Evidence and the hierarchy of mathematical theories

Charles Parsons

It is a well-known fact of mathematical logic, by now developed in considerable detail, that formalized mathematical theories can be ordered by relative interpretability, and the "strength" of a theory is indicated by where it stands in this ordering. Mutual interpretability is an equivalence relation, and what I call an ordering is a partial ordering modulo this equivalence. Of the theories that have been studied, the natural theories belong to a linearly ordered subset of this ordering.

In a lecture in 1934, when the study of these matters was in its infancy, Paul Bernays gave a classic discussion of different foundational standpoints that ordered them roughly according to the degree of strength of the "platonist" assumptions they embodied.[1] This does not coincide anything like exactly with the hierarchy based on interpretability, as is shown by the fact, which was a quite new result when Bernays mentioned it in his lecture, that classical first-order arithmetic PA is interpretable in intuitionistic HA. Since HA is a subsystem of PA, it follows immediately that they are mutually interpretable.

Bernays's scheme was set forth somewhat more systematically by Hao

---

[1] Paul Bernays, "Sur le platonisme dans les mathématiques," *L' enseignement mathématique* 34 (1935), 52-69, reprinted with an introduction by me and an English translation in Bernays, *Essays on the Philosophy of Mathematics*, Wilfried Sieg, W. W. Tait, Steve Awodey, and Dirk Schlimm (eds.), Chicago and La Salle, Ill.: Open Court, forthcoming.

Wang in a paper of 1958,[2] although his use of the term 'platonism' differs in a significant but unacknowledged way from the original of Bernays.

In recent years questions about evidence in mathematics and many questions about realism center on higher set theory. It is an interesting fact that in the great foundational debates of the inter-war period, higher set theory was hardly discussed. This is somewhat ironic in view of the fact that it was Cantor's development of set theory and the application of set-theoretic methods in mathematics that touched off the foundational debates of the early twentieth century.

But especially when one takes in set theory, Bernays' classification does relate to the interpretability hierarchy. In what follows, I will attempt to synthesize his and Wang's version and expand the resulting scheme to take in some considerations that have come to the fore in more recent years, in particular, but not only, attending more to higher set theory. The latter is the major respect in which the Bernays-Wang classification needs to be updated, but not the only one. So the situation is almost opposite to what prevailed in the inter-war period.

In the latter part of the paper, I will discuss some rather general epistemological approaches to mathematics in the light of this hierarchy and specifically of the problems of set theory.

---

[2] Hao Wang, "Eighty years of foundational studies," *Dialectica* 12 (1958), 466-497. Appropriately, this issue of *Dialectica* was a Festschrift for Bernays's 70th birthday.

I

Here is the hierarchy as Wang presents it:

1. *Anthropologism*. This is what is usually called strict finitism, a term that had already been coined by Kreisel. The basic idea is that the only number-theoretic functions that are admitted to be total are those that are feasibly computable.

Bernays had not really proposed or registered this as a foundational standpoint. However, he does question whether it is really intuitively evident that a number described by stacked exponentiation can be represented by an Arabic numeral. He then remarks"

> One could introduce the notion of a "practicable" procedure, and implicitly restrict the import of recursive definitions to practicable operations. To avoid contradictions, it would suffice to abstain from applying the principle of the excluded middle to the notion of practicability.

He insists that he is "far from recommending" such a restriction but is "concerned only to show that intuitionism takes as its basis propositions which one can doubt and in principle do without."[3]

Strict finitism has, of course, been explored in the post-war period, and in particular it is one of the motivations of Edward Nelson's work in foundations, although strict predicativity (see below) is another. Bernays's remarks already

---

[3] "Sur le platonisme," pp. 61-62.

suggest that a strict finitist standpoint will not admit exponentiation as a total function.

2. *Finitism*. This is clearly to be roughly the domain of mathematics intended by the Hilbert school as appropriate for proof theory. There have been differences of opinion both about what the Hilbert school intended and about what would be a proper analysis of finitism in some abstraction from the historical question. But there is agreement to the effect that on either reading it is naturally represented by a free variable formalism and that the strength of this formalism would fall short of, or barely reach, that of PA. I will assume for the sake of argument that the finitistically provable statements of arithmetic are exactly the theorems of PRA, as William Tait argued some years ago.[4] So finitism would be distinguished from strict finitism in allowing arbitrary primitive recursions.

4. *Intuitionism*. On any account of finitism, intuitionism differs from it in allowing logical combinations of quantified statements, so that intuitionistic first-order logic can have the same language as classical. The intuitionistic first-order arithmetic HA that thus arises naturally is, as noted above, mutually interpretable with PA.

Intuitionistic mathematics as understood by Brouwer naturally allows richer means of expression, in order to accommodate analysis. Analysis for

---

[4] W. W. Tait, "Finitism," *The Journal of Philosophy* 78 (1981), 524-546, reprinted in his *The Provenance of Pure Reason* (Oxford University Press, 2005). (See also his "Remarks on finitism," also reprinted in *The Provenance*, and "Appendix to chapters 1 and 2" in that volume.)

Brouwer is closely intertwined with his theory of ordinals, since the "bar theorem" is essential to the proof of one of his fundamental results, that every function everywhere defined on a closed interval is uniformly continuous. Different axiomatizations of intuitionistic analysis exist, but in terms of interpretability all are much weaker than classical analysis (second-order arithmetic). Intuitionistic analysis does interpret the theory of one generalized inductive definition (ID_1) and so is stronger than predicative analysis as understood by Feferman and Schütte.

5. *Predicativism*.[5] For Bernays, an essential mark of "platonism" is accepting the law of the excluded middle in the form that a predicate 'F' is either true of some element of the domain (whatever it is) or false of all of them.[6] The points of view described so far have a reasonable claim to avoid platonism in this sense.

The minimal contentful platonism would admit excluded middle to quantification over natural numbers. If we think of sets of natural numbers as constructed step by step from predicates that define them, then at each level we have only countably many predicates, so that we arguably do not introduce any additional "platonistic" assumptions. This was the point of view of the first attempt to construct a predicative analysis, that of Weyl in *Das Kontinuum* (1918).

---

[5] Wang divides this category into two, but so far as I can see his reason for doing this is to set forth certain technical results, and we have no reason to follow him. His division does not correspond to anything in Bernays.

[6] "Sur le platonisme," pp. 52-53.

The version of predicative analysis that would be easiest to understand would be a ramified second-order theory. Exploration of the limits of predicativity led to the study by several logicians of extensions into the transfinite of the hierarchy of levels. The widely accepted analysis of predicative provability given by Feferman and Schütte implies that if a statement is predicatively provable, it is provable in a ramified theory with levels less than a certain recursive ordinal Γ_0. This analysis was shown (largely by Feferman) to be robust, in that no additional power is obtained by working with unramified theories or even with a predicative system of set theory.

The theories studied took the natural numbers, with full induction, as given. One can characterize what is analyzed as predicativity given the natural numbers.

Wang's last category is "platonism," but here he departs importantly from Bernays, for whom one can be platonist about one domain of mathematics but not about another. However, Wang does follow Bernays in an important respect, in that for both the contrast is with constructivism rather than with nominalism. Neither would use the term in such a way that any theory that admits abstract objects at all is to be called platonist,[7] a usage that is common, almost standard, in American philosophy of mathematics. But from now on I will follow Bernays rather than Wang.

---

[7] I believe this usage originated with Nelson Goodman and W. V. Quine, "Steps toward a constructive nominalism," *The Journal of Symbolic Logic* 12 (1947), 105-122. Cf. my introduction to "Sur le platonisme" cited in note 1 above.

6. Platonism about arbitrary sets of natural numbers. This point of view would motivate second-order arithmetic, that is, the extension of PA by allowing full second-order logic. This is a much more powerful theory than appears at first sight; although it does not admit arbitrary sets of real numbers, it can state and prove a lot about definable sets of reals. Full second-order arithmetic is far beyond the reach of proof theory as it now stands.

7. Iterated platonism about sets. Bernays makes only the brief comment:

> In Cantor's theories, platonistic conceptions extend far beyond those of the theory of real numbers. This is done by iterating the use of the quasi-combinatorial concept of a function and adding methods of collection. This is the well-known method of set theory.[8]

II

Clearly, from a more contemporary point of view, some additions need to be made to this scheme. I will mention several.

2.5. *Strict predicativity*. It has been argued by several writers that the concept of natural number is already impredicative. One of them, Edward Nelson, undertook to develop arithmetic in a way that avoids even this impredicativity. As is well known, he undertook to work with theories interpretable in Robinson's theory Q. It appears that the number-theoretic functions that can be seen to be total from this point of view are polynomial-time

---

[8] Ibid., p. 55.

computable and therefore do not include exponentiation. Nelson's standpoint could be offered as a version of strict finitism, although he seems not to object to the use of classical logic.[9]

An alternative approach would be to undertake to develop arithmetic in ramified type theory without the axiom of reducibility, using a ramification of the Dedekind-Frege definition of the natural numbers. This idea was explored some years ago by Thoralf Skolem and more recently by John Burgess and Allen Hazen. It would lead to the conclusion that exponentiation is admissible from a strictly predicative point of view. But their work makes it unlikely that iterated exponentiation is strictly predicative.[10]

In an earlier survey on the subject, basing myself on the work of Burgess and Hazen, I suggested that Nelson's criterion is too narrow.[11] However, I believe that he would argue, using a result of Bellantoni and Cook, that exponentiation is impredicative in a way that weaker primitive recursions are not.[12]

---

[9] Edward Nelson, *Predicative Arithmetic* (Princeton University Press, 1986).

[10] John P. Burgess and Allen Hazen, "Predicative logic and formal arithmetic," *Notre Dame Journal of Formal Logic* 39 (1998), 1-17. Skolem had already shown how to obtain addition and multiplication in this theory; see his *Abstract Set Theory*, Notre Dame Mathematical Lectures 8, 1962.

[11] "Strict predicativity," unpublished paper presented to the Eleventh International Congress of Logic, Methodology, and Philosophy of Science, Cracow, 1999. Cf. *Mathematical Thought and its Objects* (Cambridge University Press, 2008), pp. 304-07. This work will be cited as MTO.

[12] Stephen Bellantoni and Stephen Cook, "A new recursion-theoretic characterization of the polytime functions," *Computational Complexity* 2 (1992), 97-110.

4.5. *Extended intuitionism*. What was described under (4) might better have been called Brouwerian intuitionism. In particular, Brouwer's definition of ordinal provided for a constructive second number class but did not make room for higher number classes. Proof theorists began to develop this idea in the late 1960s, and soon afterward Per Martin-Löf developed his intuitionistic theory of types, a constructive theory that, in its full development, is far more powerful than any developed to formalize Brouwerian intuitionism.

Where does intuitionism thus extended stand in the interpretability hierarchy? Martin-Löf himself has argued recently that there are no constructive principles in sight that would extend the theory so that it would even be able to interpret second-order arithmetic even with comprehension restricted to Π^1_2 formulae.[13] He refers to this situation as a "second failure of the original Hilbert Program, which I cannot interpret in any other way than that we have to give up the dream of being able to establish the consistency of classical mathematics by constructive means" (ibid., p. 254). He sees Michael Rathjen's 1995 result on Π^1_2 comprehension as using an essentially different method, closer to that of set theory. This view is of course controversial. Rathjen himself argues for a bound on possible extensions of Martin-Löf's type theory in terms of set theory.[14] It is

---

[13] Per Martin-Löf, "The Hilbert-Brouwer controversy resolved?" in Mark van Atten et al. (eds.), *One Hundred Years of Intuitionism (1907-2007)* (Basel: Birkhäuser, 2008), pp. 243-256.

[14] Michael Rathjen, "The constructive Hilbert program and the limits of Martin-Löf type theory," *Synthese* 147 (2005), 81-120, §6.

not clear to me how high this bound is.[15]

Writers on foundations very often make a distinction between "ordinary" or "classical" mathematics and set theory or essentially set-theoretic mathematics. It has certainly not been made out that this is a principled distinction. Martin-Löf probably has a generous conception of what classical mathematics includes, since he is impressed by the fact that certain constructive theories (such as Peter Aczel's constructive set theory CZF) become very strong when the law of excluded middle is added. A narrower view might well allow that classical mathematics already falls within the purview of what constructive proof theory has accomplished.

The final extension of the Bernays-Wang scheme would be to make some distinctions within set theory. The strength of set theories beginning with ZF is typically calibrated by large cardinal axioms, for reasons laid out very clearly in lectures by Peter Koellner. Two distinctions suggest themselves: first, the internally set-theoretic distinction between such axioms as are and are not compatible with $V = L$, often expressed as that between "small" and "large" large cardinals, and second, the more epistemological distinction between axioms that are "implied by the concept of set," as Gödel put it, and those about which we have no idea of how to argue that they are so implied.[16] Peter Koellner has

---

[15] Peter Koellner has told me that it is beyond second-order arithmetic, but Rathjen does not explicitly say that in the article cited. Solomon Feferman, in conversation, suggested that Rathjen's bound, like Martin-Löf's, is below Π^1_2 comprehension

[16] At the moment no intermediate category has been identified, but this could change.

offered an argument, based on his results about reflection principles, that the axioms implied by the concept of set are insufficient to yield the existence of a certain small large cardinal. It follows that the second distinction draws a line at a lower point than does the first.

III

Any epistemological story about mathematics should take account of the stratification involved in the Bernays-Wang hierarchy or improvements of it. The interpretability hierarchy, together with the second incompleteness theorem, implies something uncomfortable: A justification of a theory, one would think, should at least show that it is consistent. But then if T' is a theory that can give a mathematical justification of a theory T, T' must be above T in the interpretability hierarchy. It is often said that T' must be "stronger" than T. This is a little misleading, because although T' must be stronger in an important respect, it does not follow that it must be stronger in all respects. We can illustrate this by the sort of proof of the consistency of PA first given by Gentzen. The proof can be carried out by adding to PRA a single instance of induction on a primitive recursive ordering of type ε_0. It is constructive by any reasonable standard, and the assumptions of the proof do not introduce the reasoning with nesting of quantifiers that PA admits (and as HA does as well). Opinions will differ about the reasons for accepting the induction principle involved and how much the proof accomplishes epistemologically. But at least it is clear that there is a lot of baggage that PA carries that PRA plus this induction axiom does not. But I will not go into the question to what extent this is true of proof-theoretic consistency proofs of stronger theories. I have noted above that Per-Martin-Löf has argued that no way is in sight to give such a proof for second-order arithmetic with Π^1_2 comprehension that satisfies the condition of constructivity.

This mathematical constraint would suggest that for higher domains of mathematics we have to deploy less elementary epistemological ideas. This can be illustrated by two suggestions about the epistemological basis of mathematics: intuition more or less on the Kantian model, and some form or other of empiricism.

About the first, I say "more or less on the Kantian model" because for pretty obvious reasons we cannot apply Kant's own ideas too directly. First of all, the interpretation of Kant's philosophy of mathematics and even of his conception of intuition is a complicated matter about which there are important differences of view. Second, whatever comes out of an effort of interpretation applies first of all to the mathematical situation of the late eighteenth century. In particular, Kant held the then common view that geometry is a priori knowledge about space, from which, given the state of geometry at that time, it could be concluded that space is Euclidean.

Through the nineteenth century, however, when writers on mathematics talked of intuition, they had in mind something that shares with Kant's some sort or other of reference to space and time. The insistence of writers such as Frege that arithmetic and analysis should not depend on intuition is against the background of that understanding of what the relevant conception of intuition is. The case of Frege is particularly clear because he maintained the traditional view of geometry as resting for its evidence on spatial intuition.

I argue elsewhere that Brouwer's and Hilbert's conceptions of intuition are also descended from Kant's; they certainly share this spatio-temporal character.[17] Brouwer, of course, insists that the intuition underlying mathematics is purely temporal. Following the model of Hilbert, Brouwer, and Bernays, I have set forth in several writings ideas of intuition and intuitive knowledge applicable to arithmetic.[18] The conception of intuitive knowledge is not as clear as it should be, but it seems to me to track that of the Hilbert school reasonably well. It should follow that what is intuitively known in arithmetic should coincide with finitist arithmetic. That was the opinion of Hilbert, Bernays, and also Gödel. That would imply that in a case of primitive recursion, if the given functions have been seen intuitively to be total, then the function introduced can also be seen to be total. In two writings of Bernays, there are explicit arguments to that

[17] "The Kantian legacy in twentieth century foundations of mathematics," in my *Philosophy of Mathematics in the Twentieth Century* (Cambridge, Mass.: Harvard University Press, forthcoming). The claim that their conceptions of intuition are descended from Kant's is not meant to imply that the philosophy of either is close to Kant's more globally; even in the conception of intuition there are significant differences.

[18] See especially MTO chapters 5 and 7.

effect, one specifically directed at exponentiation and one more general.[19] I was unable to see the validity of these arguments and concluded that Hilbert and Bernays did not have a non-question-begging argument for the conclusion that primitive recursive functions are finitistically admissible.[20] That the obstacle is exponentiation would suggest that on the conception of intuitive knowledge presented in my writings, intuitive arithmetic coincides at least approximately with Edward Nelson's strict finitism.

I will not pursue this direction further, because even if my claim is wrong, no one would argue that the type of intuitive evidence appealed to by Hilbert and Bernays would extend very much further than they claim. It could not be the epistemological basis for all of mathematics except on a very restrictive conception, probably more restrictive, for example, than Brouwerian intitionism. I did undertake to argue that intuition as I analyze it does play a role in getting mathematics off the ground. Even that will be contested and has been by Felix Mühlhölzer, in a review article on my book *Mathematical Thought and its Objects*.[21]

Some form of empiricism, taking its principal inspiration from W. V.

---

[19] The first is in part II §1 of "Die Philosophie der Mathematik und die Hilbertsche Beweistheorie," *Blätter für deutsche Philosophie* 4 (1930), 326=367, reprinted with Postscript in *Äbhandlungen zur Philosophie der Mathematik* (Darmstadt: Wissenschaftliche Buchgesellschaft, 1976), reprinted with the Postscript and a translation in *Essays* (see note 1), the second in Hilbert and Bernays, *Grundlagen der Mathematik* I (Berlin: Springer, 1934, 2d ed. 1968), pp. 25-27.

[20] See MTO §44. I have heard a lecture by William Tait in which he challenged this conclusion, but I have not seen his argument in writing. Tait's own analysis of finitism is quite different in that it does not rely on any notion of intuitive knowledge.

[21] Felix Mühlhölzer, "Mathematical intuition and natural numbers: A critical discussion of Charles Parsons, *Mathematical Thought and its Objects*," *Erkenntnis* 73 (2010), 265-292. I have not yet undertaken to reply to this criticism.

Quine, has almost dominated the conversation in American philosophy of mathematics. The Vienna Circle's attempt to reconcile empiricism with the apparently a priori character of mathematics, even in the subtle form given to it by Carnap, has in more recent years not found wide acceptance, in spite of the more sympathetic view that has been taken of Carnap's philosophy more generally. At any rate, I don't personally feel enough at home with the issues it poses to assess it in this context. Much more influential has been the holistic form of empiricism championed by Quine. That is the point of view underlying the much discussed "indispensability argument": Accepted mathematics must be true, and the objects it refers to must exist, because mathematics is indispensable for the natural sciences. Before one enters into any details, the question immediately arises how much mathematics is indispensable for science. In particular, Solomon Feferman has argued that the mathematics that is applicable in science can be formulated in a basically predicative theory. It has also been argued that the indispensability argument is at odds with the practice of mathematicians, who do not seek a minimal theory as the indispensability argument suggests that they should.

We do not need to pursue these issues, because it is hard to see how, at least at the present stage in the development of science, considerations of indispensability could lend justification to much of higher set theory. Considerations of smoothness of theory might well lead to a significantly stronger theory than what Feferman describes. But why should it lead to an inexhaustible ascent into higher and higher reaches of the cumulative hierarchy? Gödel speculated years ago that strong axioms of infinity might find application in science, but the actual development has not led in this direction. It would seem more natural from this point of view to close off the ascent at some point high enough for any scientific needs and still yielding a reasonably simple theory. Quine indeed suggests this in at one point advocating adopting $V = L$ as an axiom.

The conclusion I draw from these rather hastily sketched considerations is that some level of rationalism is unavoidable in the epistemology of mathematics, at least if we admit that what is justified includes a significant amount of higher set theory. Once one admits that, there is at the very least a gain in intuitive naturalness in admitting it short of higher set theory, even at the beginning of arithmetic.

All modern analysis of mathematical proof gives logical inference a central role, and the rules of first-order logic (with the possible exception of those that distinguish intuitionistic logic from classical) seem to many thinkers even to be self-evident.[22] Quine, without making that exception, regards them as "obvious."[23] He refuses to give any philosophical weight to that fact, but he does

---

[22] Even what are called logic-free formalisms involve some logic, reasoning with identity and with free variables understood as expressing generality.

[23] I don't think that Quine means that they are self-evident, at least in the sense I propose (see below).

not consider a caution voiced many years ago by Brouwer, Weyl, and Hilbert. I would put the matter in the following way: Logic is entangled with mathematics. Although first-order logic presupposes no ontology beyond the trivial presupposition of the nonemptiness of the domain of quantification, to see that an inference is *not* valid, or that a sentence is *not* logically true, one may need to construct countermodels where the domain may be arbitrarily large, up to countable infinity. Of course one may argue that a sentence is not logically true because it is not derivable, appealing to the completeness theorem. But a proof of the completeness theorem typically starts with the assumption that a formula is consistent (i.e. not refutable) and then indicates how to construct a model, which again may be countably infinite. What Brouwer, Weyl, and Hilbert argued is that the applicability of classical logical inference in reasoning about the infinite is an assumption that can be contested. Hilbert, probably following Weyl, pressed the point further than Brouwer, in that his statements suggest that the language of the most unproblematic reasoning in arithmetic would not allow constructing and reasoning with sentences involving nesting of quantifiers. The falsity of a generalization could be stated only by giving an explicit instance.

It is natural to think that if anything assumed in mathematics is rationally evident, basic logic must be. I think the critical case made by Brouwer, Weyl, and Hilbert makes it reasonable not to regard it as *self*-evident. And the entanglement of logic and mathematics implies that it does depend on the mathematical context. However, one can hardly doubt that mathematics based on classical logic has stood the test of time. Constructive mathematics has its own interest, but no case not based on some first philosophy has been made for the claim that it deserves to displace classical mathematics. A "potentialist" view of the universe of sets has been taken to imply that some form of constructive logic should be used in statements about absolutely all sets. There is a case here, but I don't think it implies that we can't safely use classical logic even with unbounded quantification in the practice of set theory.

The last paragraph should leave the reader confused: In what sense are classical logical inferences, in a mathematical context such as number theory or analysis, rationally evident? There is an intrinsic difficulty here, not with the evidence of logic itself but with disentangling two different grounds on which it can be said to be evident. The idea of self-evidence is usually taken to imply that if a proposition is self-evident, then it is evident by itself. I tried to capture this by proposing that a statement is self-evident if it is evident to anyone who exercises a sufficiently clear understanding of what it means.[24] Almost any claim of self-evidence can be disputed, but there are cases from logic that are as good as any, such as 'if A and B, then A' or the inference from 'if A then B' and 'if A then not-B' to 'not-A'. It will be claimed that instances of 'A or not-A' are self-evident if one understands them in their classical meaning. That is a way in

---

[24] MTO p. 320.

which claims of self-evidence can be contested, because an intuitionist will claim that in cases involving the infinite, the "classical meaning" is itself based on illusion.

There are, however, weaker degrees of rational evidence than self-evidence. Perhaps the weakest of all is what I have called "intrinsic plausibility." A statement is treated as intrinsically plausible if one is prepared to give it a significant amount of credence when it is not the conclusion of an argument or prompted by an external event. The latter condition implies that perceptual judgments are not intrinsically plausible, no matter how evident they may be, because their plausibility or evidence is not *intrinsic*.[25] It is reasonable to think that in some cases, particularly in mathematics, intrinsic plausibility is objective, so that it is at least in principle possible to bring about agreement about cases. Most of us would regard the axioms of classical mathematical systems as intrinsically plausible to quite a high degree. When it comes to set theory, even ZFC, the matter is a little more contested; for example I doubt that Solomon Feferman would admit that, and he is a distinguished mathematician, by no means a marginal figure. (Let us leave out figures from the past such as Brouwer.)

There is, however, a difficulty about self-evidence and its weaker cousins that is suggested by holistic views of evidence and meaning. It is posed most

---

[25] The same seems to be the case of propositions of whose truth one learns by testimony, which even for a professional mathematician includes a lot of mathematics.

sharply if we consider elementary logic and arithmetic. These branches of knowledge have had elaborate deductive development and application for centuries. We find the axioms and elementary inferences evident. But why should we be sure that their evidence is *intrinsic*? When we declare them evident, are we not appealing to this long history of development and application? Shouldn't we rather speak, as Paul Bernays suggests in some of his later writings, of *acquired* evidence? One might go further and propose, with Quine, that we cannot be sure that the development of science will not give rise to reasons for regarding the most elementary and apparently evident propositions as false.

I think there are general, one might say phenomenological, grounds for thinking of the most elementary evidences in arithmetic as intrinsic or at least as resting on significant intrinsic plausibility. One is that such steps as passing from a natural number to the next are clear to schoolchildren who have little idea of the extent to which arithmetic has been developed in the course of history and only a quite restricted idea of its application.[26] Another is the conviction that arithmetical truths are necessary, so that contingencies of the natural world or of history will not upset them. In the case of mathematical induction, when mathematicians first began to reason in a way that we would render as an

---

[26] However, one could argue that something like intuition in the sense of chapter 5 of MTO is playing a role here; I suggest that intuitive evidence based on such intuition is a borderline case with regard to intrinsicness (ibid.., p. 322). But the problem is its relation to perception rather than the holistic considerations discussed in the text.

application of induction, there simply was not a background of deductive development from a principle of induction or a least number principle. It is hard to see on what grounds other than intrinsic plausibility mathematicians could at that time have accepted inferences by induction.[27] Zermelo claimed the same about the axiom of choice when he used it to prove that every set can be well-ordered and argued that certain earlier proofs involved unacknowledged appeals to the axiom.

I do not intend to deny that once a mathematical theory has been developed and applied, its assumptions come to have a higher degree of evidence, perhaps even a much higher degree. Bernays's idea of acquired evidence can apply even to elementary arithmetic. But if that is true at any given point in history, the possibility exists that its degree of evidence will increase. That presumably means, however, that it is also possible that it will decrease or be undermined, as would be the case, for example, if Edward Nelson's claim to prove the inconsistency of PRA should be sustained.[28]

What is the situation when we come to set theory? Matters are confused

---

[27] More implicit inferences that we would render as appealing to induction surely existed much earlier. I don't think that changes the case.

[28] Nelson's scenario is evidently internal to mathematics, thus different from the Quinean speculation that natural science might develop in such a way as to falsify some part of accepted mathematics. Either version of the case is, however, reminiscent of the dispute in epistemology in the middle of the twentieth century as to whether a perceptual judgment made under optimal conditions is still at best highly probable, so that it is at least possible that evidence might accumulate that would show the judgment to be false. The latter view was held by more classical epistemologists such as C. I. Lewis, but it was questioned by "ordinary language" philosophers, particularly J. L. Austin.

by the fact that a wide variety of considerations, of very different kinds, have been offered to justify or motivate axioms of set theory. This situation is rather different from that with arithmetic, where the picture is much simpler. That raises the question whether any axiom of set theory is intrinsically plausible to more than a small degree, since arguments are offered for all of them. The arguments, however, are rarely intended to be convertible into mathematical proofs; after all, that would undermine the status of the propositions involved as axioms. Those offered in the context of the iterative conception of set seem generally intended to show that the axioms fit into the general picture of that conception, in other words to persuade us that, given that conception, we should accept the axioms as implicit in it or at least in harmony with it. Something like that was very likely intended by Gödel when he suggested that certain proposed axioms are "implied by the concept of set" and others can at least at the present stage of knowledge not be said to be so implied.

Basing themselves on remarks of Gödel, writers on set theory have distinguished between "intrinsic" and "extrinsic" justifications. A common view is that the axioms of ZF and small large cardinal axioms have an intrinsic justification, while we do not have an idea for such a justification of large large cardinal axioms. Those that are intrinsically justified are those that, in Gödel's language, are "implied by the concept of set." This appears to be a form of analyticity, closely related to "truth by virtue of meaning," one of the informal explanations of analyticity that have circulated since the heyday of logical empiricism.

For the logical empiricists, the importance of the claim that logical and mathematics are analytic was that it was a way of explicating the idea that they are tautologies, "empty," or "without factual content," so that they could play their necessary role in the system of science but not come into conflict with a generally empiricist view, in particular removing any motivation either for the compromise with rationalism we have found necessary or for appeal to intuition on the Kantian model. An interesting feature of Gödel's reflections on mathematics is that in his earlier philosophical writings he is highly critical of the positivist point of view but nonetheless maintains that there is an important sense in which mathematical propositions are analytic.[29] the relevant sense is the second of two senses he discusses near the end of 'Russell's mathematical logic" (hereafter RML):

> In a second sense a proposition is called analytic if it holds "owing to the meaning of the concepts occurring in it," where this meaning may perhaps be indefinable (i.e. irreducible to anything more fundamental). (CW II 139)

There seems to be a doubling in speaking of the "meaning of the concepts" in a proposition. However, better formulations occur in later writings. Thus in the

---

[29] I discuss this aspect of Gödel's thought in "Quine and Gödel on analyticity," in Paolo Lombardi and Marco Santambrogio (eds.), *Ôn Quine: New Essays* (Cambridge University Press, 1995), pp. 297-313. I believe there is more to be said on the matter.

Gibbs lecture he says that a mathematical proposition "is true already according to the meaning of the terms in it, irrespective of the world of real things" (CW III 320), and in version V of "Is mathematics syntax of language?" (hereafter Syntax), he says, "Mathematical propositions are true in virtue of the *concepts* occurring in them" (CW III 357, emphasis Gödel's).

Already in RML Gödel indicates that it is his conceptual realism that makes it possible for him to maintain that the axioms of mathematics (in the context of RML, those of *Principia* other than Infinity) are analytic in this sense, while mathematics consists of statements with a "real content," contrary to the view that the Vienna Circle aimed at. If a statement is true by virtue of certain concepts, then the existence (and perhaps certain relations) of these concepts will itself be a substantial assumption.

In the Gibbs Lecture of 1951, Gödel says that a mathematical proposition is "true by virtue of the meaning of the terms in it" (CW III 320) but makes clear that this meaning consists in certain concepts, which form "an objective reality of their own, which we cannot create or change, but only perceive and describe" (ibid.). A mathematical proposition "still may have a very sound objective content, insofar as it says something about the relations of concepts" (ibid.). The way this is manifested is that axioms of mathematics are not tautologies (or analytic in a narrower sense he had proposed in RML) but still follow from the meaning of the primitive terms. But he also suggests a relation with his ideas on mathematical intuition, which he develops only in later writings, beginning with "Is mathematics syntax of language?" It is rather clear that essential to mathematical intuition as he conceives it is perception of concepts. He regards it as a presupposition of an instance of analytic truth that the concepts involved "exist." Gödel treats predicates as having concepts as their referents, so that some level of perception of concepts is involved in understanding language.[30] But he makes clear that one can perceive concepts more or less clearly, and some examples he cites, such as Turing's analysis of computability, suggest that mathematical analysis and proof can lead to clearer perception of a concept. If the axioms of set theory "force themselves on us as being true," as Gödel claimed in the 1964 version of "What is Cantor's continuum Problem?" (CW II 268; this paper is cited as WCCP), then this would be described as our perceiving the concept of set with enough clarity; it is that that he has in mind in speaking of "something like a perception of the objects of set theory" (ibid.). He describes this perception as mathematical intuition. That suggests the conjecture:

> According to Gödel, a mathematical proposition is derivable from axioms knowable by intuition if and only if it is analytic.

There are a number of questions about Gödel that this conjecture suggests. A consideration that speaks against it is the fact that analyticity is not mentioned as such in the 1964 version of WCCP, the publication of his where intuition is

[30] Not just mathematical language, as examples in Syntax III indicate.

mentioned most prominently, and analyticity is not mentioned in Hao Wang's record of his conversations with Gödel.

But the indications that analyticity is still present in 1964 make the connection between Gödel's version of that notion and the notion of intrinsic justification as deployed in recent writings. Earlier in the paper he observes that the "very concept of set" on which the axioms of set theory (by which he means ZF or ZFC) are based suggests "their extension by new axioms which assert the existence of still further iterations of the operation 'set of'" (CW II 260). The examples he cites are inaccessible and Mahlo cardinals. Summarizing he says

> These axioms show clearly not only that the axiomatic system of set theory as used today is incomplete, but also that it can be supplemented without arbitrariness by new axioms *which only unfold the content of the concept of set explained above*. (CW II 260-61, emphasis mine)

By the latter concept he means the iterative conception of set, which, as I have said, is better called the iterative conception of the universe of sets.[31] It is in a footnote at that place that Gödel mentions stronger axioms of infinity, evidently those postulating large large cardinals, and remarks, "That these axioms are implied by the general concept of set in the same sense as Mahlo's has not been made clear yet" (CW II 260 n. 20).

[31] Cf. Donald A. Martin, "Gödel's conceptual realism," *The Bulletin of Symbolic Logic* 11 (2005), 207-224. Martin explores in some detail how Gödel is using the phrase "concept of set" in different passages in his writings.

Gödel's language suggests that he thinks they must *be* "implied by the general concept of set." But what we know about his thought after 1964 does not indicate either that he felt closer to making this clear or was especially optimistic about the prospect that others would accomplish this. This may be one of the reasons why analyticity does not appear in the record of the conversations with Wang. Already in RML, Gödel had suggested the idea of extrinsic justification of axioms, and later writers on set theory seem to rely on extrinsic justification for large large cardinal axioms.

I will make some comment on that option later, but first I want to comment on the notion of intrinsic justification as it has been deployed after Gödel. In the footsteps of Quine, one might ask how clear the distinction is between what is implied by the concept of set (i.e. the iterative conception) and what is not, perhaps is merely a reasonably obvious truth about the universe of sets, which may be obvious in virtue of "acquired evidence" in the sense of Bernays. Even at lower levels, there have been differences of opinion of this form; thus in 1971 George Boolos argued that the axiom of replacement does not follow from the iterative conception of set.[32] This is a somewhat simple example of a feature of a lot of earlier discussions of the iterative conception, where there was not agreement about whether it included aspects that are in fact central to

---

[32] "The iterative conception of set," *The Journal of Philosophy* 68 (1971), 215-231, reprinted in *Logic, Logic, and Logic* (Cambridge, Mass.: Harvard University Press, 1998). Later, Boolos expressed doubt about the existence of sets postulated by the axiom. I think Boolos's original claim came from a too narrow view of what the conception contains, and his later skepticism could be put in Gödelian terms as a doubt about the "existence" of the concept.

our understanding of set theory, in particular the inexhaustibility of the ascent in the cumulative hierarchy, which underlies considerations of "limitation of size," which earlier writers on the iterative conception regarded as independent.[33] Gödel's cryptic phrase "concept of set" has given rise to confusion but is naturally given the more expansive interpretation.

I will not pursue further this issue about the content of the conception but raise another question, again suggested by Gödel. Let us suppose that we have a convincing argument that a certain axiom is implied by the concept of set. Does that imply that it is true? Gödel regards it as a substantial issue whether the concept *exists*. I don't think that means merely that the terms we take to express the concept are not nonsense, that we are able to understand talk of sets in the language of set theory and make sense of the idea of the cumulative hierarchy. We are, after all, able to understand the talk of extensions (more generally, courses of values) in Frege's logic, and Frege was able to carry out quite a lot of precise reasoning with this concept. But Frege's logic proved to be inconsistent. So that propositions are "implied by the concept of extension" was not sufficient to insure the truth of Frege's Basic Law V or of many of the consequences he drew from it.

I believe that Gödel included under the existence of a concept (at least in the context of his remark in RML), there being something that answers to it, in

[33] Penelope Maddy's instructive discussion of the axioms of ZF seems to maintain this earlier understanding, but she does much to sort out the different considerations offered in favor of individual axioms. See *Naturalism in Mathematics* (Oxford: Clarendon Press, 1997), I, ch. 3.

the language of D. A. Martin, that the concept is instantiated. Martin himself is not without doubt as to whether the concept of set in the relevant sense *is* instantiated. He takes seriously an argument based on Zermelo's quasi-categoricity theorem of 1930 to the effect that there is at most one structure that instantiates the full concept of set. But in that structure CH must be true or false. So this consideration implies that CH has a determinate truth-value, a conclusion of which Martin is not convinced. So he considers it a real possibility that the concept is not instantiated.[34]

Martin has, I believe, probed Gödel's attempt to combine realism with analyticity more than have other writers. A conclusion I would draw is that being "implied by the concept of set" still leaves questions about what Gödel, rather misleadingly, called the "existence" of the concept, what in Martin's language is the concept's being instantiated.

However, Martin evidently thinks of the concept as involving two maximality properties: at each level having the full power set, and having an absolute infinity of levels.[35]

A weaker condition would be consistency or coherence. I would understand the latter as something more than consistency, although exactly what

[34] Cf. "Gödel's conceptual realism," p. 221. In correspondence in August 2010, Martin wrote that he was "open-minded" about the question whether CH has a determinate truth-value.

[35] Martin does not believe that it is especially important for pure mathematics that the concept of set should be instantiated. That implies a somewhat different view about the significance of the existence of objects from what I myself have advocated and indeed from what is typical in the literature, maybe especially that part that takes nominalism seriously.

the more is has not been made clear either by me or by others. Structuralists, including myself, have suggested that coherence of the theory of a structure should be sufficient for being in a position to make true statements about objects in the structure. What is the "more than consistency" that is demanded for coherence? I am not at all sure, but one suggestion is that it should be the consistency of extensions of the theory that are motivated in a natural way. If we take the simpler case of the theory of natural numbers, we would expect the consistency to be preserved by allowing instances of induction not formulable in the language of first-order arithmetic. This is a little tricky, because such instances will typically involve notions with their own theoretical baggage. An inconsistency could arise for that reason; evidently the coherence of the notion of natural number would have to imply that, if that does happen, the inconsistency is traceable to the additional notions. But what is said here about coherence is sketchy in a fundamental way, because of the requirement that extensions be "motivated in a natural way." Beginning with ZF, we might make the case that some extensions are "implied by the concept of set," but to all appearances that falls short of being the case for large large cardinal axioms, which are certainly an integral part of set theory as currently practiced.

Consistency is a syntactical notion, and in their deductive development, second-order theories do not differ essentially from first-order theories. If my suggestion about coherence is on anything like the right track, the coherence of set theory does not imply that the "concept of set" as Martin understands it is instantiated. Furthermore, it is not evident a priori that there could not be different coherent set theories, possibly offering different verdicts concerning CH. On the structuralist view, we would be entitled to assert existence of the sets that such a theory gives rise to. I will leave for later the question of the implication of this view for the question whether CH has a determinate truth-value.

We have emphasized the problem that various conclusions might be implied by a concept such as that of the universe of sets, while it seems that, for us to accept these conclusions as true, something more is required, either instantiation as Martin understands it or coherence of the resulting theory. That might suggest that on this conception of intrinsic justification, it is rather badly off. I would suggest that things are not as bad as they seem, because in practice the working out of such justifications clarifies the concepts in such a way as to make at least coherence more likely.

What is the relevance to this discussion of the ideas about rational evidence that we discussed earlier? The connection Gödel suggests between mathematical intuition and analyticity would suggest that, if we see that something is implied by the concept, it should be rationally evident, probably intrinsically so. However, putting things in that way makes the conception excessively foundationalist, in particular probably more so than Gödel would have countenanced. But if we allow that intuition according to Gödel is fallible, since our perception of the concepts involved might be more or less clear, then if we replace 'evident' by 'plausible to a high degree', we obtain a more defensible stance. But in a particular case there may be a lot of working out of consequences, and connections made with reasoning before this particular theory was formulated. The idea of power set is built into the iterative conception, so that doubts one might have about that axiom get shifted onto the issue of instantiation. I'm not sure that replacement can be convincingly argued to be implied by the concept. Elsewhere I propose that both axioms rest to a significant extent on considerations that would belong to extrinsic justification (MTO §23).

IV

I don't propose to go into extrinsic justification either fundamentally or at any length. I will limit myself to a brief comment on an instance that many set theorists consider an outstanding instance of such justification. That is the satisfying completion of the theory of projective sets of real numbers that was obtained by assuming Projective Determinacy (PD), the deduction of PD and its generalization $AD^{L(R)}$ from large cardinals, and the construction of models for the large cardinals involved from $AD^{L(R)}$. Many more results have been cited to reinforce this picture. If very strong axioms of infinity can be justified by the theory that follows from them, then it is hard to think of a more convincing instance.

In the discussion after his talk in this series, Solomon Feferman offered an objection that I think is worth discussing. No doubt alluding to Gödel's remark that a mathematical theory might be confirmed as well as a well-established physical theory, Feferman claimed that in this situation there is no analogue of observation; one just had a lot of connections among statements at a rather high theoretical level.[36]

There is one reply that a set theorist might very well make. The results in descriptive set theory following from PD include generalizations of theorems that the classical descriptive set theorists had proved by arguments that can be rendered in ZFC, for example that analytic sets are Lebesgue measurable, have the property of Baire, and have the perfect set property (and thus cannot be counterexamples to CH). These classical theorems could be treated as analogues of observations, in that they are results that mathematicians at large accept and have accepted for some time. Although there is some merit in this reply, it would not persuade Feferman, in view of his general skepticism about higher set theory.

But let us concede for the sake of the argument that there is no analogue of observation that is applicable in this case, except totally non-controversial statements to the effect that a certain proposition is provable from certain

[36] I can't be sure I have got Feferman's point right. I would be happy to be corrected, but in any case my remarks will apply (again correctly or not) to what I take Feferman to have been arguing.

premises. Epistemologists have often argued that claims might be justified on the basis of coherence, where the sense of that word is not quite what it was in the last section; it means rather that a body of putative knowledge "hangs together," not just internally but also in its relation to other knowledge and possibly plausible conjectures. A pure holism about the justification of theories would be an example of such justification by coherence. An example is what Quine sketches in the last section of "Two dogmas of empiricism," which does not include his later introduction of observation sentences more directly keyed to stimulation. Even with that, there is a lot of coherence left in Quine's view, because of his claim of the underdetermination of theory by even possible observation.

We could think of the set theorist as starting with ZFC and noting that the theorems the classical descriptive set theorists tried and failed to prove are not provable in ZFC, naturally assuming it consistent. One might gradually develop the picture of projective sets yielded by the much more powerful axioms, at each stage noting that the picture "hangs together" in the required way.[37]

V

What, if anything, does the above discussion indicate about whether CH has a determinate truth-value, or whether the same is true of other statements

---

[37] I think it might be argued that the justification offered for both the assumptions (PD and large cardinals) and the picture of descriptive set theory that results fit John Rawls's conception of reflective equilibrium. But I have not yet seen how to work this out.

that might be claimed not to have one? I remarked that on a certain structuralist view extending what I have defended at some length in previous writings,[38] one cannot conclude a priori that CH or similar statements must be determinate.

Apart from that particular view, I don't see any way of arguing a priori that CH must have a determinate truth-value. Gödel appeals to his own realistic views, I believe going beyond the modest realism that many hold about mathematics toward a transcendental realism. It might be worth recalling that when Cohen first proved the independence of CH, many were inclined to view the situation as like that in geometry after the discovery and development of non-Euclidean geometries. Probably a picture like that has not been clearly refuted in the case of CH, but research in set theory has surely shown that it cannot be adopted too quickly, so that every independence result leads to a bifurcation of the concept.

What has in fact convinced us that a question posing conceptual difficulties, as the continuum problem undoubtedly does, has a unique determinate answer has been actually finding the answer by a combination of mathematical proof and conceptual analysis. That is what, arguably at least, has been accomplished in the case of descriptive set theory sketchily discussed in the last section. It is what Hugh Woodin has been attempting for some years now in the case of CH. One conclusion from these facts should be noncontroversial: It is

---

[38] See MTO chs. 2-4 and "Structuralism and metaphysics," *Philosophical Quarterly* 54 (2004), 56-77.

not for mere philosophers to answer these questions, not only the direct question of CH but the metaquestion whether it has a determinate truth-value.