of one. (If there is not, then there may be little to recommend FBP.) But there is another epistemological challenge which is closely related to the challenge to show that our mathematical beliefs are safe. Let me turn to that now.

2.4 Sensitivity

Despite his remarks on mathematical pluralism, Field typically suggests that his challenge is to show that our mathematical beliefs are *counterfactually persistent*—i.e., that *had the mathematical truths been (arbitrarily) different (or had there been no existentially quantified such truths at all), our mathematical beliefs would have been correspondingly different* (if we still formed our mathematical beliefs using the method we actually used to form them). For example, Field writes,

> The Benacerraf problem […] seems to arise from the thought that we would have had exactly the same mathematical […] beliefs even if the mathematical […] truths were different […] and this undermines those beliefs.
>
> Field (2005: 81)[20]

Notice that FBP does nothing to answer this challenge. Balaguer is "doubtful that mathematical theories are necessary in any interesting sense" (Balaguer 1998: 317), and he concedes that "[i]f there were never any such things as [mathematical] objects, the physical world would be exactly as it is right now" (Balaguer 1999: 113). Given the supervenience of the intentional on the physical, it follows that had there been no existentially quantified mathematical truths, our mathematical beliefs would have failed to be correspondingly different.

Would the apparent in principle impossibility of showing that our mathematical beliefs are counterfactually persistent in this sense undermine them? Surely not. We cannot even show that our perceptual beliefs are counterfactually persistent in *this* sense. As skeptics argue, the closest worlds in which the perceptual truths are sufficiently different is a world in which we are deluded.[21]

The reasonable challenge in the neighborhood is to show that our mathematical beliefs are *sensitive*—i.e., that *had the contents of our mathematical beliefs been false, we would not still have believed them* (using the method we actually used to form our mathematical beliefs). Although we cannot show that had the perceptual truths been (arbitrarily) different, our perceptual beliefs would have been correspondingly different, Nozick (1981) pointed out that we can, it seems, show that had I, e.g., not been writing this paper, I would not have believed that I was. The closest world in which I am not writing this paper is still a world in which my perceptual faculties deliver true beliefs.

But there is a problem with Field's challenge under this construal. Contra Balaguer, mathematical truths seem to be metaphysically necessary. If they are,

then our mathematical beliefs are vacuously sensitive on a standard semantics. As David Lewis writes,

> [I]f it is a necessary truth that so-and-so, then believing that so-and-so is an infallible method of being right. If what I believe is a necessary truth, then there is no possibility of being wrong. That is so whatever the subject matter […] and no matter how it came to be believed.
>
> Lewis (1986: 114–115)[22]

Field might respond that, though we are granted the *actual* truth (and defeasible justification) of our mathematical beliefs (and presumably also that the mathematical truths could not have easily been different), we are not granted the necessity of their contents. Unlike the claim that the contents of our mathematical beliefs are true, the claim that their contents are necessary requires argument. But our belief that the mathematical truths are necessary is commonly thought to enjoy a similar status to our belief that they are true. Both beliefs are commonly regarded as default-justified positions. If Field were merely trying to undermine our mathematical beliefs *under the assumption that our belief that they are necessary is not itself (defeasibly) justified*, then the interest of his challenge would be greatly diminished.[23]

Of course, it is arguable, contra Lewis, that some conditionals with metaphysically impossible antecedents are not vacuously true. For a variety of purportedly metaphysically necessary truths, we seem to be able to intelligibly ask what would have been the case had they been false. As Field writes,

> If one [says] "nothing sensible can be said about how things would be different if the axiom of choice were false," it seems wrong: if the axiom of choice were false, the cardinals wouldn't be linearly ordered, the Banach-Tarski theorem would fail and so forth.
>
> Field (1989: 237–238)

However, we seem to be equally unable to show that our relevantly uncontroversial beliefs are non-vacuously sensitive. For example, we seem to be unable to show that our belief in "bridge laws" that link supervenient properties to subvenient ones are so sensitive. Had—as a matter of metaphysical impossibility—atoms arranged car-wise failed to compose cars (as "ontological nihilists" allege), we still would have believed that they did.[24]

It might be thought that Field could simply accept that his challenge is at least as serious for realism about truths that link supervenient properties to subvenient ones. After all, such truths are metaphysically necessary, and, again, the Benacerraf Problem is often thought to arise for realism about such truths. The problem with this response is that it would not just require rejecting realism about necessary truths. It would at least *prima facie* require rejecting realism about the truths of ordinary perception. If our belief about the composition conditions of cars is undermined, then so too, presumably, is our belief that we are not sitting in one.[25] If Field's challenge generalized this wildly, then it would no longer point to an epistemic difference between our mathematical beliefs and our beliefs about the objects of ordinary perception.

I will shortly mention a way to argue that our mathematical beliefs are even *better* off than our beliefs about the composition conditions of cars with respect to sensitivity. But the above demonstrates that if Field's challenge is to show that our mathematical beliefs are sensitive, then, again, it either does not appear in principle impossible to answer, or this appearance does not plausibly undermine those beliefs.

2.5 Indispensability

Those familiar with "Evolutionary Debunking Arguments" (to be discussed) are likely to think that I have failed to consider the most obvious analysis of Field's challenge. The challenge, it might be thought, has nothing immediately to do with the safety and sensitivity of our mathematical beliefs. It has to do with the explanation of our having those beliefs. The challenge is to show *that the contents (or truth) of our mathematical beliefs figure into the best explanation of our having them*.

This proposal is related to a causal constraint on knowledge. But it is intuitively weaker. Rather than requiring that the subject matter of our beliefs from an area *F* helps to *cause* our having the *F*-beliefs that we have, it is required that the contents of our *F*-beliefs help to *explain* our having those beliefs. *Prima facie*, our beliefs may have a causally inert subject matter, though their contents figure into the best explanation of contingent states like our having those beliefs.

But there is an obvious problem with Field's challenge under this analysis. It simply does not appear in principle impossible to answer. Insofar as mathematics appears to be indispensable to empirical science, it appears impossible to show that the contents of our mathematical beliefs do *not* figure into the best explanation of our having them. As Mark Steiner writes,

> [S]uppose that we believe […] the axioms of analysis or of number theory. […] [S]omething is causally responsible for our belief, and there exists a theory — actual or possible, known or unknown — which can satisfactorily explain our belief in causal style. This theory, like all others, *will contain the axioms of number theory and analysis*.
>
> Steiner (1973: 61)

The point is that mathematics, like logic, seems to be assumed by all of our empirical theories.[26] If it is, however, then mathematics is a background assumption of the theory that best explains our having the mathematical beliefs that we have. In particular, for any typical (e.g. not higher-set-theoretic) mathematical proposition *p*, *p* is a background assumption of the best explanation of our having the belief that *p*.

One might respond to this problem by arguing that an explanation in the pertinent sense would show, not just that the contents of our mathematical beliefs figure into the best explanation of our having those beliefs, but that they do so in an "explanatory way." But, setting aside the obscurity of the quoted locution, it would still not appear in principle impossible to explain the reliability of our mathematical

beliefs. The key point of the "improved" indispensability argument pressed lately by Alan Baker and others is that mathematical hypotheses seem to figure into the best explanation of empirical phenomena in just such a way.[27] This argument does not show that the contents of our mathematical beliefs figure into the best explanation *of our having them* in an explanatory way. But the latter claim is *prima facie* plausible. Consider, for instance, Sinnott-Armstrong's suggestion that "[p]eople evolved to believe that 2 + 3 = 5, because they would not have survived if they had believed that 2 + 3 = 4" (Sinnot-Armstrong 2006: 46). The question arises: why would people not have survived if they had believed that 2 + 3 = 4? According to Sinnott-Armstrong "the reason why they would not have survived then is that it is true that 2 + 3 = 5" (Sinnot-Armstrong loc. cit.). Sinnott-Armstrong appears to be suggesting that 2 + 3 = 5 explains our ancestors' coming to believe that 2 + 3 = 5. It does not merely "figure into" the explanation of their doing this. Whether this is true is certainly debatable. The point, however, is that it does not appear "in principle impossible" to show that the contents of our mathematical beliefs figure into the best explanation of our having them—even "in an explanatory way."[28]

Field is under no illusions on this point. When discussing his challenge as it applies to logic, he explicitly rejects the above analysis of his challenge (Field 1996: 372n13). Field even notes that the realist can appeal to the explanatory role of mathematics in an effort to bolster the conclusion of the previous section. He notes, in effect, that one can argue from explanatory considerations to the *non-vacuous* sensitivity of our mathematical beliefs. Field writes,

> [O]ne can try to invoke indispensability considerations […] in the context of explaining reliability. One could argue […] that if mathematics is indispensable to the laws of empirical science, then *if the mathematical facts were different, different empirical consequences could be derived from the same laws of (mathematicized) physics*.
>
> Field (1989: 28)

Such an argument for the sensitivity of our beliefs about the composition conditions of ordinary objects would not be plausible. When p is a truth predicating a property—such as the property of being a car—which is not the "postulate" of any special science, one cannot argue from the explanatory indispensability of p to the sensitivity of our belief that p, as above. This is true even if the property is supervenient on properties which are the "postulates" of special sciences.[29]

What should we think of the above argument? I need offer no final assessment here. But Field's reason for doubt is not compelling. He objects that "the amount of mathematics that gets applied in empirical science… is relatively small" (Field 1989: 29). But, as with safety, there is a "bootstrapping" argument from the sensitivity of our belief in applicable mathematics to the sensitivity of our belief in the rest of it. Again, such an argument may fail to "decide" certain abstract hypotheses. But this would merely seem to confirm the standard view that we ought to remain agnostic about their truth-values.

Suppose, however, that it *did* appear in principle impossible to show that the contents of our mathematical beliefs figure into the best explanation of our having

them. Would this appearance undermine those beliefs? It is hard to see how it could. I have argued that we may be able to show that our mathematical beliefs are both safe and sensitive, given their (actual) truth (and defeasible justification), in the sense in which we can show that our uncontroversial beliefs are. We may be able to argue that we could not have easily had false mathematical beliefs, given their truth (and that the mathematical truths could not have easily been different), because we could not have easily had different ones. Thus, our mathematical beliefs are safe. And we may be able to argue that, since mathematical truths would be metaphysically necessary, had the contents of our mathematical beliefs been false, we would not have believed those contents (in the sense that we can argue that, had the contents of our explanatorily basic ordinary object beliefs been false, we would not have believed those contents). Thus, our mathematical beliefs are (vacuously) sensitive as well. Notice that this argument does not assume, even implicitly, that the contents of our mathematical beliefs figure into the best explanation of our having them. The problem, then, is this. How could the observation that the contents of our mathematical beliefs fail to figure into the best explanation of our having them undermine them *without giving us some reason to doubt that they are both safe and sensitive*? This obscurity points to a basic problem with the Benacerraf Problem.

2.6 In Search of a Problem

What is the Benacerraf Problem? It is not to show that our mathematical beliefs are safe or sensitive, since we may be able to show that they are safe and sensitive in the sense that we can show that our uncontroversial beliefs are. Nor is it to show that the contents of our mathematical beliefs figure into the best explanation of our having them, since we may be able to show that too—and, anyway, it is unclear how the apparent in principle impossibility of showing this could undermine our mathematical beliefs. Is there some *other* sense of "explain the reliability" in which it appears in principle impossible to explain the reliability of our mathematical beliefs and which is such that this appearance plausibly undermines those beliefs?

It is sometimes said that what is wanted is a *unified* explanation of the correlation between our mathematical beliefs and the truths. Without this, that correlation seems like a "massive coincidence."[30] But what does this worry amount to?

It cannot be that the reliability of our mathematical beliefs would be *improbable*. Either the probability at issue is epistemic or it is "objective." If it is epistemic, then the suggestion is question-begging. It effectively amounts to the conclusion that Field's argument is supposed to establish—that our mathematical beliefs are not justified. Suppose, then, that the probability is objective. Then for any mathematical truth p, presumably $\Pr(p) = 1$, given that such truths would be metaphysically necessary.[31] Moreover, as we have seen, it may be that Pr (we believe that p) ≈ 1, because the probability of our having the mathematical beliefs that we have is high.[32] But then, Pr (p & we believe that p) ≈ 1, by the probability calculus. Since

(p & we believe that p) implies that (our belief that p is true), it may be true that Pr(our belief that p is true) ≈ 1.

Nor can the request for a unified explanation be the request for a *common cause* of the mathematical truths and of our mathematical beliefs. Such a request blatantly assumes a causal constraint on knowledge, which Field's challenge is supposed to avoid. (Nor, again, can the request for a unified explanation be a request to show that the contents of our mathematical beliefs help to explain—even if not to cause—our having them. Again, the contents of our mathematical beliefs may well do that.)[33]

The problem is not just that it is hard to state an analysis of Field's challenge that satisfies the constraints that he places on it. It is not clear that there *could be* a problem that satisfies those constraints, given what I argued in Sects. 2.3 and 2.4.

To see this, let me introduce a condition on information E if it is to undermine our beliefs of a kind F. The condition is:

> *Modal Security*: If information E undermines all of our beliefs of a kind F, then it does so by giving us reason to doubt that our F-beliefs are both sensitive and safe.[34]

Modal Security states a *necessary* condition on underminers. It does not say that if information E gives us reason to doubt that our F-beliefs are both safe and sensitive, then E obligates us to give up all of those beliefs. It says that if E does not even do this, then E cannot be thought to so obligate us.[35]

The key idea behind Modal Security is that there is no such thing as a "non-modal underminer." (Of course, there is such a thing as a non-modal defeater—namely, a rebutter, i.e. a "direct" reason to believe the negation of the content of our defeated belief.[36]) If there were such a thing as a non-modal underminer, then information could undermine our beliefs "immediately." It could undermine them, but not by giving us reason to doubt that they are modally secure. The "by" is needed, since everyone should agree that if E undermines our F-beliefs, then E gives us reason to doubt that those beliefs are safe—i.e. that they could not have easily been false. If our F-beliefs are actually false, then they could have easily been.

Paradigmatic underminers seem to conform to Modal Security. If E is that we took a pill that gives rise to random F-beliefs, for instance, then E seems to undermine them by giving us reason to believe that we could have easily had different F-beliefs, and, hence (given that the F-truths do not counterfactually depend on our F-beliefs), that our F-beliefs are not safe. Something similar could perhaps be said of evidence E that there is widespread disagreement on F-matters (though whether such evidence is undermining is of course debatable). On the other hand, when the F-truths are contingent, and E is that we were "bound" to have the F-beliefs that we do have because of some constraining influence, such as a tendency to overrate one's self, then E seems to undermine our F-beliefs by giving us reason to doubt that had the contents of our F-beliefs been false, we would not still have believed those contents—i.e., by giving us reason to doubt that our F-beliefs are sensitive.

Nevertheless, it might be thought that Modal Security cannot handle necessary truths that we were "bound" to believe. Suppose that a machine enumerates sentences, deeming them validities or invalidities. Independent investigation has confirmed its outputs prior to the last five. The last five outputs are "validity." We defer to the machine's last five outputs only, and have no prior metalogical beliefs. Today a trusted source tell us that the machine was "stuck" in the last five instances. Call this evidence *E*. Then *E* does not seem to give us "rebutting" reason to believe that the last five outputs are invalidities. Nor does it seem to give us undermining reason to doubt that had, as a matter of metaphysical possibility, the last five outputs—those sentences—been invalidities, the machine would not have called them validities. (*E* does not seem to give us reason to believe that there is a metaphysically possible world in which those sentences are invalidities, and the counterfactual in question can only be false with respect to such worlds if there is.) Nor does *E* seem to give us undermining reason to believe that the machine could have easily called the last five sentences invalidities. That was the point of calling it "stuck." But *E* seems to undermine all of our metalogical beliefs. Does not this show that *E* may undermine all of our beliefs of a kind but not by giving us reason to doubt that they are both sensitive and safe?[37]

It does not. *E* may not be evidence that, for any metalogical proposition *p that we believe*, we could have easily had a false belief as to whether *p*. It does not follow that *E* is not evidence that we could have easily had a false metalogical belief. *E* is evidence that, even if the machine had considered an invalidity last, it still would have called the sentence a validity. But it is not just this fact which seems undermining. If we know that the only worlds in which it considers an invalidity last are "distant," then it is hard to see how evidence that, had we been in one, we would have had false metalogical beliefs, could undermine all of our metalogical beliefs. It must apparently be added that we know that such worlds are "similar" to the actual one.[38] But if this is added, then *E* is evidence that we could have easily had different metalogical beliefs, had the machine considered different sentences last, and hence, given the necessity of the metalogical truths, that we could have easily had false metalogical beliefs—i.e., that those beliefs are not safe.[39]

This response depends on the assumption that not just any grouping of beliefs counts as a "kind." If we could let *F* be $\{[b] = \{x: x = b\}\}$, for some belief formed by the machine *b*, then *E* could undermine all of our *F*-beliefs despite giving us no reason to believe that we might have easily had false *F*-beliefs. Intuitively, however, metalogical beliefs, like moral beliefs, are kinds, while "the last metalogical belief formed by the machine" is not. If there is no principled argument for this, however, then Modal Security may not get off the ground. The problem is similar to the "generality problem" for process reliabilism.[40]

Of course, if we could explicate "stuck" in such a way that learning that the machine was stuck in the last five instances undermines all of our metalogical beliefs, but has no modal implications at all, then the example above would be a

counterexample to Modal Security. But it is hard to see how we could do this. We might say that the machine is "stuck" in that it is not "detecting," "tracking," or "sensitive to" the metalogical truths—it is not generating its outputs "because" they are true. But what do these locutions mean? They do not mean that the truth of the machine's outputs is not implied by their best explanation. Had we imagined instead a machine that outputs only logical truths themselves, then, trivially, the machine would output such truths "because" they were true, since every logical truth is a consequence of every explanation at all. We might explicate "stuck" in terms of hyperintensional ideology like constitution or ground. But why exactly should give up beliefs which we learn are not "constituted by" or "grounded in" their truth?[41]

The relevance of Modal Security to the Benacerraf Problem is as follows. If it is true, and what was said in Sects. 2.3 and 2.4 is correct, then there could not be a problem that plays the epistemological role that the Benacerraf problem is supposed by Field to play. Even if there is a sense in which it appears in principle impossible to "explain the reliability" of our mathematical beliefs, this is not a sense which gives us reason to doubt that they are both safe and sensitive (with respect to metaphysically possible worlds). Field does not even pretend to give us ("rebutting") reason to think that they are actually false, and he does not seem to give us any ("undermining") reason to doubt that if they are true, then they are both safe and sensitive. Hence, even if there is a sense in which it appears in principle impossible to "explain the reliability" of our mathematical beliefs, this is not a sense which undermines them, if Modal Security is true.

To be sure, the conclusions of Sects. 2.3 and 2.4 are subject to amendment. As more is learned about the genealogy of our mathematical capacities, it may come to appear impossible to show that our "core" mathematical beliefs are inevitable. But if we can show that they are, and if Modal Security is true, then we can "explain their reliability" in every sense which is such that the apparent in principle impossibility of explaining their reliability undermines them.

Note that this is *not* to say that our mathematical beliefs are in good epistemic standing—anymore than it is to say that the theological beliefs of a theological realist who can argue both that the theological truths would be metaphysically necessary if true, and that she could not have easily had different theological beliefs, are in good epistemic standing. For all that has been said, our mathematical beliefs could be false and unjustified. Field *grants* for the sake of argument that our mathematical beliefs are (actually) true and (defeasibly) justified, in order to generate a *dialectically effective* argument against realists. But if Modal Security is true, then such an argument may not, in general, be possible. Modal Security implies that Field overreaches.

Let me illustrate the above reasoning with reference to two further analyses of Field's challenge. When discussing the Lewisian reply to the challenge to show that our mathematical beliefs are sensitive, Field proposes an alternative. He writes,

> If the intelligibility of talk of "varying the facts" is challenged […] it can easily be dropped without much loss to the problem: there is still the problem of explaining the *actual* correlation between our believing "*p*" and its being the case that *p*.
>
> Field (1989: 238)

The problem is this. Even if there is some hyperintensional sense of "explanation" according to which one can intelligibly request an explanation of the "merely actual correlation" between our mathematical beliefs and the truths, it is unclear how the apparent in principle impossibility of offering that could undermine those beliefs—given that we may still be able to show that they are safe and sensitive. If we can show that our mathematical beliefs are safe and sensitive, given their truth, then we can show that they were (all but) *bound* to be true.

Schechter suggests another response to Lewis on behalf of Field. He writes,

> Lewis is correct […] that the reliability challenge for mathematics […] is subject to a straightforward response, so long as the challenge is construed to be that of answering [the question of how our mechanism for mathematical belief works such that it is reliable] […]. [But] [t]here remains the challenge of answering [the question of how we came to have a reliable mechanism for mathematical belief].
>
> Schechter (2010: 445)

Schechter claims that the question of *how we came to have* a reliable mechanism for mathematical belief may remain open even under the assumption that it is unintelligible to imagine the mathematical truths being different (even if the question of *how that mechanism works* such that it is reliable may not). But, first, this appears incorrect. Schechter is explicit that the question of how we came to have a reliable mechanism for mathematical belief is different from the question of how we came to have the mechanism for mathematical belief *that we actually came to have* (since the latter question is clearly answerable in principle). However, in order to decide whether we were, say, selected to have a reliable mechanism for mathematical belief, as opposed to being selected to have a mechanism for mathematical belief with property *F which is in fact reliable*, we would seem to need to have to decide what mechanism it would have benefited our ancestors to have had *had the mathematical truths been different*.[42] Second, even if Schechter were correct, it is hard to see how the apparent in principle impossibility of explaining the reliability of our mathematical beliefs in *his* sense could undermine them.[43] For all that has been said, we might still be able to show that our mathematical beliefs are safe and sensitive.

Whether Modal Security is true requires in-depth treatment. But while successfully defending it would suffice to deflate the Benacerraf Problem, successfully challenging it would not suffice to reestablish that problem. The question would remain: *in what sense of "explain the reliability" is it plausible both that it appears in principle impossible to explain the reliability of our mathematical beliefs and that the apparent in principle impossibility of explaining their reliability undermines them?* Insofar as there seems to be no satisfactory answer to this question, the force of the Benacerraf Problem seems lost. Of course, it does not follow that there

are no mysteries surrounding mathematical knowledge. The point is that no such mystery can play the role that the Benacerraf problem is supposed to play.

Let me now turn to the broader relevance of this conclusion.

2.7 Broader Relevance

The difficulties surrounding the Benacerraf Problem are actually very general. They infect formulations of it aimed at moral realism, modal realism, logical realism, and philosophical realism. But they infect much more than this. They infect any argument which grants the (actual) truth and (defeasible) justification of our beliefs from an area and seeks to undermine those beliefs, so long as the area *F* meets two conditions. Those conditions are:

1. The *F*-truths would be metaphysically necessary.
2. There is a plausible explanation of our having the *F*-beliefs that we have which shows that we could not have easily had different ones.

Many arguments which are not supposed to be variations on the Benacerraf Problem have these features. Consider Evolutionary Debunking Arguments, which are now influential epistemological challenges themselves. Richard Joyce offers a canonical formulation:

> Nativism [the view that moral concepts are innate] offers us a genealogical explanation of moral judgments that nowhere […] presupposes that these beliefs are true […]. My contention […] is that moral nativism […] might well […] render [moral beliefs] unjustified […]. In particular, any epistemological benefit-of-the-doubt that might have been extended to moral beliefs […] will be neutralized by the availability of an empirically confirmed moral genealogy that nowhere […] presupposes their truth.
>
> Joyce (2008: 216)[44]

What is Joyce's argument? Taken at face value, it is that our moral beliefs (realistically conceived) are undermined on the mere ground that their contents fail to figure into the evolutionary explanation of our having them. But we have seen that such an argument must be too quick. In the first place, it is arguable that the contents of our moral beliefs do figure into the evolutionary explanation of our having them, just as it is arguable that the contents of our mathematical beliefs do.[45] But set this possibility aside. In order for the information to which Joyce alludes to undermine our moral beliefs, it would seem prima facie to have to give us ("direct") reason to doubt that our moral beliefs are both safe and sensitive. But, on its own, the observation that the contents of our moral beliefs fail to figure into the evolutionary explanation of our having them does not do this. If this observation has any epistemological force, it is apparently to help undercut what is arguably the only dialectically effective argument *for* the contents of our moral beliefs (realistically conceived). As Gilbert Harman writes,

> Observation plays a part in science it does not appear to play in ethics, because scientific principles can be justified ultimately by their role in explaining observations […]. [M]oral principles cannot be justified in the same way.
>
> Harman (1977: 10)

If the contents of our moral beliefs did figure into the evolutionary explanation of our having those beliefs, then those contents *could* be justified by their role in explaining observations.[46]

Perhaps, then, Joyce's argument is that the evolutionary considerations to which he alludes give us reason to doubt that our moral beliefs are safe. But, on the contrary, if anything, those considerations seem to give us reason to believe that our moral beliefs *are* safe. The whole point of Evolutionary Debunking Arguments is often taken to be that it would have been advantageous for our ancestors to have the "core" moral beliefs that we have (such as that killing our offspring is bad) "independent of their truth." If this is right, then we could not have easily come to have different such beliefs, in which case they are safe (assuming that the explanatorily basic moral truths are the same in nearby worlds—more on this below). As before, we may be able to "bootstrap up" from the safety of our "core" moral beliefs to the safety of our moral theories. Note the irony. A tentative sign that a realist about an area *F* can establish the safety of her beliefs is *that there is an Evolutionary Debunking Argument aimed at F-realism*.[47]

Perhaps, then, rather than giving us reason to believe that our moral beliefs are not safe, Joyce takes evolutionary considerations to give us reason to doubt that our moral beliefs are sensitive. This interpretation is in harmony with standard formulations of the Evolutionary Debunking Argument, due to Walter Sinnott-Armstrong, Sharon Street, Michael Ruse, and others. It would have benefited our ancestors to believe that killing our offspring is wrong even if killing our offspring were right—or, indeed, even if there were no (atomic or existentially quantified) moral truths at all. Thus, had the contents of our moral beliefs been false, we still would have believed those contents. In an earlier book, Joyce writes,

> Suppose that the actual world contains real categorical requirements — the kind that would be necessary to render moral discourse true. In such a world humans will be disposed to make moral judgments […] for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all — nothing to make moral discourse true. In such a world, humans will *still* be disposed to make these judgments… for natural selection will make it so.
>
> Joyce (2001: 163)

But this argument is also fallacious. Even setting aside the prospect of arguing for sensitivity via explanatory indispensability, the explanatorily basic moral truths —the truths that fix the conditions under which a concrete person, action, or event satisfies a moral predicate—are widely supposed to be metaphysically necessary. But, if they are, then our corresponding beliefs are vacuously sensitive on a standard semantics. Of course, as before, those beliefs may not be non-vacuously sensitive. But, then, neither are our relevantly uncontroversial beliefs, such as the belief that atoms arranged car-wise compose cars.

As in the mathematical case, it follows that we may be able to show that our moral beliefs are both safe and sensitive, given their actual truth. This affords them extraordinary intellectual security. To the extent that Joyce's evolutionary speculations give us no reason to doubt their modal security, it is hard to see how those speculations could undermine our moral beliefs.

2.8 The Benacerraf Problem and the Gettier Problem

What is the Benacerraf Problem? There does not seem to be a satisfying answer. There does not seem to be a sense of "explain the reliability" in which it is plausible *both* that it appears in principle impossible to explain the reliability of our mathematical beliefs and that the apparent in principle impossibility of explaining their reliability undermines them. The problem is quite general, infecting all arguments with the structure of the Benacerraf Problem meeting two conditions.

It remains open how this conclusion bears on our claim to knowledge in mathematics and related areas. Unlike Benacerraf's challenge, both Field's challenge and Evolutionary Debunking Arguments focus on whether our beliefs are justified, not on whether they qualify as knowledge. This is a virtue. The correct analysis of knowledge is notoriously controversial. Moreover, if our beliefs are justified, and we can relevantly explain their reliability, then it is hard to see why we should give them up—*even if* they fail to qualify as knowledge. Perhaps the most interesting feature of Field's formulation of the Benacerraf Problem and of Evolutionary Debunking Arguments is that they purport to give realists reason to *change their views*.

Nevertheless, the argument offered here may suggest that we have knowledge in mathematics and related areas. I have argued that our mathematical and related beliefs may be both safe and sensitive, given their (actual) truth (and defeasible justification). (In the case of mathematics, I also argued that the contents of our beliefs may figure into the best explanation of our having them.) But many philosophers would hold that a justified true belief which is both safe and sensitive qualifies as knowledge (and even more would hold that a justified true belief which is both safe and sensitive and such that its content figures into the best explanation of our having it so qualifies).

Perhaps the present discussion helps to explain why. "Gettiered" beliefs—justified and true beliefs which fail to qualify as knowledge—are plausibly beliefs whose truth is coincidental in a malignant sense. What is that sense? It is arguably precisely the sense in which learning that the truth of one's beliefs is coincidental would undermine them. If this is correct, then there is a "translation scheme" between the claim that it is impossible to relevantly explain the reliability of our F-beliefs, given their truth, and the claim that those beliefs are Gettiered.

Notes

1. See Benacerraf (1973: 675).
2. *Ibid*: 673.
3. See also Mackie's epistemological "argument from queerness" in Chap. 1 of Mackie (1977), as well as Alan Gibbard's discussion of "deep vindication" in Sect. 13 of his Gibbard (2003).
4. Similarly, O'Leary-Hawthorne writes, "The relevant difficulties are not, of course, peculiar to modal metaphysics. Our dilemma re-enacts Paul Benacerraf's famous dilemma for the philosopher of mathematics" (O'Leary-Hawthorne 1996: 183).
5. See also Resnik (2000).
6. Adam Pautz suggests that the Benacerraf Problem arises for realism about phenomenal properties. He writes, "If propositions about resemblances among properties report acausal facts about abstracta, then how can we explain the following regularity: generally, if we believe *p*, and *p* is a such a proposition, then *p* is true? […] [T]his problem arises for all accounts of the Mary case [of the "Mary's Room" argument] […]. [And it] resembles the Benacerraf-Field problem about mathematics" (Pautz 2011: 392–3). (The Mary's Room argument purports to show that physicalism is false on the grounds that someone—Mary—could know all the physical facts about color vision and yet learn something upon seeing colored things for the first time.).
7. The only contemporary formulation of a causal constraint on knowledge of which I am aware seems hopelessly ad hoc. Colin Cheyne contends that "[i]f *F*s are noncomparative objects, then we cannot know that *F*s exist unless our belief in their existence is caused by: (a) an event in which *F*s participate, or (b) events in which each of the robust constituents of *F*s participate, or (c) an event which proximately causes an *F* to exist" (Cheyne 1998: 46).
8. For overviews, see Liggins (2006, 2010) and Linnebo (2006).
9. I am not actually a mathematical realist (in a common sense of that phrase). But since it will be convenient to frame things in terms of what "we" can explain, I will often identify as one.
10. I will generally fail to qualify discussion of mathematical truths or of our mathematical beliefs with "realistically conceived" in what follows. But let me emphasize that this will always be what I intend. The Benacerraf Problem threatens our mathematical beliefs *under a realist construal*. This hardly detracts from its importance, however, for it is notoriously difficult to give a satisfactory non-realist construal of the subject. I clarify the relevant sense of "mathematical realism" at the end of the present section.
11. These explanatory tasks are distinguished in Schechter (2010).
12. For more on the "non-skeptical" character of Field's challenge, see Balaguer (1995).
13. But see Sect. 3 of Clarke-Doane (2014).
14. See, for example, Shapiro (1995) or Leng (2008).

15. Joshua Schechter, following Field (1998), suggests that what matters for Field's formulation of the Benacerraf Problem is not the ontology of mathematics, but its *objectivity*. He has in mind the contrast between, say, standard platonism, according to which there is a "unique" universe of mathematical objects that our (fundamental) mathematical theories aim to describe, and Balaguer's "Full-Blooded Platonism," to be discussed below, according to which any "intuitively consistent" theory that we might have come to accept would have been true. A given domain of truths F is objective in the relevant sense, if and only if "not just any [$F-$] practice counts as correct." Schechter writes, "[t]he root of the trouble is not the ontology but the apparent *objectivity* of mathematics […]. If mathematics […] were to turn out not to have an ontology, but the relevant truths were nevertheless objective, our reliability would remain puzzling" (Schechter 2010: 439). We will see in Sect. 3 that this thought, as promising as it may appear, could not be correct. (For more in-depth treatment of this issue, see Clarke-Doane [manuscript]).
16. The parenthetical qualification is needed in order to distinguish FBP, which is highly controversial, from the Completeness Theorem, which is not. FBP does not just say that every consistent mathematical theory has a model. It says that every such theory has an *intended* model. Sometimes critics of FBP accuse Balaguer of merely advocating the Completeness Theorem. See, for instance, Burgess (2001).
17. I assume the standard view that "had it been the case that p, it would have been the case that q" is true only if q is true at the closest worlds in which p is true.
18. While I am not using "safe" in exactly the sense of Pritchard (2005) or Williamson (2000), the idea is similar. It is unobvious how to spell out the pertinent sense of "easily." See Hawthorne (2004: 56) for a complication. I will mostly ignore methods of belief-formation in what follows.
19. I am not suggesting that our having *true "core" mathematical beliefs* per se might be evolutionary inevitable. I am suggesting that our having the "core" mathematical beliefs that we have, *which are actually true*, might be. These claims are very different. See Field (2005) and Clarke-Doane (2012).
20. Field alludes to the parenthetical antecedent in the following: "[W]e can assume, with at least some degree of clarity, a world without mathematical objects…" (Field 2005: 80–81). Sometimes Field describes the problem of showing that our mathematical beliefs are counterfactually persistent as that of offering a *unified* explanation of their reliability. In a discussion of the Benacerraf Problem for logical realism, Field writes "The idea of an explanation failing to be "unified" is less than crystal clear, but another way to express what is unsatisfactory about [a bad explanation] is that it isn't *counterfactually persistent* […], it gives no sense to the idea that if the logical facts had been different then our logical beliefs would have been different too." (Field 1996: 371). I will discuss another sense of "unified explanation" in Sect. 6.
21. This is not incontrovertible. Given an "external" theory of reference, one can argue that, had we been brains in vats, we would have believed that we were. See Putnam (1981). This argument relies on a causal theory of reference.

22. See also Pust (2004).
23. There *are* arguments that the mathematical truths would be necessary, though I am not sure how compelling they are. One such argument is that mathematical truths concern abstract objects like numbers, sets and tensors. Such objects are neither spatial nor temporal and participate in no physical interactions. "Hence" they do not depend on the way that the contingent world happens to be. So, if mathematical objects satisfy the standard axioms in the actual world, then they do so in all possible worlds. See Shapiro (2000: 21–24) for something like this argument.
24. *See* Korman (2014): Sect. 4.2 for relevant discussion.
25. This assumes a closure principle which could conceivably be questioned. But even if it were, an analogous point would hold. See Clarke-Doane (2016), Sect. 2.2.
26. Field, of course, hopes ultimately to show that mathematics is not assumed by (the best formulations of) our empirical theories. But he does not deny that it *seems* to be, and, anyway, appears to hold that the Benacerraf Problem arises even if it is. See Field (2005): Sect. 2.5, and Field (1989: 262).
27. See, for instance, Baker (2009), and Lyon and Colyvan (2008).
28. For more on evolutionary examples like this, see Braddock et al. (2012). I argue that, despite appearances, the contents of our mathematical beliefs do not explain our having them in Clarke-Doane (2012): Sect. III.
29. For more on this, see Sect. 2.3 of Clarke-Doane (2014).
30. Field makes gestures in the direction of this position in Field (1996), but then explicates "unified explanation" in terms of sensitivity (see *supra* footnote 22). Sharon Street makes claims like this with respect to our moral beliefs in Street (2006, 2008).
31. What if we only assign objective probability 1 to contents which are necessary in an even stronger sense—e.g. "conceptually necessary"? Then, unless one can argue that "ontological nihilism" is not just false but *conceptually impossible*, the contents of our (explanatorily basic) "ordinary object beliefs" would seem to have equal claim to being objectively improbable. See Sect. 2.4.
32. I am not assuming that if our mathematical beliefs are safe, then the objective probability that they are true is high. I am pointing out that, for all that has been said, our mathematical beliefs may be both safe and objectively probable.
33. Perhaps it amounts to the request to show that our mathematical beliefs are "grounded in" or "constituted by" the corresponding state of affairs (see Bengson (2015) for something like this proposal)? Such hyperintensional ideology does not seem to me to be more perspicuous than the quoted phrase itself. But, even if it were, this proposal would not seem to serve Field's purposes, as will become clear shortly.
34. Safety and sensitivity plausibly need to be relativized to methods of belief formation, as indicated towards the end of Sect. 2.3 and at the beginning of Sect. 2.4, respectively and reason is shorthand for "direct" "reason". (When the area F is not logic, "all of our F-beliefs" refers to all of our non-logical F-beliefs. Not even Field would deny that we are justified in believing, e.g., that

either it is the case that there are perfect numbers greater than 1,000,000 or it is not the case that there are.).

35. The next few paragraphs closely follow parts of Sect. 2.4 of my "(Debunkingand Dispensability)." Thanks to Neil Sinclair for permission to reprint them.
36. See Pollock (1986: 38–39) for the distinction between underminers (or "undercutters") and rebutters.
37. Thanks to David James Barnett for pressing me with something like this example.
38. Compare to the discussion of Full-Blooded Platonism in Sect. 2.3.
39. This assumes that we believe of at least one sentence that it is invalid. But it is hard to see how we could believe of any sentence that it is valid while failing to believe of any other that it is invalid.
40. See Connie and Feldman (1998). We could alternatively individuate kinds by methods of belief. Modal Security would then say that if *E* obligates us to give up all of our beliefs formed via *M*, then it does so by giving us reason to doubt that our beliefs formed via *M* are both sensitive and safe. Given plausible assumptions, this formulation is strictly stronger than Modal Security as I have interpreted it.
41. There is another kind of problem case. Suppose that *E* is evidence that a false theory of justification is true according to which our *F*-beliefs are not justified. It might be thought that *E* could undermine all of our *F*-beliefs, but not by giving us reason to doubt their sensitivity or safety. But, on inspection, this seems bizarre. Suppose that *F* includes only propositions for which we have excellent evidence, and *E* is evidence for the view that a belief is justified only if it is infallible. Perhaps we are students in a Philosophy 101 class, for example, and *E* is an apparently strong argument for the view. To give up our *F*-beliefs on the basis of *E*—when *E* is neither "rebutting" nor "direct" reason to doubt that our *F*-beliefs are modally secure–seems to be to give up those beliefs "for the wrong kind of reason" (Barnett, [unpublished manuscript]).
42. See Field (2005) and Clarke-Doane (2012).
43. Strangely, Schechter appears to grant something like this point in his discussion of Nagel in Schechter (2010: 447–448). See also, Chap. 4 of Nagel (1997). My own view is that the interest of the Benacerraf Problem is greatly reduced if the apparent in principle impossibility of answering it is not supposed to undermine our mathematical beliefs (realistically construed).
44. In addition to Joyce, see Greene (2007), Griffiths and Wilkins (forthcoming), Levy (2006), Lillehammer (2003), Ruse (1986), Sinnot-Armstrong (2006), and Street (2006).
45. See Brink (1989), Boyd (2003a, b) for relevant discussion.
46. I argue that debunkers have confused the challenge to empirically justify our moral beliefs with the challenge to explain their reliability in Clark-Doane (2015).
47. I am not saying that Joyce and Street is committed to holding that we could not have easily had different explanatorily basic moral beliefs. I am saying that the

view that we could not have is consistent with, and on some presentations, even suggested by, their evolutionary speculations. As a result, those speculations certainly do not seem to give us reason to believe that we could have easily had different explanatorily basic moral beliefs.

References

Baker, A. (2009). Mathematical explanation in science. *The British Journal for the Philosophy of Science, 60*(3), 611–633.

Balaguer, M. (1995). A platonist epistemology. *Synthese, 103*(3), 303–325.

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. Oxford: Oxford UP.

Balaguer, M. (1999). Review of Michael Resnik's mathematics as a science of patterns. *Philosophia Mathematica, 7*(3), 108–126.

Bealer, G. (1999). A theory of the a priori. *Philosophical Perspectives, 13*, 29–55.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy, 70*, 661–679.

Bengson, J. (2015). Grasping the third realm. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 5, pp. 1–34). Oxford: Oxford UP.

Boyd, R. (2003a). Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part I. *Philosophy and Phenomenological Research, 66*(3), 505–553.

Boyd, R. (2003b). Finite beings, finite goods: The semantics, metaphysics and ethics of naturalist consequentialism, part II. *Philosophy and Phenomenological Research, 67*(1), 24–47.

Braddock, M., Mortensen, A., & Sinnot-Armstrong, W. (2012). Comments on Justin Clarke-Doane's. 'Morality and mathematics: The evolutionary challenge'. *Ethics Discussions at PEA Soup.* http://peasoup.typepad.com/peasoup/2012/03/ethics-discussions-at-pea-soup-justin-clarke-doanes-morality-and-mathematics-the-evolutionary-challe-1.html

Brink, D. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge UP.

Burgess, J. P. (2001). Review of platonism and anti-platonism in mathematics. *The Philosophical Review, 110*, 79–82.

Carnap, R. [1950] (1983). Empiricism, semantics and ontology. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics—Selected readings* (2nd ed., pp. 241–257). Cambridge: Cambridge UP.

Casullo, A. (2002). A priori knowledge. In P. Moser (Ed.), *Oxford handbook of epistemology* (pp. 95–143). Oxford: Oxford UP.

Cheyne, C. (1998). Existence claims and causality. *Australasian Journal of Philosophy, 76*(1), 34–47.

Clarke-Doane, Justin. (Manuscript) "Mathematical Pluralism and the Benacerraf Problem".

Clarke-Doane, J. Debunking and Dispensability (2016). Neil Sinclair and Uri Leibowitz (eds.), Explanation in Ethics and Mathematics: Debunking and Dispensability. Oxford: Oxford University Press.

Clarke-Doane, J. (2012). Morality and mathematics: The evolutionary challenge. *Ethics, 122*(2), 313–340.

Clarke-Doane, J. (2014). Moral Epistemology: The Mathematics Analogy. *Noûs, 48*(2), 238–255.

Clarke-Doane, J. (2015). Justification and Explanation in Mathematics and Morality. Russ Shafer-Landau (Ed.), Oxford Studies in Metaethics, Vol. 10. New York: Oxford University Press.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies, 89*(1), 1–29.

Field, H. H. (1989). *Realism, mathematics, and modality*. Oxford: Basil Blackwell.

Field, H. H. (1996). The a prioricity of logic. *Proceedings of the Aristotelian Society New Series, 96*, 359–379.
Field, H. H. (1998). Mathematical objectivity and mathematical objects. In C. MacDonald & S. Laurence (Eds.), *Contemporary readings in the foundations of metaphysics* (pp. 387–403). Oxford: Basil Blackwell.
Field, H. H. (2005). Recent debates about the a priori. In T. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 1, pp. 69–88). Oxford: Clarendon Press.
Gibbard, A. (2003). *Thinking how to live*. Cambridge, Mass: Harvard UP.
Goldman, A. I. (1967). A causal theory of knowing. *The Journal of Philosophy, 64*(12), 357–372.
Greene, J. (2007). The secret joke of Kant's soul. In W. Sinnot-Armstrong (Ed.), *Moral Psychology, Vol. 3: The neuroscience of morality: Emotion, disease, and development* (pp. 35–80).
Hamkins, J. (2012). The set-theoretic multiverse. *Review of Symbolic Logic, 5*, 416–449.
Harman, G. (1977). *The nature of morality: An introduction to ethics*. Oxford: Oxford UP.
Hart, W. D. (1977). Review of mathematical knowledge. *The Journal of Philosophy, 74*(2), 118–129.
Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford UP.
Huemer, M. (2005). *Ethical intuitionism*. New York: Palgrave Macmillan.
Joyce, R. (2001). *The myth of morality*. Cambridge: Cambridge UP.
Joyce, R. (2008). Précis of the evolution of morality [and reply to critics]. *Philosophy and Phenomenological Research, 77*(1), 213–218 [pp. 218–267 for the "Reply to Critics"].
Korman, D. Z. (2014). Ordinary objects. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/spr2014/entries/ordinary-objects/
Leng, M., Paseau, A., & Potter, M. (Eds.). (2008). *Mathematical knowledge*. Oxford: Oxford UP.
Levy, N. (2006). Cognitive scientific challenges to morality. *Philosophical Psychology, 19*(5), 567–587.
Lewis, D. (1986). *On the plurality of worlds*. Oxford: Wiley-Blackwell.
Liggins, D. (2006). Is there a good epistemological argument against platonism? *Analysis, 66* (290), 135–141.
Liggins, D. (2010). Epistemological objections to platonism. *Philosophy Compass, 5*(1), 67–77.
Lillehammer, H. (2003). Debunking morality: Evolutionary naturalism and moral error theory. *Biology and Philosophy, 18*(4), 567–581.
Linnebo, Ø. (2006). Epistemological challenges to mathematical platonism. *Philosophical Studies, 129*(3), 545–574.
Linsky, B., & Zalta, E. (1995). Naturalized platonism versus platonized naturalism. *The Journal of Philosophy, 92*(10), 525–555.
Lyon, A., & Colyvan, M. (2008). The explanatory power of phase spaces. *Philosophia Mathematica, 16*(2), 227–243.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin.
Nagel, T. (1997). *The last word*. Oxford: Oxford UP.
Nozick, R. (1981). *Philosophical explanations*. Cambridge, Mass: Harvard UP.
O'Leary-Hawthorne, J. (1996). The epistemology of possible worlds: A guided tour. *Philosophical Studies, 84*(2–3), 183–202.
Pautz, A. (2011). Can disjunctivists explain our access to the sensible world? *Noûs* (*Supplement: Philosophical Issues, Epistemology of Perception*), *21*, 384–433.
Peacocke, C. (1999). *Being known*. Oxford: Oxford UP.
Pollock, J. (1986). *Contemporary theories of knowledge*. Lanham, Maryland: Rowman and Littlefield.
Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford UP.
Pust, J. (2004). On explaining knowledge of necessity. *Dialectica, 58*(1), 71–87.
Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge UP.
Putnam, H. [1980] (1983). Models and reality. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics—Selected readings* (2nd ed., pp. 421–444). Cambridge: Cambridge UP.
Resnik, M. (2000). Against logical realism. *History and Philosophy of Logic, 20*(3–4), 181–194.

Ruse, M. (1986). *Taking Darwin seriously*. Amherst: Prometheus Books.
Schechter, J. (2010). The reliability challenge and the epistemology of logic. *Noûs* (*Supplement: Philosophical Perspectives*), *24*, 437–464.
Schechter, J. (2013). Could evolution explain our reliability about logic? In J. Hawthorne & T. Szabò (Eds.), *Oxford studies in epistemology* (Vol. 4, pp. 214–239). Oxford: Oxford UP.
Shapiro, S. (1995). Modality and ontology. *Mind, 102*(407), 455–481.
Shapiro, S. (2000). *Thinking about mathematics: Philosophy of mathematics*. Oxford: Oxford UP.
Sinnott-Armstrong, W. (2006). *Moral skepticisms*. Oxford: Oxford UP.
Stalnaker, R. (1996). On what possible worlds could not be. In *Ways a world might be: Metaphysical and anti-metaphysical essays* (pp. 40–54). Oxford: Oxford UP.
Steiner, M. (1973). Platonism and the causal theory of knowledge. *The Journal of Philosophy, 70* (3), 57–66.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*(1), 109–166.
Street, S. (2008). Reply to copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues (Interdisciplinary Core Philosophy), 18,* 109–166.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford UP.