Lieber Herr Bernays!, Lieber Herr Gödel!

Gödel on finitism, constructivity and Hilbert's program

Solomon Feferman

1. Gödel, Bernays, and Hilbert.

The correspondence between Paul Bernays and Kurt Gödel is one of the most extensive in the two volumes of Gödel's collected works devoted to his letters of (primarily) scientific, philosophical and historical interest. It ranges from 1930 to 1975 and deals with a rich body of logical and philosophical issues, including the incompleteness theorems, finitism, constructivity, set theory, the philosophy of mathematics, and post-Kantian philosophy, and contains Gödel's thoughts on many topics that are not expressed elsewhere. In addition, it testifies to their life-long warm personal relationship. I have given a detailed synopsis of the Bernays Gödel correspondence, with explanatory background, in my introductory note to it in Vol. IV of Gödel's *Collected Works*, pp. 41-79.[1] My purpose here is to focus on only one group of interrelated topics from these exchanges, namely the light that it—together with assorted published and unpublished articles and lectures by Gödel—throws on his perennial preoccupations with the limits of finitism, its relations to constructivity, and the significance of his incompleteness theorems for Hilbert's program.[2] In that connection, this piece has an important subtext, namely the shadow of Hilbert that loomed over Gödel from the beginning to the end of his career.

---

[1] The five volumes of Gödel's *Collected Works* (1986-2003) are referred to below, respectively, as *CW I, II, III, IV* and *V*. *CW I* consists of the publications 1929-1936, *CW II* of the publications 1938-1974, *CW III* of unpublished essays and letters, *CW IV* of correspondence A-G, and *CW V* of correspondence H-Z. References to individual items by Gödel follow the system of these volumes, which are either of the form *Gödel 19xx* or of the form **Gödel 19xx* with possible further addition of a letter in the case of multiple publications within a given year; the former are from *CW I* or *CW II*, while the latter are from *CW III*. Thus, for example, *Gödel 1931* is the famous incompleteness paper, while *Gödel 1931c* is a review that Gödel wrote of an article by Hilbert, both in *CW I*; *Gödel °1933o* is notes for a lecture, "The present situation in the foundations of mathematics," to be found in *CW III*. Pagination is by reference to these volumes, e.g. *Gödel 1931, CW I,* p. 181, or simply, *CW I*, p. 181. In the case of correspondence, reference is by letter number and/or date within a given body of correspondence, as e.g. (Gödel to Bernays) letter #56, or equivalently 2 Dec. 1965, under Bernays in *CW IV*. When an item in question was originally written in German, my quotation from it is taken from the facing English translation. Finally, reference will be made to various of the introductory notes written by the editors and colleagues that accompany most of the pieces or bodies of correspondence.

[2] William Tait, in his forceful and searching essay review (2006) of *CW IV* and *V,* covers much the same ground but from a different perspective; see, especially, sec. 6 below.

Let me explain. Hilbert and Ackermann posed the fundamental problem of the completeness of the first-order predicate calculus in their logic text of 1928; Gödel settled that question in the affirmative in his dissertation a year later.[3] Also in 1928, Hilbert raised the problem of the completeness of arithmetic in his Bologna address; Gödel settled *that* in the negative in 1930 in the strongest possible way by means of his first incompleteness theorem: no consistent formal axiomatic extension of a system that contains a sufficient amount of arithmetic is complete. Both of these deserved Hilbert's approbation, but not a word passed from him in public or in writing at the time. In fact, there are no communications between Hilbert and Gödel and they never met. Perhaps the second incompleteness theorem on the unprovability of consistency of a system took Hilbert by surprise. We don't know exactly what he made of it, but we can appreciate that it might have been quite disturbing, for he had invested a great deal of thought and emotion in his finitary consistency program which became problematic as a result. There is just one comment, of a dismissive character, that he made about it four years later; I will return to that in the following.

The primary link between Gödel and Hilbert was Bernays, Hilbert's assistant in Göttingen from 1917 to 1922 and then his junior colleague until 1934 when Bernays was forced to leave Germany because of his Jewish origins. It was in this period that the principal ideas of Hilbert's consistency program and of his *Beweistheorie* to carry it out were developed, and later exposited in the two volume opus by Hilbert and Bernays, *Grundlagen der Mathematik*, whose preparation was due entirely to Bernays. It was Bernays who first wrote Gödel in 1930 complimenting him on the completeness theorem for first order logic and asking about his incompleteness theorems. And it was Bernays who did what apparently Hilbert did not, namely puzzle out the proofs and significance of the incompleteness theorems through the landmark year of 1931. And then it was Bernays for whom the results were later decisive in the preparation of volume II of Hilbert and Bernays.

[3] Hilbert introduced first order logic and raised the question of completeness much earlier, in his lectures of 1917-18. According to Awodey and Carus (2001), Gödel learned of this completeness problem in his logic course with Carnap in 1928 (the one logic course that he ever took!).

I will elaborate. But first, some biographical information about Bernays.[4] He was born in 1888 in London, from where the family soon moved to Paris and a little later to Berlin. There, at the university, Bernays began his studies in mathematics with Landau and Schur; he then followed Landau to the University of Göttingen where he also studied with Hilbert, Weyl and Klein. Bernays completed a doctorate on the analytic number theory of quadratic forms under Landau's direction in 1912. From Göttingen he moved later that year to the University of Zürich where he wrote his Habilitationschrift on analytic function theory and became a Privatdozent. In 1917 Hilbert came to Zürich to deliver his famous lecture, "Axiomatisches Denken". Having resumed his interest in foundational problems, Hilbert invited Bernays to become his assistant in Göttingen to work with him on those questions. In 1918 Bernays began his work in logic with a second habilitation thesis, on the completeness of the propositional calculus and the independence of its axioms.

As Hilbert's assistant in Göttingen from 1917-1922, Bernays was significantly involved in helping Hilbert develop and detail his ideas about mathematical logic and the foundations of mathematics; by the end of that period these explorations had evolved into Hilbert's program for finitary consistency proofs of formal axiomatic systems for central parts of mathematics. At Hilbert's urging, Bernays was promoted to the (non-tenured) position of Professor Extraordinarius at Göttingen in 1922; he held that until 1933, when he—as a "non-Aryan"—was forced by the Nazis to give up his post, but Hilbert kept him on at his own expense for an additional six months. In the spring of 1934 Bernays returned to Zürich, where he held a position at the ETH from then on; treated as a temporary position for many years, that finally turned into a regular position—albeit only as a Professor Extraordinarius—in 1945. Following the 1934 move from Göttingen, Bernays had virtually no communication with Hilbert but continued his work on the two volumes of the *Grundlagen der Mathematik* (1934 and 1939). Bernays was completely responsible for their preparation; it was the first full exposition of Hilbert's finitary consistency program and beyond in a masterful, calm and unhurried presentation.[5]

[4] This material is mainly drawn from Bernays' short autobiography (1976).
[5] For an excellent introduction online to Hilbert's program with a guide to the literature, see the entry by Richard Zach in the Stanford Encyclopedia of Philosophy (2003). A more extended exposition is to be

Alongside that, beginning by 1930, Bernays was developing his axiomatic system of sets and classes as a considerable improvement of the axiomatization due to von Neumann; this was eventually published in *The Journal of Symbolic Logic* in seven parts, stretching from 1937 to 1954.

Bernays' distinctive voice in the philosophy of mathematics began to emerge early in the 1920s with a pair of articles on the axiomatic method and an early version of Hilbert's program; the long paper Bernays (1930) is on the significance of Hilbert's proof theory for the philosophy of mathematics. But in a number of pieces from then on he distanced himself from Hilbert's strictly finitist requirements for the consistency program and expressed a more liberal and nuanced receptiveness to alternative foundational views, including a moderate form of platonism.[6] This was reinforced through his contact at the ETH in Zürich with Ferdinand Gonseth, who held an "open" philosophy that rejected the possibility of absolute foundations of mathematics or science. With Gonseth and Gaston Bachelard he founded the journal *Dialectica* in 1945.

Bernays visited the Institute for Advanced Study in Princeton in 1935-36 and again in 1959-60, during which period he had extensive contact with Gödel. As it happens, it was my good fortune to be at the Institute that same year and to make his acquaintance then. I was two years out from a PhD at Berkeley with a dissertation on the arithmetization of metamathematics, and—like so many logicians of those days—had been drawn there by the chance to confer directly with Gödel and benefit from his unique insights. I had by then gone on to establish my main results on transfinite progressions of theories, extending Turing's earlier work on ordinal logics.[7] Among other visitors that same year were Kurt Schütte and Gaisi Takeuti (my office mate at the Institute), to whom—along with Georg Kreisel in his Stanford days—I am indebted for my way into proof theory. But that's another story.

Bernays was the most senior of the visitors in logic that year; he was then aged 71, compared to Gödel's 53. I knew little of his work at the time, other than as the co-author of the *Grundlagen der Mathematik*, and for the development of his elegant theory

---

found in Part III of Mancosu (1998), as an introductory note to a number of major articles by both Hilbert and Bernays in English translation.

[6] See Parsons (200?) for an examination of Bernays' later philosophy of mathematics.

[7] See Turing (1939) and Feferman (1962).

of sets and classes. A gentle, modest man, he did not advertise the range of his thoughts and accomplishments; it was only later that I began to really appreciate his place in our subject. And it was only much more recently, when working on his correspondence with Gödel for Volume IV of the *Collected Works*, that I learned of the depth of the personal and intellectual relationship between the two of them.

In addition to his visit at the Institute in 1959-60, Bernays also paid visits to Gödel during three stays that he had as a visiting professor at the University of Pennsylvania between 1956 and 1965. He died in Zürich in September 1977, just one month shy of his 89th birthday and four months before the death Gödel in January 1978.

2. 1931: The incompleteness theorems and Hilbert's ω-rule.

The correspondence between Bernays and Gödel begins with a letter from Bernays dated Christmas eve 1930, complimenting Gödel on the completeness theorem for first order logic and then asking to see his "significant and surprising results" in the foundations of mathematics—namely the incompleteness theorems—that he had heard about from Courant and Schur. Gödel sent Bernays one set of proof sheets forthwith. The first four items of correspondence between them in 1931 are largely devoted to Bernays' struggles to understand the incompleteness theorems against the background of ongoing work on the consistency program in the Hilbert school.[8] Earlier in the 20s Ackermann and then (in revised form) von Neumann had supposedly given a finitary proof of the consistency of a formal system Z for classical first-order arithmetic, nowadays called Peano Arithmetic (PA). It was not yet realized that their proof succeeds only in establishing the consistency of the weak subsystem of Z in which induction is restricted to quantifier-free formulas.

Part of Bernays' perplexities with Gödel's incompleteness theorems had to do with the roughly concurrent work of Hilbert (1931 and 1931a) in which a kind of finitary version of the infinitary ω-rule extending Z to a system Z* was proposed as a means of overcoming the incompleteness of Z. Roughly speaking, the rule allows one to adjoin a sentence ∀x A(x) (A quantifier-free) to the axioms of Z* for which it has been shown

[8] The fifth, and last item of 1931, from Bernays to Gödel, deals with one final point on this matter; it is mainly devoted to a preliminary presentation of Bernays' theory of sets and classes.

finitarily that each instance A(n) for $n \in \omega$ is already provable in Z*. Hilbert claimed to have a finitary consistency proof of Z*, relying mistakenly on the work of Ackermann and von Neumann for Z. There is no reference to Gödel's incompleteness theorem in these articles. What would have led Hilbert to try to overcome the incompleteness of Z if he was not already aware of it? True, he was already lecturing on the proposed extension of Z in December of 1930, before he and Bernays even saw the proof sheets of Gödel's paper. On the other hand Hilbert could well have learned of Gödel's incompleteness theorem before that from von Neumann, who heard about it in September of that year; others who might have communicated the essence to him were Courant and Schur. This is not a place to go into all the relevant details.[9] Bernays' own two reports on the matter don't quite jibe: in his survey article (1935) on Hilbert's foundational contributions for Hilbert's collected works, Bernays said that even before Gödel's incompleteness theorems, Hilbert had given up the original form of his completeness problem and in its place had taken up the ideas for an extension of Z by a finitary ω-rule. But in a letter thirty years later to Hilbert's biographer, Constance Reid, Bernays wrote that Hilbert was angry about doubts that he (Bernays) had already expressed about the conjectured completeness of Z and was then angry about Gödel's results.[10]

In his correspondence with Bernays, Gödel points out things that he did not mention in his review (*1931c*) of Hilbert's article on Z*, namely that even with the proposed ω-rule, the system is incomplete. He went on to write that:

> I do not think that one can rest content with the systems Z*, Z** [a variant proposed by Bernays for "aesthetic reasons"] as a satisfactory foundation of number theory …, and indeed, above all because in them the very complicated and problematical concept of "finitary proof" is assumed (in the statement of the rule for the axioms) without having been made mathematically precise in greater detail. (*CW IV*, p. 97)

[9] In any case, I have discussed the evidence at length in my introductory note to Gödel's review *1931c* of Hilbert's work on Z*; cf. *CW I*, pp. 208-213.
[10] Cf. Reid (1970), pp. 198-199.

This resonates with Gödel's statement to Carnap in May of 1931 that he viewed the move to Z* as a step compromising Hilbert's program.[11] However, he did not make this fundamental criticism in his review of Hilbert (1931) itself, as he might well have.

Gödel's own concerns with determining the limits of Hilbert's finitism are there from the beginning. In his watershed paper on formally undecidable propositions, after stating and sketching the proof of the second incompleteness theorem on the unprovability of consistency of systems by their own means, Gödel writes:

> I wish to note expressly that [this theorem does] not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finitary means of proof is used, and it is conceivable that there exist finitary proofs that *cannot* be expressed in the formalism of [our basic system]. (*Gödel 1931, CW I*, p. 195)

Von Neumann—who was the first to grasp what Gödel had accomplished in his brief announcement of the first incompleteness theorem at the Königsberg conference in September of 1930 and who independently realized the second incompleteness theorem—had been urging the opposite view on him. At any rate, Gödel came around to von Neumann's viewpoint by the end of 1933, when he delivered a lecture at a meeting in Cambridge, Massachusetts during his first visit to the United States. That is the first of a series of three remarkable lectures that Gödel gave between 1933 and 1941 in which his thoughts about finitism, constructivity and Hilbert's program took more definite form; these are the subjects of our next two sections. (All three lectures appeared in print for the first time in *CWIII* as **1933o*, **1938a* and **1941*, respectively.)

In the same period following 1931 there is a long break in the extant correspondence between Gödel and Bernays. It does not resume again until 1939 and is then devoted almost entirely to set theory, especially Gödel's proofs of the consistency of the Axiom of Choice and the Generalized Continuum Hypothesis with the axioms of set

[11] Cf. *CW I*, p. 212. Olga Taussky-Todd writes in her reminiscence (1987), p.40, that Gödel "lashed out against Hilbert's paper 'Tertium non-datur' [Hilbert 1931a] saying something like, 'how can he write such a paper after what I have done?' Hilbert in fact did not only write this paper in a style irritating to Gödel, he gave lectures about it in Göttingen in 1932 and other places. It was to prove Hilbert's faith."

theory. Initially presented in terms of the system of Zermelo-Fraenkel, for expository purposes Gödel later adopted the system of sets and classes that Bernays had communicated to him in 1931.

Of incidental personal note is the change in salutations that also took place in 1931: where, in the first few letters, Bernays was addressed as "Sehr geehrter Herr Professor!" and Gödel as "Sehr geehrter Herr Dr. Gödel!", these now became "Lieber Herr Bernays!" and "Lieber Herr Gödel!", respectively, and so remained throughout their correspondence thenceforth.

3. 1933: The Cambridge lecture.

On December 30, 1933, Gödel gave a lecture entitled "The present situation in the foundations of mathematics" for a meeting of the Mathematical Association of America in Cambridge, Massachusetts. Gödel's aim in this lecture is clearly announced in the first paragraph:

> The problem of giving a foundation for mathematics (…[i.e.,] the totality of methods of proof actually used by mathematicians) can be considered as falling into two different parts. At first these methods of proof have to be reduced to a minimum number of axioms and primitive rules of inference, which have to be stated as precisely as possible, and then secondly a justification in some sense or other has to be sought for these axioms… (*Gödel *1933o*, *CW III*, p. 45)

He goes on to assert that the first part of the foundational problem has been solved in a completely satisfactory way by means of formalization in the simple theory of types when all "superfluous restrictions" are removed. As he explains, that is accomplished via axiomatic set theory à la Zermelo, Fraenkel and von Neumann with its underlying cumulative type structure admitting a simple passage to transfinite types. Nevertheless, set theory is said to have three weak spots: "The first is connected with the non-constructive notion of existence. … The second weak spot, which is still more serious, is… the so-called method of impredicative definitions [of classes]. …The third weak spot in our axioms is connected with the axiom of choice… .” (*ibid.,* pp. 49-50) Consideration of these lead Gödel directly to the following stunning pronouncement:

> The result of the preceding discussion is that our axioms, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent. (*Ibid.,* p. 50)

Not that it’s likely that the system is inconsistent, Gödel says, since it has been developed in so many different directions without reaching any contradiction. Given that, one might hope to prove the consistency of the system when treated in exact formal terms. But not *any* proof will do: Gödel says that “it must be conducted by perfectly unobjectionable [constructive] methods; i.e., it must strictly avoid the non-constructive existence proofs, non-predicative definitions and similar things, for it is exactly a justification for these doubtful methods that we are now seeking.” And even with this, he says, the nature of such a proof is not uniquely determined since there are different notions of constructivity and, accordingly, “different layers of intuitionistic or constructive mathematics.” (*Ibid.*, p. 51)

Concerning the use of the word ‘intuitionistic’ in this last quote it should be noted that, according to Bernays (1967), p. 502, the prevailing view in the Hilbert school at the beginning of the 1930s equated finitism with intuitionism.[12] Within a few years, finitism was generally distinguished from intuitionism in the sense of the Brouwer school, in part through Heyting’s formalization of intuitionistic arithmetic and Gödel’s 1933 translation of the classical system of Peano Arithmetic (PA) into Heyting Arithmetic (HA). In this lecture, however, Gödel never uses the words ‘finitary’ or ‘finitistic’. He also does not speak explicitly about the Hilbert consistency program except for one indirect reference below.

Gödel delineates the lowest level of constructive mathematics, that he calls the system A, in the following terms:

[12] Gödel himself, in the Postscript to his remarks at the Königsberg conference, speaks of “finitary, (that is, intuitionistically unobjectionable) forms of proof” (*1931a, CW I,* p. 205). Cf. also Bernays (1930), Part II, sec. 2 and Sieg (1990), p. 272.

> 1. The application of the notion of "all" or "any" is to be restricted to those infinite totalities for which we can give a finite procedure for generating all their elements (as we can, e.g., for the totality of integers by the process of forming the next greater integer and as we cannot, e.g., for the totality of all properties of integers).
>
> 2. Negation must not be applied to propositions stating that something holds for all elements, because this would give existence propositions. … Negatives of general propositions (i.e. existence propositions) are to have a meaning in our system only in the sense that we have found an example but, for the sake of brevity, do not wish to state it explicitly. I.e., they serve merely as an abbreviation and could be entirely dispensed with if we wished.
>
> From the fact that we have discarded the notion of existence and the logical rules concerning it, it follows that we are left with essentially only one method for proving general propositions, namely, complete induction applied to the generating process of our elements.
>
> 3. And finally, we require that we should introduce only such notions as are decidable for any particular element and only such functions as can be calculated for any particular element. Such notions and functions can always be defined by complete induction, and so we may say that our system [A] is based exclusively on the method of complete induction in its definitions as well as its proofs. (*Gödel *1933o, CW III*, p. 51).

Gödel did not spell out formally these conditions on the system A, and there has been considerable discussion about exactly how to interpret it. It is pretty clear that the formulas of A should be taken as universal generalizations of quantifier-free formulas built up from decidable atoms by the propositional operations. When dealing with the provable formulas one can just as well take them to be quantifier-free by disregarding the initial universal quantifiers. Thus—at first sight—it seems that A should be interpreted as a form of Primitive Recursive Arithmetic (PRA), the quantifier-free system which has as axioms the usual ones for zero and successor, the defining equations for each primitive recursive function—where the step from each to the next is given either explicitly or by a "complete induction" (ordinary recursion on one numerical variable)—and finally with a rule of induction on the natural numbers. Initially, Wilfried Sieg, Charles Parsons, William Tait and I all took this interpretation of A for granted. However, both Sieg and Tait subsequently changed their minds, though for different reasons.[13] The main point raised by Sieg of what is at issue has to do with Gödel's statement on p. 52 of **1933o* that the most far-reaching consistency result obtained by methods in accord with the principles of A is that due to Herbrand (1931). Herbrand's theorem is for a (somewhat open-ended) system that goes beyond PRA by including such functions as that due to Ackermann, given by a double nested recursion. Nevertheless, it seems to me that it is still possible to construe the system A as being PRA if one interprets Gödel's remark as applying to the general form of Herbrand's argument rather than to the specific statement of his theorem.

At any rate, what Gödel has to say in **1933o* about the potential reach of the system A in pursuit of constructive consistency proofs at least puts an upper bound to its strength:

> This method possesses a particularly high degree of evidence, and therefore it would be the most desirable thing if the freedom from contradiction of ordinary non-constructive mathematics could be proved by methods allowable in this system A. And, as a matter of fact, all the attempts for a proof for freedom from contradiction undertaken by Hilbert and his disciples tried to accomplish exactly that. But unfortunately the hope of succeeding along these lines has vanished entirely in view of some recently discovered facts [namely, the incompleteness theorems]. …Now all the intuitionistic proofs complying with the system A which have ever been constructed can easily be expressed in the system of classical analysis and even in the system of classical arithmetic, and there are reasons for believing that this will hold for any proof which one will ever be able to construct.

[13] See Sieg's introductory note to Gödel's correspondence with Herbrand in *CW V*, p. 9, ftn. s, and Tait (2006), pp. 98-105 in his essay review of *CW IV-V*.

> … So it seems that not even classical arithmetic can be proved to be non-contradictory by the methods of the system A. (*Gödel *1933o, CW III,* pp. 51-52)

This is the sole reference to Hilbert and his program in the Cambridge lecture*,* but there is no discussion there of Hilbert's finitist criterion for consistency proofs, let alone of Hilbert's conception of finitism.[14] And this also returns us to the question about how to interpret A, for if it were exactly PRA, Gödel would no doubt have recognized that its consistency could be proved in PA, given his 1931 definition of the primitive recursive functions in arithmetic and the arguments of Herbrand (1931). At any rate, whatever the exact reach of the methods provided by A, it evidently appeared hopeless to Gödel by 1933 to carry out that program for the system PA of arithmetic, let alone for analysis and set theory.

Thus in the final part of **1933o*, Gödel takes up the question whether stronger constructive methods than those provided by the system A ought to be admitted to consistency proofs, in particular "the intuitionistic mathematics as developed by Brouwer and his followers." That would take one at least up to arithmetic, given the reduction of PA to Heyting's system HA of intuitionistic arithmetic. The essential formal difference of A from HA is that in the latter, all first order formulas are admitted to the language. But Gödel is not at all satisfied with intuitionistic logic in pursuit of the consistency program, for the meaning of its connectives and quantifiers is explained in terms of the (in his view) vague and unrestricted concept of constructive proof. In particular, a proof of an implication $p \rightarrow q$ is supposed to be provided by a construction which converts *any* proof of p into a proof of q, and a proof of $\neg p$ is supposed to be given by a construction which converts *any* proof of p into an absurdity, or contradiction. This justifies, for example, $p \rightarrow \neg\neg p$ in intuitionistic logic, since p and $\neg p$ constitute a contradiction. Intuitionistic logic does meet condition 2 above to the extent that a proof of $\exists x\, P(x)$ is supposed to be a construction which provides an instance t for which P(t) holds. But

[14] What Hilbert meant by finitism has been a subject of extensive discussion in the literature; the trouble is that he made no specific effort to delineate it and instead relied on informal explanations and examples. A strong case has been made by Zach (1998, 2001, 2003) among others that Hilbert's conception of finitism definitely goes beyond PRA to include Ackermann's function and other functions obtained by multiply nested recursions, together with certain forms of transfinite recursion (in view of Hilbert 1926).

condition 2 is *not* met in its *stated* form since negation may be applied to universal statements in Heyting's formalism and since $\neg\forall x\, P(x)$ is not intuitionistically equivalent to $\exists x\, \neg P(x)$, and hence cannot be considered as an abbreviation for the latter. However, Gödel's main objection to intuitionistic logic is that it does not meet condition 1, since "the substrate on which the constructions are carried out are proofs instead of numbers or other enumerable sets of mathematical objects. But by this very fact they do violate the principle, which I stated before, that the word 'any' can be applied only to those totalities for which we have a finite procedure for generating all their elements. ... And this objection applies particularly to the totality of intuitionistic proofs because of the vagueness of the notion of constructivity. Therefore this foundation of classical arithmetic by means of the notion of absurdity is of doubtful value." (*Ibid.,* p. 53)

4. Transitions: 1934-1941

Hilbert was unaffected by any of the reconsiderations of the possible limits to finitary methods in pursuit of his consistency program that had been stimulated by the second incompleteness theorem. In his preface to Volume I of the *Grundlagen der Mathematik*, and in his sole reference anywhere to Gödel or his incompleteness theorems, Hilbert writes:

> This situation of the results that have been achieved thus far in proof theory at the same time points the direction for the further research with the end goal to establish as consistent all our usual methods of mathematics.
>
> With respect to this goal, I would like to emphasize the following: the view, which temporarily arose and which maintained that certain recent results of Gödel show that my proof theory can't be carried out, has been shown to be erroneous. In fact that result shows only that one must exploit the finitary standpoint in a sharper way for the farther reaching consistency proofs.
>
> (Hilbert, preface to Hilbert and Bernays 1934)[15]

[15] "Dieser Ergebnisstand weist zugleich die Richtung für die weitere Forschung in der Beweistheorie auf das Endziel hin, unsere üblichen Methoden der Mathemattick samt und sonders als widersprchsfrei zu erkennen.

Nevertheless, the need for a modified form of Hilbert's consistency program began to become generally recognized by others in his school by the mid-30s, among them his very collaborator on the *Grundlagen*. As Bernays wrote years later, "it became apparent that the '*finite Standpunkt*' is not the only alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of a restriction to finitist methods of reasoning, it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inference." (Bernays 1967, p. 502)

One striking new specific way forward was provided by Gerhard Gentzen in his 1936 paper in which the consistency of arithmetic is proved by transfinite induction up to Cantor's ordinal ε_0, otherwise using only finitary reasoning. In Hilbert's Preface to volume II of the *Grundlagen der Mathematik* (1939) in which he thanks Bernays for carrying out the exposition and development of his ideas for proof theory and the consistency program, no mention is made of Gödel or Gentzen, even though the volume contains an extended exposition of Gödel's incompleteness theorems and a description of Gentzen's work in a section entitled *Überschreitung des bisherigen methodischen Standpunktes der Beweistheorie*.

Gödel reported on the issues of a modified consistency program in a remarkable though sketchy presentation that he made to Edgar Zilsel's seminar in Vienna in 1938, the notes for which were reconstructed from the *Gabelsberger* shorthand script as *Gödel *1938a, CW III*. He begins by pointing out that one can only deal with the consistency of partial systems of mathematics as represented in formal systems T, and this must be accomplished by other systems S. Moreover,

> [the choice of such a system S] has …an epistemological side. *After all we want a consistency proof for the purpose of a better foundation of mathematics (laying*

"Im Hinblick auf dieses Ziel möchte ich hervorheben, dass die zeitweilig aufgekommene Meinung, aus gewissen neueren Ergebnissen von Gödel folge die Undurchfürbarkeit meiner Beweistheorie, als irrtümlich erwiesen ist. Jenes Ergebnis zeigt in der Tat auch nur, dass man für die weitergehenden Widerspruchsfreiheitsbeweise den finiten Standpunkt in einer shärferen Weise ausnutzen muss, als dieses bei der Betrachtung der elementaren Formalismen erforderlich ist."

the foundations more securely), and there can be mathematically very interesting proofs that do not accomplish that. A proof is only satisfying if it either

(i) *reduces to a proper part* or

(ii) *reduces to something which, while not a part, is more evident, reliable, etc., so that one's conviction is thereby strengthened.*

(*Gödel *1938a, CW III*, p. 89, italics in the original)

Though a reduction of kind (i) may be preferred because of its objective character while (ii) by comparison involves subjective judgments, historically the latter is the route taken through the reduction of non-constructive to constructive systems. But since the concept of constructivity is hazy, it is useful to make a "framework definition, which at least gives necessary if not sufficient conditions" (*ibid.,* p. 91). That is provided by conditions like those for the system A in the Cambridge lecture, here broken up into four parts, of which the last is that "objects should be surveyable (that is denumerable)." But now the system at the lowest level of constructivity is specifically referred to as finitary number theory and Gödel raises the question, "how far do we get, or fail to get with finitary number theory?" (*ibid.,* p.93) His conclusion is that "transfinite arithmetic" (presumably PA) is not reducible to A.[16] It should be noted that the disagreements as to how the system A of **1933o* should be interpreted have not been raised for Gödel's informal system of finitary number theory in the 1938 lecture; instead, all who have considered it agree that it may be formalized as PRA.

In order to go beyond what can be treated by finitary number theory, the fourth condition—that the objects be denumerable—is jettisoned in all three of the further constructive approaches considered at Zilsel's seminar. These are (not in the order presented by Gödel): the "modal-logical" route, the route of induction on transfinite ordinals, and the route of constructive functions of finite type. By the "modal-logical" route is meant the intuitionistic logic of Brouwer and Heyting, for which a foundation of the notion of constructive proof underlying it is sought in terms of a modal-like operator B (for 'Beweisbar'). After considering several possible conditions on B, Gödel's

[16] Hilbert (1926) refers to the axioms for quantifiers in the context of arithmetic as the "transfinite axioms" because the law of excluded middle for quantified formulas requires the assumption of the "completed infinite".

conclusion is that there is no reasonable way carry this out and that “this [route] is the worst of the three ways” (*ibid.,* p. 103). The second route follows Gentzen’s proof of consistency of PA by transfinite induction up to ε_0, concerning which Gödel indicates a new and more intuitive way why the proof works; however, few details are given.[17] As to the ordinal route itself, Gödel questions it on several grounds, the main one being its lack of direct evidence for the principle of transfinite induction up to ε_0, let alone for the ordinals that would be needed to establish the consistency of still stronger systems. The final route, by means of constructive functions of finite (and possibly transfinite) type, is barely indicated, but it is that approach which Gödel himself was to explore in depth, as we shall see shortly.

The Zilsel lecture notes conclude with the following quite interesting general assessment that is worth quoting at length:

> I would like to return to the historical and epistemological side of the question and then ask (1) whether a consistency proof by means of the three extended systems has a value in the sense of laying the foundations more securely; (2) what is closely related, whether the Hilbert program is undermined in an essential respect by the fact that it is necessary to go beyond finitary number theory.
>
> To this we can say two things: (1) If the original Hilbert program could have been carried out, that would have been without any doubt of enormous epistemological value. … (a) Mathematics would have been reduced to a very small part of itself … (b) Everything would have been reduced to a concrete basis, on which everyone must be able to agree. (2) As to the proofs by means of the extended finitism, the first [i.e., (a)] is no longer the case at all … [while] the second [i.e., (b)] (reduction to the concrete basis, which means increase of the degree of evidence) obtains for the different systems to different degrees, thus for the modal-logical route not at all, for the higher function types the most, [and] for the transfinite ordinal numbers … also to a rather high degree. (*Ibid.,*p. 113)

[17] The idea is elaborated in Tait (2005a).

In April 1941, Gödel delivered a lecture entitled "In what sense is intuitionistic logic constructive" (**1941*, *CW III*) at Yale University. He had by then become established as a member of the Institute for Advanced Study, having fled Austria (and its threatened conscription) with his wife at the last minute at the beginning of 1940. In the Yale lecture he gave the first public account of what was later to be called the *Dialectica* interpretation of HA in a quantifier-free system of functionals of finite type—the route which he considered to have the greatest degree of evidence (beyond that of finitism) of the three considered at Zilsel's. His stated aim there is that "if one wants to take constructivity in a really strict sense [then] the primitive notions of intuitionistic logic cannot be admitted in their usual sense. This however does not exclude the possibility of defining in some way these notions in terms of strictly constructive ones and then proving the logical axioms which are considered as self-evident by the intuitionists. It turns out that this can actually be done in a certain sense, namely, not for intuitionistic logic as a whole but for its applications in definite mathematical theories (e.g. number theory)." (*Gödel *1941*, *CW III*, p. 191) The functional interpretation proposed as the means to carry this out is the subject of the final body of correspondence between Gödel and Bernays, taken up in the next section.

As already mentioned, set theory was the main subject of the correspondence between them in a flurry of letters that began in 1939 and continued for three years. The last one from Bernays in this period is dated 7 September 1942. In it, no mention is made of the war except to remark on its mild effects on the Swiss. Then Bernays—a bachelor—charmingly concludes his letter as follows:

> Hopefully you are now well settled in Princeton, and married life and the domesticity associated with it is quite salubrious for your physical and emotional health, and thereby also for your scientific work.
>
> Would you please convey my respects to your wife, even though I am not personally acquainted with her.
>
> Friendly greetings to you yourself.

After 1942 the correspondence between the two lapsed for fourteen years; no doubt the difficulties of transmission during the war was the initial reason. But once resumed in 1956, it was to continue in a steady stream until their final exchange in 1975. The date 1956 marked the first postwar visit of Bernays to the US and his first stay at the University of Pennsylvania, during which he was able to come to Princeton to renew his personal contact with Gödel.[18]

5. The *Dialectica* interpretation

In 1958, in honor of Bernays' 70th birthday, Gödel published in the journal *Dialectica* his interpretation of intuitionistic number theory—and thereby, classical number theory—in a quantifier-free theory of primitive recursive functions of finite type; this spelled out the notions and results of what had been presented in the Yale lecture in 1941, and it subsequently came to be known as Gödel's *Dialectica* interpretation (*Gödel 1958, CW II*). The change in title from the Yale lecture, "In what sense is intuitionistic logic constructive?", to that of the *Dialectica* paper, "On a hitherto unitilized extension of the finitist standpoint" [Über eine noch nicht benützte Erweiterung des finiten Standpunktes], indicates both a change of focus and a more precise attention to the bounds for finitism, at least in Hilbert's sense. That is reinforced by the opening paragraph of the *Dialectica* piece:

> P. Bernays has pointed out on several occasions that, since the consistency of a system cannot be proved using means of proof weaker than those of the system itself, it is necessary to go beyond the framework of what is, in Hilbert's sense, finitary mathematics if one wants to prove the consistency of classical mathematics, or even that of classical number theory. Consequently, since finitary mathematics is defined as the mathematics in which evidence rests on what is *intuitive,* certain *abstract* notions are required for the proof of the consistency of number theory.... In the absence of a precise notion of what it means to be evident, either in the intuitive or in the abstract realm, we have no

[18] Dana Scott, who was a student in Princeton at the time, recalls a "conversation" with Church, Gödel, Bernays, Kreisel and a couple of graduate students at Church's home. (Personal communication.)

strict proof of Bernays' assertion; practically speaking, however, there can be no doubt that it is correct … (*Gödel 1958*, *CW II*, p. 241)

Seven years after the publication of the *Dialectica* paper, Bernays told Gödel of a plan to publish in the same journal an English translation that had been made of it by Leo F. Boron. However, Gödel was not too happy with certain aspects of both the original and its translation, and set out to revise it. A year later he changed his mind and decided instead to improve and amplify the original by means of a new series of extensive footnotes. Preparation of these dragged on for four more years and it was only after much help and encouragement by Bernays and Dana Scott that a revised manuscript was sent to the printer in 1970. However, when the proof sheets were returned, Gödel was again dissatisfied, especially with two of the added notes. Though he apparently worked on rewriting these until 1972, the paper was never returned in final form for publication. The corrected proof sheets found in his *Nachlass* were reproduced for the first time in volume II of the *Collected Works*, where they appear as *1972*. The full story of the vicissitudes of that paper is told by A. S. Troelstra in his introductory note to *Gödel 1958 and 1972* in that volume.

The question of the bounds on Hilbert's finitism comes up repeatedly in the correspondence from this period; in letter #61 from Gödel to Bernays of January 1967 he wrote:

My views have hardly changed since 1958, except that I am now convinced that ε_0 is a bound on Hilbert's finitism, not merely in practice [[but]] in principle, and that it will also be possible to prove that convincingly. (*CW IV,* p. 255)

During the same period, Bernays had been working on a second edition of the *Grundlagen der Mathematik.* Its first volume eventually appeared in 1968 and the second in 1970. As it happens, the latter was to contain a new supplement with an exposition of proofs due to Kalmár and Ackermann—subsequent to Gentzen's—of the consistency of the system of Peano arithmetic by means of transfinite induction on the natural ordering of order-type ε_0. In connection with that, Bernays developed a new and (on the face of it) more perspicuous proof of induction up to ε_0 to be included in the supplement, and he sent that proof along with his letter (#67) to Gödel of January 1969. As Bernays puts it there, what he establishes with this proof is "the weak form of induction, which says that every decreasing sequence comes to an end after finitely many steps", i.e. that the relation is well-founded.[19] Gödel became quite excited about Bernays' proof and in July of 1969 prepared a draft of a letter (#68a), in which he wrote:

> You undoubtedly have given the most convincing proof to date of the ordinal-number character of ε_0 ... Since functions of the first level [i.e., sequences of ordinals] can be interpreted as free choice sequences and that concept is obviously decidable, a statement of the form 'For all free choice sequences..." contains no intuitionistic implication, and you have consequently completely eliminated the intuitionistic logic. If one reckons choice sequences to be finitary mathematics, your proof is even finitary. … I now strongly doubt whether what was said about the boundaries of finitism [at the beginning of *1958*] is really right. For it now seems to me, after more careful consideration, that choice sequences are something concretely evident and therefore are finitary in Hilbert's sense, even if Hilbert himself was perhaps of another opinion. (*CW IV*, pp. 269)

Then Gödel included a draft remark to a footnote of the revised version of *1958* to this effect, adding:

> Hilbert did not regard choice sequences ... as finitary, but this position may be challenged on the basis of Hilbert's own point of view. (*Ibid.*)[20]

---

[19] As it happens, Bernays' proof as it stands is mistaken, or at best incomplete, as shown by Tait (2006), pp. 89-91; see footnote 21 below for the reason.

[20] The version of the footnote that did appear in the proofs for *Gödel 1972 (*ftn. c, *CWII*, p. 272) reads: "A closer approximation to Hilbert's finitism can be achieved by using the concept of free choice sequence rather than 'accessibility'."

Following a remark of Kreisel, Mark van Atten (2006, p. 26) suggests that Gödel, *qua* Platonist, considered choice sequences to belong to the ontological realm of "our own 'constructions' or choices" and hence not to pure mathematics, but whose theory could well be considered as a matter of applied mathematics.

Indeed, the idea of "[free] choice sequences" used here is that due to Brouwer in a form that is quintessentially intuitionistic and generally regarded as non-finitary in nature. This is one of Gödel's letters that is marked "*nicht abgeschickt"*.

In the letter of 25 July 1969 that Gödel *actually* sent (#68b), he changed his mind about the significance of Bernays' proof, though he still regarded it as "extraordinarily elegant and simple":

> At first one also has the impression that it comes closer to finitism than the other proofs. But on closer reflection that seems very doubtful to me. The property of being 'well-founded' contains two quantifiers after all, and one of them refers to all number sequences (which probably are to be interpreted as choice sequences). In order to eliminate the quantifiers …one would use a nested recursion… But nested recursions are not finitary in Hilbert's sense (i.e. not intuitive) ... Or don't you believe that? (*CW IV,* p. 271)[21]

What is at issue in all this for Gödel goes back to his letter of 1931 to Bernays in which he said that "the complicated and problematical concept of 'finitary proof'" needs to be made mathematically precise in order to decide such questions. In 1969, two such characterizations of finitism were on offer, the first due to Georg Kreisel some ten years prior to that, in terms of an ordinal logic whose limit is exactly ε_0, i.e., the strength of Peano Arithmetic (Kreisel 1960). Gödel had discussed this with both Kreisel and Bernays and given it serious consideration, but was equivocal about the conclusion. A second proposed characterization that arrived at primitive recursive arithmetic PRA as

[21] Bernays answered the question concerning nested recursion, saying that the verschränkte rekursion of vol. I of *Grundlagen der Mathematik* appear to him to be finitary in the same sense as the primitive recursions (see *CW IV,* p. 277); cf. also ftn. 14 above. Tait (2006), pp. 91-92, critically examines the questions concerning nested recursion and choice sequences; he argues that these go beyond finitism as it ought to be understood (which may well differ from the way Hilbert understood it). In addition (op. cit., pp. 89-91) he usefully spells out Bernays' proof of induction up to ε_0, considered as the limit of the ordinals $\omega[n]$, where $\omega[0] = \omega$ and $\omega[n+1] = \omega^{\omega[n]}$. His analysis reveals that Bernays' prima facie inductive argument to show for each n that there are no infinite sequences descending from $\omega[n]$ actually only reduces that property for $\omega[n+1]$ to the assumption of nested recursion on $\omega[n]$; so the proof as it stands is mistaken or at best incomplete without the additional claim that the no descending sequence property on an ordinal justifies nested recursion on that ordinal. The situation here is related to the earlier result of Tait (1961) that reduces, for any ordinal α, ordinary recursion on ω^α to nested recursion on $\omega \times \alpha$, a fact, as it happens, known to Gödel but not invoked in his enthusiastic response to Bernays' supposed proof.

the upper bound of finitism, was sketched by William Tait in his 1968 article "Constructive reasoning" (and later spelled out in his article "Finitism" (Tait 1981)). As we have seen, that is the formal interpretation of the system A of finitary number theory indicated by Gödel at the Zilsel seminar in 1938, but there is no reference to Tait's proposal in the correspondence with Bernays.

In any case, only the first of these could be considered to be an explication of finitism in Hilbert's sense, which unquestionably went beyond PRA. In letter #40 of August 1961, Gödel wrote Bernays that he had had interesting discussions with Kreisel about his work and that "[H]e now really seems to have shown in a mathematically satisfying way that the first ε-number is the precise limit of what is finitary. I find this result very beautiful, even if it will require a phenomenological substructure in order to be completely satisfying." (*CW IV*, p. 193) The characterization of finitist proof in Kreisel (1960) was given in terms of a transfinite sequence of proof predicates for formal systems $\sum_\alpha$, or ordinal logics in the sense of Turing (1939), under the restriction that the ordinal stages α to which one may ascend are controlled autonomously—i.e., there must be for each such α a recognition at an earlier stage β that the iteration of the process α times is (finitarily) justified. Kreisel's main results there are that the least non-autonomous ordinal is ε_0, and the provably recursive functions of the union of the $\sum_\alpha$ for $\alpha < \varepsilon_0$ are exactly the same as those of PA. Both the description of the ordinal logic for the proposed characterization and the proofs of the main results are very sketchy (as acknowledged by Kreisel, due to limitations of space); full details, though promised, were never subsequently published.[22] In addition, Kreisel offered little in the way of convincing arguments to motivate his proposed explication of the informal concept of finitist proof; this was perhaps the reason for Gödel's statement that more would be needed to make it "completely satisfying."[23]

[22] A variant formulation of the proposed ordinal logic that is a little more detailed was presented in Kreisel (1965), sec. 3.4 (pp. 168 ff).

[23] In Kreisel (1965), p. 169, it is said that the concept being elucidated is "of proofs that one can *see* or *visualize*. ... our primary subject is a *theoretical* notion for the actual visualizing, not that experience itself." The matter was revisited in Kreisel (1970), where the project is to determine "[W]hat principles of proof ... we recognize as valid once we have understood ... certain given concepts" (p. 489), these being in the case of finitism, "the concepts of ω-sequence and ω-iteration." (p. 490).

Kreisel's 1960 proposal had actually been made at the 1958 meeting of the International Congress of Mathematicians, and Gödel had already been cognizant of it at that time. It is referred to in footnote 4 to the 1958 version of the *Dialectica* paper, where Gödel says that a possible extension of the "original finitary standpoint ... consists in adjoining to finitary mathematics abstract notions that relate, in a combinatorially finitary way, only to finitary notions and objects, and then iterating this procedure. Among such notions are, for example, those that are involved when we reflect on the content of finitary formalisms that have already been constructed. A formalism embodying this idea was set up by G. Kreisel." Just what Gödel has in mind here by the "original finitary standpoint" is not clear; but whether he intends it to mean a system like PRA or a system for Hilbert's finitism in practice, either would put its strength well below that of PA.

These matters are revisited in *Gödel 1972*; in its opening paragraph*,* finitary mathematics in Hilbert's sense is now defined as "the mathematics of *concrete intuition,*" instead of "the mathematics in which evidence rests on what is *intuitive*," as it appeared in *1958*. And the relevant footnote 4 now reads:

> Note that an adequate proof-theoretic characterization of concrete intuition, in case this faculty is *idealized* by abstracting from the practical limitation, will include induction procedures which *for us* are *not* concretely intuitive and which could very well yield a proof of the inductive inference for ε_0 or larger ordinals. Another possibility of extending the original finitary viewpoint for which the same comment holds consists in considering as finitary any abstract arguments which only reflect ... on the content of finitary formalisms constructed before, and iterate this reflection transfinitely, using only ordinals constructed in previous stages of this process. A formalism based on this idea was given by G. Kreisel [1960]. (*Gödel 1972, CW II*, p. 274)

But there is now a *further* footnote f to this in *1972*, in which Gödel says that "Kreisel wants to conclude from [the fact that the limit of his procedure is ε_0] that ε_0 is the exact limit of idealized concrete intuition. But his arguments would have to be elaborated further to be fully convincing."

6. What, really, were Gödel's views on finitism and the consistency program?

In the preceding I have concentrated primarily on summarizing the available evidence concerning Gödel's views on finitism, Hilbert's program and its constructive extensions without critically examining those views themselves. But just what were those views and what was their significance for Gödel? There are two main questions, both difficult: First, were Gödel's views on the nature of finitism stable over time, or did they evolve or vacillate in some way? Second, how do Gödel's concerns with the finitist and constructive consistency programs cohere with his platonistic philosophy of mathematics that he supposedly held from his student days? Our problem is that the evidence is relatively fragmentary and, except for the published version of the *Dialectica* article, for the most part comes from lectures and correspondence that Gödel himself did not commit to print.

With respect to the first of these questions, in my introductory note to the correspondence with Bernays, I spoke of "Gödel's unsettled views over the years as to the exact upper bound of finitary reasoning", but this characterization has been challenged. In Tait's trenchant essay review of *CW IV-V* he takes me and others to task about such judgments, in contrast to his own interpretation of the available evidence:

> In any case, the alternative to this reading of the situation is to attribute to [Gödel] an unreasonable fluctuation in his views about "finitism". I feel that there has been rather too much easy settling for obscurity or inconsistency on Gödel's part in discussions of his works, especially—but not exclusively—his unpublished work, appearing in volumes III-V of the *Collected Works*. Of course, one can reasonably suppose that the wording in those works, by their nature—lecture notes, letters written in a day, *etc.*—did not receive the same care that he devoted to wording in his published papers. Nevertheless, it seems all the more reasonable that, in such cases, one should look for an interpretation of what he wrote, against the background of all of his writings, that has him saying something sensible. (Tait 2006, p. 93)

Tait refers to this as the McKeon principle,[24] but as he himself remarks, one can carry it too far: "In McKeon's hands, it degenerated into the less salubrious view that the great philosophers were never wrong: one only needed to discover the right principle of translation." I agree with both the principle and the caveat, so let's proceed with caution.

Another caveat: in his writings on finitism, reprinted as Chs. 1 and 2 with an Appendix in *The Provenance of Pure Reason,* Tait (2005) has rightly emphasized the need to distinguish between the proper characterization of finitism, and historical views of it, especially Hilbert's.[25] Further, in his essay reviews of *CW III* (2001) and *CW IV-V* (2006), this has been extended to Gödel's views of both the proper characterization of finitism and what he took Hilbert's view to be. In the last of these, pp. 92-98, Tait argues that Gödel's conception of finitism *was* stable and is represented by what he has to say about the system of finitary number theory for Zilsel's seminar, that we have seen is interpreted as PRA. He sets aside the problematic interpretation of the system A in the Cambridge lecture since that was not referred to by Gödel as being finitist, but rather as being at the lowest level of a hierarchy of constructive systems. Tait says that the ascription of unsettled views to Gödel in the correspondence and later articles "is accurate only of his view of *Hilbert's* finitism, and the instability centers around his view of whether or not there is or could be a precise analysis of what is 'intuitive'." (*ibid.,* p. 94). So, if taken with that qualification, my ascription of unsettled views to Gödel is not mistaken. As to Gödel's *own* conception of finitism, I think the evidence offered by Tait for its stability is quite slim, but that's nothing I'm concerned to make a case about, one way or another.

I would only add to all this a speculation concerning Gödel's vacillating views of Hilbert's finitism: perhaps he wanted it seen as one of the values of his work in *1958* and *1972* that the step to the notions and principles of the system for primitive recursive functionals of finite type would be just what is needed to go beyond finitary reasoning in the sense of Hilbert in order to capture arithmetic. So, for that it would be important to

---

[24] Tait describes this principle in a footnote, *loc. cit.* as follows: "Richard McKeon…preached…this very salubrious view in connection with the interpretation of the great philosophers, such as Plato: If your interpretation of them makes them fools…then it is likely you have more work to do." But one would hardly say that ascribing changing views to Gödel makes him out to be a fool.

[25] This obvious need has been emphasized by other commentators as well; cf. e.g. Zach (2003) and the further references there.

tie down Hilbert's conception precisely to have limit ε_0, by means, for example, of a characterization of the sort proposed by Kreisel.

The second question raised above is the more difficult one and to me the most intriguing of the whole affair: how does Gödel's well-known Platonic realism cohere with his engagements in the constructive consistency program and his claims for its necessity? One way is to deny that there is a genuine incoherence by denying the seriousness of that engagement. That is the position taken by Kreisel, who is utterly dismissive of it, first in his biographical memoir for the Royal Society (1980) and later in his piece, "Gödel's excursions into intuitionistic logic" (1987). In the former, he writes:

> [Gödel's] last self-contained publication, [*1958*]…was presented as a consistency proof. Between 1931 and 1958…he studied other such proofs… Very much in contrast to the break with traditional aims, advocated throughout this memoir, Gödel continued to use traditional terminology. For example, the original title of Spector (1962), extending Gödel (1958), did not contain the word 'consistency'; it was added for posthumous publication at Gödel's insistence. He knew only too well the publicity value of this catchword, which—contrary to his own view of the matter—had made the second incompleteness theorem more spectacular than the first. (Kreisel 1980, p. 174)

The reference is to Spector's posthumously published 1962 article, in which a functional interpretation of a system of 2nd order number theory was obtained by means of the so-called bar-recursive functionals. The "full story" of the title is recounted in Kreisel (1987):

> Spector's simple title was: *Provably recursive functionals of analysis.* Gödel did not find this exciting, and proposed the addition: a consistency proof of analysis. … Of course, I appreciated his flair for attracting attention, but my views about the sham of the consistency business have remained uncompromising. So, to water down his addition, I proposed the further qualification 'by an extension of

> principles formulated in current intuitionistic mathematics', to which Gödel agreed, albeit reluctantly. (Kreisel 1987, p. 161)

Kreisel also writes (*ibid.*, p.144) more generally that in later years "Gödel used crude, hackneyed formulations that had proved to have popular appeal (and had put me off…), [though] in his very early writing he was more austere." Given the accumulated evidence that has been surveyed here of Gödel's serious engagement with the constructive consistency program from the beginning to the end of his career, it seems to me this tells us more about Kreisel than about Gödel.[26]

A second way to deal with the *prima facie* incoherence of Gödel's Platonism with his engagement in the constructive consistency program is to question his retrospective claims of having held the Platonistic views back to his student days, at least if understood in their latter-day form; the case for their development prior to 1944 is made by Charles Parsons (1995). Also, through extensive quotation, Martin Davis (2005) has pointed out that they were by no means uniform. This could account for such things as the Cambridge lecture and the seminar at Zilsel's. But that doesn't account for the work on the *Dialectica* interpretation well past the point where he had made plain his adherence to full-blown set-theoretical realism.

A third possible way for a confirmed Platonist, such as Gödel, to make an engagement in the constructive consistency program is to recognize the additional epistemological value of success under that program. That is, even if the Platonist is convinced of the truth of his axioms, and hence their consistency, he could still appreciate the additional evidence of a different nature, that a constructive consistency proof would give.[27] In Gödel's case, this is supported by the final section of his Zilsel presentation that was quoted in full in section 4 above, whether or not he was a Platonist at the time. While that can be argued in principle, it seems to me curious that Gödel's own engagement in the program never went beyond arithmetic, where he can hardly have

---

[26] According to Gödel's communications to Hao Wang (1984, p. 654), an additional piece of evidence for the seriousness of his engagement with Hilbert's consistency program from the very beginning is that he was led to the incompleteness theorems through an aborted attempt to carry out the program for analysis (2nd order number theory).

[27] This possibility has been emphasized by Jeremy Avigad and Mark van Atten in personal communications.

thought that a constructive proof would make its consistency more evident than what is provided by the intuitive conception. But perhaps he believed that that was only a test case, and that a good start with arithmetic, such as that which he essayed with the *Dialectica* interpretation, could take the program much farther where intuition is no longer so reliable.

Let me venture a psychological explanation instead that goes back to what I suggested at the outset: Gödel simply found it galling all through his life that he never received the recognition from Hilbert that he deserved. How could he get satisfaction? Well, just as (in the words of Bernays) "it became Hilbert's goal to do battle with Kronecker with his own weapon of finiteness",[28] so it became Gödel's goal to do battle with Hilbert with his own weapon of the consistency program. And when engaged in that, he would have to do so—as he did—with all seriousness. This explanation resonates with the view of the significance of Hilbert for Gödel advanced in Ch. 3 of Takeuti (2003)[29]: he concludes that "Hilbert's existence had tremendous meaning for Gödel" and that his "academic career was molded by the goal of exceeding Hilbert." But Takeuti says that Gödel did not do any work subsequent to 1940 that was comparable to the work on completeness, incompleteness and the continuum hypothesis, perhaps, among other things, because "there was no longer the challenge to excel Hilbert." In that we differ: in my view, the challenge remained well into his last decade for Gödel to demonstrate decisively, if possible, why it is necessary to go beyond Hilbert's finitism in order to prosecute the constructive consistency program.[30]

Appendix

Here, in full, is a touching letter that Bernays wrote to Gödel on 24 April 1966, testifying to the highest intellectual and personal esteem in which Bernays held him:

Dear Mr. Gödel,

[28] Quoted in Reid (1970) p. 173.
[29] That was brought to my attention by Mark van Atten after reading a draft of this article.
[30] I wish to thank Jeremy Avigad, Charles Parsons, Wilfried Sieg, William Tait, and Mark van Atten, for their helpful comments on a draft of this article.

I feel quite embarrassed, since I had not prepared anything suitable for the *Festschrift* in honor of your 60th birthday—all the more so, as you made such a significant contribution to the *Dialectica* issues for my 70th birthday.

In any case, I would like now to express my cordial good wishes on your [just] completed decade of life.

In view of the situation in foundational investigations, you can certainly ascertain with much satisfaction that the discoveries and methods you brought to metamathematics are dominant and leading the way in the research of today. May it be granted to you also in the future to influence the direction of this research in a way that is frutiful.

Yet the foundations of mathematics are of course only one of the concerns of your research; and I would also like to wish that your philosophical reflections may turn into such results that you are induced to publish them.

Last [but] not least, I wish in addition that your general state of health in the coming years be as satisfying as possible. Hopefully you can celebrate your 60th birthday quite beautifully and take pleasure in the certainly impressive statement of the general appreciation of your intellectual work.

With very cordial greetings, also to your wife,

Yours truly,

Paul Bernays (*CW IV*, p. 251)

References

Awodey, Steven and André W. Carus (2001), Carnap, completeness and categoricity: the *Gabelbarkeitssatz* of 1928*, Erkenntnis 54*, 145-171.

Bernays, Paul (1930), Die Philosophie de Mathematik und die Hilbertsche Beweistheorie, *Blätter für deutsche Philosophie 4*, 326-367; reprinted in Bernays (1976a), 17-61; english translation in Mancosu (1998), 234-265.

_____________ (1935), Hilberts Untersuchungen über die Grundlagen der Mathematik, in Hilbert (1935), 196-216.

____________ (1967), Hilbert, David, in (P. Edwards, ed.) *Encyclopedia of Philosophy* vol. 3, Macmillan, New York, 496-504.

____________ (1976), Kurze Biographie, in (G. Müller, ed.) *Sets and Classes: On the work of Paul Bernays* (North-Holland, Amsterdam), xiv-xvi; English translation, op. cit., xi-xiii.

____________ (1976a), *Abhandlungen zur Philosophie der Mathematik* (Wissenschaftliche Buchgesellschaft, Darmstadt).

Davis, Martin (2005), What did Gödel believe and when did he believe it?, *Bull. Symbolic Logic* 11, 194-206.

Feferman, Solomon (1962), Transfinite recursive progressions of axiomatic theories, *J. Symbolic Logic 27*, 259-316.

Gentzen, Gerhard (1936), Die Widerspruchsfreiheit der reinen Zahlentheorie, *Mathematische Annalen 112*, 493-465; English translation in Gentzen (1969), 132-213.

_______________ (1969), *The Collected Papers of Gerhard Gentzen* (M. E. Szabo, ed.), (North-Holland, Amsterdam).

Gödel, Kurt (1931), Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik* 38, 173-198; reprinted with facing English translation in Gödel (1986), 144-195.

__________ (1931a), Diskussion zur Grundlegung der Mathematik, *Erkenntnis* 2, 147-151; reprinted with facing English translation in Gödel (1986), 200-205.

__________ (*1933o), The present situation in the foundations of mathematics, in Gödel (1995), 45-53.

__________ (*1938a), Vortrag bei Zilsel, in Gödel (1995) with facing English translation, 86-113.

__________ (*1941), In what sense is intuitionistic logic constructive?, in Gödel (1995), 189-200.

__________ (1958), Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica* 12, 280-287; reprinted with facing English translation in Gödel (1990), 240-251.

__________ (1972), On an extension of finitary mathematics which has not yet been used, in Gödel (1990), 271-280.

__________ (1986), *Collected Works, Vol. I: Publications 1929-1936* (S. Feferman, et al., eds.), (Oxford University Press, New York).

__________ (1990), *Collected Works, Vol. II: Publications 1938-1974* (S. Feferman, et al., eds.), (Oxford University Press, New York).

__________ (1995), *Collected Works, Vol. III: Unpublished Essays and Lectures* (S. Feferman, et al., eds.), (Oxford University Press, New York).

__________ (2003), *Collected Works, Vol. IV: Correspondence A-G* (S. Feferman, et al., eds.), (Clarendon Press, Oxford).

___________ (2003a), *Collected Works, Vol. V: Correspondence H-Z* (S. Feferman, et al., eds.), (Clarendon Press, Oxford).

Herbrand, Jacques (1931), Sur la non-contradiction de l'arithmétique, *J. für die reine und angewandte Mathematik 166,* 1-8; English translation in van Heijenoort (1967), 618-628.

Hilbert, David (1926), Über das Unendliche, *Mathematische Annalen 95*, 161-190; English translation in van Heijenoort (1967), 367-392.

___________ (1931), Die Grundlegung der elementaren Zahlenlehre, *Mathematische Annalen 104*, 485-494; English translation in Mancosu (1998), 266-273.

____________ (1931a), Beweis des tertium non datur, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Math.-Physik. Klasse,* 120-125.

____________ (1935), *Gesammelte Abhandlungen*, vol. 3 (Springer, Berlin).

Hilbert, David and Paul Bernays (1934), *Grundlagen der Mathematik,* Vol. I (Springer, Berlin); 2nd edition, 1968 (Springer, Berlin).

____________________________ (1939, *Grundlagen der Mathematik*, Vol. II (Springer, Berlin); 2nd edition, 1970 (Springer, Berlin).

Kreisel, Georg (1960), Ordinal logics and the characterization of informal concepts of proof, *Proc. International Congress of Mathematicians, 14-21 August 1958* (Cambridge University Press, Cambridge), 289-299.

____________ (1965), Mathematical logic, in (T. L. Saaty, ed.) *Lectures on Modern Mathematics,* vol. 3 (Wiley, New York), 95-195.

____________ (1970), Principles of proof and ordinals implicit in given concepts, in (J. Myhill, et al., eds.) *Intuitionism and Proof Theory* (North-Holland, Amsterdam), 489-516.

____________ (1980), Kurt Gödel, 28 April 1906-14 January 1978, *Biographical Memoirs of Fellows of the Royal Society 26*, 148-224; corrections, *ibid. 27*, 697 and *28*, 718.

____________ (1987), Gödel's excursions into intuitionistic logic, in (P. Weingartner and L. Schmetterer, eds.) *Gödel Remembered* (Bibliopolis, Naples), 65-186.

Mancosu, Paolo (1998), (ed.) *From Brouwer to Hilbert. The debate on the Foundations of Mathematics in the 1920s* (Oxford, New York).

Parsons, Charles (1995), Platonism and mathematical intuition in Kurt Gödel's thought, *Bull. Symbolic Logic 1,* 44-74.

_____________ (200?), Paul Bernays' later philosophy of mathematics, in (C. Dimitracopoulos et al., eds.) *Logic Colloquium 2005, Lecture Notes in Logic 28* (forthcoming).

Reid, Constance (1970), *Hilbert* (Springer, New York).

Spector, Clifford (1962), Provably recursive functionals of analysis. A consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, in (J. C. E. Dekker, ed.) *Recursive Function Theory* (American Mathematical Society, Providence), 1-27.

Tait, William (1961), Nested recursion, *Mathematische Annalen 143*, 236-250.

__________ (1968), Constructive reasoning, in (B. van Rootselaar and J. F. Staal, eds.), *Logic, Methodology and Philosophy of Science III* (North-Holland, Amsterdam), 185-199.

___________ (1981), Finitism, *The Journal of Philosophy 78*, 524-546; reprinted as Ch. 1 in Tait (2005).

___________ (2001) Gödel's unpublished papers on foundations of mathematics, *Philosophia Mathematica* (*series III*) *9*, 87-126; reprinted as Ch. 12 in Tait (2005).

___________ (2002), Remarks on finitism, in (W. Sieg et al., eds.), *Reflections on the Foundations of Mathematics. Essays in honor of Solomon Feferman*, Lecture Notes in Logic 15, 410-415; reprinted, with an Appendix, as Ch. 2 in Tait (2005).

___________ (2005), *The Provenance of Pure Reason* (Oxford University Press, New York).

___________ (2005a), Gödel's reformulation of Gentzen's first consistency proof for arithmetic: The no-counterexample interpretation, *Bull. Symbolic Logic 11*, 225-238.

___________ (2006), Kurt Gödel, Collected Works. Vol. IV, Selected Correspondence A-G, Vol. V, Selected Correspondence, H-Z, *Philosophia Mathematica* (*series III*) *14*, 76-114.

___________ (2006a), Gödel's interpretion of intuitionism, *Philosophia Mathematica* (*series III*) 14, 208-228.

Takeuti, Gaisi (2003), *Memoirs of a Proof Theorist. Gödel and other logicians* (World Scientific, Singapore).

Taussky-Todd, Olga (1987), Remembrances of Kurt Gödel, in (P. Weingartner and L. Schmetterer, eds.) *Gödel Remembered* (Bibliopolis, Naples), 29-48.

Turing, Alan (1939), Systems of logic based on ordinals, *Proc. London Mathematical Society* (ser. 2) 45, 161-228; reprinted in (M. Davis, ed.) *The Undecidable. Basic papers on undecidable propositions, unsolvable problems and computable functions* (Raven Press, Hewlett, NY, 1965).

van Atten, Mark (2006), *Brouwer meets Husserl. On the phenomenology of choice sequences*, Synthese Library vol. 335 (Springer, Berlin).

van Heijenoort, Jean (1967), *From Frege to Gödel: A source book in mathematical logic, 1879-1931* (Harvard University Press, Cambridge).

Wang, Hao (1981), Some facts about Kurt Gödel, *J. Symbolic Logic 46*, 653-659.

Zach, Richard (1998), Numbers and functions in Hilbert's finitism, *Taiwanese J. for Philosophy and History of Science 10*, 33-60.

____________ (2001), *Hilbert's Program. Historical, philosophical and metamathematical perspectives,* PhD Dissertation, University of California, Berkeley.

______________ (2003), Hilbert's Program, Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/hilbert-program/ .