

# Gödel's correspondence on proof theory and constructive mathematics

W. W. Tait \*

The volumes of Gödel's collected papers under review consist almost entirely of a rich selection of his philosophical/scientific correspondence, including English translations face-to-face with the originals when the latter are in German. The residue consists of correspondence with editors (more amusing than of any scientific value) and five letters from Gödel to his mother, in which explains to her his religious views. The term "selection" is strongly operative here: The editors state the total number of items of personal and scientific correspondence in Gödel's *Nachlass* to be around thirty-five hundred. The correspondence selected involves fifty correspondents, and the editors list the most prominent of these: Paul Bernays, William Boone, Rudolph Carnap. Paul Cohen, Burton Dreben, Jacques Herbrand, Arend Heyting, Karl Menger, Ernest Nagel, Emil Post, Abraham Robinson, Alfred Tarski, Stanislaw Ulam, John von Neumann, Hao Wang, and Ernest Zermelo. The correspondence is arranged alphebetically, with A-G in Volume IV. The imbalance results from the disproportionate size of the Bernays correrspondence: 85 letters are included (almost all of them), spanning 234 pages) including the face-to-face originals and translations). Each volume contains a calendar of all the items included in the volume together with separate calendars listing all known correspondence (whether included or not) with the major correspondents (seven in Volume IV and ten in Volume V).

Let me recommend to the reader the review of these same volumes by Paolo Mancosu in the *Notre Dame Journal of Formal Logic* 45 (2004):109-125. This essay very nicely describes much of the correspondence in terms of broad themes relating, especially, to the incompleteness theorems—their origins in Gödel's thought, their reception, their impact on Hilbert's program,

---

\*Charles Parsons read part of an early draft of this review and made important corrections and suggestions.

etc. I will do something less synoptic but perhaps complementary to Mancosu's essay: I will concentrate more on small details in the correspondence that seem to me worthy of further discussion; although there are many of these and I will not take up all or even most of them. In fact, I will restrict myself to the correspondence concerned with constructivity and constructive foundations of mathematics, and so, in particular, will not comment on the fairly large correspondence on set theory.

The introductory essays are a valuable part of these two volumes. Indeed, this is true of the preceding volumes of the collected works as well, but because of the more or less off-the-cuff nature of correspondence, I think that the interpretive and evaluative elements of these introductions take on more weight. For this reason, I will be more inclined to draw the introductions into my discussion of the texts than in my earlier critical review [Tait, 2001] of Volume III of these collected works.

## 1 Herbrand's Consistency Proof

Gödel's correspondence with Jacques Herbrand consists of a letter from Herbrand dated 7.4.31 and a reply dated 26.7.31 (one day before Herbrand died), both of which are included in Volume V.

The most striking thing about Herbrand's letter has to do with Gödel's mistaken memory of it. In [1934, p. 26, footnote 34], Gödel refers to a letter from Herbrand, presumably this one, which served as inspiration for his definition of the general recursive functions as functions computable from systems of equations. Herbrand also considers systems of defining axioms for functions. He places two conditions on the defining equations which Gödel seems not to have remembered: One, call it the *computability condition*, which does belong in the definition of the notion of general recursive function, is the requirement that the functions be computable from the defining axioms (without specifying, however, any set of rules of computation).<sup>1</sup> The second, call it the *provability condition*, which does not belong, is a requirement that satisfaction of the computability condition be intuitionistically provable. There is a valuable discussion of this in the introductory note by Wilfried Sieg. However, one thing puzzling about Gödel's memory of the letter is that, as he mentions in his reply to Herbrand, he had also read [Herbrand,

---

<sup>1</sup>Herbrand's defining axioms can be propositional combinations of equations, but these can always be resolved into sets of equations.

1931], which contains essentially the same description of the conditions on the defining axioms as in the letter. Had he forgotten about that paper? Or had he misremembered that as well?

Herbrand's letter for the most part summarizes [Herbrand, 1931]. The main result of that paper is this: Let the formulas be built up from equations involving 0, the successor function  $'$ , and finitely or infinitely many other function symbols  $f_0, f_1, \dots$ . Let  $\Gamma$  be a set of quantifier-free axioms, including the axioms for 0 and successor and defining axioms for the  $f_n$ , where the defining axioms for  $f_n$  contains  $f_k$  only if  $k \leq n$ . Let  $F$  denote the set of functions defined by these axioms and let  $\mathcal{T}(\Gamma)$  ( $1 + 2' + 3F$  in his notation) denote the theory in first-order logic with identity whose axioms are  $\Gamma$  and instances of the axiom schema

$$(1) \quad \phi(0) \wedge \forall x[\phi(x) \longrightarrow \phi(x')] \longrightarrow \forall x\phi(x)$$

of complete induction, where  $\phi$  is quantifier-free. Suppose that there is a deduction in this theory of the quantifier-free formula  $\theta$ . Then we may construct from it a deduction of  $\theta$  from  $\Gamma \cup \Delta$  which contains no quantifiers. The formulas in  $\Delta$  consist, for each instance (1) of complete induction, the definition by primitive recursion in  $\phi(x)$  of a function  $\epsilon(x)$  of  $x$  (and presumably the other parameters in  $\phi(x)$ ) which is 0 up to the first  $x$ , if any, such that  $\neg\phi(x)$  and has the value  $x$  thereafter, together with the axiom

$$\epsilon(x) = y' \longrightarrow \epsilon(y') = y' \wedge \epsilon(y) = 0$$

In fact, in sketching the deduction of  $\forall x\phi(x)$  from  $\phi(0)$  and  $\forall x[\phi(x) \longrightarrow \phi(x')]$ , Herbrand assumes that every number is 0 or a successor, and so to carry out his deduction, we need also to assume that  $\Delta$  contains the axioms

$$\text{pred}(0) = 0 \quad \text{pred}(x') = x \quad x \neq 0 \longrightarrow x = \text{pred}(x)'$$

These new axioms allow us to eliminate all instances of mathematical induction. The result then follows simply from Herbrand's fundamental theorem.

If we replace the axiom (1) by the equivalent rule of induction

$$\phi(0), [\phi(x) \longrightarrow \phi(x')] \Rightarrow \phi(t)$$

(with  $\phi(x)$  still quantifier-free) then the formulas  $\Delta$  are not needed and there is the straight forward result that  $\mathcal{T}(\Gamma)$  is conservative over its quantifier-free part  $\mathcal{T}(\Gamma)^*$ , rather easily obtained from Herbrand's fundamental theorem or

Gentzen’s (later) *Hauptsatz*. Whichever variant of the result we choose, it can be formalized in primitive recursive arithmetic *PRA*. Herbrand himself simply states that it is intuitionistically provable. In what follows, assume that  $\mathcal{T}(\Gamma)^*$  contains the rule of induction or, alternatively, just assume that  $\Gamma$  includes  $\Delta$ .

Herbrand assumes that the axioms  $\Gamma$  are intuitionistically valid and, in particular, that the defining axioms satisfy the provability condition. He argues that the intuitionistic validity of the axioms  $\Delta$  follows and, consequently, so does that of  $\theta$ . Thus, the intuitionist is free to apply first-order logic and complete induction for quantifier-free formulas to any set  $\Gamma$  of formulas he recognizes to be intuitionistically valid: any quantifier-free formula he derives in this way will be intuitionistically valid. In particular, there follows an intuitionistic consistency proof for  $\mathcal{T}(\Gamma)$ .

He gives three examples of systems  $\Gamma$  to which he believes to be intuitionistically valid and so to which his intuitionistic consistency theorem applies: (i) the defining axioms for addition and multiplication, (ii) the defining axioms for all primitive recursive functions, and (iii) the defining axioms for a class of functions containing non-primitive recursive functions such as the Ackermann function.

Herbrand goes on to explain why his consistency theorem does not conflict with Gödel’s second incompleteness theorem, but his explanation, both in the published paper and in his letter, is a bit awkward and Gödel did not entirely understand it. The explanation is that, in order to formalize his consistency proof for  $\mathcal{T}(\Gamma)$  in a theory of the form  $\mathcal{T}(\Gamma')$ , we need functions which are not explicitly definable from those in  $F$ , i.e.  $\Gamma' \not\subseteq \Gamma$ . He mentions, first, that in order to arithmetize the consistency proof, we must introduce a certain number of primitive recursive functions, beyond those explicitly definable from addition and multiplication, the functions in example (i), in order to arithmetize syntax. But in any case, we need to add more functions to  $F$ —he says that “this can be proved precisely: it is easy”, as indeed it is. In Herbrand’s argument, the consistency of  $\mathcal{T}(\Gamma)$  is reduced (in *PRA*) to that of  $\mathcal{T}(\Gamma)^*$ , and the argument that the latter is consistent is that its consequences are obtained by intuitionistically valid inferences from  $\Gamma$ , and  $\Gamma$  is intuitionistically valid. But to formalize the latter, “we have to number all objects occurring in proofs” [Herbrand, 1931] ([Herbrand, 1971, p. 296]). The most reasonable interpretation of this is that we need to enumerate the functions defined by terms of  $\mathcal{T}(\Gamma)$ ; in other words,  $\mathcal{T}(\Gamma')$  must contain an evaluation function for the terms in  $\mathcal{T}(\Gamma)$ , i.e. a function  $f$  of two variables

such that, for each term  $t$  of  $\mathcal{T}(\Gamma)$  containing the variables  $\vec{x}$ ,  $f([t], \langle \vec{x} \rangle) = t$ , where  $[t]$  is the Gödel number of  $t$ ,  $\vec{x}$  is a list of the distinct variables in  $t$  (in some standard order), and  $\langle \vec{x} \rangle$  is the sequence number of  $\vec{x}$ . As Herbrand points out in [1931], the diagonal argument shows that  $f$  cannot be explicitly defined from  $F$ .

Unfortunately, he can be understood to be saying that Gödel's result itself also depends upon there being an enumerating function in  $\mathcal{T}(\Gamma)$ . That is, he can be understood in that way if one ignores the fact that, in the paper, he had just gone (schematically) through Gödel's argument that in  $\mathcal{T}(\Gamma)$  one can prove that its consistency implies its Gödel sentence, without pointing to any difficulty in it.<sup>2</sup> Gödel, in his reply, stated that he did not completely understand what Herbrand meant, but then goes on himself to make the point: that the new functions are needed only to prove the consistency of  $\mathcal{T}(\Gamma)$ , not to prove the second incompleteness theorem. Perhaps he suspected that that is what Herbrand meant (Herbrand had apologized in his letter for his limited command of German): he suggested that in future correspondence, they each write in his mother tongue. (The response to this did indeed come in French: it was a letter from Herbrand's father, informing Gödel of his death. See Volume V, p. 24, fn. d.)

Here are some corrections of typos in the Herbrand correspondence: “or” for “oder” in two occurrences on p. 17 and “Principia Mathematica” for “Principia mathemtica” in four occurrences on p. 23.

## 2 Herbrand's Error and the Dreben Correspondence

The correspondence with Burton Dreben consists of five letters and a letter written by Dana Scott for Gödel. Two letters by Dreben, # 2 (5.24.66) and # 5 (4.15.70), one by Gödel, # 3 (7.19.66), and the letter by Scott, (Volume V, Appendix A, pp. 568-570), concern Gödel's claim that his decision procedure for satisfiability for the class of prenex formulas with prenex of the form  $\forall\forall\exists$  and without identity, the so-called *Gödel class*, can be extended to this class with identity. As we know (some of us quite painfully: the reviewer once

---

<sup>2</sup>He seems to have noticed that the second incompleteness theorem can be applied to systems such as *PRA* in which mathematical induction is restricted to quantifier-free formulas.

had a beautiful proof of solvability of the Gödel class with identity), Gödel was wrong here: Warren Goldfarb, who wrote the introductory note for the Dreben correspondence, proved that in [1984a].

Also of interest is #4 (12.30.69), Dreben's request for clarification from Gödel of a remark in his "Russell's mathematical logic" [Gödel, 1944] that (to quote Dreben) "Russell's definition of the truth functions as applied to propositions containing quantifiers 'proved its fecundity in a consistency proof for arithmetic.' " Gödel did not answer this letter, but Goldfarb observes that, at the end of lecture notes he wrote for a lecture in Princeton in 1941 on the *Dialectica* interpretation, he had added a note which answers Dreben's question. Namely, that interpretation, too, involves precisely the problem of interpreting the propositional connectives when applied to formulas containing quantifiers. Gödel's remark in the paper on Russell is in the context of discussing a certain on-again-off-again constructivist thread in *Principia Mathematica*. But Russell's interpretation is just an application of the De Morgan laws and so is not constructive in the ordinary sense. But one element that Russell's and Gödel's interpretations share is that they are both interpreting the propositional connectives as proposition-valued functions of propositions. The very formulation of Dreben's question, in which he refers to the "truth functions," may have precluded him from understanding Gödel's remark: The truth functions are defined by truth tables.

But I bring up the correspondence with Dreben in the context of the Herbrand correspondence because of letter # 1 (3.6.63), which concerns an error in the proof of a lemma ((3.3) in [Herbrand, 1930]) to Herbrand's fundamental theorem ("Herbrand's Theorem") in his thesis. Gödel seems to have noticed it in the early 1940s and to have found a satisfactory correction of the lemma. However, he did not announce his discovery and there was no general awareness of the error until it was rediscovered in 1962. Goldfarb mistakenly attributes the rediscovery to Dreben (p. 389) and the attribution is repeated in his introductory note (p. 306) to the van Heijenoort correspondence. The error was indeed rediscovered in 1962, but it was discovered by Peter Andrews, then a graduate student at Princeton. An exchange of letters and a conversation convinced Dreben that there was indeed an error in the proof. Shortly thereafter, Dreben sent Andrews a counterexample, and Stål Aanderaa, then a graduate student of Dreben, soon produced a simpler family of counterexamples, which are presented in [Dreben, Andrews and S.Aanderaa, 1963]. Given that Dreben had been working on Herbrand's thesis for some years prior to Andrew's discovery, and that, once discovered,

the production of counterexamples followed in quite short order, it would seem that the discovery itself was of sufficient significance to have merited explicitly crediting Andrews with it.<sup>3</sup>

In his thesis, Herbrand gives an argument that the lemma (3.3) implies Lemma 3 (5.3), which has as an easy consequence that, given a deduction of height  $m$  of a sequent in the classical sequent calculus, there would be a cut-free deduction of it of a height  $h$  which is polynomial in  $m$ . Now [Gentzen, 1935] did not give bounds in the proof of his *Hauptsatz* (that cuts can be eliminated in predicate logic), but an understanding of his proof would certainly lead one to believe that  $h$  is exponential in  $m$ ; and it seems quite likely to me that this is how Gödel came to discover Herbrand's error. In any case, the proof of cut-elimination in [Schütte, 1951] for Peano arithmetic with the  $\omega$ -rule applies directly to predicate logic, and gives a bound for  $h$  which is exponential in  $m$ —namely  $h \leq f^n(m)$ , where  $n$  is the least number greater than the logical complexity of the cut formulas and  $f(x) = 2^x$ . Actually, there is the slightly better bound  $h \leq g^n(m)$ , where  $g(0) = 0$  and  $g(x + 1) = 2^x$ . I don't know whether this bound is optimal, but it is known that the bound must in general be exponential in  $m$ .

The simplest proof of this I know consists of finite versions of Gentzen's proofs of the derivability in first-order arithmetic of induction up to each  $\omega_{(k)}$  on the one hand and of the non-derivability of induction up to  $\epsilon_0$  on the other, and parallels an analogous result for the system of natural deduction in [Orevkov, 1979]. Let  $\Gamma$  consist of the axioms of identity, the universal closure of recursion equations for addition and exponentiation to base 2 ( $2^0 = 1$  and  $2^{x'} = 2^x + 2^x$ ), and the axioms  $N\bar{0}$  and  $\forall x[Nx \rightarrow Nx']$ . It is easy to show that there is a cut-free proof of the sequent  $\Gamma \Rightarrow N\bar{k}$  for each  $k$  and that the least height of such a proof is linear in  $k$  but  $\geq k$ . There are also deductions of  $\Gamma \Rightarrow Nx \wedge N0 \rightarrow Nx + 0$  and  $\Gamma, [Nx \wedge Ny \rightarrow Nx + y] \Rightarrow Nx \wedge Ny' \rightarrow Nx + y'$ . So by mathematical induction,  $\Gamma \Rightarrow \forall xy[Nx \wedge Ny \rightarrow Nx + y]$ . Using this, we obtain a deduction of  $\Gamma \Rightarrow Nx \rightarrow N2^x$  by mathematical induction. Substituting the numeral  $\bar{k}$  for  $x$ , the mathematical inductions involved reduce to finite sequences of at most  $k$  cuts, and we have a deduction in predicate logic of  $\Gamma \Rightarrow N\bar{k} \rightarrow N\overline{2^k}$  of height polynomial in  $k$ . So, by a cut, there is a proof of  $\Gamma \Rightarrow N\overline{2^k}$  of height polynomial in  $k$ ; but there is no

---

<sup>3</sup>An account of the discovery, including Andrews' correspondence with Dreben, can be found in [Andrews, 2003], which is also obtainable from his website [<http://gtps.math.cmu.edu/andrews.html>].

cut-free proof of it of height  $< 2^k$ . (Iterating this, we even obtain deductions of  $\Gamma \Rightarrow N\overline{2^{2^k}}, \Gamma \Rightarrow N\overline{2^{2^{2^k}}}$ , etc., of height polynomial in  $k$ .) Eliminating individual constants and function constants in favor of a purely relational language is routine and changes bounds only linearly.<sup>4</sup>

In any case, after Schütte’s result, at least, and certainly by the late 1950’s, anyone reading Herbrand’s Lemma 3 should have been suspicious. Probably the error escaped public notice for so long simply because Herbrand’s paper itself, important though it was at that time, was badly written and seems not to have been widely studied. Alternative proofs of (versions of) Herbrand’s Theorem were soon available, using either the epsilon theorems or Gentzen’s *Hauptsatz*.<sup>5</sup>

### 3 The two meanings of “Intuitionism”

There are three questions arising from the Herbrand correspondence concerning the term “intuitionistic,” which Herbrand uses throughout his letter as well as in his 1931 paper to indicate the constructive character of his results. One question is: What did Herbrand mean by the term? Another is: What did Gödel understand it to mean? A third question is: How was it generally understood at that time?

In the final remark (a) of [1931] (p, 296-297 in [Herbrand, 1971]), Herbrand writes that “it seems to us almost certain that every intuitionistic argument can . . . be carried out in an arithmetic [of the form  $\mathcal{T}(\Gamma)^*$ ].” In particular, it would follow that mathematical induction is restricted to quantifier-free formulas in intuitionistic arguments. On pages 2-3 of letter # 1 (p. 17), he repeats this statement about intuitionistic proofs without any expression of uncertainty. But this is clearly not the intuitionism whose logic (admitting all propositions built up from atoms by means of propositional connectives and quantifiers) had already been formalized in [Heyting, 1930]. In [Herbrand, 1931a], Herbrand relates intuitionism to the position of Brouwer, but then describes it “in its extreme form” in terms that, as Gödel noted in his correspondence with von Heijenoort (letter # 11), are closer to finitism. In fact, Gödel states that “taken in the sense that Herbrand probably had

---

<sup>4</sup>Although I haven’t carefully checked, these examples also seem closely related to Aanderaa’s family of counterexamples to Herbrand’s lemma (3.3).

<sup>5</sup>See Goldfarb’s discussion of this in his introduction to [Herbrand, 1971].



in mind, it is the correct demarcation of finitism within intuitionism.”<sup>6</sup>

Herbrand’s letter was addressed from Berlin, where he was working with, among others, von Neumann. In [von Neumann, 1927], the latter also uses the term “intuitionistic” in connection with Hilbert’s proof theory and he states explicitly that he is using it in the sense of Brouwer and Weyl. On the other hand, when he begins in §C1 to discuss consistency proofs, he states that all the inferences should be “intuitionistic (i.e. *finite*)” (my emphasis). So it seems reasonably clear that Herbrand was using the term “intuitionism” in his letter and in [Herbrand, 1931] to refer to the “extreme form” of intuitionism and that he got this usage from von Neumann, a conclusion that is endorsed by van Heijenoort in his letter # 14 to Gödel. It also seems reasonable to conclude that they both intended this usage to coincide with what Hilbert referred to as finitary reasoning, whether or not that is in fact the case.

But why did von Neumann use the term in this way? Some confusion over this question has been caused by van Heijenoort’s statement in his introductory note to [Herbrand, 1931] in [von Heijenoort, 1967] that “the identification of ‘intuitionistic; with ‘finitist’ was then current among members of the Hilbert school” and, also published in 1967, Bernays’ statement in his *Encyclopedia of Philosophy* article on Hilbert, in reference to Gödel’s and Gentzen’s interpretation of Peano arithmetic  $PA$  in Heyting arithmetic  $HA$ , that “In this way it appeared that intuitionistic reasoning is not identical with finitist reasoning, contrary to the prevailing views at that time.” (Presumably these two statements are independent of each other.) The only example van Heijenoort cites for his assertion is von Neumann’s use of “intuitionistic” in [von Neumann, 1927] and, as far as I know, no member of Hilbert’s school (other than von Neumann and Herbrand) used the term in this way. The date of publication of Gödel’s interpretation of  $PA$  in  $HA$  (Gentzen’s was not published) is 1932. But by that time, Hilbert had long rejected quantified propositions as finitist and Heyting had already published his formalization of intuitionistic logic admitting quantifiers. How can it be that people were not aware of the difference?

---

<sup>6</sup>Strangely, Gödel infers from this that, in his letter and in [1931], Herbrand is referring to intuitionism in Brouwer’s sense—presumably because otherwise he would have modified his use of the term with “in its extreme form”. But in both documents (with or without some expression of uncertainty) Herbrand asserts that every intuitionistic argument is formalized in some  $\mathcal{T}(\Gamma)^*$ . Surely this indicates that he was thinking of intuitionism in the narrow sense.

I believe that it was not a case of identification, but of *ambiguity*—that there were two conceptions of arithmetic reasoning at that time, recognized as distinct, one more restrictive than the other, and both called intuitionistic by von Neumann and Herbrand.

Like Herbrand, Gödel was entirely clear about the distinction between the wider sense of the term ‘intuitionism’ and the narrow sense. In his reply to Herbrand’s letter he seems to be leaving it open how to interpret the latter’s usage. For example, as we will see below, he questions Herbrand’s assertion that intuitionistic proofs are always proofs in some  $\mathcal{T}(\Gamma)^*$ . In a footnote he points out that Brouwer and Heyting admit proofs which go beyond what can be formalized in such systems (presumably because they would admit arbitrary quantification over numbers and, in particular, mathematical induction applied to formulas containing numerical quantifiers), and then goes on to doubt Herbrand’s assertion even if “intuitionistic” is interpreted more narrowly.

A name surprisingly missing from much of the discussion of this issue is that of Weyl. In the early 1920s, he was converted to many of the themes of Brouwer’s intuitionism, such as Brouwer’s analysis of the continuum and the rejection of excluded middle. (See [Weyl, 1921].) But in fact his view was more radical than Brouwer’s. (See [van Dalen, 1995] and [Mancosu, 1996, pp. 76-79].) In particular, his view leads to the conclusion that all genuine propositions in arithmetic or analysis are purely universal propositions about numbers or choice sequences, respectively, i.e. they can be expressed by quantifier-free formulas. In fact, these universal propositions are themselves propositions only in the limiting sense that their assertion stands for the claim to have a rule  $r$  which, applied to an arbitrary system  $\vec{n}$  of values of the free variables, yields a proof  $r(\vec{n})$  of the corresponding instance; and so they cannot be combined by means of quantifiers or propositional connectives into other propositions. (Both the functions that occur in the equations and the rules  $r$  are to be obtained by explicit definition and primitive recursion.) Existential propositions are also not genuine propositions: the proposition  $\exists y\phi(x, y)$ , for example, stands for the claim to have a function  $f$  and a proof of an arbitrary instance of  $\phi(xf(x))$ . In the context of arithmetic, this certainly agrees with the position stated in [Hilbert, 1926]. In fact, concerning arithmetic, Weyl’s intuitionism would seem to coincide with what can be formalized in *PRA* and so to fall within the scope of Hilbert’s finitism. (His analysis, involving the use of choice sequences and the principle of bar recursion, has rather the proof-theoretic strength of first-order arithmetic.

See §7 below.)

Von Neumann, for the three years preceding his appointment in Berlin, had been at ETH in Zurich, where he interacted with Weyl. Although Weyl's views on arithmetic diverged from the intuitionism of Brouwer, as formalized by Heyting, he identified his views *as* intuitionism. It seems quite likely, then, that von Neumann, when speaking of intuitionism, was simply adopting Weyl's terminology—although, in as much as he accepted intuitionistic methods that go beyond *PRA*, I don't believe that his ideology was the same as Weyl's.

## 4 Finitary Number Theory

Sieg's introductory note to the Herbrand correspondence refers to the example (iii) to show that Herbrand takes the concept of finitist function to go beyond the primitive recursive functions. It does so indeed, assuming (as I believe we can) the equivalence of finitism and intuitionism in the 'extreme' sense for him in this context. But the example (ii) already shows that. As noted above, following Herbrand, in order for his consistency proof to apply to example (ii), the enumeration function for the primitive recursive functions must be regarded as intuitionistic.

In connection with this, a remark is needed about Herbrand's assertion that every intuitionistic proof can be carried out in some  $\mathcal{T}(\Gamma)^*$  and that, conversely, every such proof is intuitionistic. Clearly, in this characterization of intuitionism, the provability condition on  $\Gamma$  cannot be in play. Otherwise, in order to prove anything intuitionistically in the language of some  $\Gamma$ , we would have first to have proved intuitionistically that we can compute unique values from the defining axioms of the evaluation function for the terms of  $\Gamma$ —and we would be in a circle. In his reply to Herbrand's letter (p. 3) Gödel remarks:

For even if we admit that every intuitionistic proof can be carried out in one of the systems  $[\mathcal{T}(\Gamma)^*]$  (which seems not at all obvious to me), the question still always remains open whether the intuitionistic proofs that are required in each case to justify the unicity of the recursion axioms are all formalizable in *Principia Mathematica*.

So Herbrand's characterization of intuitionistic reasoning in arithmetic as

free-variable reasoning from systems of defining axioms for functions, from which unique values of the functions are computable cannot carry with it (on pain of circularity) any condition that we have a proof of this computability. This situation illustrates the conceptual difficulty in analyzing the notion of finitist or (in the narrow sense) intuitionistic reasoning when one wishes to take this notion to go beyond *PRA*.

The latter system itself, to which Gödel refers in [Gödel, 1938a] as *finitary number theory*, is characterized by two moments. One is that the basic principle of definition of functions, aside from explicit definition, is definition by iteration or primitive recursion. This principle follows from the idea of an arbitrary number. It is simply the principle that we can uniquely transfer the finite iteration given by the arbitrary number  $n$  from the successor function starting at 0 to any operation  $\Phi$  on a domain  $D$  of objects starting at  $a$ , yielding  $\Phi^n(a)$ , where  $a$  is any object in  $D$ . (Mathematical induction is just primitive recursion applied to a domain of proofs. The reduction of the general form of primitive recursion, including mathematical induction, to pure iteration is discussed in [Tait, 2005b, pp. 57-58]. and in [Tait, 2005a]). The usual axioms of 0 and successor (which are derivable when we take  $\neg\phi$  to abbreviate  $\phi \rightarrow 0 = 1$ ) simply express the fact that the arbitrary iteration given by  $n$  is ‘free’ (i.e., contains no loops), so that it can operate as an iterator of an arbitrary operation. It seems reasonable to hold, with Weyl,<sup>7</sup> that definition by iteration, together with 0 and the successor operation, simply express what we mean by number, and itself needs no foundation. Thus, in the case of primitive recursion definition, there is no question of proving that the recursion equations define a function, in the sense that a unique value is determined for each argument. These equations merely express the fact that we can uniquely construct  $\Phi^n(a)$  for arbitrary  $n$ ; and this is contained in the very idea of an arbitrary number.

The other moment of *PRA* is that we apply iteration or primitive recursion only to operations on domains  $D$  of finite objects—call them *finitary domains*. (Since proofs of closed numerical equations are computations and these are finite objects, proof by mathematical induction in *PRA* is justified on these grounds.) These two moments determine *PRA*, and this explains its special ‘minimal’ role and why it deserves the title “finitary” (independently of any historical uses of this term): it contains the principle that defines our

---

<sup>7</sup>See [Weyl, 1921], the end of Part II §1 (“The Basic Ideas”) and the first paragraph of §2a (“*Functio Discreta*”), and also [Weyl, 1949, p. 33].

concept of number, but applies it only to operations on finitary domains.

*PRA* has interesting subsystems such as the system based on the Kalmar elementary functions, but they all involve restrictions on iteration to some special class of operations on finitary domains and so, at least from our point of view, are arbitrary.

Extensions of *PRA* are of two kinds: in one, we admit iteration applied to operations on domains which contain infinite objects—call them *infinitary domains*. For example, we may apply iteration to operations on the domain of functions of some finite type (over the domain  $\mathbb{N}$  of numbers), obtaining Gödel’s theory  $T$  of impredicative primitive recursive functions of finite type, or on the domains of objects of type  $\phi$  (in the sense of the Curry-Howard theory of propositions-as-types), where  $\phi$  is a sentence of *HA*, obtaining the proofs of *HA*. These extensions still follow from our basic conception of number as arbitrary free finite iteration, but they give up all pretense to finiteness.

The other kind of extension of *PRA*—and this is the kind relevant to the present discussion of intuitionism in the narrow sense or finitism—retains the restriction to finitary domains, but admits principles of definition of functions on them that cannot be obtained solely by iteration of operations on them. The main examples in the literature are functions (such as the Ackermann function) defined by manifold nested recursions. But on what grounds do we admit such definitions? In the case of manifold nested recursion, we can obtain the function  $f$  in question by applying iteration to operations on either one of Gödel’s finite types or on the Curry-Howard type  $\phi$  for suitable arithmetic sentence  $\phi$ . It is also possible to reduce  $k + 1$ -fold nested recursion ( $k \geq 0$ ) to ordinary (i.e. un-nested) recursion on (a natural ordering of  $\mathbb{N}$  of order type)  $\omega^{(\omega^k)}$ . But then, as with the case of the original definition by nested recursion, we may ask on what grounds should we accept such definitions by transfinite recursion: to obtain them by iteration, we again need to iterate operations on infinitary domains.

To the response that we just see intuitively that this or that is a valid means of defining functions, that we can see that a value can always be computed from the definition, it seems justified to demand that the intuition be transformed into a proof. For the basis of the intuition must be the principles that define our conception of number, 0 and successor and iteration. To see intuitively the validity of a particular definition of a numerical function therefore must mean to see intuitively that it follows from these principles. But what can “follows from” mean other than that “it can be proved from”?

When such a proof has been given to show that a certain non-primitive recursive function is defined and computable, we will see that it involves applying iteration to operations on infinitary domains. This demand for proof applies not only to definition by manifold nested recursions, but to *any* candidate for an intuitionist or finitist function defined by a system of equations. If it is not defined by iteration and explicit definition, then we should demand a proof that the equations are satisfied. Otherwise, the question of what, beyond the primitive recursive functions, is to be admitted as a finitist or intuitionistic function would seem to be entirely subjective, depending upon one’s personal intuitions. I will argue that this is precisely the position that Gödel attributes to Hilbert.

As a matter of fact, in letter #69 (7 January 1970) Bernays himself manifests this conception of finitism when he suggests that nested recursion on  $\omega^2$  is finitary in the same sense that ordinary recursion on  $\omega$  is “i.e., if one views them as a description of computation procedures for which it can be seen that the function determined by the specific procedure satisfies the recursion equations . . .” But the nature of his “it can be seen” becomes clear when he then goes on to make the insight more concrete by regarding the computation of the function  $f(m, y)$  defined by nested recursion on  $\omega^2$  (where the argument  $(m, y)$  represents the ordinal  $\omega \cdot m + y$ ) as the step-by-step computation of the functions  $f_m$ , where  $f_m(y) = f(m, y)$  and  $f_{m+1}$  is defined by primitive recursion from  $f_m$ . But the insight that  $f(x, y)$  can be computed for arbitrary  $x$  and  $y$ , is really the insight that  $f_{x+1}(y)$  can be computed if  $f_x(z)$  can be computed for arbitrary  $z$ . Thus the insight becomes a non-finitist construction of the function  $f$ : namely,  $f_0$  is a given function and  $f_{x+1} = G(x, f_x)$ , where  $G$  is a function of the *infinitary* type  $\mathbb{N} \rightarrow [(\mathbb{N} \rightarrow \mathbb{N}) \rightarrow (\mathbb{N} \rightarrow \mathbb{N})]$ . (This is discussed further in [Tait, 2002] and in the Appendix to [Tait, 2005b].) In this case, what is ‘seen’ is, in outline, a *proof*. But by suppressing the details of the proof, one avoids noticing exactly what is involved in it. We will encounter another case of this in §10 in connection with Gödel’s contention that we see immediately that his primitive recursive functions of finite type are computable (in his sense).

Somehow, in connection with questions of constructivity, although the very term should serve as a warning, the question of what can be computed has been substituted for the proper question of what can be constructed.<sup>8</sup>

---

<sup>8</sup>Once one takes construction as the central notion, one sees why (or at least one has a sense in which) Markov’s principle is not constructive: If  $F(x)$  is a decidable arithmetic

Extensions of *PRA* which are ‘finitary’ only in the sense that they are obtained by adding only defining equations from which non-primitive recursive numerical functions can be computed, such as manifold nested recursive functions, without explicitly adding iteration of operations on infinitary domains, are *foundationally incomplete*: We may “see” that the new functions defined by the equations are computable, but they lack the means of defining them from the basic principles.

I should emphasize that this argument involves a *conceptual* analysis of what finitist *should* mean, in terms of whether or not iteration is applied (implicitly or explicitly) only to operations on finitary domains. It does not confront the historical question of what this or that person meant by “finitism” at this or that time. But it should be noted that in [Weyl, 1921] the description of the foundations of arithmetic would seem to limit intuitionistic arguments in this field to precisely those formalized in *PRA*. Likewise, what Gödel refers to as ‘finitary number theory’ in [Gödel, 1938a] is, as we have already noted, precisely *PRA*.

## 5 Nested Recursion

Interestingly, there is some indication that von Neumann thought in 1930 that all intuitionist functions (in the narrow sense) can be obtained by manifold nested recursions—an apparently stronger thesis than Herbrand’s. At least, in his letter # 2 to Gödel (11.29.30), he writes

I believe that every intuitionistic consideration can be formally copied, because the “arbitrarily nested” recursions of Bernays-Hilbert are equivalent to ordinary transfinite recursions up to appropriate ordinals of the second number class. This is a process that can be formally captured, unless there is an intuitionistically definable ordinal of the second number class that could not be defined formally—which is in my view unthinkable. Intuitionism clearly has no finite axiom system, but that does not prevent its being a part of classical mathematics that does have one.

---

predicate and one knows that  $\neg\forall xF(x)$  is true, then we can compute an  $n$  such that  $\neg F(n)$ ; but from a proof of  $\neg\forall xF(x)$  we cannot construct such an  $n$ . See [Tait, 2005a] for further discussion of this.

But he overestimates the power of the “arbitrarily nested” recursions.<sup>9</sup> Determining that was precisely the problem investigated in [Tait, 1961]. Nested or ordinary (i.e. un-nested) recursion on  $k$  variables can be reduced to nested or ordinary recursion, respectively, on (the natural well-ordering of  $\mathbb{N}$  of type)  $\omega^k$ . (Oddly, Bernays, in his letter # 69 to Gödel (1.7.70), seems not to have understood this point.) Nested recursion<sup>10</sup> on  $\alpha$  can be reduced to ordinary recursion on  $m^\alpha$ , where  $m < \omega$  is effectively determined from the syntactical form of the definition.<sup>11</sup> Thus, the  $k + 1$ -fold nested recursions in [Hilbert and Bernays, 1934] are all reducible to ordinary recursion on  $\omega^{(\omega^k)}$  ( $k \geq 0$ ). Conversely, ordinary recursion on  $\omega^\alpha$  can be reduced to nested recursion on  $\omega \times \alpha$ . Thus, ordinary recursion on ordinals less than  $\omega^{(\omega^\omega)}$  would seem to characterize von Neumann’s conception of intuitionism in his letter. Obviously he would not have been happy with that bound.

## 6 Well-Foundedness of Exponentiation

Another oddity concerning this subject in the Bernays correspondence is that Gödel, in letter 68b, states that [Tait, 1961] shows that “one gets up to  $\epsilon_0$  with nested recursions”. I had thought (and still think) that, as we just noted in connection with von Neumann, it established the bound  $\omega^{(\omega^\omega)}$  for what one can get up to with manifold nested recursions. From the context in Gödel’s letter, namely a discussion of a very nice proof by Bernays that every descending sequence of ordinals from  $\omega^\alpha$  is finite, using (when it is spelled out) nested recursion on  $\omega \times \alpha$ , it seems that he must be referring to the (less elegant) proof of that in [Tait, 1961]—it was the crux of the reduction of ordinary recursion on  $\omega^\alpha$  to nested recursion on  $\omega \times \alpha$ . Bernays’ proof

---

<sup>9</sup>I am assuming here that he is referring to  $k$ -fold nested recursion for some  $k$ . I do not know the reference to Bernays and Hilbert in the letter, and it is conceivable that he is referring in general to systems of equations that define what we old-fashion folks would call “general recursive functions”. But, in that case, I am not sure how the assignment of ordinals to the systems goes.

<sup>10</sup>The function  $f$  is defined by nested recursion on  $\alpha$  if the definiens of  $f(\vec{x}, y)$  contains  $f$  only in contexts of the form  $f(\vec{s}, [t]_y)$ , where  $[z]_y$  is defined to be  $z$  if  $z$  is less than  $y$  in the standard ordering of  $\mathbb{N}$  of order type  $\alpha$  and is 0 otherwise. Here the terms in  $\vec{s}$  and the term  $t$  may themselves contain occurrences of  $f$ . This would seem to be the most general form of nested recursion on an ordinal, and certainly includes the so-called  $k$ -fold nested recursions as nested recursions on  $\omega^k$ .

<sup>11</sup>In my paper, I stated that it reduced to ordinary recursion on  $\omega^\alpha$ , but in the proof, this can be replaced by  $m^\alpha$  for suitable  $m$ .



(a slight simplification of which I present below) is in the second edition of [Hilbert and Bernays, 1939] (pp. 533-535). Both Bernays and Gödel take it as a proof that descending sequences of ordinals  $< \epsilon_0$  are always finite; but that is a mistake. For setting  $\omega_{(0)} = \omega$  and  $\omega_{(n+1)} = \omega^{\omega_{(n)}}$  (so that  $\epsilon_0 = \lim_n \omega_{(n)}$ ), the result only shows that descending sequences from  $\omega_{(n+1)}$  are finite by appealing to *nested* recursion on  $\omega_{(n)}$  (or on  $\omega^2$  when  $n = 0$ ). Now, from the fact that every descending sequence from  $\omega_{(n)}$  is finite, we can derive in *PRA* the principle of *ordinary* recursion on  $\omega_{(n)}$ , but not the principle of nested recursion (of the appropriate kind). Therefore, the implicit complete induction step, assumed by both Bernays and Gödel, from no infinite descending sequences from  $\omega_{(n)}$  to no infinite descending sequences from  $\omega_{(n+1)}$ , is not established by Bernays' (or my, or any) argument. Gödel explicitly notes in the letter that Bernays' construction seems to involve a nested recursion; but his concern with this is only that nested recursions "are not finitary in Hilbert's sense (i.e., not intuitive)". It seems not to have occurred to him that the need for a nested recursion invalidates the proof as a proof of the well-foundedness of  $\epsilon_0$ .

Let me stop here to present Bernays' proof: It is simpler even than his own exposition makes it out to be. The (Cantor) normal form for  $\alpha < \epsilon_0$  is

$$\alpha = \omega^{\alpha_1} \cdot a_1 + \cdots + \omega^{\alpha_n} \cdot a_n$$

where

$$\alpha > \alpha_1 > \cdots > \alpha_n \quad 0 < a_1, \dots, a_n < \omega.$$

Let  $\mathbf{T}(\gamma)$  be the set (call it a "spread" if you like) of all  $f : \omega \rightarrow \gamma$  (call them "choice sequences", if you like) such that  $f(n) > 0 \rightarrow f(n+1) < f(n)$ . Call a function  $F : \mathbf{T}(\gamma) \rightarrow \omega$  a *well-founding function* for  $\mathbf{T}(\gamma)$  iff  $f(F(f)) = 0$  for all  $f \in \mathbf{T}(\gamma)$ .

The crux of Bernays' proof is simply this: If  $f \in \mathbf{T}(\omega^\alpha \cdot (n+1))$ ,  $n > 0$ , let  $f'$  be the element of  $\mathbf{T}(\omega^\alpha \cdot n)$  defined by

$$f'(m) = f(m) - \omega^\alpha \cdot n$$

In other words, if  $\omega^\alpha \cdot n$  is the leading term in the normal form of  $f(m)$ , then it is dropped to obtain  $f'(m)$ . If  $f(m) < \omega^\alpha \cdot n$ , then  $f'(m) = 0$ . Suppose that  $f'(m) = 0$ . Then in any case  $f(m) < \omega^\alpha \cdot n$ , so that  $\lambda x f(m+x) \in \mathbf{T}(\omega^\alpha \cdot n)$ .<sup>12</sup> For  $0 < \alpha < \epsilon_0$  and  $n < \omega$ , we define a well-founding function  $F_{\alpha,n}$  for  $\mathbf{T}(\omega^\alpha \cdot (n+1))$  by nested recursion on  $\omega \cdot \alpha + n$ :

<sup>12</sup>In carrying out his argument, Bernays restricts the functions  $f \in \mathbf{T}(\gamma)$  to ones with

- $F_{1,0}$  is to be a well-founding function for  $\mathbf{T}(\omega)$ . Let  $G(f, n) = 1 + G(\lambda x f(1 + x), f(1))$  if  $n = f(0) > f(1)$ , and  $G(f, n) = 0$  otherwise. Then  $F_{1,0}(f) = G(f, f(0))$ . So  $F_{1,0}(f)$  is primitive recursive in  $f$ .
- $F_{\alpha,n+1}(f) = F_{\alpha,n}(f') + F_{\alpha,n}(\lambda n f(F_{\alpha,n}(f') + n))$ .  
Notice that there is a nested recursion in this case.
- For  $\alpha > \omega$

$$F_{\alpha,0}(f) = F_{\delta,n+1}(f)$$

where  $\omega^\delta \cdot n$  is the leading term in the Cantor normal form of  $f(0)$ .

A similar argument proves well-foundedness of  $\gamma^\alpha$  by nested recursion on  $\alpha \times \gamma$  (using Cantor's normal form to the base  $\gamma$  in the nontrivial case that  $\gamma > 1$ ).

## 7 Choice Sequences

In his earlier letter #68a (7.69) to Bernays, in connection with the latter's 'proof' of the well-foundedness of  $\epsilon_0$ , Gödel is more positive about the proof and suggests that the proof uses the notion of choice sequence. He means that  $\mathbf{T}(\gamma)$  may be regarded as a spread of choice sequences in Brouwer's sense. If one takes as the mark of the distinction between reasoning about numerical functions and reasoning about choice sequences that functions defined on a spread of the latter should satisfy the bar theorem, i.e., their unsecured sequences should be well-ordered,<sup>13</sup> then Bernays' definition of a well-founding  $f(0)$  of the form  $\omega^\alpha \cdot a$  and such that, if

$$f(m) = \omega^{\alpha_1} \cdot a_1 + \cdots + \omega^{\alpha_n} \cdot a_n$$

(in normal form) then either

$$f(m+1) = \omega^{\alpha_1} \cdot a_1 + \cdots + \omega^{\alpha_i} \cdot a_i$$

or

$$f(m+1) = \omega^{\alpha_1} \cdot a_1 + \cdots + \omega^{\alpha_i} \cdot a_i + \omega^\beta \cdot b$$

(in normal form) for some  $i < n$ . He notes that any  $g \in \mathcal{T}(\gamma)$  as we have defined it can be transformed into one satisfying this condition by interpolating some steps before each step of  $g$ , so that a well-founding function for  $\mathcal{T}(\gamma)$  in this restricted sense is one in our sense. But this does not really simplify the argument.

<sup>13</sup>I would argue that, unless one injects entirely subjective ideas into mathematics, it is the *only* mark. See [Tait, 2005b, Introduction §4].

function for  $\mathbf{T}(\omega^\alpha)$  by nested recursion on  $\omega \times \alpha$  satisfies this condition. Gödel goes on to say “If one reckons choice sequences to be finitary mathematics, your proof is even finitary.”<sup>14</sup> He asks whether Bernays would agree with this observation and suggests including it in a footnote to the revised and translated *Dialectica* paper. In the event, the observation does not appear in [Gödel, 1990b]. Gödel returns to the subject of choice sequences in letter # 68b (7.25.69) and asks Bernays “*Hilbert, I presume, didn’t want to permit choice sequences?*” (emphasis Gödel’s). Bernays replies in letter # 69 that, as far as he knew, Hilbert never took a position at all on choice sequences.

Remembering that we are restricted to quantifier-free formulas, bar induction can only be the principle that, if we have constructed a numerical-valued function  $F$  of numerical function arguments, then we may define new functions by recursion on the ordering of its unsecured sequences. But [Tait, 1965] proves that if we consider the second-order extension of *PRA* by allowing numerical function (‘choice sequence’) variables as parameters in definitions of functions, then adding this form of bar recursion yields the same numerical functions as adding definition by  $\alpha$ -recursion for each  $\alpha < \epsilon_0$ . (Incidentally, it is because of this result and because I believe that Weyl’s intuitionistic arithmetic is *PRA* that I asserted in §3 that his analysis is proof-theoretically equivalent to first-order arithmetic.)

## 8 Gödel’s Conception of Hilbert’s Finitism

Let me say straight off that, contrary to what I wrote in [Tait, 2001], I believe one should distinguish between Gödel’s pronouncements about finitism, which he usually refers to as Hilbert’s finitism, and what he himself calls finitary number theory in [Gödel, 1938a] and what, both in that lecture and in the earlier one [Gödel, \*1933o], he refers to as the lowest level of constructivity or finitary mathematics. (As we will see, there is a puzzling difference between what he describes as the lowest level in these two papers.) At the time of writing [Tait, 2001], I was vividly aware of the distinction between the conceptual question of what *ought* to be called finitism and the historical

---

<sup>14</sup>He also writes that the proof in [Tait, 1961] does not use choice sequences. But the difference is that the proof there shows that a hypothetically *given* descending sequence of ordinals from  $\omega^\alpha$  is finite. If one replaced the given sequence by a variable for a numerical function or choice sequence, the conceptual basis for the proof would be exactly the same as for Bernays’ proof.

question of what Hilbert had in mind; but it had not occurred to me that Gödel also had, at least implicitly, made this distinction. But we will see that, given his conception of Hilbert’s finitism, there is strong evidence that he rejected it himself as an adequate foundation for proof theory. In any case, the alternative to this reading of the situation is to attribute to him 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. Allow me to refer to this as the *McKeon Principle*.<sup>15</sup>

Both Herbrand’s letter and von Neumann’s letters # 2 and #3 are concerned with Gödel’s statement in [Gödel, 1931, p. 197] that his second incompleteness theorem did not necessarily show the impossibility of Hilbert’s program, i.e. to obtain finitist consistency proofs for axiomatizations of mathematics. As we see from the above quote in §4 from p. 3 of his reply to Herbrand, Gödel persisted in 1931 in his belief that there could be finitist proofs that escape formalization in *Principia Mathematica* or other formalisms, such as first- or second-order number theory or set theory; and so the possibility remained for each of these systems that there could be a finitist proof of its consistency. Gödel seems also to have rejected von Neumann’s argument to the contrary in at least one of two missing letters; for in letter #3 (12.1.31) von Neumann thanks Gödel for the letters and later goes on to write “I absolutely disagree with your view on the formalizability of intuitionism.”

---

<sup>15</sup>Of course, one can carry the principle too far. Richard McKeon, who lectured on philosophy at the University of Chicago in the years 1935-1974, preached (in his earlier years) this very salubrious view in connection with the interpretation of the great philosophers, such as Plato: If your interpretation of them makes them fools (even in their own time), then it is likely that you have more work to do. Unfortunately, this doctrine never took hold—in the case of Plato scholarship at least, among contemporary scholars of either ancient philosophy or philosophy of mathematics—and, 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.

A reasonable conclusion to be drawn from Gödel's writings is that this view of finitism remained constant into the 1960's: It was someone else's ideology, namely Hilbert's, and he was unhappy with the terms in which it was formulated, namely in terms of intuition. Thus, a short while before he received the letter from Herbrand, in letter #3(4.2.31) to Bernays, he discusses Hilbert's extension  $Z^*$  and Bernays extension  $Z^{**}$  of  $Z$  ( $= PA$  with the least number operator) obtained by adding as axioms the universal closures of finitistically valid formulas of  $Z$  which, respectively, contain no bound variables and are arbitrary. He sketches the proof that, under a reasonable assumption, these systems also have undecidable sentences. Then he writes

By the way, I don't think that one can rest content with the systems  $[Z^*, Z^{**}]$  as a satisfactory foundation of number theory (even apart from their lack of deductive closure), and indeed, above all because in them the complicated and problematical concept "finitary proof" is assumed (in the statement of the rule for axioms) without having been made mathematically precise.

As late as 1968, we have quoted him in letter # 68b to Bernays as stating that "nested recursions are not finitary in Hilbert's sense (i.e. not intuitive)". In the postscript to that letter, he again writes "Hilbert's finitism (through the requirement of being "intuitive") has a quite unnatural boundary."

In his introductory note to the Bernays correspondence, Solomon Feferman refers (e.g. on p.55) to Gödel's "unsettled views" of finitism. But that 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'. Beginning with letter #40 (8.11.61) to Bernays, he entertains the possibility that there is such an analysis. He writes

[Kreisel] now really seems to have shown in a mathematically satisfying way that the first  $\epsilon$ -number is the precise limit of what is finitary.

The reference is to [Kreisel, 1960]. (Later on he refers to [Kreisel, 1965]. There is also an earlier mention of Kreisel's work in this connection in letter # 23 (9.30.58), but without the strong claim for it.) That he is referring to Hilbert's conception is confirmed by the continuation of this passage:

I find this result very beautiful, even if it will perhaps require a phenomenological substructure in order to be completely satisfying.

In letter # 61 (1.21.67) he writes, without reference to Kreisel

I am now convinced that  $\epsilon_0$  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.

Note that the bound  $\epsilon_0$  here is not "the precise limit" of the previous quote. But perhaps the reason is that, although it might be a precise limit in principle, he does not want to commit to the assertion that we can in practice obtain (well-foundedness of all the ordinals  $<$ )  $\epsilon_0$ . In letter # 64 (5.16.68) he again mentions Kreisel's analysis:

I would still like to ask you whether in the new edition of the book on foundations [namely of the *Grundlagen der Mathematik* I and II] you have taken account of Kreisel's derivation of induction up to  $\epsilon_0$ . [The reference is to [Kreisel, 1965].] It seems to me that it comes much closer to being finitary than your earlier proof.

The 'earlier proof' is, of course, the fallacious proof discussed above, which *was* included. However, Bernays exercised better judgment in the case of Kreisel's argument. The latter is perfectly valid as a sketch of how, in an autonomous hierarchy of quantifier-free formal systems including primitive recursive definition generated by a certain reflection principle, the well-foundedness of each ordinal  $< \epsilon_0$ , but not that of  $\epsilon_0$ , can be derived. I have already explained in [Tait, 1981] why the reflection principle even at the lowest stage goes beyond what "finitism" *should* mean. But I confess surprise that either Kreisel or Gödel should have believed that what is accessible by 'intuition' is closed under this principle. Suppose that we agree that we can intuitively see through the computation of each given primitive recursive function—essentially that we can intuitively envision the tree of all descending sequences of ordinals less than  $\alpha$  for each  $\alpha < \omega^\omega$ . The crucial first step in Kreisel's argument (§3.42, with reference back to §3.412, pp. 170-171, in his paper) is that we can then see through the computation of an evaluation function for the primitive recursive functions. (I.e. we can envision all descending sequences of ordinals less than  $\omega^\omega$ .) Why in the world would anyone find that convincing? My concrete insight into the structure of the computations of any given primitive recursive function may depend upon the definition of that particular function in a non-uniform way, in no way yielding an intuition of the structure of the computations of an arbitrary

such function. The situation is analogous to that  $\omega$ -incompleteness, where, say in  $HA$ , one may have, for each  $n$ , a proof of  $\phi(\bar{n})$  without having a proof of  $\forall x\phi(x)$ . Of Kreisel's analysis, Gödel remarks later on in footnote *f* of [1990b] that "his arguments would have to be elaborated further to be fully convincing". Nevertheless, Gödel seems to be accepting the possibility of  $\epsilon_0$  being at least an upper bound of the ordinals on which recursion is finitary in the unidealized sense of concrete intuition that he takes to be Hilbert's conception of finitism; and so this is a modification of his earlier view that, *in principle*, no formal bound could be placed on finitism in Hilbert's sense. On the other hand, as we see from footnote *f*, he remained unconvinced that such a bound had actually been established. indeed, in [1990b, p. 273], he repeats essentially what he wrote in [1958]:

Due to the lack of a precise definition of either concrete or abstract evidence there exists, today, no rigorous proof for the insufficiency (even for the consistency proof of number theory) of finitary mathematics.

But, again, notice that he is not now contesting even the *possibility* of establishing a bound.

Why did Gödel change his mind in the 1960's and '70's about the possibility of an upper bound on finitary reasoning in Hilbert's sense? One possible explanation, gaining some corroboration from the reference quoted above to phenomenology, would be his study of Husserl's writings, which according to [J. W. Dawson, 1997, p. 218], he began in 1959. For in transcendental phenomenology he may have felt one could find a basis for the requisite precise analysis of intuition.

One could very well question Gödel's understanding of what Hilbert meant by finitary reasoning: As far as I know, he cites only Hilbert's discussion in "Über das Unendliche" [1926]. But then in [1990b, p. 272, footnote b] he writes

"Concrete intuition", "concretely intuitive" are used as translations of "Anschauung", "anschaulich". The simple terms "concrete" or "intuitive" are also used in this sense in the present paper. 'What Hilbert means by "Anschauung" is substantially Kant's space-time intuition confined, however, to configurations of a finite number of discrete objects. Note that it is Hilbert's

insistence on *concrete* knowledge that makes finitary mathematics so surprisingly weak and excludes many things that are just as incontrovertibly evident to everybody as finitary number theory. E.g., while any primitive recursive definition is finitary, the general principle of primitive recursive definition is not a finitary proposition, because it contains the abstract concept of function. There is nothing in the term "finitary" which would suggest a restriction to concrete knowledge. Only Hilbert's special interpretation of it introduces this restriction.

Gödel's distinction between concrete objects and abstract objects coincides extensionally with our distinction between finite objects and infinite objects. In particular, his examples of abstract objects, namely functions of finite type over the natural numbers and proofs of arithmetic sentences, are in general infinite. But surely there is something in the term "finitary" that suggests a restriction to finite objects. Bernays, in his essay [1930–31], develops the notion of a *formal object* as the underpinning of Hilbert's finitism, and he argues that finiteness is "an essential characteristic of formal objects."

It is incidentally noteworthy that, in his letters to Bernays, when Gödel is writing about Hilbert's finitism, there is never the suggestion that he is equally writing about *Bernays'* finitism. It may be that, since there tends to be a critical edge to his remarks in such cases, he is simply avoiding direct criticism of Bernays. (There is a remarkably civil—indeed friendly and evolving towards affectionate—tone to their correspondence in general.) But perhaps, too, he did not share what I take now to be a generally held opinion (that I share) that the philosophical underpinning of Hilbert's proof theory owed much to Bernays.

I don't have a clear understanding of Bernays's notion of a formal object. In whatever sense a natural number is a formal object, it also is applicable as an iterator of an arbitrary operation on a domain. But it seems clear that the role played by intuition in this conception is a very different thing from the way in which Gödel conceives it. There is one idea of the role of intuition in arithmetic, which seems to have been Weyl's (in his intuitionistic phase) and arguably (although by no means certainly) was Hilbert's. It may also be possible to trace it to Kant's notion of pure inner intuition. See [Friedman, 1992, chapters 1 and 2]. On this conception, it is intuition that provides us with the operation of finite iteration (in contrast with the second-order logical construction of this operation by Frege and Dedekind). However, once



we have satisfactorily analyzed that intuition, by introducing 0 and successor and the principle of iteration, intuition plays no more role: it gives content to arithmetic; but it is not directly a source of new truths.<sup>16</sup> Gödel, on the contrary, when he is speaking of Hilbert’s finitism, takes intuition to refer to direct insight into arithmetic truth in individual cases. (That is why it has no ‘natural boundary’.) Thus, in [1990b, p. 273] he writes

Recursion for  $\epsilon_0$  could be proved finitarily if the consistency of number theory could. On the other hand the validity of this recursion can certainly not be made *immediately* evident, as is possible for example in the case of  $\omega^2$ . That is to say, one cannot grasp at one glance the various structural possibilities which exist for decreasing sequences, and there exists, therefore, no *immediate* concrete knowledge of the termination of every such sequence.

In the early 1960’s at Stanford, someone (I am no longer sure who it was—or maybe I just don’t want to say) proposed as the measure of mathematical IQ the bound on the ordinals  $\alpha$  for which one could “grasp at one glance the various structural possibilities which exist for decreasing sequences” from  $\alpha$ . The inventor, as I recall, claimed the math IQ of  $\epsilon_0$ . I had to keep very quiet in discussions of this, since, although I convinced myself that my math IQ was  $\geq \omega_{(2)}$ , I couldn’t get it higher than that. I later realized that some of those ‘ $\dots$ ’s in my head getting me up even to  $\omega_{(2)}$  represented iterations of operations, not on  $\mathbb{N}$ , but on infinitary domains, and that my math IQ was, like everyone else’s, at most  $\omega^\omega$ .<sup>17</sup> I hasten to add, though, that I don’t think that this is a good measure of mathematical IQ. Why shouldn’t the latter also take account of the ability to avoid such combinatorics and invent concepts, such as that of accessibility, which lead to easy insight into, e.g., the well-foundedness of each ordinal  $< \epsilon_0$ ? I think that this question is closely related to Gödel’s dissatisfaction, expressed in footnote *b* (quoted above), with Hilbert’s finitism. From the point of view of direct insight into truth, the restriction to the concrete is too restrictive. But then, this was not the point of view of Hilbert. Hilbert’s aim was to provide the *content* to ‘contentual’ mathematics.

---

<sup>16</sup>It is analogous to another pair of myths: God, the prime mover, versus a God who intervenes at every step in the workings of the world.

<sup>17</sup>Since recursion on  $\omega^n$  for each  $n$  is reducible to primitive recursions, the ‘ $\dots$ ’s in this case can always represent iterations of numerical operations.

## 9 Gödel’s System $A$

In [Gödel, 1938a] Gödel speaks of finitary number theory in a sense which does not refer to Hilbert’s finitism: It is not defined by reference to intuition, but by a description of objective rules of definition and proof—and it is precisely the system  $PRA$ . He describes it as the lowest level of a hierarchy of ‘finitary’ systems, which also includes the introduction of the primitive recursive functions of finite type, intuitionistic logic, and transfinite induction. His use of the term “finitary” to describe all these extensions as well is odd. (Perhaps he really didn’t think that the term had anything to do with being finite.) On the other hand, we are speaking here of very rough lecture notes, and maybe we shouldn’t worry too much about this choice of terminology. In the lecture notes [Gödel, \*1933o], he speaks instead of a hierarchy of *constructive* systems. There the lowest level is given the name “system  $A$ ”. In the introductory note to [Gödel, 1938a], Charles Parsons and Sieg identified that system (along with the system of finitary number theory), with  $PRA$ —a view which I echoed in [Tait, 2001].

In his introductory note to the Herbrand correspondence, Sieg correctly rejects that identification—but for the wrong reasons. He writes that Gödel changed his views significantly from July 1931, the time of his reply to Herbrand, to late December 1933, when he delivered the lecture [Gödel, \*1933o]. He claims that in these lecture notes Gödel “sharply distinguishes intuitionist from finitist arguments, the latter constituting the most restrictive form of constructive mathematics.” Here Sieg is identifying finitism (as Gödel then understood it) with the system  $A$ , and his claim is, further, that this system extends beyond  $PRA$  and, in particular encompasses functions such as the Ackermann function which are defined only with the aid of manifold nested recursions. The same point of view is expressed in Sieg’s introductory note to the von Neumann correspondence as well as in [Sieg, 2005], which contains a more extended discussion of the Herbrand-Gödel correspondence.

First, let me note this: It may be that the word “finitism” or “finitist” occurs somewhere in [\*1933o], but I haven’t yet found an occurrence, and it certainly does not occur in the passages in which Gödel is describing the layers of constructivity. What he does write is

Now, what remains of mathematics if we discard these methods [and retain only things that can be constructed and operations that can actually be carried out] is the so-called intuitionistic

mathematics, and the domain of this intuitionistic mathematics is by no means so uniquely determined as it may seem at first sight. For it is certainly true that there are different notions of constructivity and, accordingly, different layers of intuitionistic or constructive mathematics.

We see here a distinction similar to that which he implicitly drew in the letter to Herbrand between the narrow conception of intuitionism and the wider. He again refers the latter conception to Brouwer and Heyting. I think that Sieg is right that Gödel draws a sharper distinction in [\*1933o], but it confuses matters to say that it is between intuitionism and finitism; it is between the intuitionism of Brouwer and Heyting and reasoning in conformity with his system *A*. That he might, at that time at least, have been willing to identify *A* as finitary number theory (as he did *PRA* in [Gödel, 1938a]) seems quite reasonable. But there is no reason to think that he had changed his mind about what *Hilbert* meant by finitism. He has changed the subject, not his mind, and is pursuing the question of constructive foundations for classical mathematics in a way more congenial to himself.

According to Sieg's view, Gödel changed his conception of finitism between 1931 and 1933, then changed it again in 1938, when finitism became identified with *PRA*, and then changed it again, when finitist reasoning again became founded on intuition. Add to this the cases in which Gödel really did seem to vacillate, namely in connection with the question of whether there is an in principle bound on intuition, and it is no wonder that Feferman wrote of "unsettled views." I would prefer to do some McKeonizing here.

What is system *A*? In footnote *q* on p. 8 of his introductory note to the Herbrand correspondence, Sieg writes

The restrictive characteristics of the system *A* are formulated on pp. 23 and 24 of [Gödel, \*1933o]: (i) universal quantification is restricted to totalities whose elements can be generated by a "finite procedure"; (ii) negation cannot be applied to universal statements; (iii) notions have to be decidable and functions must be calculable. As to condition (iii), Gödel claims, "such notions and functions can always be defined by complete induction;" . . . .

Requirement (i) is that the *totality* be generated by a finite procedure. The effect of (ii) (he has in mind propositions built up from atoms by means of  $\wedge, \vee, \neg$  and  $\forall$ : if implication were included, then negation would sneak in as

a defined operation) is that each statement is  $\Pi_1^0$ , so that the statement can be expressed by a quantifier-free formula. In (iii), one should note that definition and proof by complete induction can be generalized to other systems of objects “generated by a finite procedure”, providing we assume that each object has a unique generation. But Sieg has dropped off the last—and quite important—part of the quoted passage from Gödel concerning (iii):

and so we may say that our system [I will call it  $A$ ] is based exclusively on the method of complete induction in its definitions as well as in its proofs. (The brackets are Gödel’s.)

*This* of course is not a ‘claim’: it is a *baptism*. With it, and if we restrict ourselves to arithmetic (i.e. where the objects generated by a finite procedure are the natural numbers generated by successor), the ‘restrictive characteristics’ become a definition of  $A$ : it is *PRA*.

Nevertheless, Gödel’s words, quoted in connection with (iii), if they *did* constitute a claim, would be puzzling: He certainly knew that there are computable functions, such as the Ackermann function, that cannot be defined by complete induction. But rather than saddling Gödel with a claim that he knew was false, it is more reasonable to suppose that he is simply making more specific what he meant by (iii). Exactly what definitions of computable functions are to be admitted in  $A$ ? (As we will have occasion to note later on, even a year later in [Gödel, 1934, footnote 34], he did not regard the notion of a computable function to be adequately explicated.) His answer is that the functions must be introduced by complete induction. I would agree that Gödel’s notes do not make this point cleanly. But remember that they are notes *to himself* for a lecture and, in such a case particularly, the McKeon principle ought again to apply: look for a reading that makes sense and is commensurate with our knowledge of the author. The alternative, that he is making a claim, makes his position incoherent.

Sieg argues (in footnotes  $q$  and  $s$ ), primarily on the basis of Gödel’s response to Herbrand’s letter, that  $A$  includes the Ackermann function. But surely we should begin with Gödel’s own description of  $A$ . The issue boils down to this: To reject the view that Gödel’s system  $A$ , restricted to the domain of natural numbers, is *PRA*, aside from implying a discontinuity with his reference to the lowest level of constructivity in [1938a], is to hold that he was using the expression “the method of complete induction in its definitions as well as in its proofs” to refer to something other than proof by mathematical induction and definition by primitive recursion.

The earliest use of the expression “complete induction” that I know is in [Dedekind, 1887]. There, followed in parentheses by “inference from  $n$  to  $n + 1$ ”, it forms the heading of §80, in which Dedekind proves mathematical induction. The only other occurrences of its use that I have found (although I have by no means attempted an exhaustive search) are in one or more works by one of Poincaré, Brouwer, Weyl, Hilbert, Gentzen and other works of Gödel. With one exception, these referred only to a principle of proof, and in each case, that principle is mathematical induction. Primitive recursive definition was generally called “definition by induction” (as in Dedekind’s essay), “definition by recursion”, “recursive definition”, etc. (all of these instances preceding the use of the term “recursive” to refer to more general kinds of recursive definition). The one exception that I found was [Weyl, 1949, p. 33], which uses the term “complete induction” to refer both to a principle of proof and to one of definition. The former is, again, mathematical induction and the latter, primitive recursive definition. (Weyl also paraphrased “complete induction” by “inference from  $n$  to  $n + 1$ “.) Nor does it seem likely that complete induction in the context of proof would mean mathematical induction, i.e. proof of  $\phi(n + 1)$  induced by proof of  $\phi(n)$ , but that in the context of definition of functions, it would mean something other than value of  $f$  at  $n + 1$  induced by its value at  $n$ . It therefore seems quite unlikely that Gödel was using “complete induction” to refer to anything other than proof by mathematical induction and definition by primitive recursion.

I have so far been discussing  $A$  only with respect to arithmetic, i.e., where the only totality involved is that of the natural numbers. Of course, even if the totalities also include the domain of all words over some finite alphabet, nothing will change, since the words can be coded by numbers in such a way that definition and proof by complete induction with respect to words reduces to complete induction with respect to numbers. (The requirement of uniqueness of generation, needed in order to apply definition by complete induction, means that the totality may be represented by a set of words over a finite alphabet.) But it would seem that he had in mind something more; otherwise the following passage on p. 26 ([Gödel, 1995, p. 52]) would make no sense:

Now all the intuitionistic proofs complying with the requirements of 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.

If we assume that only the totality of natural numbers or the totality of words over some finite alphabet is in question, then the first part of this is reasonable. The syntactical objects can be coded by numbers and so the proof can be formalized in *PRA*. But Dedekind's proof of the principle of definition by recursion easily translates into a proof in classical analysis (second-order number theory) of the unique existence of each primitive recursive function and Gödel showed that in fact they could in fact be explicitly defined in *PA* itself. It is the second part of the quote that is puzzling. If the only totality involved were essentially that of the numbers, there would surely be more than "reasons for believing" that all such arguments can be expressed in *PA*.

The only explanation of what Gödel had in mind here that occurs to me with any plausibility at all takes its hint from his use of the expression "finite procedure". Suppose we consider a totality  $W$  of words over some finite alphabet containing the null word, but where the question of whether or not the unit extension  $w * a$  of the word  $w \in W$  by the symbol  $a$  from the alphabet is in  $W$  is made to depend upon some condition. It is reasonable to suppose that Gödel would not call the procedure of generating these words "finite" if the condition were not decidable or at least effective (in the sense that there is an effective enumeration of words such that, for  $w \in W$ ,  $w * a \in W$  iff it occurs in the enumeration). This notion of an effective enumeration involves that of a 'finite procedure', namely for determining the  $n$ th word in the enumeration as a function of  $n$ . In [Gödel, 1934, page 3], Gödel speaks of computability "by a finite procedure" and remarks in footnote 3 that the notion of finite computation is not defined. On the other hand, he seems to be saying there that it serves as a good heuristic principle to equate computability by a finite procedure with general recursiveness. Now, with that equation, the set of Gödel-numbers of the words in  $W$  is defined by a predicate  $\exists y \phi(x, y)$ , where  $\phi$  is a formula of *PRA*. Then proof by complete induction with respect to  $W$  reduces to complete induction with respect to  $\mathbb{N}$  applied to  $\Sigma_1^0$  formulas, and each function defined by complete induction over  $W$  is simply the restriction of a primitive recursive function to arguments in  $W$ . Thus, even when system  $A$  is extended beyond *PRA* to include such domains  $W$ , under this assumption, every proof in  $A$  will be formalizable in *PA*. Indeed, it is fairly easy to see that complete induction applied to  $\Sigma_1^0$  formulas is conservative over *PRA* [Parsons, 1972]. So, it seems reasonable to

suppose that Gödel understood that, under this ‘heuristic principle’, system  $A$  does not go beyond  $PRA$ .

If this is the correct explanation for Gödel’s lack of complete conviction that every intuitionistic proof in conformity with system  $A$  is formalizable in  $PA$ , then it also may explain something else: Gödel expresses in [Gödel, \*193?] his belief that the “gap” between the informal notion of computability and a precise mathematical definition “has been filled by Herbrand, Church and Turing.” Certainly he is referring to work of Church and Turing in 1936. If we assume that he had that belief when he wrote his lecture notes for [Gödel, 1938a], this will explain why, in that work, he restricted the lowest level of finitary mathematics to  $PRA$ , i.e., to the case in which the only totality involved is  $\mathbb{N}$ . This bit of Gödel’s vacillation will in this way be explained as an entirely reasonable adjustment of his views in the light of further progress.

On the issue of whether or not Gödel, in [Gödel, \*1933o], intended system  $A$  to encompass more general forms of recursion, such as nested manifold recursions, the foregoing would seem to be entirely conclusive. But let’s look at Sieg’s argument to the contrary. On pp. 26-27 of [\*1933o] Gödel states that the methods of  $A$  are insufficient to prove the consistency of  $PA$ ; but he then goes on to state that interesting partial results have been obtained and mentions Herbrand’s result, which he formulates in this way:

If we take a theory which is constructive in the sense that each existence assertion made in the axioms is covered by a construction, and if we add to this theory the non-constructive notion of existence and all the logical rules concerning it, e.g., the law of excluded middle, we shall never get into any contradiction.

Sieg (footnote *s*) takes this passage (and its context) as evidence that  $A$  is stronger than  $PRA$  and, in particular, admits the Ackermann function:

In Gödel’s judgment, Herbrand had given a finitist consistency proof for a theory of arithmetic with quantifier-free induction and a large class  $F$  of calculable functions that included the Ackermann function; . . .

Moreover, as we have already noted, he understand’s Gödel to equate finitism with what can be proved in system  $A$ . The sole evidence for the latter is that Gödel “insists that all known finitist arguments given by ‘Hilbert and his

disciples’ can be carried out in  $[A]$ .” (p. 8.) But, first of all, as already noted, the term “finitist” does not occur in Gödel’s text. Gödel said simply that the methods so far used in the Hilbert school for attempting consistency proofs are all formalizable in  $A$ : There is no claim in [\*1933o] that, *in principle*, all finitist arguments in the sense of Hilbert are formalizable in  $A$ . Only such a claim would contradict his earlier statements about Hilbert’s finitism.<sup>18</sup>

As for the quoted passage concerning Herbrand’s result, there is no explicit reference either to the system  $A$  or to finitism, only a reference to constructivity; and Gödel has already distinguished different layers of constructivity, of which  $A$  is only the lowest. Nevertheless, although the passage by no means has to be read in this way, its context might well suggest that what Gödel means is that, although  $PA$  cannot be proved consistent in  $A$ , by Herbrand’s methods, we are able to prove the consistency of certain fragments of  $PA$  in  $A$ .

But it should also be noted that Gödel singles out no particular such fragments. In particular, he does not refer to Herbrand’s example (iii) in which the axioms define the Ackermann function—nor the example (ii) in which they define all the primitive recursive functions. It is only by supposing that Gödel means both that the theorem yields a consistency proof in  $A$  and that it applies to example (iii) [or (ii)] that one can infer that  $A$  is stronger than  $PRA$ . Herbrand’s theorem itself does not depend upon any decision about what constitutes an intuitionistically valid axiom set  $\Gamma$ , it just begins with the assumption that we have one. Independent of that assumption, the reduction of a deduction of a quantifier-free  $\theta$  in  $\mathcal{T}(\Gamma)$  to a quantifier-free deduction of  $\theta$  in  $\mathcal{T}(\Gamma)^*$  can be formalized in  $PRA$ . I believe that anyone at that time with an interest in proof theory would have been impressed with *that* result—that quantifiers can be eliminated from the proofs (provided that they don’t occur in axioms or formulas to which mathematical induction is applied), independently of any particular view about what constitutes finitary reasoning, what Hilbert thought constituted finitary reasoning, or the lowest level of constructive reasoning. (The earlier arguments of Ackermann and von Neumann for the finitary consistency of  $PA$  were known to be fallacious by that time.)

At the end of the day, the question to be answered is this: Is it more likely

---

<sup>18</sup>The same confusion between an historical remark and a claim about the compass of what Gödel took to be finitist occurs in [Sieg, 2005, p. 180], where Sieg writes “Gödel’s reasons for conjecturing that  $A$  contains all finitist arguments are not made explicit.”



that Gödel meant that Herbrand’s result, applied to either example (ii) or (iii), could be proved in  $A$  or that he was using the expression “complete induction” in its customary sense? Both cannot be true.

## 10 The *Dialectica* Interpretation

We have noted that Gödel first mentioned the use of functions of finite type (over  $\mathbb{N}$ ) as an extension of finitary number theory in [Gödel, 1938a]. In 1941 he lectured on his interpretation of Heyting arithmetic  $HA$  in his theory  $T$  of primitive recursive functions of finite type both at IAS in Princeton and at Yale. ([Gödel, \*1941] contains the notes for the latter lecture.) In these lectures, he presented the interpretation as giving a constructive content to intuitionistic arithmetic, whose interpretation by Heyting he regarded as lacking in clear constructive meaning. In the *Dialectica* paper, he presented it as a consistency proof for  $HA$ —and so for the corresponding classical system  $PA$ —relative to  $T$ .

The earliest mention of the functional interpretation in the correspondence with Bernays is in letter # 17 (2.6.57), in which he mentions his earlier lecture in Princeton on the subject. He raises the question of whether the interpretation can be extended to analysis by admitting definition by recursion over transfinite (but still constructive) ordinals—a question that he repeats in letter # 27 (1.7.59), after the publication of the *Dialectica* paper. If we leave aside the question of constructivity, Spector’s result [Spector, 1962] constitutes such an interpretation of analysis, since even his higher type bar recursion may be understood as recursion over a well-ordering (though not in general of numbers). No mention is made in Gödel’s correspondence of Girard’s extension of the functional interpretation to analysis and simple type theory using functions in a hierarchy of generalized types [Girard, 1971]. Likewise, there is no reference to the Curry-Howard propositions-as-types theory.

Anyone who fails to discern an element of saintliness in Bernays’ character needs only to read the correspondence with Gödel on the subject of the publication of a translation of Gödel’s *Dialectica* paper. It begins a bit rocky. In letter #53 (9.17.65) from Bernays:

That Mr. Leo Boron, in association with William Howard, is translating your beautiful p[aper in the *Dialectica* volume of 1958 into English, and that this translation was submitted to *Dialectica*

for publication is surely known to you. If I correctly understood Mr. Boron in what he wrote, you have agreed to the publication of his translation.

In letter # 54 (9.27.65) from Gödel:

Above all, I thank you for having made me aware of its intended publication. . . . But in any case I would like to ask you, in the event the translation really is supposed to appear in print, to have a copy of the manuscript sent to me.

And it doesn't get any smoother. In letter # 56 (12.5.65) Gödel indicates the need for some small revisions and Bernays responds mentioning the plan to have the manuscript in the hands of the printer towards the end of January. There ensues an exchange in which, from 25 January 1966 until 26 December 1972, Gödel wrote ten letters explaining why he was delayed or why he was not yet satisfied with the manuscript and needed to make further changes. Bernays' responses: "It would be nice if you could send me the manuscript soon" (#62 (10.22.67); "Now you can probably send me the paper for *Dialectica* very soon" #67 (1.6.69); "Mr. Gonseth [an editor of the volume in which the paper was to appear] would certainly be very pleased to receive it" # 69 (1.7.70). In # 70 (7.12.70), Bernays writes "It was very pleasing to me to receive your rewritten version . . .". But then, in # 74 (12.22.70) from Gödel: "Please wait with the page layouts." In # 75 (12.31.70) Bernays: "I'm pleased to hear that you have now brought the text of the notes to your *Dialectica* paper into a form that is satisfying to you, and I look forward with great interest to the receipt of the corrected version of your paper." —And so on. Bernays finally just dropped the subject.

Why couldn't Gödel let the paper go into print? The problem was with the notion of a computable function of finite type, which he took as the foundation of his theory  $T$ . A computable function of type  $A \rightarrow B$  is defined in [Gödel, 1990b] to be a well-defined mathematical procedure for which it is constructively evident that, applied to every computable object of type  $A$ , it yields a computable object of type  $B$ . (Here one takes the numbers, or better, the numerals to be the computable objects of type  $N$ .) Thus, the logical complexity of the notion of a computable object of type  $A$  grows with the complexity of the type  $A$ , itself. Now Gödel himself suggests that this notion of a computable function may not be sufficiently clear; but he

asserts that “there can be no doubt” that the functions introduced in his theory, namely the primitive recursive functions of finite type, are computable [Gödel, 1958, p. 283, footnote 5]. No doubt that is true; but consider a case of the function  $f$  of type  $\mathbb{N} \rightarrow A$  defined by iteration  $fn = g^n(a)$ , where  $a$  is a computable object of type  $A$  and  $g$  is a computable function of type  $A \rightarrow A$ . We indeed don’t doubt the computability of  $f$ ; but that is because we immediately accept mathematical induction applied to the concept ‘ $fx$  is a computable object of type  $A$ ’, a concept with an unboundedly high logical complexity, depending on  $A$ . The correspondence with Bernays leaves one with the impression that they both (but perhaps especially Bernays) were seeking a way to formulate the notion of computability and, in particular, the condition of ‘constructive evidence’ that would avoid this problem. (In an earlier version of the 1972 manuscript, Gödel had written “intuitionistically demonstrable” in place of “constructively evident.” Bernays had suggested in letter # 70 (7.12.70) that “the reader could well be taken aback.” by the former wording.) But, except in politics, spin has its limits and the problem wouldn’t go away.

There are in fact two problems, one local and one global. The former, which we just mentioned, is that the very notion of computable function of higher type, with the condition of ‘constructive evidence’, seems to invoke non-trivial logic, which is just what Gödel is attempting to give constructive content to. The global problem arises from the fact that the theorems of intuitionistic propositional logic need to be theorems of  $\mathcal{T}$  (since the *Dialectica* interpretation of a quantifier-free formula of  $HA$  is just itself). But Gödel’s complaint against intuitionistic logic concerns propositional logic—in particular, Heyting’s explanation of implication. So Gödel requires that the logic of  $\mathcal{T}$  be truth-functional. But, since  $\mathcal{T}$  is to be a constructive theory, this means that the atomic formulas of  $\mathcal{T}$  must be decidable. The atomic formulas of  $\mathcal{T}$  are equations between terms of like type, which, on the usual extensional meaning of equality of functions, is not decidable. Gödel replaces this notion of equality by the intensional notion of *definitional equality*. This notion really has no precise sense beyond some specific rules of definition, such as those provided by  $\mathcal{T}$  itself. But within such a boundary, we may understand two terms of like type to be definitionally equal if sequences of replacements of definiendum by definiens in the two terms yields a common term. However, proof of the decidability of equations even between closed numerical terms of  $\mathcal{T}$  requires all of intuitionistic arithmetic. (I.e., the proof of  $t = \bar{0} \vee t \neq \bar{0}$  requires mathematical induction applied to formulas of arbi-

trarily high logical complexity, depending on the complexity of the types of function constants occurring in  $t$ .)

In footnote  $h$  of [Gödel, 1990b, p. 275-6] there is an attempt to solve at least the local problem by replacing the requirement of constructive evidence with that of *reductive proof*, which is a more restrictive kind of proof than that formalized in Heyting’s logic and isn’t open to Gödel’s objections to the latter. I have been unable to find a coherent reading of the footnote. But Gödel seems to have, at least sometimes, remained convinced that the idea of reductive proof solves at least the local problem. (I’m not sure how clearly he understood the global one.) In his unsent reply to Fredrick Sawyer’s letter (2.1.74), Gödel refers to his footnote  $h$  in [Gödel, 1990b] as a means of avoiding the difficulty in the notion of a computable function of finite type; although he writes that “the matter, probably, cannot be explained convincingly in a footnote.”

A coherent foundation for the theory  $T$  is available, however, as we indicated in §reffinitary number theory. It merely requires that we take the notion of a function of a given finite type over  $\mathbb{N}$  as primitive. Iteration applied to finite types yields the principle of primitive recursion and certainly explicit definition of functions is implicit in the concept of a function. It is not the restriction to computable functions that makes a theory such as  $\mathcal{T}$  constructive: rather, it is the restriction to objects that we can construct. On the other hand, this same point of view provides a constructive foundation for the Curry-Howard theory of types and so for  $HA$  directly. Thus, from the point of view of providing constructive meaning for the logical constants as employed in  $HA$ , the *Dialectica* interpretation is a failure in two respects: First, it is superfluous and, second, it interprets as constructive propositions like instances of Markov’s principle that are not constructive.

## 11 Impredicativity of the Number Concept

In letter #74 (12.22.70), in explaining his delay with the *Dialectica* paper, Gödel refers to complications with his footnote  $h$  (there called “note  $k$ ”) arising from the “unavoidable ‘self-reflexivities’ ”. Here he is referring to the impredicativity of the definitions of functions of higher type in  $\mathcal{T}$ , where the definitions of functions of lower type essentially involve functions of higher type. In letter #76 (3.16.72) Bernays has an interesting response: He writes that it indeed seems unavoidable, but that the concept of number already

itself involves a certain impredicativity, which therefore is already present in finitary number theory. This impredicativity has to do with the two faces of numbers: numbers as objects, arising from 0 by applications of the successor operation, and numbers as (type-free) iterators of operation on any domain, *including the domain  $\mathbb{N}$  of numbers, itself*. Essentially the same point about the impredicativity of the number concept, although in somewhat different terms, appears in some work of Edward Nelson, who goes on in [Nelson, 1986] to develop arithmetic on the basis of the separation of these two faces of number. In Nelson’s discussion, this impredicativity is tied to the impredicativity in the familiar sense of the definition of the set of numbers as the intersection of all sets containing 0 and closed under successor. Bernays’ remark reveals that, even rejecting that foundation of arithmetic, which a finitist must in any case do, the trace of that impredicativity remains in the most elementary reasoning about numbers.

## References

- Andrews, P. [2003]. Herbrand award acceptance speech, *Journal of Automated Reasoning* **31**: 169–187.
- Bernays, P. [1930–31]. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie, *Blätter für deutsche Philosophie* **4**: 326–367. Reprinted in [Bernays, 1976]. A translation by P. Mancosu appears in [Mancosu, 1998], pp. 234–265.
- Bernays, P. [1976]. *Abhandlungen zur Philosophie der Mathematik*, Darmstadt: Wissenschaftliche Buchgesellschaft.
- Dedekind, R. [1872]. *Stetigkeit und irrationale Zahlen*, Braunschweig: Vieweg. in [Dedekind, 1932]. Republished in 1969 by Vieweg and translated in [Dedekind, 1963].
- Dedekind, R. [1887]. *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg. In Dedekind (1932). Republished in 1969 by Vieweg and translated in [Dedekind, 1963].
- Dedekind, R. [1932]. *Gesammelte Werke, vol. 3*, Braunschweig: Vieweg. Edited by R. Fricke, E. Noether, and O. Ore.

- Dedekind, R. [1963]. *Essays on the Theory of Numbers*, New York: Dover. English translation by W.W. Berman of [Dedekind, 1872] and [Dedekind, 1887].
- Dekker, J. (ed.) [1962]. *Recursive Function Theory, Proceedings of Symposia in Pure Mathematics*, Vol. 5, Providence: American Mathematical Society.
- Dreben, B., Andrews, P. and S.Aanderaa [1963]. False lemmas in herbrand, *Bulletin of the Americal Mathmetical So0ciety* **69**: 699–706.
- Friedman, M. [1992]. *Kant and the Exact Sciences*, Cambridge: Harvard University Press.
- Gentzen, G. [1935]. Untersuchungen über das logische Schliessen I, II, *Mathematische Zeitschrift* **39**: 176–210,405–431.
- Girard, J.-Y. [1971]. Une extension de l’interprétation de Gödel à l’analyse, et son application à l’élimination des coupures dans l’analyse et la théorie des types, in J. E. Fenstad (ed.), *Proceedings of the Second Scandanavian Logic Symposium*, Amsterdam: North-Holland, pp. 63–92.
- Gödel, K. [\*193?]. Undecidable diophantine propositions, in *Collected Works, Vol. III* [Gödel, 1995], pp. 164–175.
- Gödel, K. [1931]. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I, *Monatshefte für Mathematik und Physik* **38**: 173–198.
- Gödel, K. [\*1933o]. The present situation in the foundations of mathematics, in *Collected Works, Vol. III* [Gödel, 1995], pp. 45–53.
- Gödel, K. [1934]. On undecidable propositions of formal mathematical systems, pp. 346–372. Mimeographed lecture notes by S. Kleene and J. B. Rosser reprinted in [Gödel, 1986].
- Gödel, K. [1938a]. Lecture at Zilsel’s, in *Collected Works, Vol. III* [Gödel, 1995], pp. 87–113.
- Gödel, K. [\*1941]. In what sense is intuitionistic logic constructive?, in *Collected Works, Vol. III* [Gödel, 1995], pp. 189–201.

- Gödel, K. [1944]. Russell's mathematical logic, [*Schilpp, 1944*], pp. 123–53. Reprinted in [Gödel, 1990a, 119-143].
- Gödel, K. [1958]. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica* **12**: 280–287. Reprinted with an English translation in [Gödel, 1990a, 240-252]. [Gödel, 1990b] is a revised version.
- Gödel, K. [1986]. *Collected Works, Vol. I*, Oxford: Oxford University Press.
- Gödel, K. [1990a]. *Collected Works, Vol. II*, Oxford: Oxford University Press.
- Gödel, K. [1990b]. On an extension of finitary mathematics which has not yet been used, in *Collected Works, Vol. II* [Gödel, 1990a], pp. 271–280. Revised version of [Gödel, 1958].
- Gödel, K. [1995]. *Collected Works, Vol. III*, Oxford: Oxford University Press.
- Goldfarb, W. [1984a]. The unsolvability of the gödel class with identity, *The Journal of Symbolic Logic* pp. 1237–1252.
- Herbrand, J. [1930]. *Recherches sur la théorie de la démonstration*, PhD thesis, University of Paris. English translation in [Herbrand, 1971].
- Herbrand, J. [1931]. Sur la non-contradiction de l'arithmétique, *Journal für die reine und angewandte Mathematik* **166**: 1–8. English translation in [Herbrand, 1971, 282-298].
- Herbrand, J. [1931a]. Unsigned note on herbrand's thesis, written by herbrand himself, *Annales de l'Université de Paris* **6**: 186–189. English translation in [Herbrand, 1971, 272-276].
- Herbrand, J. [1971]. *Collected Works*, Harvard University Press, Cambridge. Warren D. Goldfarb, editor.
- Heyting, A. [1930]. Die formalen Regeln der intuitionistischen Mathematik, *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse* pp. 57–71.
- Hilbert, D. [1926]. Über das Unendliche, *Mathematische Annalen* **95**: 161–90. Translated by Stefan Bauer-Mengelberg in [von Heijenoort, 1967, 367-92].

- Hilbert, D. and Bernays, P. [1934]. *Grundlagen der Mathematik I*, Berlin: Springer-Verlag. The second edition was published in 1968.
- Hilbert, D. and Bernays, P. [1939]. *Grundlagen der Mathematik II*, Berlin: Springer-Verlag. A second edition was published in 1970.
- J. W. Dawson, J. [1997]. *Logical Dilemmas: The Life and Work of Kurt Gödel*, Wellesley, Mass.: A K Peters.
- Kreisel, G. [1960]. Ordinal logics and the characterization of informal notions of proof, *Proceedings of the International Congress of Mathematicians, Edinburgh* pp. 289–299.
- Kreisel, G. [1965]. Mathematical logic, in Saaty [1965], pp. 95–195.
- Mancosu, P. [1996]. *Philosophy of Mathematics and Mathematical Practice in the Seventeenth Century*, Oxford: Oxford University Press.
- Mancosu, P. (ed.) [1998]. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920's*, Oxford: Oxford University Press.
- Nelson, E. [1986]. *Predicative Arithmetic*, Princeton University Press.
- Orevkov, V. [1979]. Lower bounds for the lengthening of proofs after cut-elimination (russian), *Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Ordena Lenina Matematicheskogo Instituta imeni V.A. Steklova Akademii Nauk SSSR (LOMI)* **88**: 137–162, 242–243. Translation in *Journal of Soviet Mathematics*, 20 (1982), 2337–2350.
- Parsons, C. [1972]. On  $n$ -quantifier induction, *jsl* **37**: 466–482.
- Saaty, T. (ed.) [1965]. *Lectures on Modern Mathematics*, New York: Wiley.
- Schilpp, P. (ed.) [1944]. *The Philosophy of Bertrand Russell*, New York: Tudor Publishing Co.
- Schütte, K. [1951]. Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie, *Mathematische Annalen* **122**: 369–389.
- Sieg, W. [2005]. Only two letters: The correspondence between Herbrand and Gödel, *The Bulletin of Symbolic Logic* **11**(2): 172–184.



- Sieg, W., Sommer, R. and Talcott, C. (eds) [2002]. *Reflections on the Foundations of Mathematics: Essays in Honor of Solomon Feferman*, Lecture Notes in Logic 15, Association for Symbolic Logic.
- Spector, C. [1962]. Provably recursive functionals of analysis: a consistency proof of analysis by an extension of the principles formulated in current intuitionistic mathematics, *in* Dekker [1962], pp. 1–27.
- Tait, W. [1961]. Nested recursion, *Mathematische Annalen* **143**: 236–250.
- Tait, W. [1965]. Functionals defined by transfinite recursion, *The Journal of Symbolic Logic* **30**: 155–174.
- Tait, W. [1981]. Finitism, *Journal of Philosophy* **78**: 524–556.
- Tait, W. [2001]. Gödel’s unpublished papers on foundations of mathematics, *Philosophia Mathematica* **9**: 87–126.
- Tait, W. [2002]. Remarks on finitism, , pp. 410–19.
- Tait, W. [2005a]. Gödel’s interpretation of intuitionism, *Philosophia Mathematica* .
- Tait, W. [2005b]. *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History*, Oxford: Oxford University Press.
- van Dalen, D. [1995]. Hermann Weyl’s intuitionistic mathematics, *Bulletin of Symbolic Logic* **1**(2): 145–169.
- von Heijenoort, J. (ed.) [1967]. *From Frege to Gödel: A Source Book in Mathematical Logic*, Cambridge: Harvard University Press.
- von Neumann, J. [1927]. Zur Hilbertschen Beweistheorie, *Mathematische Zeitschrift* **26**: 1–46.
- Weyl, H. [1921]. Über die neue Grundlagenkrise der Mathematik, *Mathematische Zeitschrift* **10**: 39–79. Translated by P. Mancosu in [Mancosu, 1998].
- Weyl, H. [1949]. *Philosophy of Mathematics and Natural Science*, Princeton: Princeton University Press.