## Introduction

The *Twin Primes Conjecture* says that there are infinitely many prime numbers, $p$, such that $p + 2$ is also prime.[*] As for this writing, it is an open question whether this conjecture is true or false, although most experts believe that it is true. If the question is settled, it will be settled by proof.

What is a proof? It is basically an argument that convinces experts of the claim proved. More fully, it is an argument that convinces experts that there exists a *formal* proof of that claim.[1] A formal proof of *S* is a finite sequence of sentences, each of which is either an axiom or follows from the previous sentences by a rule of formal inference, the last line of which is *S* itself.

Whether a proof is sound is widely supposed to be a mind- and language-independent matter. It is supposed to be independent of us whether the logic used is correct and whether the argument is valid in that logic. This is *logical realism*. *Mathematical realism* is the view that it is independent of us whether the (nonlogical) axioms are true (and that some are non-vacuously true). Their combination says that we do not make up the logical or mathematical facts.

Mathematical realism raises a question: How do we know that the axioms are true?[2] Even if knowledge of the logical axioms is intelligible (a matter to which we return in Section 4), mathematical axioms are not just (first-order) logical truths. Consider the *Axiom of Choice* (*AC*). This says that if t is a disjointed set not containing the empty set, $\varnothing$, then there exists a subset of $\cup$t whose intersection with each member of t is a singleton. In symbols: $(t)[(x)[x \in t \rightarrow (\exists z)(z \in x) \& (y)(y \in t \& y \neq x \rightarrow \sim(\exists z)(z \in x \& z \in y))] \rightarrow (\exists u)(x)(x \in t \rightarrow (\exists w)(v)[v = w \longleftrightarrow (v \in u \& v \in x)])]$. It is consistent with standard mathematics, minus *AC*, that *AC* is false if standard mathematics is consistent. Universes in which *AC* fails are studied and deeply understood – unlike, say, universes in which squares are circles. So there is nothing "unintelligible" about choiceless mathematics, in any ordinary sense. How, then, do mathematicians know that *AC* is true?

---

[*] Chapters 1 and 3 draw on Chapters 2 and 5, respectively, of my book, *Morality and Mathematics* (Oxford University Press, 2020). For overviews of much of the technical material discussed here, see the following additional Cambridge Elements: *Set Theory* by John Burgess, *Gödel's Theorems* by Juliette Kennedy, and *Foundations of Quantum Mechanics* by Emily Adlam.

[1] Although this is the standard view, it can be questioned. See De Toffoli [2021]. (Of course, no one should be under the illusion that mathematics *as practiced* simply consists of deducing theorems from axioms. See Harris [2015] for a lovely portrayal of the experience of pure mathematical research.)

[2] Logical realism bears on the more general question of how proofs supply mathematical knowledge. For instance, how do we reliably determine that there exists a formal proof on the basis of the (informal) arguments that convince experts (setting aside the question of how we know that the axioms of feasible length are true)? This is not obvious because formal proofs of mainstream theorems are typically too lengthy to be humanly comprehensible. See Gaifman [2012, 506].

The difficulty can sound generic. There is also the question of how scientists know their laws. How do physicists know the *Dirac Equation* or geologists know the theory of plate tectonics? There is nothing unintelligible about the failure of these claims either. The difference is that in these cases, we have the beginning of an answer. Electrons and the Earth's crust leave marks on the world to which our nervous systems respond. We bear no relevant physical relations to the likes of sets. So no story like this suggests itself in connection with our knowledge of *AC*.

To be clear, it is not that mathematical cognition is beyond the reach of science. It is an established subject in cognitive psychology.[3] But we must distinguish the study of mathematical belief acquisition from the study of the *correlation* between our mathematical beliefs and the facts. Science is illuminating our beliefs about sets (and, especially, natural numbers). But it has been conspicuously silent on how they relate to the mathematical facts.

It is tempting to dismiss the problem as stemming from an unwarranted "platonism" about mathematical entities. If platonism is just mathematical realism – that is, the view that the *Twin Primes Conjecture* is either true or false, independent of what anybody says or believes – then the problem does stem from this. But platonism in this sense is difficult to discharge. Our best theories of the physical world are up to their ears in mathematics (Section 2). So absent a way to "factor out" those theories' mathematical commitments, the view that the mathematical truths depend on us would seem to imply that the physical truths do too. For example, the (time-dependent) Schrödinger Equation of quantum mechanics tells us how the state vector of a physical system changes with time. How could this express an independent fact if there are not any independent facts about vectors? Or consider the banal claim that some of our scientific theories are at least *consistent*, that is, do not (classically) imply a contradiction, independent of what anyone says or believes. This claim turns out to be a simple arithmetic claim whose negation (for typical theories) is consistent if elementary arithmetic itself is consistent.[4]

So the question of how humans acquire knowledge of independent mathematical facts is pressing. This Element clarifies the problem, sketches a solution, and discusses its import for philosophy more generally, including modal metaphysics, (meta)logic, and normative theory.

[3] See Butterworth [1999], Carey [2009], De Cruz [2006], Dehaene [1997], Pantsar [2014], and Relaford-Doyle and Núñez [2018] for work on the psychology of number concepts. See Marshall [2017] and Opfer *et al*. [2021] for a critical discussion of the relevance of work like this to the problem of mathematical knowledge.

[4] Technically, the claim is a $\Pi_1$ claim, that is, a claim of the form "for all natural numbers, $\Phi$," where $\Phi$ is a formula with only bounded quantifiers.

# 1 Self-evidence, Analyticity, and Intuition

Knowledge is justified and non-coincidentally true belief (where specifying the relevant sense of "coincidence" is the so-called Gettier Problem). So the problem of explaining our knowledge of the axioms, partitions into two. First, there is the problem of explaining the (defeasible) justification of our belief in the axioms, what I call the *justificatory challenge*. Second, there is the problem of explaining our belief's non-coincidental truth, that is, the *reliability challenge*. Let us begin with the first.

## 1.1 Two Kinds of Axiom

What are the axioms of mathematics? There are two varieties. On the one hand, there are axioms that just speak of their class of models. These are *structural* axioms. For example, a mathematical *group* is any set that is closed under a binary operation satisfying the axioms of associativity, identity, and invertibility. We may also stipulate that the operation is commutative. In that case, the group is said to be *Abelian*. But there is no nonverbal question as to whether the *Axiom of Commutativity* itself is true. It is true of Abelian groups and false of the others.

The situation is prima facie different with *foundational* axioms, like those of set theory, type theory, category theory, and arithmetic. Foundational axioms are, roughly, those on the basis of which one can carry out metatheoretic reasoning. For instance, already in (first-order) *Peano Arithmetic* (*PA*), one can formulate claims about the consistency of theories and prove *relative consistency* results. One can prove, say, that if *Zermelo–Fraenkel* (*ZF*) set theory is consistent, then so is *ZFC* + Cantor's *Continuum Hypothesis* (*CH*); where *ZFC* is *ZF* plus *AC*, and *CH* says that there is a bijection between every uncountable subset of the real numbers and all of them, or, equivalently, given *AC*, that the cardinality of the real numbers is the next greatest after that of the natural numbers. This is written: $PA \vdash \text{Con}(ZF) \rightarrow \text{Con}(ZFC + CH)$. Arithmetic axioms do not seem to be structural – just about their class of models – because there seems to be a nonverbal question as to whether metatheoretic claims like the aforementioned are true. Indeed, Gödel's *Second Incompleteness Theorem* says that if *PA* is consistent, then it cannot prove that it is, written $PA \nvdash \text{Con}(PA)$. Nor, thankfully, do we have that $PA \vdash {\sim}\text{Con}(PA)$. So $PA + \text{Con}(PA)$ and $PA + {\sim}\text{Con}(PA)$ are both consistent if *PA* is – just like group theory with the Axiom of Commutativity and group theory with the negation of that axiom. But the question of whether arithmetic is consistent cannot be dismissed like the question of whether the Axiom of Commutativity for groups is true! There either is or is not a natural number that codes a proof (in classical logic) of "0 = 1", from the axioms of *PA*. Or so it seems. If this were not the case – if, for instance, $PA + \sim \text{Con}(PA)$ were really analogous to Abelian group theory – then there would also be no nonverbal question as to what counts as finite (since a model of $PA + \sim \text{Con}(PA)$ disagrees with us about this), what a formula is, and even what a theory, like *PA*, consists in.

Such considerations only take us so far. They do not show that there is a nonverbal question of whether characteristic axioms of set theory, like *AC*, are true, for example. Indeed, a view to which we return in Section 3.5 says, roughly, that nonstructural foundational axioms are limited to those of (first-order) arithmetic [Weaver 2014, Ch. 30]. But many realists deny that we should draw the line at arithmetic (Koellner [2014], Van Atten and Kennedy [2009], Woodin [2010]). Is it really just a verbal question whether, for any disjointed set not containing $\varnothing$, there is a subset of $\cup$x whose intersection with each member of x is a singleton? What about the claim that there is a so-called *Inaccessible Cardinal (IC)*? This implies new arithmetic results. $ZFC \nvdash \text{Con}(ZFC)$ (if $Con(ZFC)$), but $ZFC + IC \vdash \text{Con}(ZFC)$. So, arguably, belief in *IC* is presupposed by belief in *ZFC*.[5]

Whatever we include among the structural and foundational axioms, there are axioms that cleanly qualify as neither. Tarski's axioms for first-order geometry do not have the flavor of the Axiom of Commutativity for groups (Tarski [1959]). Prima facie, they have an intended subject, Euclidean space, of which they could be wrong. But those axioms are also not foundational, in that one cannot carry out metatheoretic reasoning in the theory.[6] The *Parallel Postulate*, which says, informally, that two straight lines intersecting another so as to make less than a 180° angle on one side intersect on that side, will serve as a key example in Sections 3 and 4. A debate over it would be misconceived, like a debate over the Axiom of Commutativity for groups. But, unlike group theory, this is not because geometry is about its class of models. It is because, if geometric reality exists, it is rich enough to afford an *intended model* of the postulate and its negation.[7]

---

5 "Arguably" because the assumption of *IC* is stronger than the assumption that there is a model of *ZFC* (which is equivalent, by *Soundness* and *Completeness* to Con(*ZFC*)). That is, $ZF + IC$ is stronger than $ZF + \text{Con}(ZFC)$.

6 The theory is decidable and complete and so, by Gödel's theorems, cannot even interpret *Robinson Arithmetic* (i.e., *PA* minus all instances of the *Induction Schema*).

7 The distinction between structural and foundational axioms is similar to Shapiro's distinction between algebraic and non-algebraic ones, although he appears to think that the distinction is exhaustive. See Shapiro [1997, 41 and 50]. Likewise, Balaguer [2001] distinguishes between mathematical domains in which our intentions are exhausted by the (first-order) axioms that we adopt from those in which they are not. This is different from the distinction above if our intentions about a domain can transcend any recursive axiomatization while failing to interpret arithmetic.

## 1.2 Self-evidence

Despite being a relatively fringe area of pure mathematics, set theory is of special philosophical interest. While it has only one nonlogical predicate, $\in$, the claims of all other branches of mathematics can be *interpreted* in it. Those claims can be understood as claims about sets in disguise. It does not follow that all mathematical entities are *really* sets (Benacerraf [1965]). It follows that if the axioms of set theory are consistent, then so are our other mathematical theories.[8]

Could the axioms of set theory, and of all other areas of mathematics, including arithmetic, be consistent but false (or vacuous)? Not if consistency is understood standardly, as a claim about proofs, or set-theoretic models. One could take the notion of consistency as primitive (an idea to which we return in Section 2.3). But on what basis might we believe that, say, set theory is thus consistent? Perhaps the standard answer is: on the basis that it is true, and truth implies consistency (Frege [1980/1884, 106]; Woodin [2004, 31])! But this answer, in tandem with the assumption that mathematical claims are true independent of us, implies mathematical realism.

What, then, explains the justification of our belief that the axioms are true? In other words, why is it *rational* or *reasonable* for us to believe those axioms? A common answer outside of the philosophy of mathematics is that "[a]xioms are mathematical statements that are *self-evidently* true" [Greene 2013, 184, italics in original]. This is perhaps defensible in rudimentary cases.[9] Consider the *Axiom of Extensionality*, which says that if "two" sets have the same members, then they are really one and the same (the converse is a logical truth in first-order logic with identity). In symbols: $(x)(y)(z)[(z \in x \longleftrightarrow z \in y) \rightarrow (x = y)]$. Set theory without Extensionality has been explored (Friedman [1973]; Hamkins [2014]; Scott [1961]). But this axiom is often taken to be some kind of truism about sets. Similarly, the *Axiom of Pairing* says that for any "two" (perhaps not distinct) sets, there is another containing just those two. That is: $(x)(y)(\exists z)(w)[w \in z \longleftrightarrow (w = x \vee w = y)]$. This is also difficult to deny – though it is unclear that any existential statement, even conditional on the existence of other objects, could be self-evident.

8 Whether set theory, rather than another theory, or no theory, can serve as a "foundation" for mathematics in any of the myriad senses that term have been discussed will be irrelevant. It is certainly not unique in interpreting mathematics (see, e.g., Tsementzis & Haverson [2018]). However, it is canonical in this respect, so I focus on it for concreteness.

9 Authors rarely say exactly what they mean by "self-evident." But the idea seems to be that *P* is self-evident when, if one understands *P*, one is thereby (defeasibly) justified in believing *P* (the "thereby" would require explication).

However, Extensionality and Pairing do not imply the existence of a single set![10] Set theory gets going with the *Axiom of Infinity*, written: $(\exists y)(((\exists x)(x \in y \,\&\, (z)(z \notin x) \,\&\, (x)(x \in y \rightarrow (\exists z)(z \in y \,\&\, (w)(w \in z \longleftrightarrow (w \in x v w = x)))))$.[11] This says that there is an inductive set – that is, a set that (according to the usual definitions) includes *0* and includes the number $n + 1$, whenever it includes $n$. It is hard to see the point of calling the claim that something *infinite* exists "self-evident" (Mayberry [2000, 10]).[12] Other axioms are still more doubtful. Consider the *Axiom of Replacement*. This is a schema, not a single axiom. It says that for any set, z, and *any formula*, Φ, such that, for every $t \in z$, there is exactly one x with $\Phi(t, x)$, there exists a set that contains just those things, x, for which $\Phi(t, x)$ holds for some $t \in z$. Formally: $(a)[(u)(v)(w)(u \in a \,\&\, \Phi(u, v) \,\&\, \Phi(u, w) \rightarrow u = w) \rightarrow (\exists y)(x)(x \in y \longleftrightarrow (\exists t)(t \in a \,\&\, \Phi(t, x)))]$, where u, v, w, and y are not free in $\Phi(t, x)$. This has important consequences for set theory, like the *Reflection Principle* (to which it is actually equivalent in the context of the other axioms), which says that if a formula is true of the set-theoretic universe, *V*, then it is already true in an initial segment, $V_\alpha$, of it. The Axiom of Replacement is even needed to prove that the number $\omega + \omega$ exists. But it also implies the existence of outrageously huge sets (though they are tiny for set theory!). Of a relatively small such set, $\kappa$, Boolos, laments: "Let me try to be as accurate, explicit, and forthright about my belief about the existence of $\kappa$ as I can … I … think it probably doesn't exist" [1999, 121].

Finally, consider, again, *AC*. In the context of the other axioms, *AC* is equivalent to the claim that every set is well-orderable (totally orderable so that every non-empty subset of it contains a least element). Thus, *AC* ensures that the set of real numbers, *R*, has a well-order. But what is that order? It cannot be the standard order, since there is no least real number in any open subset of real numbers, like (0, 1). In fact, it is consistent with *ZFC* (if that is consistent!) that there is no *definable* well-order on *R* at all – that is, no well-order specified by a formula, no matter how lengthy and baroque. Even if *AC* is *true*, it is not self-evident!

Needless to say, if typical axioms like Infinity, Replacement, and Choice, are not self-evident, then neither are speculative extensions of them, contra the

---

[10] It is a classical logical truth that there is an $x$ such that $x = x$, since domains are *defined* to be non-empty. But Extensionality and Pairing give us nothing beyond this, an assumption which can, anyway, be dropped by adopting a free logic.

[11] *PA* and *ZF* minus Infinity plus its negation are actually bi-interpretable (if the *Axiom of Foundation*, to be discussed, is stated as a scheme of $\epsilon$-induction). So Infinity is essential to set theory, as opposed to arithmetic. See https://math.stackexchange.com/questions/315399/how-does-zfc-infinitythere-is-no-infinite-set-compare-with-pa

[12] The claim that there is an inductive (infinite) set must be clearly distinguished from the claim that there are infinitely many things. Set theory, minus Infinity, proves the latter, but not the former. The former proves Con(*PA*).

rhetoric of some set theorists. Consider Gödel's *Axiom of Constructibility,* $V = L$. Let $P_{Def}(A)$ refer to the set of all subsets of $A$ definable in the structure $\langle A, \in \rangle$ by first-order formulas with parameters in $A$. Then $V = L$ says that every set lies in the following hierarchy obtained by transfinite recursion on the ordinals: $L_0 = \emptyset, L_{\alpha+1} = P_{Def}(L_\alpha)$ and $L_\gamma = \cup_{\alpha<\gamma} L_\alpha$ for limit $\gamma$ (Gödel [1990/1938]). Is $V = L$ true? The dominant narrative, originating with Gödel himself (see his 1947 work), is that $V = L$ must be false because it settles undecidables – especially "large" large cardinal axioms – in the wrong way (Maddy [1997, Pt II, § 4]; Magidor [2012]; Woodin [2010, 1]). But Fontanella points out that Gödel's "feeling that [$V$=$L$'s] consequences would be implausible is not unanimously shared" [2019, 32]. Indeed, Jensen writes, "I personally find [$V = L$] a very attractive axiom" [1995, 398]. He continues, "I do not understand … why a belief in the objective existence of sets obligates one to seek ever stronger existence postulates [large cardinal axioms]. Why isn't Platonism compatible with the mild form of Ockham's razor …?" [1995, 401].[13] Devlin thinks that $V = L$ "is … a natural axiom, closely bound up with what we mean by 'set' …. [and] tends to decide problems in the 'correct' direction" [1977, 4]. And Eskew queries, "The axiom $V = L$ … settles 'nearly all' mathematical questions …. [I]t can be motivated by constructivist views that are still widely held today …. [A] wealth of powerful combinatorial principles … follow from … $V = L$ …. [So] why hasn't there been … a stronger push to adopt it as a[n] … axiom for mathematics?" [2019).][14]

## 1.3 Analyticity

So appeal to self-evidence does not afford a satisfying answer to the justificatory challenge. How else might we explain the justification of belief in the axioms? Another prominent proposal is that the axioms are *analytic*, "a system of tautologies, the basic elements of which are true by virtue of the meanings of

[13] See Arrigoni [2011] for an explication and defense of Jensen's position.

[14] Consequences of $V = L$ that are said to be particuarly counterintuitive (besides that there does not exist a so-called *Measurable Cardinal*) include there is a definable but nonmeasurable set of reals, and the *Diamond Principle* holds. Gödel [1947] contains further arguments against the axiom. On the other hand, Fraenkel, Bar-Hillel, and Levy [1973, 108–109] contain additional arguments supporting $V = L$, and Simpson compares skepticism about large cardinals (the larger of which imply that $V = L$ is false) to (rational) religious skepticism in his [2009]. Friedman quips, "[some s]et theorists say that $V = L$ has implausible consequences … [and] claim to have a direct intuition which allows them to view these as so implausible that this provides 'evidence' against $V = L$. However, mathematicians [like me] disclaim such direct intuition about complicated sets of reals. Many … have no direct intuition about all multivariate functions from $N$ into $N$" [2000]! Arrigoni and Friedman emphasize that criteria of success and intuitiveness vary, and that "$ZFC + V = L$ … is fruitful in consequences, furnishes powerful methods for solving problems and introduces the concept of *constructability*, important throughout set theory" (Arrigoni & Friedman [2012, 1361, italics in original]).

the terms used" [Singer 1994, 8]. In light of Quine [1951a], most philosophers are careful to distinguish epistemic from metaphysical versions of this proposal.[15] The *metaphysical* version says that the meaning of the term "$\in$" somehow *makes it the case* that the axioms are true. This is hard to even understand. How could a meaning make a fact? The *epistemic* version says that it is "part of the concept of $\in$" that standard axioms hold, and those of us with that concept are, therefore, defeasibly justified in believing those axioms (at least assuming that we are justified in believing that there are any sets at all).

Supposing for the moment that the notion of epistemic analyticity is in good order, it is doubtful that standard axioms are so analytic. First, it is hard to imagine a compelling argument that it is just "part of the concept of $\in$" that standard axioms hold, given that some theorist actually denies them. Consider the *Axiom of Foundation* (or *Regularity*). This schema says that for any formula, Φ, if there is a set that satisfies Φ, then there is a *minimal* x that does – an x such that Φ and no y$\in$x such that Φ. In symbols: $(\exists x)\Phi \rightarrow (\exists x)[\Phi \,\&\, (y)(y \in x \rightarrow \sim \Phi^*)]$ (where Φ does not contain y and Φ* is just Φ but contains y wherever Φ contains free occurrences of x). This is equivalent to a *Principle of Set-theoretic Induction*, according to which, if Φ is a formula such that, whenever all members of x satisfy Φ, x does too, then every set satisfies Φ. *Foundation* is widely alleged to be the foremost example of a nontrivial analytic axiom (Boolos [1971, 498], Shoenfield [1977, 327]). It is just part of what we mean by "$\in$" that every set is formed at some stage of a transfinite generation process via the powerset and union operations, beginning with $\varnothing$ – so that no set contains itself, and there are no infinitely descending chains of membership, for example. This "platitude" is equivalent to Foundation, given the other axioms. It says, if $V \neq L$, that all sets lie in a liberalized version of Gödel's $L$, the *Cumulative Hierarchy*: $V_0 = \varnothing$, $V_{\alpha+1} = P(V_\alpha)$, and $V_\gamma = \cup_{\beta<\gamma} V_\beta)$, for limit $\gamma$, where $P(x)$ is the *ordinary* powerset operation (i.e., the set of *all* subsets of the set x, even those that are not definable in the structure $\langle x, \in \rangle$). But far from being beyond dispute, many doubt the *coherence* of the resulting "iterative conception of set" (*ICS*)! What, after all, could "formation" and "generation" mean when these terms are applied to the likes of (pure) sets (Ferrier [Forthcoming], Potter [2004, § 3.3])? Rieger complains: "[*ICS*] does not embody a philosophically coherent notion of set. There is a coherent constructivist position …. There is also a coherent anti-constructivist

[15] See Boghossian [2003] for the distinction.

position …. But [*ICS*] is an uneasy compromise between these two: it pays lip-service to constructivism without really meaning it" ([2011, 17–18].[16]

Even if it were just "part of the concept of ∈" that standard axioms hold, however, epistemic analyticity is a suspect idea. If we were worried that some sets fail to occur at any $V_\alpha$, then under the assumption that it is "part of the concept of ∈" that all sets do, we should just worry *that our concept of set is not satisfied*. Maybe instead of sets, there are only set-like things, which are similar to sets except that some fail to live in any $V_\alpha$ (because they are, say, self-membered). Epistemic analyticity makes justification too cheap. For any claim of interest, *S*, consistent with the other claims that we believe, we could be justified in believing *S* simply by enriching our concepts! Of course, if every consistent concept of set – or, more carefully, theory in the language of first-order set theory – were satisfied (in a class model, under a face-value Tarskian satisfaction relation), then we might be able to rule out the worry that ours is not. But, if that were the case, then the whole project of seeking out the "true" set-theoretic axioms would be misconceived. *Every consistent set-theoretic sentence that was not a logical truth would be like the Parallel Postulate* (understood as a claim of pure mathematics). By Gödel's Second Incompleteness Theorem, this includes (a coding of) the claim that *PA* is consistent, if it is.

## 1.4 Reflective Equilibrium

So the axioms of set theory seem to be neither self-evident nor analytic in a useful sense. Is there any other way to explain the justification of our belief in them? Russell proposes what is perhaps the canonical way. He writes, "We tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises …. But the inferring of premises from consequences is the essence of induction; thus the method in investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science" [1973/1907, 273–4]. Russell's proposal is that, first, the *epistemic* priority of mathematical principles is opposed to their *logical* priority. Although we deduce theorems from axioms, we are justified in believing the axioms because we are justified in believing the theorems that they imply, rather

[16] See also Azcel [1988, Introduction]. Advocates of the so-called logical conception of set, such as Quine [1937] and [1969], reject the Axiom of Foundation too. Quine's *New Foundations* (*NF*) for mathematical logic proves the existence of a universal set, which contains itself. (The relative consistency of *NF* is still officially an open problem. However, experts appear to be converging on the view that it is consistent even relative to quite weak theories. See: https://mathoverflow.net/questions/132103/the-status-of-the-consistency-of-nf-relative-to-zf)

than the other way around. Second, the theorems that justify us in believing the axioms need not be self-evident or analytic. They need only be initially plausible, or *intuitive*.

Russell's method prefigures *reflective equilibrium*, championed by Goodman and Rawls. Rawls writes, "Although … various judgments are viewed as firm enough to be taken provisionally as fixed points, there are no judgments of any level of generality that are … immune to revision" [1974, 8].[17] An attraction of the method is that it analogizes the epistemology of mathematics to that of empirical science, which is better understood. Gödel stresses that "the axioms need not … be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these 'sense perceptions' to be deduced" [1990/1944, 121]. But neither Russell nor Gödel distinguishes the justificatory and reliability challenges. The analogy at most holds for the former. Benacerraf complains, "there is a *superficial* analogy …. [W]e 'verify' axioms by deducing consequences from them concerning areas in which we seem to have more direct 'perception' (clearer intuitions). But we are never told how we know even these, clearer, propositions" [1973, 674, italics in original]. Field clarifies that "we [can] grant … that there may be positive reasons for believing in [select theorems]. These … might involve … initial plausibility …. But Benacerraf 's challenge … is to … *explain how our beliefs about these remote entities can so well reflect the facts about them*" [1989, 26, my emphasis].

We discuss the reliability challenge in detail in Section 3. For the present, even the idea that the *justification* of our mathematical beliefs can be explained in analogy with the justification of our empirical scientific ones is tendentious. The problem is that there is *disagreement over the data to be accounted for* in the mathematical case that has no apparent analog in the empirical one.[18]

Consider a paradigmatic disagreement over an empirical scientific theory, the theory of dark matter. Those who reject the hypothesis of dark matter, like Milgrom [2002, 45], and propose amendments to Newtonian gravity do so in order to account for *the same data*.[19] They do not disagree over *it*. But disagreement in the foundations of mathematics seems *characteristically* to

[17] See also Goodman [1955, 63–64].

[18] This is why comments like the following are too quick. "Many realists … take the epistemological challenge to be one about … epistemic justification … .And they reply in the obvious ways … by showing that their favorite theory of epistemic justification in general nicely applies to the case of [mathematical] beliefs … .[T]his is not a promising way of understanding … the epistemological challenge … . [W]hatever your theory of epistemic justification, it is hard to see any special difficulties applying it to [mathematical] beliefs [Enoch 2009, 2]." (Enoch is actually talking here about normative beliefs, although the more general context is both normative and mathematical ones.)

[19] For details, see Milgrom's online overview here: http://ned.ipac.caltech.edu/level5/Sept01/Milgrom2/paper.pdf See Merritt [2020] for a philosophical discussion of Milgrom's program.