Feferman on the Indefiniteness of CH

Peter Koellner

> I can on no way agree to taking 'intuitively clear' as a criterion of truth in mathematics, for this criterion would mean the complete triumph of subjectivism and would lead to a break with the understanding of science as a form of social activity.
>
> Markov (1962)

Feferman is arguably the foremost critic of set theory since Weyl. His paper "Is the continuum hypothesis a definite mathematical problem?" provides a useful overview of his negatie answer, drawing on many papers over many years. The main thesis is that the continuum hypothesis (CH) is an "indefinite" statement and that the underlying reason is that the "concept of arbitrary set [which is essential to the formulation of CH] is vague or underdetermined" and "there is no way to sharpen it without violating what it is supposed to be about." (1)

In what follows I am will exposit and extend Feferman's critique, argue that each component fails, and conclude that when the dust settles the entire case rests on the claim that the concept of natural number is clear while the concept of arbitrary sets of natural numbers is not clear. My stance on this final resting point is captured in the above quotation from Markov.

1 Overview

The paper “Is the Continuum Hypothesis a definite mathematical problem?” gives an account of his main reasons for thinking that CH is not a definite mathematical problem. The arguments are of two kinds—direct and indirect (or circumstantial). The direct arguments are based on (a) the meta-theory of set theory and (b) a philosophical view of the nature of mathematics (in particular, an anti-platonist view). This is nicely summarized in “Philosophy: 5 Questions”:

> I came to the conclusion some years ago that CH is an inherently vague problem (see, e.g., the article (2000) cited above). This was based partly on the results from the metatheory of set theory showing that CH is independent of all remotely plausible axioms extending ZFC, including all large cardinal axioms that have been proposed so far. In fact it is consistent with all such axioms (if they are consistent at all) that the cardinal number of the continuum can be anything it ought to be, i.e. anything which is not excluded by Konigs theorem. The other basis for my view is philosophical: I believe there is no independent platonic reality that gives determinate meaning to the language of set theory in general, and to the supposed totality of arbitrary subsets of the natural numbers in particular, and hence not to its cardinal number.

The circumstantial argument is based on a thought experiment involving the Millenium Prize Problems. Feferman claims that this thought experiment shows that the mathematical community has implicitly endorsed his claim that CH is an indefinite statement. This is nicely brought out in the continuation of the above quoted passage:

> Incidentally, the mathematical community seems implicitly to have come to the same conclusion: it is not among the seven Millennium Prize Problems established in the year 2000 by the Clay Mathematics Institute, for which the awards are \$1,000,000 each; and this despite the fact that it was the lead challenge in the famous list of unsolved mathematical problems proposed by Hilbert in the year 1900, and one of the few that still remains open.

This structure is also present in the paper "Is the Continuum Hypothesis a definite mathematical problem?". There are three main sections: (1) The first section presents the indirect (or circumstanial) argument involving the Millenium Prize Problems. (2) The second section presents the direct argument, covering both (a) the metamathematical argument and (b) the argument pertaining to the nature of mathematics. (3) The third section presents a framework in which one can articulate claims of the form "φ is indefinite" for various statements φ.

We can quickly dispense with the third section, where Feferman provides a framework for articulating statements of the form "φ is indefinite". This work is of interest in its own right but for two reasons it does not have bearing on our present concerns: First, the work could at most be used to *articulate* the claim that CH is indeterminate but not *argue for* that claim. Second, and more to the point, even by Feferman's own admission, the notion of formal definiteness that he provides ia "a very crude criterion of definiteness" and we "need more refined notions of definiteness/indefiniteness to throw light in whether CH is a definite statement"; in fact, as far as the results are concerned, it is not even known if one of the systems that Feferman provides is suitable for *expressing* the claim (let alone arguing for it) that CH is indeterminate.

I should begin by laying my cards on the table: I am agnostic on the question of whether CH is a mathematical problem and I have pursued frameworks—multiverse conceptions, the approach via incompatible Ω-complete theories, the approach via incompatible ultimate inner models—on which the indefiniteness of CH might be maintained. I also have enormous respect for Feferman and his work and have learned more from it than from practically any other philosophically oriented logician. Nevertheless, I think that the results that Feferman cites and the arguments that he gives in this paper do not provide any reasons for believing that CH is an indefinite statement. I will make this case as forcefully as I can and I hope that the reader appreciates that in presenting the criticisms in so direct a manner my goal is only to place the fundamental issues in the sharpest possible light.

2 The Indirect Argument

The thought experiment involving the Millenium Prize Problems is supposed to provide "considerable circumstantial evidence to support the view that CH is not definite".

Here is the background: The Clay Mathematics Institute of Cambridge, Massachusetts established seven Millenium Prize Problems. The prizes were announced in Paris on May 24, 2000. One of the problems on the list—namely, the Riemann Hypothesis—also appears on the famous list of twenty-three problems that Hilbert presented on August 9, 1900.

Feferman notes that CH (which *was* on Hilbert's list) is *not* on the prize problem list and he imagines a discussion between the prize problem advisory board and a group of set theorists to determine whether CH is suitable for inclusion on the list. The set theorists explain some of the advances in the program for large cardinal axioms—in particular, the work of Martin, Steel, and Woodin, which showed (among other things) that large cardinal axioms settled the classic undecided statements of second-order arithmetic (largely through implying PD) and statements somewhat further up in the complexity hierarchy (namely, statements about $L(\mathbb{R})$). The panel expresses concerns over whether large cardinal axioms are acceptable and, setting this aside, inquire about the situation with CH. The set theorists then explain that (in contrast to the case of PD) large cardinal axioms (of the variety discovered so far) do not settle CH and they go on to give a (mistaken) account of Woodin's approach to settling CH.[1] The board ends up concluding that CH is not suitable for inclusion. What does this show?

Feferman takes this as evidence that mathematical community has implicitly accepted his conclusion that CH is not a definite mathematical problem. But the whole issue is tied up with an ambiguity involved in the word 'problem'. One must draw a distinction between a *statement* (say an open problem) being indefinite and a *task* (say settling an open problem) being indefinite. Feferman is claiming that the *statement* (the open problem) CH is indefinite. But the facts about the scientific advisory boards conjectured behaviour shows at most that the *task* of settling CH is insufficiently delimited to warrant placing one million dollars on the implementation of the

[1]Feferman writes: " "Some of the experts think that one of the most promising avenues is that being pursued by Woodin (2005a, 2005b) via his strong Ω-logic conjecture which, if true, would imply that the cardinal number of the continuum is $\aleph_2$." This is mistaken. First, the Strong Ω Conjecture does not logically imply anything about the size of the continuum; rather, assuming it there is a very involved argument (not a proof) to the effect that CH is false. Second, the argument is that CH is false, not that the continuum has size $\aleph_2$. To get the stronger conclusion one need a different argument, one involving maximality considerations.

task. The task "Settle CH" is not sufficiently definite to warrant inclusion on the list since it is known that any resolution of CH is going to involve subtle issues surrounding the justification of new axioms. For this reason, *no set theorist (even a set theorist who was firmly convinced that* CH *is a definite mathematical problem) would dream of making a case that* CH *should be included on the list.* Once one is clear on this ambiguity one sees that the scientific advisory board's conjectured behaviour does not lend support to the claim that CH is no a definite statement.[2] Let me elaborate on this with three points:

First, the panel wants problems where the solution is going to be relatively uncontroversial. This is entirely reasonable since there is a million dollars riding on the implementation of the task. The panel is thus looking for "a complete mathematical solution to one of the problems" and there is a systematic procedure for determining this—"a proposed solution must be published in a refereed mathematics publication of worldwide repute"; "it must also have general acceptance in the mathematics community two years after"; etc. It is clear that CH does not meet these criteria and the reason it does not meet these criteria is independent of whether it is a definite mathematical problem or not.

Second, imagine a parallel case involving a strong statement of arithmetic. Harvey Friedman has constructed statement of arithmetic that have attracted the attention of number theorists. The statements have the look and feel of statements that can be resolved in a weak system like PA but they actually require strong assumptions (as measured by the large cardinal hierarchy) for their resolution. Suppose a Friedman-like statement catches the interest of number theorists and becomes a central problem and it has the feature that it is known to require strong assumptions but it is not known which way the answer will come out when one adds those assumptions. Suppose further that Friedman forgets whether or not the statement in question required strong assumptions for its resolution.

[2]Of course, certain members of the board or the mathematical community (like Feferman) might believe that stronger statement. But using the behaviour of the board—which would also be the behavior of a board of set theorists who believed that CH is a definite statement—to draw such a sociological conclusion is to rest a case on an ambiguity. (Besides, even if this was a prevalent sociological fact what should we takes its significance to be? It used to be a sociological fact that irrational numbers (negative numbers, complex numbers, curves that were not constructible via Cartesian mechanical process) did not exist. But the significance of such sociological facts is rather fleeting.

Suppose Friedman takes such a statement and gets a number theorist interested. The number theorist works on it over the weekend and is unsuccessful. a number theorist takes such a problem and spends the weekend working on it only to learn later that it is not resolvable using the assumptions being employed—it requires much stronger assumptions; just as in the case of PD it requires statements of strong (large cardinal) consistency strength. Should the panel learn of the metamathematical facts they would certainly not include these problems on the list. And this would not indicate that these arithmetical statements are not definite. It would simply reflect the fact that any proposed solution would have to rest on the case for large cardinal axioms and that case will be subtle and controversial.

Third, there are related problems that one *could* put on the list, namely, one could take a conjecture of the form "φ is provable in T" and put that conjecture on the list, where φ is one of the above statements of arithmetic or CH and T is a strong system. Here is an example of such a conjecture: Let T be the theory discussed in the final section of my paper "On the Question of Absolute Undecidability". (It involves ZFC, the large cardinal assumption I_0, and the structural axioms (A) and (B).) The question is: Assume T. Does CH hold? That would actually be a remarkable result! And the resolution of it is clear enough to warrant inclusion on the list.

In short, what changed between 1900 (when Hilbert presented his list) and 2000 (when the Clay Institute presented their list) is that independence in arithmetic and set theory became a reality. From that time onward it became clear that unless a statement was expected to be resolvable in one of the standardly accepted systems—like PA or ZFC—then one had better render the *task* (not the *statement of the problem*) more precise by conditioning on the conjectured necessary background assumptions. The problems on the Millenium Prize list *are* expected to be resolvable in the standard systems. In the case of CH we are even at a loss for interesting theorems of the form "Assume T. Then CH." That alone is reason for not including CH—even in a conditional form—on the list. To answer the question of whether the *statement* CH is not definite—as opposed to whether the *task* "Resolve CH" is not sufficiently definite for inclusion on the list—one is going to have to dig deeper, even if one is only looking for circumstanial evidence.

3 The Direct Argument

The direct argument has two components. The first component concerns the nature of mathematics. Here Feferman argues that platonism is untenable and in its place he advocates what he calls "conceptual structuralism". The position of conceptual structuralism is supposed to support the main thesis that CH is an indefinite statement. The second component concerns the metamathematics of set theory. Here Feferman argues that the metamathematical situation in set theory is distinctively different than that in first-order arithmetic and, moreover, in a way that supports the thesis that CH is an indefinite statement.

3.1 The Nature of Mathematics: "Platonism" Versus Conceptual Structuralism

There are two opposing extreme views concerning the nature of mathematics. One extreme maintains that the mathematical realm is completely independent of our practices—it is outside of the space-time manifold, lying eternal in a third realm. This view is often called *platonism* (with grave injustice to Plato) and the challenge for this view has been to explain what by its lights would appear to be a miracle, namely, how our practice of proving things ("down here") can manage to track the nature of things ("up there"). In other words, the challenge for this view is to overcome the *alienation problem*, that is, the problem that on this view truth is (in Tait's words) alienated from proof. The other extreme maintains that the mathematical realm is really just a projection of our practice. This certainly overcomes the alienation problem but it leads to the *objectivity problem*, that is, the problem that on this view mathematics would seem to be like fiction, an area where we have more control than we appear to have in the case of mathematics.[3]

It appears that when discussing platonism it is the above extreme version that Feferman has in mind and accordingly it is "[t]he well-known difficulties of platonism have left it with few if any adherents" (9). It should be borne in mind that few if any people ever held such a view and that there are many platonist today. For example, for recent versions of platonism—versions that

[3]For what it is worth my own view is that each of these pictures is exceedingly naive and that when one tries to cash out the metaphors one finds that they fall through one's fingers. The truth lies in between.

do not lead to the alienation problem—see Tait and Maddy. In any case, if this is the above extreme view is indeed the target, then Feferman is certainly right to be drawn in the other direction.

But I fear that Feferman is drawn too far. For in drawing away from the alienation problem it appears that he embraces the other extreme and so faces the objectivity problem.

> My claim is that the basic conceptions of mathematics and their elaboration are also social constructions and that the objective reality that we ascribe to mathematics is simply the result of intersubjective objectivity about those conceptions and not about a supposed independent reality in any platonistic sense. (13)

It would take us too far afield to enter a detailed discussion of Feferman's conceptual structuralism. But it will be helpful to discuss the view briefly and separate the *central tenants of the view* from *various applications of it.*

The centerpiece of the view is that "The basic objects of mathematical thought exist only as mental conceptions".[4] These "[b]asic conceptions differ in their degree of clarity. One may speak of what is true in a given conception, but that notion of truth may only be partial. Truth in full is applicable only to completely clear conceptions." "What is clear in a given conception is time dependent, both for the individual and historically." "The objectivity of mathematics lies in its stability and coherence under repeated communication, critical scrutiny and expansion by many individuals often working independently of each other." These are the central tenets of the view.

The first point I wish to make is that the central tenants of conceptual structuralism are consistent with the claim that CH is definite. Of course, conceptual structuralism is flexible enough to incorporate the idea that CH is indefinite.[5] The point is that it is also flexible enough to incorporate the idea that CH is definite (just as, in Feferman's particular application of it, it is

[4]I suspect that this is not literally what Feferman means. Take the case of number theory. I don't think that he intends to say that *every individual number* exists as a mental construction can't since that would equip us with highly idealized conceptual powers (consider very large numbers, much larger than (current estimate of) the number of fundamental particles in the universe). He seems to be saying that *the concept of natural number* (with 0, successor, addition, etc.) exists as a mental construction. For more on this kind of view see Tony Martin's "Gödel's Conceptual Realism".

[5]"[T]his view of mathematics does not require total realism about truth values. That is, it may simply be undecided under a given conception whether a given statement in

flexible enough to embody the idea that statements of first-order arithmetic are definite).[6]

Now, Feferman does go further. He applies the framework in such a way that it applies asymmetrically to number theory and set theory. The concept of natural number is deemed in such a way that there are no truth-value gaps—it is sufficiently robust to ensure that every statement of first-order number theory is definite (in that it has a determinate truth value). The concept of arbitrary subset of natural number is deemed in such a way that there are gaps—it is insufficiently robust to ensure that statements of second-order or third-order number theory are definite; in particular, CH comes out indefinite. How does this come about?

At the general level Feferman maintains that "The objectivity of mathematics is a special case of intersubjective objectivity that is ubiquitous in social reality." (12) Here I think that there is already a problem. The assimilation of mathematical objectivity with intersubjective objectivity in social reality is a mistake. To see this it is helpful to consider counterconventional conditionals (in the sense of Iris Einheuser): Compare: "If the political structure wer different in such a way that there were no countries then PK would not be a Canadian citizen." "If there were no humans (and hence no human concepts) then there would not be infinitely many prime numbers". In the case of the latter statement there are two ways of reading it. On the first reading we imagine how things look in the counterfactual situation (where there are no humans); in that situation there would (by hypothesis) be no people to talk about the prime numbers. Nevertheless, on a different reading, we can use the conceptual apparatus here to talk about that situation and, using that apparatus, we can still say that there are *still* infinitely many prime numbers. That fact—the infinitude of the prime numbers—is not "injured" by the absence of people. The situation with the other statement—the one about citizenship—is entirely different. The second reading cannot get a foothold. There is no way of thinking of me being a Canadian citizen (or there being any Canadian citizens) in the counterfactual situation: citizenship is

the language of that conception has a determinate truth value, just as, for example, our conception of the government of the United States is underdetermined as to the presidential line of succession past a certain point." (13)

[6]For example, the central tenets of conceptual structuralism are quite close to those of Tony Martin's version of conceptual realism and Daniel Isaacson's structuralism. Martin's view CH is a definite problem (provided that the concept of set is non-vacuous). Isaacson's view is that CH is a definite problem.

too intimately tied to the political structures. Take away the structures and you take away citizenship. In short, there is an asymmetry between the case of mathematical objectivity and social objectivity, one that is not tracked by Feferman's assimilation.

But all of that is a bit off of our main topic. The main point I wish to stress is that conceptual structuralism does not on its own say anything about whether or not a given conception (the natural numbers, arbitrary sets of natural numbers) leads to definite or indefinite statements. We are thus provided with a flexible framework, one that is capable of being applied to views on which CH is definite and also to views on which CH is indefinite. So the framework alone is not going to illuminate the question of definiteness. For that we are going to have to turn to other considerations, like those addressed in the next section.

3.2 The Meta-Mathematical Arguments

The meta-mathematical arguments concern the difference between first-order arithmetic (where Feferman maintains that all statements are definite) and second- and third-order arithmetic and set theory (where Feferman maintains that many statements are indefinite). In this section I want to go beyond Feferman's critique and examine many more aspects of the meta-mathematical situation. In each case I will cite results showing that there is a parallel between the case of first-order arithmetic and second- and third-order arithmetic, thereby undermining the asymmetrical stance.

(A) Conceptions of the Continuum. Feferman notes that there are many conceptions of the continuum, some physical, some having to do with intuition, some purely mathematical. Even narrowing down on the purely mathematical there are still many conceptions—Euclidean, Hilbertian, Dedekindian, set-theoretic.[7] But the point is that none of this is a direct relevance to the question at hand since CH is a statement of set theory and set theorists are quite clear (internal to set theory) on what question they are asking.[8]

[7] Incidentally, there is a parallel in the case of number theory; for example, there is the ordinal conception and the cardinal conception.

[8] It is true that there are several different formulations—for example, you can take the reals to be 2^ω, ω^ω, $\mathscr{P}(\omega)$, etc.—but they are all (easily proved to be) equivalent and set theorists pass back and forth between the formulations with so ease and without confusion.

(B) Full SO Categoricity. Some have maintained that the quasi-categoricity results in set theory ensure that CH is a definite problem. For example, Kreisel and, more recently, Isaacson, maintain this. It is thus claimed that there is a parallel between first-order arithmetic and second-order arithmetic (and third-order arithmetic and any "height-definite" fragment of set theory) in that categoricity secures definiteness in both realms.

Feferman rejects these results since he maintains that they beg the questions at the metalevel through their employment of full second-order logic, a logic that is entangled with set theory in an intimate manner. I agree with much of this. (See my "Strong Logics of First and Second Order.")

Feferman concludes that first-order arithmetic and second-order arithmetic (and third-order arithmetic and "height-definite" fragments of set theory) *are* indeed parallel with regard to categoricity but that that is because in each case the categoricity results have no (or little) bearing on the question of definiteness. I agree. We are going to have to look deeper if we are going to find the crucial asymmetry between the two cases.

(C) Schematic Categoricity.[9] It is worth mentioning that there is another version of categoricity, one that does not employ full second-order logic, namely, what might be called *schematic* or *internal* categoricity. The idea dates back to Parson's 1990 Iyunn paper and is pursued further in his recent book *Mathematical Thought and Its Objects"*. The idea is that in being committed to the natural numbers we are committed to accepting induction for any predicate that we come to accept. Suppose two people have their respective number systems, $\langle N, 0, S, \ldots\rangle$ and $\langle N', 0', S', \ldots\rangle$. If each accepts the others number predicate (N, respectively, N' in the range of their induction schema then together they can show that the natural mapping π (sending 0 to $0'$ and $S(n)$ to $S'(\pi(n))$ is an isomorphism.[10]

I do not want to enter a discussion of the philosophical significance of such results. I just want to point out (as Tony Martin observes in his paper "Multiple Universes of Sets and Indeterminate Truth-Values" that the situation in set theory is exactly parallel. So we are going to have to dig deeper if we are going to find the crucial asymmetry between the two cases.

[9]This is a point that Feferman does not discuss.

[10]This works over weak systems like $I\Delta_0$, provided one treats the induction axiom as schematic. The idea is schematic. For example, it does not presuppose working over PA—it scales up and down as long as one remains schematic at each level.

(D) Non-Standard Models. There is an approach to finding an asymmetry between the two cases that rests on non-standard models. In the case of arithmetic when you build a non-standard model its non-standardness is immediately revealed since the standard part is a proper subset of the entire structure. Some have thought that the case is quite different in set theory.

But the case in set theory is exactly parallel. Take a case like Cohen's method of forcing where you build a non-standard model. In the case of forcing there are two standard model-theoretic ways of doing this. In the first approach one starts with a *countable* transitive model M of the theory in question (say ZFC) and one adds a generic object. The point is that all of this is done *within* set theory and it is revealed in the first step that one is dealing with a non-standard model since it is a countable object. In the second approach one builds class size models (and do the above point is no longer applicable) but one builds a *Boolean-valued* model $V^{\mathbb{B}}$ and, once again, it is immediately revealed that it is non-standard. So, once again, arithmetic and set theory are parallel in this regard. We have to keep digging.

(E) Facility in constructing models. Some (for example, Hamkins) have cited the facility with which we can manipulate models of set theory as evidence that certain statements of set theory are not determinate. For example, given a countable transitive model M of ZFC we can force to obtain an extension $M[G_0]$ that satisfies CH and then force again to obtain an extension $M[G_0, G_1]$ that satisfies $\neg$CH and then force again to obtain an extension $M[G_0, G_1, G_2]$ that satisfies CH, and so on. We can flip CH on and off like a switch. This god-like control has led some to think that it indicates that CH is not definite.

But exactly the same thing holds in the setting of first-order arithmetic and the reason is that there are Orey sentences of first-order arithmetic. Let φ be an Orey sentence for PA. (One can choose theories other than PA and one can arrange that φ is quite simple—it can be Δ^0_1 (provably over PA (or over which ever background theory one is working with)).) Such a sentence has the feature that if we take a model M of PA we can end-extend it to obtain a model M' in which φ holds and then end-extend that model to obtain a model M'' in which $\neg\varphi$ holds, and so on. If the above argument concerning CH is a good one then this god-like control should indicate that φ is not a definite statement. But Feferman (and most people) think that *all* statements of first-order arithmetic are definite. The main point is that once again, arithmetic and set theory are parallel in this regard. We have to keep digging.

(F) *Flexible Orey Sentences.* One might try to argue that set theory is different in that there are *flexible* Orey sentences like PD and CH. For example, PD has the feature that it is not just an Orey sentence over ZFC but it also remains an Orey sentence when one supplements ZFC with large cardinal axioms (measurable, strong, etc) until at a certain point (in the region of Woodin cardinals) it ceases to be an Orey sentence.

But in joint work with Woodin we showed that the situation in first-order arithmetic is exactly the same. There are even flexible Orey sentences which parallel CH. (This work was done quite some time ago and is unpublished. I have included some of it in the appendix.)

In each of the above cases (the first two of which are considered by Feferman) one has an attempt to isolate an asymmetry between the case of arithmetic and number theory—one that argues for definiteness in the first case and indefiniteness in the second. But in each case the attempt fails. The cases are parallel.

3.3 The Final Retreat: Clarity

What then is the key difference? The only thing remaining (so far as Feferman has argued and so far as I can see) has to do with clarity:

> We have a clearer conception of what it means to be an arbitrary infinite path through the full binary tree than of what it means to be an arbitrary subset of N, but in neither case do we have a clear conception of the totality of such paths, resp. sets.

I don't understand the first point. Feferman must mean that when we *first* encounter these notions the first is clearer than the second. But by his own admission clarity is something that evolves over time. And it doesn't take long for the two notions to become equally clear since after five minutes of explanation any student will come to see that one can pass from one to the other and back again with ease.

The main point that is relevant for our purposes is the second point. Feferman thinks that we do not have a clear conception of an (arbitrary) path through the binary tree (or, equivalently, the idea of an arbitrary subset of natural numbers). And yet he thinks that we have a clear conception of the natural numbers. This, I have argued, is what the entire case for asymmetrical treatment ultimately rests on.

The difficulty I have with this is embodied in the quote from Markov at the beginning of this paper. The concept of clarity is not sufficiently clear. It is too subjectivistic. Intuitions of clarity are not robust. Some people think that the concept of an arbitrary subset of natural numbers is about as clear a conception as one can have; others disagree. Whose intuitions are to count? We are hear dealing with a notion—one that is carrying all of the weight of Feferman's case, or so I have argued—where there is not even the kind of intersubjective objectivity that is had by the cases of social realism that he cites.

I certainly agree that the concept of (an arbitrary) natural number is clear and at times I am willing to agree that it is *clearer* than the concept of an (arbitrary) path through the binary branching tree. I want here to put a bit more pressure on these initial judgments.

To begin with, most students when they learn of these matters find (at least in my experience) each of the conceptions to be clear. It is only when one learns more about the metamathematical subtleties involved the situation becomes more complex and the initial judgements become more suspect.

What, for example, are we saying when we say that the concept of the natural numbers is clear? Are we just saying that the idea of starting with 0, taking successors, and continuing indefinitely is clear? And what does this idea of continuing indefinitely really mean? Is the conception clear in such a way that intrinsically justifies all of PA? If so, then it is clear in such a way that it intrinsically justifies the totality of the Ackermann function. But the totality of that function (although I am certain of it) is far from being *clear* in a sharp sense of the term.

It seems to me that when most people say that the concept of natural numbers is clear they are really talking about this pre-theoretic underlying idea of starting with 0 and iterating successor indefinitely.[11] The questions, then, of the form "Does there exist a natural number with property P?" (say, of coding up a proof of 0=1 from ZFC + "$0^{\#}$ exists") are questions that are not necessarily settled on the basis of this underlying conception.

The situation with paths through the binary branching tree is clear. When people first learn of it they take it to be quite clear. But then (just as

[11] One might even set up $I\Delta_0$ in a schematic form and articulate internal categoricity results to lend substance to the underlying idea.

in the case of number theory) there arise questions of the form "Does there exists a path coding up $0^{\#}$?" that are not necessarily settled on the basis of that underlying conception. It is clear that the simply definable paths exist (just as it is clear that the small numbers that we can count to exist) but as we climb up to stronger and stronger statements these matters become less clear and *that* fact does not (in either case) injure the idea that the initial conception with which we started is quite clear.

4 Conclusion

I have argued that the indirect argument rests on an ambiguity and that once one is clear on the difference between an indefinite statement and an indefinite task the thought experiment does not provide circumstantial evidence that the mathematical community has implicitly endorsed the view that CH is an indefinite statement. I have argued that Feferman's conceptual structuralism is flexible enough to be applied to views that CH is definite and views that CH is indefinite but, on its own, does not make a pronouncement one way or the other. I have argued that in each of the metamathematical arguments provided by Feferman (and many others provided by me and by other people) there is in fact a parallel between the case of first-order and second-order arithmetic (and, more generally, set theory). Finally, in the end, when all the dust settles the entire case rests on the claim that the concept of natural numbers is clear while the concept of arbitrary natural numbers is not clear. Here I side with Markov.

Appendix

The following material is taken from my unpublished paper "Independence in Arithmetic and Set Theory":

The questions motivating this work have to do with the attempt to find analogues of set-theoretic independence in arithmetic. This is best explained in terms of interpretability degrees.

Let us write $T_1 \leqslant T_2$ to indicate that T_1 is (relatively) interpretable in T_2. For all of the theories that we shall consider this is equivalent to saying that the Π^0_1-consequences of T_1 are a subset of the Π^0_1-consequences of T_2.[12] We shall write $T_1 < T_2$ to indicate that $T_1 \leqslant T_2$ and $T_2 \nleqslant T_1$; and we shall write $T_1 \equiv T_2$ to indicate that $T_1 \leqslant T_2$ and $T_2 \leqslant T_1$.

In terms of interpretability there are three possible ways in which a statement φ can be independent of a theory T.

(1) (Single Jump) Only one of φ or $\neg\varphi$ leads to a jump in strength, that is,

$$T + \varphi > T \quad \text{and} \quad T + \neg\varphi \equiv T$$

(or likewise with φ and $\neg\varphi$ interchanged).

(2) (No Jump) Neither φ nor $\neg\varphi$ lead to a jump in strength, that is,

$$T + \varphi \equiv T \quad \text{and} \quad T + \neg\varphi \equiv T.$$

(3) (Double Jump) Both φ and $\neg\varphi$ lead to a jump in strength, that is,

$$T + \varphi > T \quad \text{and} \quad T + \neg\varphi > T.$$

Each of these possibilities is realized. For the first it suffices to take the Π^0_1-sentence $\mathrm{Con}(T)$. (The non-trivial direction ($T + \neg\varphi \leqslant T$) is due to Feferman, building on work of Hilbert and Bernays.) For the second it is easy to see that there is no example that is Π^0_1 and, in fact, that there is no example that is Boolean Π^0_1; however, there are examples that are Δ^0_2; examples of this type of independence are called *Orey sentences*.[13] For the third kind of independence there are Π^0_1 instances. (This is a corollary of Lemma 14 on pages 128–129 of Lindström 2003.)

It is of interest to ask whether there are "natural" non-metamathematical instances of these kinds of independence in arithmetic and set theory. Let us discuss what is known. There is no known example (in either arithmetic or set theory) of the third kind of independence and it seems quite unlikely that there will ever be such an example. So our discussion will concentrate on the first two kinds of independence.

There are many natural examples of the first kind of independence in the set theoretic case. The most natural way to extend PA is by iterating

[12] We shall assume that all of our theories can interpret PA and are essentially reflexive.

[13] In these notes, unless otherwise specified, Δ^0_2 will mean provably Δ^0_2 in PA.

the powerset and thereby arriving at the theories PA_2, PA_3, etc. Continuing this into the transfinite one arrives at ZFC and its extensions by *large cardinal axioms* such as the axioms “There is a strongly inaccessible cardinal” and “There are ω many Woodin cardinals.” These theories provide a natural way of climbing the hierarchy of interpretability. Let us refer to this sequence of theories as the *coarse canonical sequence.* This sequence is canonical in that (a) it is well-ordered under interpretability and (b) it is a remarkable empirical fact that theories throughout the spectrum of theories can be compared (in terms of interpretability) by showing that each is mutually interpretable with a theory on this sequence. However, one problem with this sequence is that it is rather coarse. Even at the lowest level, there is a lot of room between PA and PA_2; for example, one has the theories $\mathrm{PA}+\mathrm{Con}(\mathrm{PA})$, $\mathrm{PA}+\mathrm{Con}^2(\mathrm{PA})$, $\ldots$ and much more.[14] For this reason it is also natural to consider the *fine canonical sequence* by which we mean the sequence of theories obtained by replacing T in the canonical sequence with $\mathrm{PA}+\bigcup_{n<\omega}(T\restriction n)$ and interpolating the finer levels of consistency strength by iterating the consistency statements.

There are two natural questions that arise. First, are there natural arithmetical statements that lie in the interpretability degrees of the theories in the coarse canonical sequence? There are some preliminary results in this direction. To begin with recall that there are some arithmetical instances of the first kind of independence, most notably the Paris-Harrington theorem. Unfortunately, this statement does not lie in an interpretability degree of the coarse canonical sequence. However, Harvey Friedman has initiated a program of finding such statements (even natural Π^0_1-statements) throughout the entire hierarchy of interpretability.

Question 1. *Are their natural arithmetical statements that lie in the interpretability degrees of the theories in the canonical sequence?*

Second, are their natural arithmetical statements that are mutually interpretable with $\mathrm{PA}+\mathrm{Con}(\mathrm{PA})$, $\mathrm{PA}+\mathrm{Con}^2(\mathrm{PA})$, $\ldots$. More generally:

Question 2. *Are there natural arithmetical statements that lie in the interpretability degrees of the theories in the fine canonical sequence?*

Let us now turn to the second kind of independence. Once again there are many natural examples of this kind of independence in the set theoretic

[14] Here $\mathrm{Con}^2(\mathrm{PA})$ is $\mathrm{Con}(\mathrm{PA}+\mathrm{Con}(\mathrm{PA}))$.

case. For example, the statement that all projective sets have a projective uniformization (PU) and the continuum hypothesis (CH) are examples of this kind of independence. However, there are no known natural examples of such Orey sentences in the case of arithmetic.

Question 3. *Are there natural examples of Orey sentences for* PA*?*

This problem appears quite difficult—-it would seem to involve the development of an analogue of set-theoretic forcing for arithmetic and for this we do not even have a clear and definite test question that would signal success.

Going further one can ask for finer parallels with the set theoretic case. For example, although both PU and CH are Orey sentences for ZFC they have additional features. To begin with the are both "flexible" Orey sentences in that they are Orey sentences not just for ZFC but also for ZFC + "There is a strongly inaccessible cardinal" and much stronger theories in the canonical sequence.

Question 4. *Are there natural examples of "flexible" Orey sentences for* PA*?*

Moreover, PU and CH are different in a key respect. For in the case of PU when one climbs the canonical sequence sufficiently high it is actually resolved; more precisely, it becomes resolved positively when one reaches the theory ZFC + "For every $n < \omega$, there are n Woodin cardinals with a measurable above them all". In contrast, CH is not resolved by any large cardinal axioms. namely, as one climbs the hierarchy of interpretability along the "canonical sequence" one eventually reaches axioms that settle PU but this does not appear to be the case for CH. Let us elaborate on this.

Question 5. *Are there natural examples of "flexible" Orey sentences for* PA *that become resolved only when one reaches a certain point in the (extended) canonical sequence?*

Question 6. *Are there natural examples of "flexible" Orey sentences for* PA *that are not resolved by any theory in the (extended) canonical sequence?*

The philosophical interest of these last two questions is this: Con(PA) and other consistency statements have the feature that if one knows that they are independent then one knows that they are true. In contrast, Orey sentences like PU and CH are provably independent of ZFC in a weak theory (assuming Con(ZFC), of course) but this knowledge of independence provides *no insight* *whatsoever* as to whether or not they are true. So one thing we wish to know is whether there are analogues of this phenomenon in arithmetic. For this purpose we do not demand that the arithmetical sentences be natural. Furthermore, it is sensible as a preliminary step to relax the demand of seeking natural statements if one is to hope to make a first step in approaching Questions 3 and 4.

...

Analogue of PD

The standard construction of an Orey sentence for PA (see below) yields a sentence which is settled in PA+Con(PA).[15] We should like an Orey sentence that resembles PU in that it is (a) a "flexible" Orey sentence in that it is not just an Orey sentence for PA but also for much stronger theories in the fine canonical sequence and (b) it is settled at a certain point by a theory in the fine canonical sequence. It turns out that for any specifiable point in the sequence one can find an Orey sentence that is "flexible" below that point and settled at that point. This follows from the following lemma:

Theorem 4.1. *Suppose $\langle \psi_i : i < \omega \rangle$ is a recursive sequence of Π^0_1-statements that are of increasing strength; that is, such that for all $i < j$, $\mathrm{PA} \vdash \psi_j \rightarrow \mathrm{Con}(\psi_i)$. There is a Π^0_2-statement φ such that for all $i < \omega$,*

$$\mathrm{PA} \vdash \text{"}\varphi \textit{ is an Orey sentence for } \mathrm{PA} + \psi_i.\text{"}$$

Proof. Notice that (working in PA) we may assume $\mathrm{Con}(\mathrm{PA} + \psi_i)$ since if this fails then every sentence is trivially an Orey sentence with respect to $\mathrm{PA} + \psi_i$.

We begin by noting that there is a uniform way for constructing an Orey sentence O_T for a recursive extension T of PA: Given T, let O_T be such that

$$\mathrm{PA} \vdash O_T \leftrightarrow \forall n \left(\mathrm{Con}(T{\restriction}n + O_T) \rightarrow \mathrm{Con}(\mathrm{T}{\restriction}n + \neg O_T)\right).$$

This example is due to Lindström. [Notice that O_T is the Gödel sentence for the Feferman provability predicate in T.] Using the fact that $T_1 \leqslant T_2$ if and only if for all n, $T_2 \vdash \mathrm{Con}(T_1{\restriction}n)$ it is straightforward to see that O_T is an Orey sentence for T.

[15] The observations in this and the next section are joint with Hugh Woodin.

Notice that PA + Con(T) also proves O_T: In PA + Con(T) we can prove that O_T is independent of T. Thus, in PA + Con(T) we can prove $\forall n\, \mathrm{Con}(T\restriction n + \neg\mathrm{Orey}(T))$, which implies O_T.

The statement O_T is Δ^0_2 (provably in PA). It is seen to be Π^0_2 by inspection. To show that (in PA) it can also be written as Σ^0_2 one uses the fact that

$$\begin{aligned}\mathrm{PA} \vdash \forall n \left(\mathrm{Con}(T_1\restriction n) \to \mathrm{Con}(T_2\restriction n)\right) \\ \leftrightarrow \mathrm{Con}(T_2) \vee \exists n \left(\neg\mathrm{Con}(T_1\restriction n) \wedge \forall m < n\, \mathrm{Con}(T_2\restriction m)\right).\end{aligned}$$

(See Exercise 12 of Chapter 6 of Lindström 2003).

For a recursive extension T of PA let 'Max(T)' abbreviate

$$\forall n\, \mathrm{Con}(T\restriction n) \wedge \neg\mathrm{Con}(T).$$

Let Tr be the Σ^0_2 truth predicate for Σ^0_2 statements and let Tr$'$ be the Π^0_2 truth predicate for Π^0_2 statements.

By the Diagonal Lemma, let φ be such that

$$Q \vdash \varphi \leftrightarrow \forall i < \omega \left(\mathrm{Tr}(\ulcorner \mathrm{Max}(\mathrm{PA} + \psi_i)\urcorner) \to \mathrm{Tr}'(\ulcorner O_{\mathrm{PA}+\psi_i}\urcorner)\right).$$

The statement φ is Π^0_2.

Now fix $i < \omega$. Work in PA and assume Con($\mathrm{PA} + \psi_i$). We claim that φ is an Orey sentence for $\mathrm{PA} + \psi_i$.

For convenience we shall actually give a model-theoretic proof. It is straightforward (using the Hilbert-Bernays arithmetization of the completeness theorem (as we have above)) to transform this model-theoretic proof into a formal proof in PA.

To show that φ is an Orey sentence for $\mathrm{PA} + \psi_i$ it suffices to show that (i) any model of $\mathrm{PA} + \psi_i$ can be end extended to a model of $\mathrm{PA} + \psi_i + \varphi$ and (ii) any model of $\mathrm{PA} + \psi_i$ can be end extended to a model of $\mathrm{PA} + \psi_i + \neg\varphi$. Let M be a model of $\mathrm{PA} + \psi_i$. Let M' be an end-extension of M such that

$$M' \models \mathrm{PA} + \psi_i + \mathrm{Max}(\mathrm{PA} + \psi_i).$$

Since the antecedents of φ are mutually exclusive,

$$M' \models \forall j\, (j \neq i \to \neg(\mathrm{Max}(\mathrm{PA} + \psi_j)).$$

Thus, the truth of φ in M' or any of its end extensions hinges on whether Orey($\mathrm{PA} + \psi_i$) holds. But since this is an Orey sentence, we can toggle its truth-value by shifting to end extensions of M'. We can thereby toggle the truth-value of φ. □

Analogue of CH

Theorem 4.2. *For each $i < \omega$, there is a Π^0_{i+2}-sentence φ such that for all Π^0_i-sentences ψ*

$$\mathrm{PA} \vdash \mathrm{Con}(\mathrm{PA} + \psi) \to \big(\mathrm{Con}(\mathrm{PA} + \psi + \varphi) \wedge \mathrm{Con}(\mathrm{PA} + \psi + \neg\varphi)\big).$$

Proof. For each $i < \omega$, let '$\mathrm{Prf}_i(x, \ulcorner\varphi\urcorner)$' abbreviate

$$\exists\psi \in \Pi^0_i \left(\mathrm{Tr}_i(\ulcorner\psi\urcorner) \wedge \mathrm{Prf}_{\mathrm{PA}+\psi}(x, \ulcorner\varphi\urcorner)\right),$$

where 'Tr_i' is the Π^0_i truth predicate, and let '$\mathrm{Prf}^{\mathrm{R}}_i(x, \ulcorner\varphi\urcorner)$' abbreviate

$$\mathrm{Prf}_i(x, \ulcorner\varphi\urcorner) \wedge \forall\bar{x} \leqslant x\, (\neg\mathrm{Prf}_i(x, \ulcorner\neg\varphi\urcorner)).$$

Thus, $\exists x\, \mathrm{Prf}_i(x, \ulcorner\varphi\urcorner)$ asserts that φ is provable from some Π^0_i-truth adjoined to PA and $\exists x\, \mathrm{Prf}_i(x, \ulcorner\varphi\urcorner)$ is the associated "Rosser variant" of this statement.

Fix $i < \omega$. By the Diagonal Lemma, let φ be such that

$$Q \vdash \varphi \leftrightarrow \neg\exists x\, \mathrm{Prf}^{\mathrm{R}}_i(x, \ulcorner\varphi\urcorner).$$

Let ψ_0 be Π^0_i. Work in PA and assume $\mathrm{Con}(\mathrm{PA} + \psi_0)$. We claim that φ is independent of $\mathrm{PA} + \psi_0$.

For convenience we shall actually give a model-theoretic proof. It is straightforward (using the Hilbert-Bernays arithmetization of the completeness theorem (as we have above)) to transform this model-theoretic proof into a formal proof in PA.

Step 1. Assume for contradiction that $\mathrm{PA} + \psi_0 \vdash \varphi$. Fix $n < \omega$ such that

$$Q \vdash \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner\varphi\urcorner).$$

Let $M \models \mathrm{PA} + \psi_0$. So $M \models \varphi$ and $M \models \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner\varphi\urcorner)$.

Claim 1. $M \models \forall\bar{n} \leqslant n\, (\neg\mathrm{Prf}_i(\bar{n}, \ulcorner\neg\varphi\urcorner))$.

Proof. Assume for contradiction that this fails; so

$$M \models \exists\bar{n} \leqslant n\, \exists\psi \in \Pi^0_i\, (\mathrm{Tr}_i(\ulcorner\psi\urcorner) \wedge \mathrm{Prf}_{\mathrm{PA}+\psi}(\bar{n}, \ulcorner\neg\varphi\urcorner)).$$

Fix such an $\bar{n} \leqslant n$. Since n is standard, $\bar{n}$ is standard. Thus, M is correct in thinking

$$\mathrm{Prf}_{\mathrm{PA}+\psi}(\bar{n}, \ulcorner\neg\varphi\urcorner).$$

Since

$$M \models \mathrm{PA} + \psi$$

(as φ comes from the Π^0_i oracle) it follows that

$$M \models \neg\varphi,$$

which is a contradiction. $\square$

We now claim that $M \models \neg\varphi$, that is,

$$M \models \exists x \, \mathrm{Prf}_i(x, \ulcorner \varphi \urcorner) \wedge \forall \bar{x} \leqslant x \, (\neg \mathrm{Prf}_i(\bar{x}, \ulcorner \neg\varphi \urcorner)),$$

as witnessed by $x = n$ and $\psi = \psi_0$. We have that $M \models \psi_0 \wedge \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner \varphi \urcorner)$; so $M \models \mathrm{Prf}_i(n, \ulcorner \varphi \urcorner)$. And by the claim we have that $M \models \forall \bar{n} \leqslant n \, (\neg \mathrm{Prf}_i(\bar{n}, \ulcorner \neg\varphi \urcorner))$. Thus, $M \models \neg\varphi$, which is a contradiction.

Step 2. Assume for contradiction that $\mathrm{PA} + \psi_0 \vdash \neg\varphi$. Fix $n < \omega$ such that

$$Q \vdash \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner \neg\varphi \urcorner).$$

Let $M \models \mathrm{PA} + \psi_0$. So $M \models \neg\varphi$ and $M \models \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner \neg\varphi \urcorner)$. In M fix witnesses x and ψ for the existential quantifiers in $\neg\varphi$.

Claim 2. $x < n$.

Proof. We have that

$$M \models \psi_0 \in \Pi^0_i \wedge \psi_0 \wedge \mathrm{Prf}_{\mathrm{PA}+\psi_0}(n, \ulcorner \neg\varphi \urcorner).$$

But

$$M \models \forall \bar{x} \leqslant x \, \neg \exists \psi \in \Pi^0_i \, (\mathrm{Tr}_i(\ulcorner \psi \urcorner) \wedge \mathrm{Prf}_{\mathrm{PA}+\psi}(\bar{x}, \ulcorner \neg\varphi \urcorner).$$

So ψ_0 witnesses that $x < n$. $\square$

Thus, x is standard. To emphasize this let us write $x = \bar{n}$. We have

$$M \models \psi \wedge \mathrm{Prf}_{\mathrm{PA}+\psi}(\bar{n}, \ulcorner \varphi \urcorner).$$

Since $\bar{n}$ is standard, M is correct in thinking

$$\mathrm{Prf}_{\mathrm{PA}+\psi}(\bar{n}, \ulcorner \varphi \urcorner).$$

Since

$$M \models \mathrm{PA} + \psi$$

it follows that

$$M \models \varphi,$$

which is a contradiction. $\square$