Bernays Project: Text No. 6

Appendix to Hilbert's lecture *The foundations of mathematics* (1927)

Paul Bernays

(Zusatz zu Hilbert's Vorlesung *Die Grundlagen der Mathematik*, 1927.)

Translation by: *Stefan Bauer-Mengelberg and Dagfinn Føllesdal*

Comments:

Revisions 1/7/2005 by Richard Zach

‖[89] 1. To supplement the preceding paper [[by Hilbert]] let me add some more detailed explanations concerning the consistency proof by Ackermann that was sketched there.

First, as for an upper bound on the number of steps of replacement in the *case of embedding*, it is given by 2^n, where n is the number of ε-functionals distinct in form. The method of proof described furnishes yet another, substantially closer bound, which, for example, for the case in which there is no embedding at all yields the upper bound $n + 1$.[1]

2. Let the argument by which we recognize that the procedure is finite in the *case of superposition* be carried out under simple specializing assumptions.

[1] [[See in *Hilbert and Bernays 1939*, pp. 96–97, how this bound is obtained.]]

The assumptions are the following: Let the ε-functionals occurring in the proof be

$$\varepsilon_a \mathfrak{A}(a, \varepsilon_b \mathfrak{K}(a, b))$$

and

$$\varepsilon_b \mathfrak{K}(\mathfrak{a}_1, b), \varepsilon_b \mathfrak{K}(\mathfrak{a}_2, b), \ldots, \varepsilon_b \mathfrak{K}(\mathfrak{a}_n, b),$$

where $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ may contain $\varepsilon_a \mathfrak{A}(a, \varepsilon_b \mathfrak{K}(a, b))$ but no other ε-functional.

The procedure now consists in a succession of "total replacements"; each of these consists of a function replacement $\chi(a)$ for $\varepsilon_b \mathfrak{K}(a, b))$, by means of which $\varepsilon_a \mathfrak{A}(a, \varepsilon_b \mathfrak{K}(a, b))$ goes over into $\varepsilon_a \mathfrak{A}(a, \chi(a))$, and a replacement for $\varepsilon_a \mathfrak{A}(a, \chi(a))$, by means of which $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ go over into numerals $\mathfrak{z}_1, \ldots, \mathfrak{z}_n$ and the values

$$\chi(\mathfrak{z}_1), \ldots, \chi(\mathfrak{z}_n)$$

are obtained for

$$\varepsilon_b \mathfrak{K}(\mathfrak{a}_1, b), \ldots, \varepsilon_b \mathfrak{K}(\mathfrak{a}_n, b).$$

$\|^{90}$ We begin with the function $\chi_0(a)$, which has the value 0 for all a ("zero replacement"), and accordingly also replace the terms

$$\varepsilon_b \mathfrak{K}(\mathfrak{a}_1, b), \ldots, \varepsilon_b \mathfrak{K}(\mathfrak{a}_n, b)$$

by 0.

Holding this replacement fixed, we apply to

$$\varepsilon_a \mathfrak{A}(a, \chi_0(a))$$

the original testing procedure, which after two steps at most leads to the goal; that is, all the critical formulas corresponding to

$$\varepsilon_a \mathfrak{A}(a, \chi_0(a))$$

then become correct.

Thus we obtain one or two total replacements, $\mathfrak{E}_0$, or $\mathfrak{E}_0$ and $\mathfrak{E}_0'$, respectively. Now either $\mathfrak{E}_0$ (or $\mathfrak{E}_0'$) is final or one of the critical formulas corresponding to $\varepsilon_b\mathfrak{K}(\mathfrak{a}_1, b), \varepsilon_b\mathfrak{K}(\mathfrak{a}_2, b), \ldots$ becomes false. Assume that this formula corresponds to, say $\varepsilon_b\mathfrak{K}(\mathfrak{a}_1, b)$ and that $\mathfrak{a}_1$ goes into $\mathfrak{z}_1$. Then we find a value $\mathfrak{z}$ such that

$$\mathfrak{K}(\mathfrak{z}_1, \mathfrak{z})$$

is correct. Now that we have this value, we take as replacement function for

$$\varepsilon_b\mathfrak{K}(a, b)$$

not $\chi_0(a)$, but the function $\chi_1(a)$ defined by

$$\begin{aligned} \chi_1(\mathfrak{z}_1) &= \mathfrak{z} \\ \chi_1(a) &= 0 \quad \text{for } a \neq \mathfrak{z}_1. \end{aligned}$$

At this point we repeat the above procedure with $\chi_1(a)$, the values of the $\varepsilon_b\mathfrak{K}(\mathfrak{a}_\nu, b)$ $(\nu = 1, \ldots, n)$ now being determined only after a value has been chosen for

$$\varepsilon_a\mathfrak{A}(a, \chi_1(a)),$$

||[91] and thus we obtain one or two total replacements, $\mathfrak{E}_1$, or $\mathfrak{E}_1$ and $\mathfrak{E}_1'$.

Now either $\mathfrak{E}_1$ (or $\mathfrak{E}_1'$) is final or for one of the ε-functionals that result from

$$\varepsilon_b\mathfrak{K}(\mathfrak{a}_1, b), \ldots, \varepsilon_b\mathfrak{K}(\mathfrak{a}_n, b),$$

by the previous total replacement we again find a value $\mathfrak{z}'$, such that for a certain $\mathfrak{z}_2$

$$\mathfrak{K}(\mathfrak{z}_2, \mathfrak{z}')$$

is correct, while

$$\mathfrak{K}(\mathfrak{z}_2, \chi_1(\mathfrak{z}))$$

is false. From this it directly follows that $\mathfrak{z}_2 \neq \mathfrak{z}_1$.

Now, instead of $\chi_1(a)$ we introduce $\chi(a)_2$ as replacement function by means of the following definition:

$$\begin{aligned} \chi_2(\mathfrak{z}_1) &= \mathfrak{z} \\ \chi_2(\mathfrak{z}_2) &= \mathfrak{z}' \\ \chi_2(a) &= 0 \quad \text{for } a \neq \mathfrak{z}_1, \mathfrak{z}_2. \end{aligned}$$

The replacement procedure is now repeated with this function $\chi_2(a)$.

As we continue in this way, we obtain a sequence of replacement functions

$$\chi_0(a), \chi_1(a), \chi_2(a), \ldots,$$

each of which is formed from the preceding one by addition, for a new argument value, of a function value different from 0; and for every function $\chi(a)$ we have one or two replacements, $\mathfrak{E}_p$, or $\mathfrak{E}_p$ and $\mathfrak{E}'_p$. The point is to show that this sequence of replacements terminates. For this purpose we first consider the replacements

$$\mathfrak{E}_0, \mathfrak{E}_1, \mathfrak{E}_2, \ldots.$$

In these,

$$\varepsilon_a \mathfrak{A}(a, \varepsilon_b \mathfrak{K}(a, b))$$

is always replaced by 0; the $\varepsilon_b \mathfrak{K}(\mathfrak{a}_\nu, b)$ $(\nu = 1, \ldots, n)$ therefore always go over into the same ε-functionals; for each of these we put either 0 or a numeral different from 0, $\|^{92}$ and this is then kept as a final replacement. Accordingly, at most $n+1$ of the replacements $\mathfrak{E}_0, \mathfrak{E}_1, \mathfrak{E}_2, \ldots$ can be distinct.[2] If, however, $\mathfrak{E}_k$ is identical with $\mathfrak{E}_l$, then neither one has, or else each has, a successor replacement $\mathfrak{E}'_k$, or $\mathfrak{E}'_l$, and in these

$$\varepsilon_a \mathfrak{A}(a, \varepsilon_b \mathfrak{K}(a, b))$$

is then in both cases replaced by the same number found as a value, so that, for both replacements, the $\varepsilon_b \mathfrak{K}(\mathfrak{a}_\nu, b)$ $(\nu = 1, \ldots, n)$ also go over into the same ε-functionals.

Accordingly, of the replacements $\mathfrak{E}'_l$ for which $\mathfrak{E}_l$ coincides with a fixed replacement $\mathfrak{E}_k$ again at most $n+1$ can be distinct.

Hence there cannot be more than $(n+1)^2$ distinct $\mathfrak{E}_p$, or $\mathfrak{E}_p$ and $\mathfrak{E}'_p$ altogether. From this it follows, however, that our.procedure comes to an end at the latest with the replacement function $\chi_{(n+1)^2}(a)$. For, the replacements associated with two distinct replacement functions $\chi_p(a)$ and $\chi_q(a)$, $q > p$, cannot coincide completely, since otherwise we would by means of $\chi_q(a)$ be led to the same value $\mathfrak{z}^*$ that has already been found by means of $\chi_p(a)$, whereas this value is already used in the definition of the replacement functions following $\chi_p(a)$, hence in particular also in that of $\chi_q(a)$.

3. Let us note, finally, that in order to take into consideration the axiom of complete induction, which for the purpose of demonstrating the consistency may be given in the form

$$(\varepsilon_a A(a) = b') \rightarrow \bar{A}(b),$$

we need only, whenever we have found a value $\mathfrak{z}$ for which a proposition $\mathfrak{B}(a)$ holds, go to the least such value by seeking out the first correct proposition

[2] [[See footnote 1.]]

in the sequence

$$\mathfrak{B}(0), \mathfrak{B}(0'), \ldots, \mathfrak{B}(\mathfrak{z})$$

of propositions that have been reduced to numerical formulas.[3]

---

[3][[In *1935*, p. 213, end of footnote 1, Bernays writes that this last paragraph, on mathematical induction, should be deleted. In 1927 Hilbert and his collaborators had not yet gauged the difficulties facing consistency proofs of arithmetic and analysis. Ackermann had set out (in *1924*) to prove the consistency of analysis; but, while correcting the printer's proofs of his paper, he had to introduce a footnote, on page 9, that restricts his rule of substitution. After the introduction of such a restriction it was no longer clear for which system Ackermann's proof establishes consistency. Certainly not for analysis. The proof suffered, moreover, from imprecisions in its last part. Ackermann's paper was received for publication on 30 March 1924 and came out on 26 November 1924. In 1927, received for publication on 29 July 1925 and published on 2 January 1927, von Neumann criticized Ackermann's proof and presented a consistency proof that followed lines somewhat different from those of Ackennann's. The proof came to be accepted as establishing the consistency of a first-order arithmetic in which induction is applied only to quantifier-free formulas. When he was already acquainted with von Neumann's proof, Ackermann communicated, in the form of a letter, a new consistency proof to Bernays. This proof developed and deepened the arguments used in Ackermann's 1924 proof, and, like von Neumann's, it applied to an arithmetic in which induction is restricted to quantifier-free formulas. It is with this proof of Ackermann's that Hilbert's remarks above (pp. 477-479) and Bernays's present comments are concerned. It was felt at that point, among the members of the Hilbert school, that the consistency of full first-order arithmetic could be established by relatively straightforward extensions of the arguments used by von Neumann or by Ackermann (see *Hilbert 1928a*, p. 137, lines 20-21; *1930*, p. 490, line 4u, to p. 491, line 2; *Bernays 1935*, p. 211, lines 4-7). These hopes were dashed by Gödel's 1931. Ackermann's unpublished proof was presented in *Hilbert and Bernays 1939*, pp. 93-130. In 1940 Ackermann gave a consistency proof for full first-order arithmetic, using a principle of transfinite induction (up to ε_0) that is not formalizable in this arithmetic.]]