The Substitution Method Revisited

W.W.Tait

It is a pleasure for me to contribute a paper in honor of Grisha Mints. In view of his interesting work involving the epsilon-substitution method, I am returning to some work I did on that topic in 1960-1. I was primarily trying to understand the *concept* behind Ackermann's consistency proof for first-order number theory (1940), which for me was too heavy on syntax and too light on ideas. I found a satisfactory treatment the basic idea of which, I think, could have been behind Ackermann's approach.[1] But, aside from presenting my work on this topic in a lecture at the Eighth Logic Colloquium held at Oxford in 1963, I put it aside in favor of trying to extend it to full second-order number theory, a project that came to an abrupt halt in the winter of 1961-2. The work on the epsilon-calculus for first-order systems of arithmetic and predicate logic finally appeared in 1965 in two somewhat bloated papers (Tait 1965a, Tait 1965b). I am happy to have this occasion to present a leaner and cleaner exposition of that work. I will end with a brief discussion of why I believed that the method I have applied for first-order number theory would not extend to second-order number theory.

1 Preliminaries

Let

$$x, y_1, y_2, \ldots$$

be a fixed list of distinct individual variables.

Definition A *matrix* and in particular an *n-matrix* is a quantifier-free formula $A(x, y_1, \ldots, y_n)$ in a first-order language, with $n \geq 0$, whose free variables are $x, y_1, \ldots, y_n$ and such that for each i

- Every term in A other than $y_1, \ldots, y_n$ contains x.
- y_i has exactly one occurrence in A.
- The occurrence of y_i is to the left of y_{i+1} for $(i < n)$. □

[1] On the other hand, Mints himself has lectured on an approach to eliminating epsilon-terms in PA which he seems to have extended to the theory of one elementary inductive definition. This method has some claim to be the development of the intuition behind Hilbert's original belief that epsilon-terms could be eliminated.

Every quantifier-free formula $B(x)$ of a first-order theory is uniquely of the form $B(x) = A(x, \mathbf{t})$, where $A(x, \mathbf{y})$ is a matrix, $\mathbf{y} = y_1, \ldots, y_n$, and the $\mathbf{t} = t_1, \ldots, t_n$ do not contain x.

Let $\mathcal{T}$ be classical first-order number theory. For simplicity, we take the non-logical axioms of $\mathcal{T}$ to be the axioms of primitive recursive arithmetic PRA. These include the axioms for the predecessor function *pred*

$$\mathit{pred}\ 0 = 0 \qquad\qquad \mathit{pred}(x+1) = x$$

The logical axioms are those for the propositional connectives and existential quantifier introduction

$$A(t) \to \exists x A(x)$$

The logical rules of inference are modus ponens and the rule of existential quantifier elimination

$$\frac{B(x) \to C}{\exists x B(x) \to C}$$

where x is not in C. Finally, there is the rule of mathematical induction:

$$\frac{B(0) \qquad B(x) \to B(x+1)}{B(s)}$$

Definition. For each $k < \omega$, the theory $\mathcal{T}^k$ is defined by induction.

$$\mathcal{T}^0 = \mathcal{T}.$$

T^{k+1} results from adding for each n-matrix $A = A(x, \mathbf{y})$ of $\mathcal{T}^k$ $(0 \leq n)$ but not of $\mathcal{T}^m$ for any $m < k$

- To the language of $\mathcal{T}^k$ a distinct new n-ary function constant f_A. f_A is called the *Skolem function constant* for A. (0-ary function constants are just individual constants.)
- To the axioms of $\mathcal{T}^k$ all substitution instances of

  $$A(x, \mathbf{y}) \to A(f_A\mathbf{y}, \mathbf{y})$$

  called the *axioms of the first kind* for f_A.
- In the case of arithmetic, we also add all substitution instances of

  $$A(x, \mathbf{y}) \to f_A\mathbf{y} \neq x + 1$$

  called the *axioms of the second kind* for f_A.
- If $A(x)$ is an axiom of $\mathcal{T}^k$ and s is a term of T^{k+1}, then $A(s)$ is an axiom of T^{k+1}. □

Set

$$\mathcal{T}^* = \bigcup_k \mathcal{T}^k.$$

Note that the axioms of each $\mathcal{T}^k$ and hence those of $\mathcal{T}^*$ are closed under substitutions. $\mathcal{T}^*$ is sometimes called the "ϵ-calculus" and the substitution method the "ϵ-substitution method" because Hilbert wrote $\epsilon x B(x)$ for $f_A\mathbf{t}$ when A is a matrix and $B(x) = A(x, \mathbf{t})$.

Definition. The *rank* of a Skolem function constant f_A in $\mathcal{T}^*$ is the least n such that f_A is in $\mathcal{T}^{n+1}$.

Thus, if f_A is of rank n, then all of the Skolem function constants in the matrix A are of rank $< n$. If, for every matrix $A(x, \mathbf{y})$ in T^*, we abbreviate

$$\exists x A(x, \mathbf{t}) := A(f_A\mathbf{t}, \mathbf{t})$$

then every axiom

$$B(s) \to \exists x B(x)$$

transforms into an axiom of the first kind

$$B(s) \to B(f_A\mathbf{t})$$

for a suitable Skolem function f_A and each inference

$$\frac{B(x) \to C}{\exists x B(x) \to C}$$

where x is not in C, becomes the substitution

$$\frac{B(x) \to C}{B(f_A(\mathbf{t})) \to C}$$

Thus, we may consider $\mathcal{T}^*$ to be a quantifier-free system.

Consider now instances of the rule of mathematical induction

$$\frac{B(0) \qquad B(x) \to B(x+1)}{B(s)}$$

From $\neg B(s)$ we can infer $\neg B(f_A\mathbf{t})$, where $\neg B(x) = A(x, \mathbf{t})$, by an axiom of the first kind. So it follows from the first premise of the induction that $f_A\mathbf{t} \neq 0$ and so $f_A\mathbf{t} = \mathit{pred}(f_A\mathbf{t}) + 1$. By the second premise of the induction then, $\neg B(\mathit{pred}(f_A\mathbf{t}))$, which contradicts an axiom of the second kind for f_A. Thus we deduce $B(s)$.

In this way, given a deduction in $\mathcal{T}$ of a formula C, we obtain a deduction of C in the quantifier-free system $\mathcal{T}^*$ which contains, along with the logical and non-logical axioms $\mathcal{T}$, closed under substitution of the new terms of $\mathcal{T}^*$, only the new axioms of the first and second kind for the Skolem functions. The only rule of inference is *modus ponens*.

2 Eliminating Skolem Functions in First-Order Number Theory

Let $\mathcal{D}$ be a deduction in $\mathcal{T}^*$ of a formula C. We can assume that all of the free variables in $\mathcal{D}$ are in C, since all others can be replaced throughout by 0. Let $\mathbf{z}$ be a list of the free variables in C. All of the axioms in $\mathcal{D}$ other than the axioms for Skolem functions are obtained from axioms of $\mathcal{T}$ by substituting terms of the form $f_A\mathbf{t}$ for variables. Let

$$f_{A_1}, \ldots, f_{A_m}$$

be all of the distinct Skolem constants occurring in $\mathcal{D}$, listed in order of non-decreasing rank, so that f_{A_j} occurs in the matrix $A_i(x, \mathbf{y}_i)$ only if $i > j$. The axioms occurring in $\mathcal{D}$ for Skolem functions are of course finite in number. We shall call these axioms the *critical formulas* of $\mathcal{D}$.

We will show that, for a certain extension $\mathcal{T}^+$ of the quantifier-free part PRA of $\mathcal{T}$, for each finite set of axioms for Skolem functions

$$f_{A_1}, \ldots, f_{A_m}$$

we can define numerical-valued functions

$$\phi_1, \ldots, \phi_m$$

of $\mathbf{z}$ such that the result of replacing each term $f_i\mathbf{t_i}$ by $\phi_i\mathbf{z}\mathbf{t_i}$ for $i = 1, \ldots, m$ transforms each of the given axioms for Skolem functions into a theorem of $\mathcal{T}^+$. It will follow that the result of this substitution in C is a theorem of $\mathcal{T}^+$.

$\mathcal{T}^+$ is the quantifier-free system $PRA^2_{\epsilon_0}$ of second-order primitive recursive arithmetic with definition by recursion on each ordinal $\alpha < \epsilon_0$. We give a brief description of this system; but our construction below of the required ϕ_i will not include a detailed formalization in $\mathcal{T}^+$.

$\mathcal{T}^+$ contains variables of types $\omega^n \to \omega$ for $n \geq 0$. (When $n = 0$, these are the numerical variables.) Functions of one or more variables ranging over these types, whose values are of one of these types, are introduced by explicit definition, primitive recursion and by recursion on some $\alpha < \epsilon_0$. The order type ϵ_0 is represented in some standard way by a primitive recursive ordering $\prec$ of ω with least element 0. I will use lower case Greek letters to denote 'ordinals', i.e. natural numbers in their role as ordinals $< \epsilon_0$. Corresponding to the addition of the ordinal numbers represented by α and β, there is the primitive recursive function $\alpha \oplus \beta$ and, corresponding to raising the ordinal represented by α to the power 2 is the primitive recursive function E^α. Define

$$[x, y] = \begin{cases} x & \text{if } x \prec y \\ 0 & \text{otherwise} \end{cases}$$

Then definition by recursion on a limit ordinal $\alpha < \epsilon_0$ has the form

$$\Phi\mathbf{g}0 = \Psi\mathbf{g}$$

and for $0 \prec x \prec \alpha$

$$\Phi\mathbf{g}x = \Xi\mathbf{g}x(\Phi\mathbf{g}[\Theta\mathbf{g}x, x])$$

and for $\alpha \preceq x$

$$\Phi\mathbf{g}x = 0_\tau.$$

Here $\mathbf{g}$ is a list of distinct variables of arbitrary types. $\Phi\mathbf{g}x$, $\Psi\mathbf{g}$, $\Xi\mathbf{g}xu$ are all of the same type $\tau = \omega^k \to \omega$ $(k \geq 0)$ when u is a variable of type τ. $\Theta\mathbf{g}x$ is of type ω and 0_τ is some standard object of type τ, say the function with constant value 0. The formulas of $\mathcal{T}^+$ are built up from equations between terms of the same type by means of the propositional connectives. The axioms of $\mathcal{T}^+$ are those of identity, zero and successor, the defining equations of the function constants of each type and the axioms of propositional logic. The rules of inference are modus ponens, the rule of mathematical induction, and the rule of substitution

$$\frac{A(s) \qquad s\mathbf{x} = t\mathbf{x}}{A(t)}$$

where s and t are terms of some type $\omega^n \to \omega$ with $n > 0$ and $\mathbf{x}$ is a list of n distinct numerical variables that occur in neither s nor t.

$PRA^1_{\epsilon_0}$ is the result of restricting the variables to numerical variables and the function constants to types $\omega^n \to \omega$. The system $PRA^2_{\alpha+1}$ is a conservative extension of $PRA^1_{2^\alpha+1}$ and, in particular, $PRA^2_{\epsilon_0}$ is conservative over $PRA^1_{\epsilon_0}$. (Tait 1965a, Theorem 4)

Remark Our construction below would remain valid if we took the original system $\mathcal{T}$ to be, not the result of adding quantification to PRA, but the result of adding quantification to $PRA^1_{\epsilon_0}$. But then, from many points of view, that is the natural system of first-order number theory. □

$f = f_{A_m}$ has the highest rank of Skolem function constants in $\mathcal{D}$. The Skolem function constants in $A(x, \mathbf{y}) = A_m(x, \mathbf{y}_m)$ are among $f_{A_1}, \ldots, f_{A_{m-1}}$. Let $\mathbf{g} = \mathbf{z}, f_{A_1}, \ldots, f_{A_{m-1}}$. The critical formulas in $\mathcal{D}$ for f then are substitution instances of formulas of the form

$$A(s_i(f, \mathbf{g}), \mathbf{y}) \to A(f\mathbf{y}, \mathbf{y})$$

or

$$A(s_i(f, \mathbf{g}), \mathbf{y}) \to f\mathbf{y} \neq s_i(f, \mathbf{g}) + 1$$

for $i = 0, \ldots p$ for some $p < \omega$. Set

$$Sf\mathbf{g} = Max_{i \leq p} s_i(f, \mathbf{g}).$$

Then it suffices to find f as a function of the $\mathbf{g}$ satisfying

$$f\mathbf{y} \simeq \mu x \leq Sf\mathbf{g}.A(x, \mathbf{y})$$

where $\simeq$ means that the two terms are equal if the right hand side is defined, i.e. if there is an $x \leq Sf\mathbf{g}$ such that $A(x, \mathbf{y})$. Call this the *principal semi-equation*.

At the first step, where f is the Skolem function constant of highest rank, the $s_i(f, \mathbf{g})$, and so $Sf\mathbf{g}$, are just terms of $\mathcal{T}^*$. But in order to proceed by induction, we need to be able to solve the principal semi-equation in a more general setting in which $Sf\mathbf{g}$ is a functional of f and $\mathbf{g}$ in $PRA^2_{\epsilon_0}$.

We do that by first defining successive approximations $\theta_n = \Theta_n\mathbf{g}$ of a solution for f as follows:

$$\theta_0\mathbf{y} = 0$$

$$\theta_{n+1}\mathbf{y} = \mu x \leq Max\{\theta_n\mathbf{y}, S\theta_n\mathbf{g}\}A(x, \mathbf{y}).$$

where the latter is understood to be 0 if no such x exists. If, as a function of $\mathbf{g}$, we can determine an n such that

$$S\theta_n\mathbf{g} = S\theta_{n+1}\mathbf{g}$$

then $f = \theta_n$ is a solution of the principal semi-equation. Call this the *principal equation.*

Remark The principal equation arises in another proof-theoretic context, namely, in deriving the so-called *no-counterexample interpretation* of B from the witnesses of the no-counterexample interpretation of A and $A \to B$, when A and B are arithmetic formulas. (Gödel 1938a) refers to a solution of it using "Souslin's schema," meaning bar recursion, but does not give details. (Kohlenbach 1999) actually carries out the derivation using an extensional form of bar recursion. This is discussed in (Tait 2005). I will sketch here the derivation given in (Tait 1965a) where bar recursion is avoided using induction non ordinals $< \epsilon_0$.

In 1962 Paul Cohen showed me a handwritten manuscript in which he presented the present procedure for eliminating the Skolem functions in first-order number theory, intuitively applying bar recursion to solve the principal equation. He had at that time no prior knowledge of the work in Hilbert's school on the ϵ-calculus. □

Suppose that f is a numerical constant, so that $\mathbf{y}$ is null. In that case, it is immediate that $\theta_2 = \theta_1$ and so $f = \theta_1$ solves the principal equation. So we may assume that f is a function constant of some type $\omega^{m+1} \to \omega$ and, by contracting the arguments, we can assume that $m = 0$. The nth *sequence number*

$$\bar{h}n$$

of h is defined to be h if h is a number and it is the usual sequence number $\langle h0, \ldots, h(n-1)\rangle^{\#}$ of $\langle h0, \ldots, h(n-1)\rangle$ of h if h is a numerical function of one variable. If it is a numerical function of m variables with $m > 1$, then $\bar{h}n = \bar{h}'n$, where h' is the function of one variable with $h'\langle x_1, \ldots, x_m\rangle^+ = hx_1 \cdots, x_m$ and $\langle x_1, \ldots, x_m\rangle^+$ denotes the standard bijection from ω^m onto ω. For sequences of numbers and numerical functions, set

$$\overline{h_0, \ldots, h_p}n = \langle \bar{h}_0 n, \ldots, \bar{h}_p n\rangle^+.$$

A numerical-valued function S in $\mathcal{T}^+$ of $f, \mathbf{g}$ can be associated with a triple consisting of:

- An ordinal $\alpha_S < \epsilon_0$
- A non-decreasing function $\Phi_S : \omega \to \omega$ in $\mathcal{T}^+$ and
- A function $\Psi_S : \omega \to \alpha_S$ in $\mathcal{T}^+$

such that the following are theorems of $\mathcal{T}^+$:

$$\Phi_S(\overline{f, \mathbf{g}}n) = 0 \Rightarrow \Psi_S(\overline{f, \mathbf{g}}(n+1)) \prec \Psi_S(\overline{f, \mathbf{g}}n)$$

and

$$\Phi_S(\overline{f, \mathbf{g}}n) > 0 \Rightarrow Sf\mathbf{g} = \Phi_S(\overline{f, \mathbf{g}}(n+m)) - 1$$

for all m. In fact when S is defined in $PRA^2_{\epsilon_0}$ using only recursions on β, then $\alpha_S < \omega^\beta$. (Tait 1965a, §5)

Let our given S be so represented by α_S, Φ_S and Ψ_S. We show how toobtain a solution

$$n = N\mathbf{g}$$

of the principal equation in $\mathcal{T}^+$. Since

$$\Phi_S(\overline{\theta_n, \mathbf{g}}k) = 0 \Rightarrow \Psi_S(\overline{\theta_n, \mathbf{g}}(k+1)) < \Psi_S(\overline{\theta_n, \mathbf{g}}k) < \alpha_S$$

it follows that

$$r_n = \mu x[\Phi_S(\overline{\theta_n, \mathbf{g}}x) > 0]$$

is definable by recursion on α_S. For each $m \geq 0$

$$S\theta_m\mathbf{g} = \Phi_S(\overline{\theta_m \mathbf{g}}r_m) - 1 \tag{1}$$

Let

$$m_{n,1} < \ldots < m_{n,p_n}$$

be all the $x < r_n$ such that $\theta_n x = 0$. Set

$$\gamma_{n,i} = \Psi_S(\overline{\theta_n, \mathbf{g}}(m_{n,i} + 1))$$

Thus

$$\gamma_{n,1} \succ \cdots \succ \gamma_{n,p_n}$$

Now set

$$\gamma_n = E^{\gamma_{n,1}} \oplus \cdots \oplus E^{\gamma_{n,p_n}} \prec E^{\alpha_S}.$$

Now assume that $S\theta_n\mathbf{g} \neq S\theta_{n+1}\mathbf{g}$. Then $\Phi_S(\overline{\theta_n, \mathbf{g}}r_n) \neq \Phi_S(\overline{\theta_{n+1}, \mathbf{g}}r_{n+1})$. So $\overline{\theta_n \mathbf{g}}r_n \neq \overline{\theta_{n+1}\mathbf{g}}r_n$, since otherwise $r_{n+1} = r_n$ and so, by (1), $S\theta_n\mathbf{g} = S\theta_{n+1}\mathbf{g}$. Hence $\overline{\theta_n}r_n \neq \overline{\theta_{n+1}}r_n$. Let x be the least number such that $\theta_n x \neq \theta_{n+1}x$. Then $x < r_n$ and $\theta_n x = 0$. I.e. $x = m_{n,i}$ for some $i = 1, \ldots, p_n$. Thus, for

$j < i, m_{n+1j} = m_{n,j}$ and so $\gamma_{n+1,j} = \gamma_{n,j}$. If there is no $m_{n+1,i}$, then clearly $\gamma_{n+1} \prec \gamma_n$. If $m_{n+1,i}$ exists, then it is $> m_{n,i}$ and so the sequence

$$\langle \overline{\theta_n \mathbf{g}} 0, \ldots, \overline{\theta_n \mathbf{g}} m_{n,i} \rangle = \langle \overline{\theta_{n+1} \mathbf{g}} 0, \ldots, \overline{\theta_{n+1} \mathbf{g}} m_{n,i} \rangle$$

is a proper initial subsequence of $\langle \overline{\theta_{n+1} \mathbf{g}} 0, \ldots, \overline{\theta_{n+1} \mathbf{g}} m_{n+1,i} \rangle$. Hence $\gamma_{n+1,i} \prec \gamma_{n,i}$ and therefore $\gamma_{n+1} \prec \gamma_n$. We have proved that

$$S\theta_n \mathbf{g} \neq S\theta_{n+1} \mathbf{g} \Rightarrow \gamma_{n+1} \prec \gamma_n.$$

Hence

$$N\mathbf{g} = \mu x S\theta_x \mathbf{g} = S\theta_{x+1} \mathbf{g}$$

can be defined by recursion on E^{α_S}. Thus our solution of the principal semi-equation is

$$\Theta_{N\mathbf{g}} \mathbf{g}.$$

Substituting this for f in the critical formulas, we have reduced the number of Skolem function constants by one.

3 Eliminating Skolem Functions in the Case of Predicate Logic

Now we consider the case in which $\mathcal{T}$ is a first-order theory whose axioms are quantifier-free and closed under substitution. $\mathcal{T}^*$ is defined exactly as above, except that there are no axioms of the second kind for Skolem functions.

Again, let $\mathcal{D}$ be a deduction of C in $\mathcal{T}^*$ which we can assume to contain no variables other than those in C, since the others can be replaced by some individual constant. (If there are none, add one, carry out the following elimination procedure and then replace the constant by a variable.) Let $\mathbf{z}$ be a list of the distinct variables in C. All of the axioms in $\mathcal{D}$ other than the axioms for Skolem functions are obtained from valid numerical formulas by substituting terms containing Skolem function constants for variables and let

$$f_{A_1}, \ldots, f_{A_m}$$

be all of the distinct Skolem function constants occurring in $\mathcal{D}$, listed in order of non-decreasing rank. Only axioms for these constants of the first kind occur in $\mathcal{D}$.

Again we define the functions

$$\phi_1, \ldots, \phi_m$$

such that replacing each term $f_i \mathbf{t_i}$ by $\phi_i \mathbf{z} \mathbf{t_i}$ for $i = 1, \ldots, m$ transforms the axioms for the f_{A_i} into theorems of a suitable quantifier-free extension $\mathcal{T}^+$ of the quantifier-free part of $\mathcal{T}$. The result of this substitution in C will then be a theorem of $\mathcal{T}^+$.

$\mathcal{T}^+$ contains variables over individuals and over individual-valued functions of n individuals ($n > 0$) and it contains constants for functions of these variables whose values may be individuals or functions from n individuals to individuals. These constants are introduced by explicit definition or by definition by cases. The remaining axioms are the quantifier-free axioms of $\mathcal{T}$, closed under substitution for individual terms. Definition by cases

$$\Phi\mathbf{z} = \begin{cases} \Psi\mathbf{z} & \text{if } A(\mathbf{z}) \\ \Xi\mathbf{z} & \text{if } \neg A(\mathbf{z}) \end{cases}$$

is obviously expressed by quantifier-free axioms. Since $B(\Phi\mathbf{z})$ is equivalent in s to

$$[A(\mathbf{z}) \to B(\Psi\mathbf{z})] \wedge [\neg A(\mathbf{z}) \to B(\Xi\mathbf{z})]$$

$\mathcal{T}^+$ is conservative over the quantifier-free part of $\mathcal{T}$.

Again, $f = f_{A_m}$ has the highest rank of Skolem function constants in $\mathcal{D}$. Let $\mathbf{g} = \mathbf{z}, f_{A_1}, \ldots, f_{A_{m-1}}$.

$$C(f, \mathbf{g}, \mathbf{y}) = \bigvee_{i=1}^{p} A_m(S_i(f, \mathbf{g}), \mathbf{y}) \to A_m(f\mathbf{y}, \mathbf{y})$$

expresses the conjunction of axioms

$$A_m(S_i(f, \mathbf{g}), \mathbf{y}) \to A_m(f\mathbf{y}, \mathbf{y})$$

for f. Every axiom for f in $\mathcal{D}$ results by substitution for the $\mathbf{y}$ in one of these axioms. Let θ be the constant function defined by

$$\theta\mathbf{y} = c$$

where c is some individual constant of $\mathcal{T}$. By repeated use of definition by cases, we define $\Phi\mathbf{g}$ to be $\theta\mathbf{y} = c$, if $C(\theta, \mathbf{g}, \mathbf{y})$ and, if $\neg C(\theta, \mathbf{g}, \mathbf{y})$, to be $S_i(\theta, \mathbf{g})$ for the least $i \leq p$ such that $A_m(S_i(\theta, \mathbf{g}), \mathbf{y})$. It follows that

$$C(\Phi\mathbf{g}\mathbf{y}, \mathbf{y})$$

is a theorem of $\mathcal{T}^+$. Substitute $\Phi\mathbf{g}$ for f in the axioms for the remaining Skolem functions $f_{A_1}, \ldots, f_{A_{m-1}}$ and now solve for this shorter list of Skolem functions.

Let $\mathcal{D}$ be a deduction in $\mathcal{T}$ of

$$C = \exists x_1 \forall y_1 \cdots \exists x_n \forall y_n A(x_i, y_j)$$

We obtain from $\mathcal{D}$ a deduction of

$$\exists x_1 \cdots x_n A(x_i, g_j x_1 \cdots x_j)$$

where the g_j are distinct new function variables, or in T^*

$$A(s_i, g_j s_1 \cdots s_j)$$

where the s_i are terms containing Skolem function constants. In $\mathcal{T}^+$ we then obtain

$$A(t_i, g_j t_1 \cdots t_j)$$

where the t_i are defined by multiple cases. Eliminating them we obtain a deduction in $\mathcal{T}$ of

$$\bigvee_{k_1<r_1} \cdots \bigvee_{k_n<r_n} A(u_{k_i}, f_j u_{k_1} \cdots u_{k_n})$$

This is the *First ϵ-Theorem* of (Hilbert & Bernays 1939). (Tait 1965b, §6.3)

4 Failure of Continuity of Third-order Computable Functions.

We note that there is a third order function

$$F : [[\omega \to \omega] \to \omega] \to \omega$$

and a computable sequence $\langle \phi_n \mid n < \omega \rangle$ such that

$$\phi_n : [\omega \to \omega] \to \omega$$

and

$$\phi_n g > 0 \to \phi_{n+1} g = \phi_n g$$

but with

$$F\phi_n \neq F\phi_{n+1}$$

for all n. This casts some doubt on whether the method used above for eliminating ϵ-terms in first-order number theory will extend to second-order number theory. Of course, to be a counterexample, we would have to show that such 'discontinuous' third-order functions actually arise in solving for Skolem functions in second-order number theory. As far as I know, that question is open. (See also (Tait 1965b, the Remark in §5).)

Define

$$f : \omega \times \omega \to \omega$$

by

$$fxy = f_x y = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{if } x < y \end{cases}$$

Let $\phi_n g$ be defined for each $n < \omega$ and $g : \omega \to \omega$ by

$$\phi_n g = \begin{cases} \mu y \leq n(gy = 0) & \text{if } \exists y \leq n(gy = 0) \\ 0 & \text{if otherwise} \end{cases}$$

Then

$$\phi_n(g) > 0 \to \phi_{n+1} g = \phi_n g.$$

But

$$\phi_n f_x = \begin{cases} x+1 & \text{if } x < n \\ 0 & \text{if otherwise} \end{cases}$$

Hence

$$\phi_n(\lambda x \phi_n f_x) = n.$$

So define the third-order function

$$F : [\omega \to \omega] \to \omega$$

by

$$F(\psi) = \psi(\lambda x \psi f_x).$$

Then for all n

$$F\phi_n = n < F\phi_{n+1}.$$

References

Ackermann, W. (1940). Zur Widerspruchsfreiheit der Zahlentheorie, *Mathematische Annalen* **117**: 162–194.

Gödel, K. (1938a). Lecture at Zilsel's, *Collected Works, Vol. III*, Oxford: Oxford University Press, pp. 87–113.

Hilbert, D. & Bernays, P. (1939). *Grundlagen der Mathematik II*, Berlin: Springer-Verlag. A second edition was published in 1970.

Kohlenbach, U. (1999). On the no-counterexample interpretation, *The Journal of Symbolic Logic* **64**: 1491–1511.

Tait, W. (1965a). Functionals defined by transfinite recursion, *The Journal of Symbolic Logic* **30**: 155–174.

Tait, W. (1965b). The substitution method, *The Journal of Symbolic Logic* **30**: 175–192.

Tait, W. (2005). Gödel's reformulation of Gentzen's first consistency proof for arithmetic, *Bulletin of Symbolic Logic* **11**(2): 225–238.